

Advanced Image and Video Resolution Enhancement Techniques

Geavanceerde resolutieverbeteringstechnieken
voor beelden en beeldsequenties

Quang H. Luong

Promotor: prof. dr. ir. W. Philips
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen

Vakgroep Telecommunicatie en Informatieverwerking
Voorzitter: prof. dr. ir. H. Bruneel
Faculteit Ingenieurswetenschappen
Academiejaar 2008 - 2009



ISBN 978-90-8578-261-2
NUR 958, 965
Wettelijk depot: D/2009/10.500/19

There are two types of PhD thesis: perfect and submitted.
—Anonymous

Members of the jury

prof. dr. ir. Rik Van de Walle (Ghent University, chairman)
prof. dr. ir. Aleksandra Pižurica (Ghent University, secretary)
prof. dr. ir. Wilfried Philips (Ghent University, supervisor)
dr. ir. Patrick De Smet (Nationaal Instituut voor Criminalistiek en Criminologie)
prof. dr. ir. Peter Schelkens (Vrije Universiteit Brussel)
dr. ir. Dimitri Van De Ville (Ecole Polytechnique Fédérale de Lausanne)
prof. dr. ir. Yves d'Asseler (Ghent University)
prof. dr. ir. Lieven Eeckhout (Ghent University)

Affiliations

Research Group for Image Processing and Interpretation (IPI)
Interdisciplinary Institute for Broadband Technology (IBBT)
Department of Telecommunications and Information Processing (TELIN)
Faculty of Engineering
Ghent University

Sint-Pietersnieuwstraat 41
B-9000 Ghent
Belgium



Acknowledgements

This work would never have been possible without the assistance and support of colleagues, friends and family. I would like to express my generous gratitude to the following people:

First and foremost, I would like to thank my advisor, prof. dr. ir. Wilfried Philips for the opportunity to conduct research in the exciting research field of image processing. Not only did he show me how to do research well but he was also of great help in finalizing this thesis. I would also like to thank him for allowing me a lot of freedom in following my interests.

I would like to thank the members of my jury, prof. dr. ir. Rik Van de Walle, prof. dr. ir. Peter Schelkens, dr. ir. Dimitri Van De Ville, prof. dr. ir. Yves d'Asseler, prof. dr. ir. Lieven Eeckhout, for being in my thesis committee, and for reading and commenting this thesis. I am very grateful to dr. ir. Patrick De Smet and prof. dr. ir. Aleksandra Pižurica for the many interesting discussions and for proofreading my papers.

Many thanks to the driving force behind TELIN: Annette, Patrick and Alice, for their administrative assistance, and Philippe and Davy, for their excellent computer-related advice and repair service. Sincere thanks to all TELIN-members for the countless chatters and social events.

This work is also the result of close collaboration with a unique team of colleagues. I would like to thank some people who contributed to this research: Alessandro and Valérie De Witte (Fuzziness and Uncertainty Modelling - FUM) for the MMint link, Bart for the many cooperations and fruitful discussions, Johan for the “quick” segmentation service, Els Fieremans (Medical Image and Signal Processing - MEDISIP) and Jan for their help with the nocturnal scan sessions in the hospital and for their clarifications about the MRI functioning, Karel De Blaere (Department of Neuroradiology) and Johan Michiels (SIEMENS) for their valuable feedback on MRI, Vladimir for the insights in video processing, Ljubomir and Tijana for unraveling Vladimir's code ;), Stefaan, Ewout and Filip for the many useful hints and discussions, Sidharta for the introduction in SAR imaging, Rudy Verhelst (Mechanical Engineering Department) for co-supervising master students, Koen for his help in video mosaicing, Werner for his assistance in setting up the cameras in the experiments, Marleen and Linda for their collaboration and their hospitality in Palo Alto, Joost for his cooperation in video deblocking and all IPI-members for giving me a broader perspective on image processing and its applications. Special thanks to my office mates, Johan, Bart, Stefaan and Bruno for their help and for creating a pleasant atmosphere at work. Thank you guys!

I wish to thank my friends, my teammates and sports friends, for the funny time and to remind me that there is a lot more than just work. My special thanks goes to my family, my mother and my two sisters and their families, for their assistance in my daily life. Finally, I am especially thankful to Valérie, for her unconditional support, for her lovely smile and her sweet love. I am looking forward to moving into our new house!

Hiệp Quang Luong
March, 2009.

Samenvatting

In vele toepassingen wenst men informatie te verkrijgen die niet expliciet gegeven wordt door metingen. Een gemakkelijke manier om dit doel te bereiken is interpolatie, een techniek die reeds teruggaat tot de Babyloniërs (ca. 300 v. Chr.). Interpolatie construeert een benaderende functie op zo een manier dat het perfect samenvalt met de gegeven meetpunten. Sinds het begin van het “informatie- en communicatietijdperk” (20ste eeuw) nam het aantal ontwikkelingen in de klassieke lineaire interpolatietheorie explosief toe, waarvan de Whittaker-Kotel’nikov-Shannon bemonsteringstheorie de belangrijkste mijlpaal is.

Lineaire interpolatie wordt in het algemeen voorgesteld als een praktische oplossing voor het verbeteren van de resolutie van beelden, waarin we proberen het vermogen om beelddetails te onderscheiden te verbeteren. Met andere woorden, we willen “tussen de pixels kijken”. Jammer genoeg creëert lineaire interpolatie onvermijdelijk visuele en storende artefacten zoals trapvorming, vervaging en ringing effecten in het hogeresolutiebeeld en daarom zijn er nieuwe technieken vereist. In het laatste decennium werden er talrijke niet-lineaire of adaptieve interpolatiemethoden voorgesteld om deze artefacten te voorkomen.

In de praktijk lijden digitale beelden onder verschillende degradaties zoals vervaging, ruis, compressie-artefacten, enz. Omdat de meetpunten of pixels niet langer nauwkeurig zijn, zullen de prestaties van interpolatietechnieken enorm verminderen. Daarom worden er beeldrestauratietechnieken geïntroduceerd om resolutieverbetering door te voeren en om deze degradatiemodellen in rekening te brengen. Onder de vele bestaande restauratiemethoden gepubliceerd over de verscheidene decennia, zijn de regularisatietechnieken en de Bayesiaanse methoden in het bijzonder zeer krachtig omdat ze in staat zijn voorkennis over de degradatie en het hypothetische ideale beeld te vervatten. In het algemeen zijn de hoeveelheid voorkennis en de keuze van de overeenkomstige priormodellen bepalend voor de restauratiekwaliteit.

Meer recent, met de groeiende interesse in videobewerking, begonnen onderzoekers het probleem van beeldrestauratie voor meerdere beelden te behandelen. Omdat de twee problemen heel nauw gerelateerd zijn, kunnen de restauratietechnieken voor meerdere beelden heel veel werk hergebruiken dat reeds ontwikkeld werd voor het restauratieprobleem van enkelvoudige beelden. Hoe dan ook, restauratietechnieken voor meerdere beelden gaan nog een stap ver-

der. Gegeven een verzameling van gelijkaardige maar niet-identieke beelden, bijvoorbeeld van een videosequentie, wordt het mogelijk om de verschillende gegevens uit beelden te gebruiken voor resolutieverbetering. Met de nieuwe gegevens wordt het mogelijk om schijnbaar verloren informatie of in technische termen de ware hoge frequentie inhoud te herstellen. Daarom wordt de nieuwe generatie van beeldrestauratie superresolutie genoemd.

Hebben we eigenlijk nood aan resolutieverbeteringstechnieken? Want hoge resolutiebeelden kunnen immers bekomen worden door optica van hoge precisie en hoogkwalitatieve camerasensoren te gebruiken. Dit gaat echter gepaard met zeer hoge kosten, wat een belangrijke belemmering vormt in vele commerciële toepassingen. Aan de andere kant worden er vandaag de dag massaal goedkope camerasensoren met lage resolutie gebruikt in bijvoorbeeld webcams, GSM's en bewakingscamera's. Er is ook een sterke bijkomende beperking op het vergroten van de spatiale resolutie door het reduceren van de pixelgrootte (dit is het toenemen van het aantal pixels per eenheid van oppervlakte) in de fabricage van de sensoren. Als de pixelgrootte verkleint, neemt de hoeveelheid opgemeten licht per pixeleenheid af. Dit veroorzaakt opnameruis die de beeldkwaliteit sterk vermindert.

Resolutieverbeteringstechnieken worden vooral gebruikt in hoge-definitie televisieschermen om standaarddefinitie videomateriaal (bijvoorbeeld van DVD's) te tonen. In forensische en beveiligingstoepassingen hebben we vaak nood aan duidelijke hoge-resolutie beelden om het gezicht van een crimineel te herkennen of om de nummerplaat van een vluchtauto te lezen. Helaas gebruiken bewakingscamera's meestal goedkope lage-resolutie sensoren, die de herkenning moeilijker maken. Zodoende kunnen resolutieverbeteringstechnieken deze beelden verbeteren en helpen de rechtszaken op te lossen.

In dit proefschrift bestuderen en ontwikkelen we verscheidene beeld- en videoresolutieverbeteringstechnieken. Als eerste grote bijdrage hebben we een nieuwe niet-lineaire beeldinterpolatietechniek ontwikkeld die ongewenste artefacten ten gevolge van lineaire interpolatie verwijdert. Het voorgestelde algoritme verscherpt randen door de isokrommen in het beeld te transformeren gebruik makend van adaptieve contrastverbeteringstechnieken met randvoorwaarden. Isokrommen worden gedefinieerd als spatiale krommen met een constant intensiteitsniveau. Om gekartelde randen te vermijden worden de isokrommen op voorhand behandeld door isophote smoothing schema's met bijkomende randvoorwaarden. Experimenten tonen verbeteringen aan zowel in de numerieke PSNR resultaten als in de visuele kwaliteit ten opzichte van andere state-of-the-art interpolatietechnieken.

Een tweede nieuwheid in dit doctoraat is de introductie van twee nieuwe beeldkleuren-priors in het Bayesiaanse restauratieraamwerk. Enerzijds wordt het adaptieve bimodale kleurenprior voorgesteld, gebaseerd op het feit dat de waarde van een randpixel een combinatie is van de kleuren van de twee

verbonden gebieden, elk met een dominante kleurendistributie. Anderzijds wordt de multimodale kleurenprior gedefinieerd voor beelden die slechts enkele dominante kleuren bezitten. Restauratieresultaten tonen de doeltreffendheid en de visuele superioriteit tegenover andere interpolatie/restauratieschema's voor beelden met een sterke kleurenmodaliteit. Beide kleurenpriors zijn zeer geschikt voor de restauratie van tekeningen en tekenfilms, logo's, kaarten, enz.

Algemene beeldrestauratietechnieken maken enkel gebruik van de spatiale redundantie in de lokale omgeving. In dit werk hebben we aangetoond dat het schatten van de pixelintensiteit kan gebaseerd worden op informatie verkregen over het volledige beeld. Hiervoor wordt de aanwezigheid van gelijkaardige patronen en eigenschappen in het beeld uitgebuit, ook wel herhalende structuren genoemd. De nieuwe aanpak wordt ook wel gerefereerd als de niet-lokale strategie, die eveneens gerelateerd is aan de voorbeeld- en fractaalgebaseerde aanpakken. Als derde belangrijke bijdrage hebben we een nieuw resolutieverbeteringsschema ontwikkeld dat deze herhalende structuren uitbuit. Dit algoritme hebben we dan ook uitgebreid en geoptimaliseerd voor toepassingen in de documentverwerking. Een verbeterde segmentatie van karakters werd geïntroduceerd om de computationele complexiteit te reduceren en een specifieke beeldprior voor tekst werd eveneens toegevoegd aan het Bayesiaanse restauratieraamwerk. Experimenten tonen aan dat de karakters zeer goed worden gereconstrueerd. Daarenboven tonen OCR resultaten significante verbeteringen in vergelijking met andere bestaande resolutieverbeteringsmethoden.

Het voorgestelde algoritme legt geen beperkingen op het lettertype of alfabet en is daardoor ook geschikt voor algemene symbolen zoals muzieknoden, hiërogliefen, wiskundige symbolen of andere geschriften. Dezelfde strategie kan worden toegepast in een voorbeeldgebaseerde zoekprogramma en een efficiënte compressieschema voor documenten, wat nieuwe mogelijkheden opent voor toekomstige toepassingen.

Superresolutie van meerdere beelden is een redelijk complex probleem dat over verschillende takken van beeldverwerking loopt, waaronder b.v. bewegingsschatting of beeldregistratie, herbemonstering van onregelmatige roosters (ook fusie genaamd), beelddeconvolutie en ontruising. Dankzij zijn relatief lage rekenvereisten en lage geheugenvereisten wordt het standaard drie-stappenparadigma van de superresolutie-aanpak aanbevolen in de meeste praktische toepassingen. Deze drie opeenvolgende stappen zijn subpixel beeldregistratie, beeldfusie en beeldrestauratie.

Als vierde hoofdbijdrage in deze thesis hebben we een zeer nauwkeurige registratie-algoritme ontwikkeld, zowel in het domein van fotometrie en geometrie. De ontwikkelde lage-resolutie-naar-hoge-resolutie gradiëntgebaseerde registratiemethode met regressie met behulp van georiënteerde kernen produceert momenteel de meest nauwkeurige subpixel informatie onder verscheidene state-of-the-art methoden. Voor het fotometrische en

gecombineerde geometrische/fotometrische registratieprobleem hebben we het gebruik van de totale kleinste-kwadratenmethode voorgesteld. De totale kleinste-kwadratenoplossing produceert in beide gevallen meer nauwkeurige en consistente registratieparameters in vergelijking met de gebruikelijke kleinste-kwadratenaanpak in de literatuur. Met hetzelfde idee hebben we het kernregressie-algoritme voorgesteld en ontwikkeld, wat reeds een state-of-the-art fusietechniek is, gebruik makend van totale kleinste-kwadraten, om positionele of registratiefouten te behandelen. Numerieke experimenten tonen aan dat de ontwikkelde methode accurater en robuuster is dan de standaard kernregressie-algoritmen.

In een uitgebreide studie in nauwe samenwerking met MEDISIP-IBBT-IBITECH en de vakgroep radiologie (Universitair Ziekenhuis Gent) hebben we enkele beperkingen uitgestippeld in de recente ontwikkelingen van MRI superresolutie reconstructie en hebben we ook aangetoond dat klassieke superresolutie niet kan worden toegepast in het Fourier-gecodeerde vlak door de volledige afwezigheid van frequentieverwarring gedurende MRI opnames.

Als vijfde vernieuwde bijdrage in dit werk hebben we een elegante manier bedacht voor de beeldresolutie bij meerdere geroteerde MRI acquisities te verhogen. We hebben een nieuwe hybride reconstructie-algoritme voorgesteld die in het beelddomein herbemonstert gevolgd door fusie van meerdere gealigneerde k -ruimte opnames. Simulaties tonen de sterkte aan van onze methode, zowel kwantitatief als kwalitatief. De resultaten tonen ook verbeteringen op echte MRI scans van een resolutiefantom en een ajuin. Analyse van de Fourier-gegevens toont aan dat we echte spatiale resolutie verworven hebben. Praktische implementaties leiden tot het gebruik van niet-vierkante voxels zoals gebruikelijk in PROPELLER MRI schema's.

Tot slot, uit ons werk kwamen 3 papers voort gepubliceerd in de Science Citation Index, waarvan 2 als eerste auteur. In totaal verschenen er 24 andere papers in de proceedings van internationale en nationale conferenties, waarvan 13 als eerste auteur.

Summary

In many applications, one wishes to acquire information that is not explicitly given by measurements. An easy way to achieve this goal is *interpolation*, which already dates back to the Babylonians (ca. 300 BC). Interpolation constructs an approximating function in such a way that it perfectly agrees with the given measurement points. Since the beginning of the Information and Communication Era (20th century), developments in classical linear interpolation theory have been explosive, of which the (Whittaker-Kotel'nikov-Shannon) *sampling theory* is the most important milestone.

Linear interpolation is commonly proposed as a practical solution for the image resolution enhancement problem, in which we try to increase the ability to distinguish image details. In other words, we would like “to see between the pixels”. Unfortunately, linear interpolation inevitably creates visual and annoying artefacts such as staircasing, blur and ringing in the high-resolution image and therefore, new techniques are required. In the last decade, numerous *non-linear* or *adaptive* image interpolation methods have been proposed to prevent these artefacts.

In practice, digital images suffer from various degradations such as blur, noise, compression artefacts, etc. Because the measurement points or pixels are no longer accurate, the performance of interpolation techniques will decrease tremendously. Therefore, image restoration techniques are then introduced to perform resolution enhancement and to take these degradation models into account. Among the numerous existing restoration methods published over several decades, regularization techniques and especially the Bayesian methods are very powerful because they are able to include prior knowledge about the degradation and the hypothetical ideal image. In general, the amount of prior knowledge and the choice of the corresponding image prior models determine the performance in terms of restoration quality.

More recently, with the growing interest in video processing, researchers started to address the problem of multi-frame image restoration. Since the two problems are closely related, much of the work in multi-frame restoration borrows from techniques developed for the single image restoration problem. However, multi-frame restoration techniques take us a step further. Given a set of similar but non-identical images, e.g. from a video stream, it becomes possible to use different image data for resolution improvement. With the new

data it becomes possible to recover apparently lost information or the *true* high frequency content in technical terms. Therefore, this new generation of image restoration techniques is referred to as *super-resolution*.

Do we actually need resolution enhancement techniques? After all, high-resolution images can be acquired using high precision optics and high quality camera sensors. Unfortunately, this is accompanied by very high costs, which is an important concern in many commercial imaging applications. On the other hand, cheap camera sensors with low resolution are used massively nowadays in for example webcams, mobile phones and surveillance cameras. Additionally, there is a strong limitation on increasing spatial resolution by reducing the sensors (i.e. increasing the number of pixels per unit area) by manufacturing techniques. If the pixel size decreases, the amount of incoming light per pixel unit also decreases. This generates shot noise that degrades the image quality severely.

Resolution enhancement algorithms are for instance used in high definition television screens to display standard definition video material (e.g. from DVD disks). In forensics and homeland security applications, we often need clean high-resolution images to recognize the face of a criminal or to read the license plate of a getaway car. Unfortunately, surveillance cameras commonly employ cheap low-resolution sensors, which make the recognition more difficult. Therefore, resolution enhancement techniques can improve these images and help to solve the cases.

In this dissertation, we study and develop several image and video resolution enhancement techniques. As the first main contribution, we have developed a novel non-linear image interpolation technique that eliminates unwanted artefacts produced by linear interpolation. The proposed algorithm sharpens edges by mapping the image level curves using constrained adaptive contrast enhancement techniques. Level curves are defined as spatial curves with a constant intensity level. To avoid jagged edges, the level curves are preprocessed by a constrained isophote smoothing scheme. Experiments show improvements in both numerical PSNR results as well as in visual quality compared to other state-of-the-art interpolation techniques.

A second novelty in this dissertation is the introduction of two new image colour priors in the Bayesian restoration framework. On the one hand, the adaptive bimodal colour prior assumes that the value of an edge pixel is a combination of the colours of two connected regions, each having a dominant colour distribution. On the other hand, the multimodal colour prior is proposed for images that normally just have a few dominant colours. Restoration results show the effectiveness and the visual superiority to other interpolation/restoration schemes for images with a strong colour modality. Both colour priors are found very suitable to the restoration of drawings and

cartoons, logos, maps, etc.

Common image restoration techniques only exploit the spatial redundancy in a local neighbourhood. In this work, we have demonstrated that the estimation of the restored pixel intensity can be based on information retrieved from the whole image, thereby exploiting the presence of similar patterns and features in the image, which we call *repetitive structures*. The new approach is referred to as the non-local strategy, which is also related to the exemplar- and fractal-based approaches. As the third important contribution, we have developed a novel resolution enhancement scheme that exploits these repetitive structures. We also have extended and optimized this algorithm for document image processing applications. An improved character segmentation scheme is introduced to reduce the computational complexity and an additional text specific image prior is included in the Bayesian restoration framework. Experiments show that characters are reconstructed very well. In addition, OCR accuracy results show significant improvements in comparison with other existing resolution enhancement methods.

The proposed algorithm is not restricted to font type or alphabet, therefore, it is also suitable to generic symbols such as musical notes, hieroglyphics or mathematical symbols. The same strategy can also be applied in an exemplar-based search engine and in an efficient document compression scheme, which opens up new possibilities in future applications.

Multi-frame super-resolution is quite a complex problem, which spans over several fields of image processing, such as motion estimation or image registration, image reconstruction from irregularly spaced samples (also called fusion), image deconvolution and denoising. Due to its relatively low computational load and low memory requirements, the standard three-step paradigm of the super-resolution approach is recommended in most practical applications. These three successive steps are subpixel image alignment, image fusion and image restoration.

As the fourth main contribution in this dissertation, we have developed very accurate registration algorithms, both in photometric and geometric domain. The proposed low-resolution-to-high-resolution gradient-based registration method with steering kernel regression fusion currently produces the most accurate subpixel information among several state-of-the-art methods. For the photometric and joint geometric/photometric registration problem, we have proposed the use of the total least square framework. The total least square solution produces in both cases more accurate and consistent registration parameters compared to the ordinary least square approach, which is commonly employed in the literature. In the same spirit, we also have proposed and developed the kernel regression algorithm, which is a state-of-the-art fusion technique, in the total least square sense to handle positional or registration errors. Numerical experiments show that the proposed method is more accurate and robust compared to the standard kernel regression algorithms.

In an extensive study in close collaboration with MEDISIP-IBBT-IBITECH and the department of radiology (Ghent university hospital), we have pointed out some limitations in the recent developments in super-resolution *magnetic resonance imaging* (MRI) reconstruction and we also have argued that classical super-resolution cannot be applied in the Fourier encoded plane because of the complete absence of frequency aliasing during MRI acquisition.

As the fifth main novelty in this work, we have introduced an elegant way to enhance the image resolution by multiple *rotated* MRI acquisitions. We have proposed a novel hybrid reconstruction algorithm that performs resampling in the image domain followed by fusion of multiple aligned k-space data. Simulations demonstrate the superiority of our method, both quantitatively and qualitatively. The results also demonstrate improvements on real MRI data of a resolution phantom and an onion. Analyzing the Fourier data reveals that we really have gained *true* spatial resolution. Practical implementations require non-squared voxel sizes as in PROPELLER MRI schemes.

Finally, our work resulted in 3 papers that have been published in the Science Citation Index, of which 2 as first author. In total, 24 other papers appeared in the proceedings of international and national conferences, of which 13 as first author.

Contents

1	General introduction	1
1.1	Applications of resolution enhancement	3
1.2	Contributions and list of publications	4
1.3	Organization of this dissertation	7
2	Linear interpolation theory	11
2.1	The sampling theorem	11
2.1.1	The Fourier transform	11
2.1.2	The sampling process	12
2.1.3	Frequency aliasing	13
2.1.4	Ideal reconstruction	16
2.1.5	Sampling methods	18
2.1.6	Modern sampling theory	19
2.2	Linear image interpolation	22
2.2.1	Sinc-approximating interpolation kernels	23
2.2.2	Piecewise polynomial-based interpolation kernels	25
2.2.3	Generalized interpolation kernels	27
2.2.4	Discussion	28
2.3	Image restoration	35
2.3.1	Adaptive interpolation	35
2.3.2	Acquisition model	36
2.3.3	Image restoration as an ill-posed inverse problem	36
2.4	Super-resolution imaging	38
2.4.1	Multi-frame restoration	39
2.4.2	Classification	40
2.5	Related problems	41
2.6	Conclusion	43
3	Level curve mapping interpolation	45
3.1	Non-linear image interpolation	46
3.1.1	Edge-directed interpolation	46
3.1.2	Example-based interpolation	47
3.1.3	Restoration-based interpolation	48
3.1.4	Discussion	49
3.2	Proposed level curve mapping interpolation	50
3.2.1	Constrained isophote smoothing	53

3.2.2	Image sharpening	56
3.2.2.1	Image sharpening via shock filters	56
3.2.2.2	Adaptive contrast enhancement	57
3.2.3	Colour image interpolation	60
3.3	Experimental results	61
3.4	Conclusion	69
4	High-resolution image restoration using colour priors	75
4.1	Image restoration techniques	76
4.1.1	Classification of image restoration techniques	76
4.1.1.1	Image denoising	77
4.1.1.2	Image deconvolution	77
4.1.1.3	Restoration of other degradations	78
4.1.2	Joint denoising and deconvolution	78
4.1.2.1	Inverse filtering and ML estimator	80
4.1.2.2	Regularization techniques	81
4.1.2.3	Non-regularized approaches	83
4.2	Regularization from a Bayesian point of view	84
4.3	Image priors	86
4.3.1	The evolution of image priors	87
4.3.2	Geometry-driven smoothness prior	91
4.3.3	Proposed colour image priors	92
4.3.3.1	Adaptive bimodal colour prior	93
4.3.3.2	Multimodal colour prior	95
4.4	Optimization	97
4.4.1	The choice of regularization parameter	97
4.4.2	Gradient-based minimization	98
4.4.2.1	Restoration with adaptive bimodal colour prior	99
4.4.2.2	Restoration with multimodal colour prior	100
4.5	Experimental results	101
4.5.1	Restoration with adaptive bimodal colour prior	102
4.5.2	Restoration with multimodal colour prior	105
4.6	Conclusion	108
5	Non-local reconstruction methods	111
5.1	Repetitive structures, examples and self-similarity	112
5.2	Non-local image denoising	113
5.2.1	The original non-local means algorithm	115
5.2.2	On the improvement of non-local means	117
5.2.2.1	Accelerating the non-local means filter	117
5.2.2.2	Enhancing the visual quality	119
5.3	Non-local resolution enhancement	121
5.3.1	Selection and registration of repetitive structures	121
5.3.2	Robust data fusion	122
5.3.3	Joint denoising and deblurring	124
5.4	Experimental results	125

5.4.1	Resolution enhancement of regular images	125
5.4.2	Resolution enhancement of text images	130
5.5	Conclusion	134
6	Application to document image processing	137
6.1	Related work	138
6.2	Robust restoration of greyscale document images	139
6.2.1	Character segmentation	140
6.2.1.1	Background modelling	140
6.2.1.2	Segmentation of lines	141
6.2.1.3	Search for non-linear segmentation paths	142
6.2.1.4	Bounding box determination	145
6.2.2	Matching and aligning similar characters	145
6.2.2.1	Reduction of search space	145
6.2.2.2	Matching characters	146
6.2.2.3	Dealing with imperfect character segmentation	146
6.2.2.4	Subpixel registration	147
6.2.3	Robust Bayesian restoration	147
6.2.3.1	High-resolution fusion	147
6.2.3.2	Joint deconvolution and denoising	148
6.3	Experimental results	149
6.4	Robust restoration of binary document images	156
6.4.1	Clustering	159
6.4.2	Subpixel alignment of binary characters	160
6.4.3	Noise smoothing	162
6.4.4	Results	164
6.5	Conclusion	168
7	Multi-frame super-resolution restoration	169
7.1	Survey of multi-frame super-resolution techniques	170
7.1.1	Frequency-domain super-resolution restoration	170
7.1.2	Spatial-domain super-resolution restoration	171
7.1.3	Motionless super-resolution restoration	175
7.1.4	Super-resolution in the temporal direction	176
7.1.5	On the performance of super-resolution techniques	177
7.2	Motion estimation	180
7.2.1	Pixel-based motion estimation	181
7.2.2	Region-based motion estimation	184
7.2.2.1	Block matching methods	185
7.2.2.2	Segment-based methods	186
7.2.2.3	Mesh-based methods	187
7.2.3	Image-based motion estimation	187
7.2.3.1	Fourier methods	187
7.2.3.2	Feature point matching	188
7.2.3.3	Gradient-based registration	191
7.2.3.4	Consistent registration	195

7.3	Photometric registration	197
7.3.1	Related work	198
7.3.2	Photometric registration in the total least square sense	201
7.4	Fusion of irregularly spaced samples	206
7.4.1	Survey of fusion techniques	206
7.4.2	Kernel regression in the ordinary least square sense	208
7.4.3	Kernel regression in the total least square sense	210
7.5	Experimental results	213
7.5.1	Super-resolution in time	213
7.5.2	Geometric image registration	215
7.5.3	Photometric image registration	222
7.5.4	Image fusion	225
7.5.5	Practical super-resolution	231
7.6	Conclusion	239
8	MRI resolution enhancement	241
8.1	Introduction to resolution enhancement in MRI applications	242
8.2	On super-resolution magnetic resonance imaging	244
8.2.1	Basic concepts of k-space	244
8.2.2	Super-resolution in MRI?	247
8.2.3	Does a k-space span limitation affect MRI resolution?	248
8.2.4	Subject shifts before acquisition	250
8.3	Multi-frame MRI reconstruction in the Fourier encoded plane	251
8.3.1	Acquisition of low-resolution images	251
8.3.2	The proposed reconstruction algorithm	253
8.3.2.1	Resampling of k-space data	254
8.3.2.2	Weighted fusion of aligned k-space data	255
8.4	Results and evaluation	256
8.4.1	Simulated experiment	256
8.4.2	MRI of a hardware resolution phantom	260
8.4.3	MRI of an onion	264
8.5	Potential future applications	266
8.6	Conclusion	266
9	Conclusion	269
9.1	Review of our contributions	269
9.2	Hints for future research	272

1

General introduction

The beginning is the most important part of the work
—Plato

In this dissertation, we study several image and video resolution enhancement techniques. Images can be seen as a set of pixels (typically denoted by its intensity or colour value) placed on a (regular) grid, while video is denoted as a sequence of images, in which time is referred to as an extra dimension.

Just to go straight to the point, we can ask ourselves: what does *resolution* mean anyway? A common fallacy is that image resolution is simply defined as the number of pixels in an image, which is actually not correct.¹ Resolution is in essence related to the ability to distinguish image details, for example: can we resolve two closely spaced point sources? In effect, the spatial resolution refers to the number of *independent* pixel values per unit length, which is often expressed in *pixels per inch* (PPI) in scanners and monitors. In remote sensing and medical applications, the spatial resolution is often *inversely* determined by the sampling interval (in m or mm). Analogously, we can for example define the temporal resolution, i.e. the ability to resolve events at different points in time, which is usually expressed in *frames per second* (FPS).

We illustrate the difference between spatial resolution and image size with some examples in Figure 1.1.² On the top left side, we show the 0.26 megapixels image with a spatial resolution of 30 cm. The other images have a lower resolution, although the images in Figures 1.1(b) and 1.1(c) have the same number of pixels as Figure 1.1(a). The blurred image also has a lower resolution because the independence of neighbouring pixels is largely violated. On the other hand, the image size do put an upper bound on the spatial resolution for a given view (approximately 150 m by 150 m) as illustrated by Figure 1.1(d).

¹By convention, when the pixel counts are referred to as resolution, the term *image size* or *pixel resolution* is used. The image size is often described by the set of two positive integer numbers (i.e. width \times height) or by the total number of pixels in the image (e.g. expressed in *megapixels*).

²These images are obtained via Microsoft Virtual Earth.

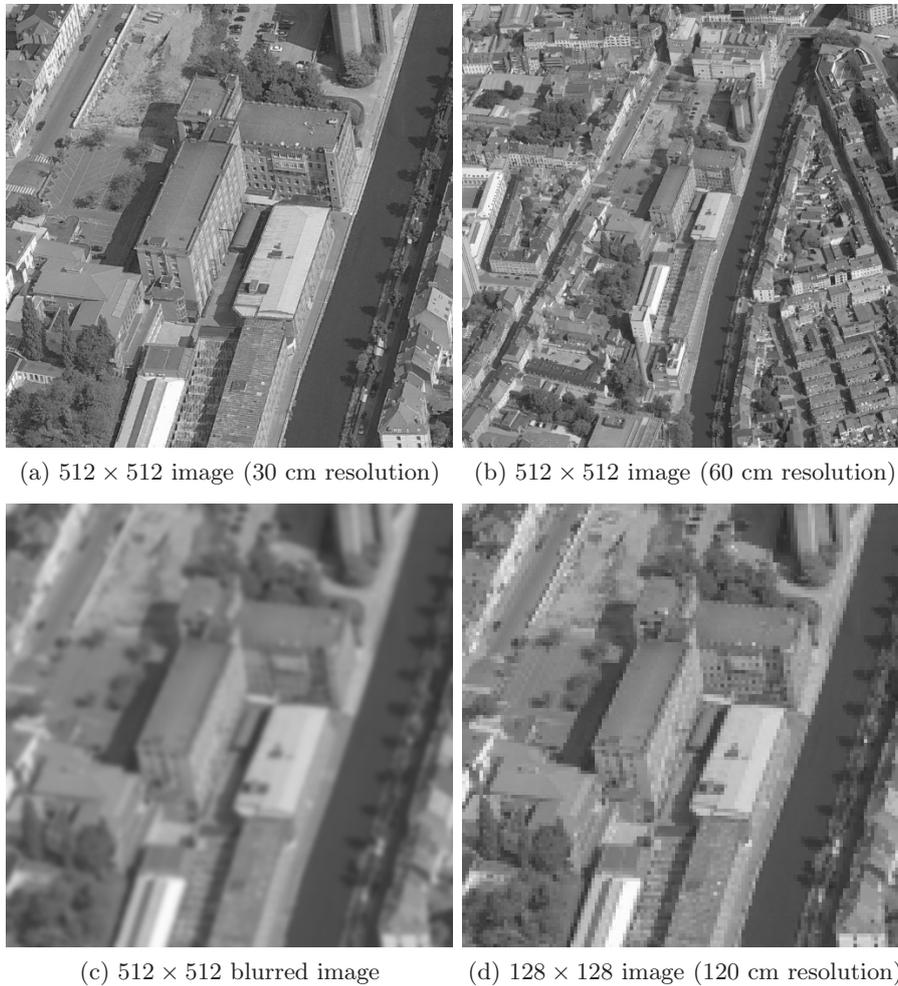


Figure 1.1: Different pictures of the Technicum building.

That is why we frequently need to enlarge the image in order to show (possible) improvements on the spatial resolution.

High-resolution images can be acquired using high precision optics and high quality camera sensors. However, this is also accompanied by very high costs, which is an important concern in many commercial imaging applications. Additionally, there is a strong limitation on increasing spatial resolution by reducing the sensors (i.e. increasing the number of pixels per unit area) by manufacturing techniques. If the pixel size decreases, the amount of incoming light per pixel unit also decreases. This generates shot noise that degrades the

image quality severely. On the one hand, a huge amount of digital images and video are thus currently acquired, compressed and stored in low resolution (e.g. by cheap off-the-shelf cameras). On the other hand, more and more high-resolution displays are sold, for example due to the growing popularity of *high definition television* (HDTV). To bridge the gap between the low-resolution material and the high-end displays, resolution enhancement techniques are needed.

Throughout this thesis, we will mainly focus on algorithms that improve the spatial resolution (or just simply resolution). These resolution enhancement techniques belong to a more general class of image and video restoration. In practice, resolution is also limited by other degradation sources such as noise, blur, compression artefacts, etc. That is why restoration techniques are proposed to solve these problems jointly. In the remainder of this introduction chapter, we give a brief overview of potential applications, we describe the outline of this thesis and we list the main novelties and contributions of our work.

1.1 Applications of resolution enhancement

Resolution enhancement techniques can be applied and are already used in various imaging domains. In this section, we will focus on the resolution enhancement aspect in two very important fields of image processing.

The first important application domain is covered by the so-called *consumer imaging*. By this we understand all applications that perform image and video format/standards conversions to fit for instance the content to the display or printer devices. Resolution enhancement (both in space and time) is needed to display *standard definition* (SD) video material on HDTV sets. These SD sources can be transmitted or stored in an interlaced format in broadcasting companies or via DVD disks. Another application is the conversion of old dirty and grainy analog video material (e.g. which are stored on tapes) to a new digital format such as DVD or Blu-Ray. Scratches and blotches can be detected and removed by dedicated inpainting algorithms.

Low-resolution document images (commonly displayed at *75 dots per inch* (DPI) on the monitor) need to be upscaled before printing on high-resolution printer devices (typically at 600 or 1200 DPI) or before displaying on the next-generation high-resolution e-papers. In general document image processing applications, resolution enhancement can improve the readability either for human beings or machines (i.e. measured in terms of *optical character recognition* (OCR) accuracy), which is for example important in increasing the performance of office automation or in digitalizing libraries.

Several resolution enhancement algorithms are already implemented in off-the-shelf digital cameras, e.g. the so-called *digital zoom* function or in the demosaicing stage (i.e. the reconstruction of full colour images from the colour-filtered CCD or CMOS samples). Resolution enhancement algorithms

can also be found in many image processing software tools such as Adobe Photoshop and GIMP.

The second important application domain is denoted by *forensics and scientific imaging*. Resolution enhancement and image restoration in general are used for instance in remote sensing and astronomy (a set of telescope or satellite images could be combined to display fine details that are not distinguishable in any of the input images), biological imaging (e.g. enhancing images from electronic and optical microscopes to improve the analysis) and medical imaging (e.g. enhancing the images from MRI, CT, PET, etc. to support medical doctors with their decisions).

Resolution enhancement and especially resampling techniques are needed to perform accurate geometric transformations and image registrations. Image registration techniques are important in for example computer-assisted alignment in modern X-ray imaging systems and in creating panoramic images that facilitate automatic annotation in soil research.

In forensics and homeland security applications, we often need clean high-resolution images to recognize the face of a criminal or to read the license plate of a getaway car. Unfortunately, surveillance cameras commonly employ cheap low-resolution sensors, which makes the recognition more difficult. Therefore, resolution enhancement techniques can improve these images and help to solve the cases.

Recently, numerous patent applications have been issued in the field of interpolation and super-resolution, see for example [Volovelsky et al., 2005, Young, 2006, Milanfar et al., 2006]. This indicates the growing interest and importance of image and video resolution enhancement techniques in the business community.

1.2 Contributions and list of publications

During the work of this dissertation, several contributions have been made. The main novelties, extensively discussed in this thesis, are:

- the development of a new non-linear interpolation technique that eliminates unwanted artefacts such as blur, staircase and ringing effects. The algorithm sharpens edges by mapping the smooth image level curves using adaptive contrast enhancement techniques (see Chapter 3). The main novelty is the combination of two existing image processing techniques (i.e. isophote smoothing and contrast enhancement), each having a different purpose, in a non-trivial way on top of the linear interpolation techniques. Additional constraints are put on these algorithms to achieve strict interpolation and to preserve the topology of the image level curves. This results in sharper high-resolution images with less disturbing artefacts compared to the state-of-the-art non-linear interpolation techniques.

This work led to an ICIP publication [Luong et al., 2005] and another conference publication [Luong and Philips, 2005].

- the introduction of two new colour image priors in the image restoration framework, namely the adaptive bimodal colour prior and the multimodal colour prior, which take advantage of local and global dominant colour distribution (see Chapter 4). The novelty is to incorporate and to adapt the bimodal black-and-white text prior for a more general class of images. The key achievements are the automatic determination of the parameters (both locally adaptive as globally) and the extension to the multimodal case. We have been able to show that the use of these priors is very powerful in the restoration of cartoons and logos, thereby outperforming other existing state-of-the-art techniques in visual quality. This work led to two publications in the Springer's *Lecture Notes on Computer Science* (LNCS) series [Luong et al., 2007, Luong and Philips, 2007b].
- the development of non-local resolution enhancement techniques that exploits repetitive structures (see Chapter 5). These techniques are adapted and optimized for greyscale and binary document image processing (see Chapter 6). The main novelty here is to use the non-local strategy, borrowed from image denoising techniques for image resolution enhancement. The blocks are selected based on the newly proposed dual matching criterion that enables a fast and robust filtering of these blocks. This work led to an ICIP publication [Luong et al., 2006c] and another publication in the LNCS series [Luong et al., 2006b] for regular image restoration, and a journal publication [Luong and Philips, 2008] and another conference publication [Luong and Philips, 2007a] for document processing. In addition, joint works with B. Goossens and A. Dauwe on improving non-local denoising led to three more conference publications [Dauwe et al., 2008, Goossens et al., 2008a, Goossens et al., 2008c].
- the derivation and development of joint geometric/photometric image registration and kernel regression in the total least square sense (see Chapter 7). The novel idea is to use the total least square metrics instead of the ordinary least square metrics, which is commonly used in the literature. In image registration applications, the total least square solution of the motion parameters is much more consistent and accurate in presence of noise compared to existing registration algorithms. In image fusion applications, the improved kernel regression algorithm appears to be much more robust to noise and registration errors than existing state-of-the-art techniques. These improvements are important for the performance of multi-frame super-resolution image reconstruction algorithms.
- the introduction of a novel multi-frame reconstruction framework for MRI images in the Fourier encoded plane (see Chapter 8). We first point out some shortcomings of the recent developments in super-resolution MRI. As the key achievement, we introduce an elegant way to enhance the

image resolution by multiple MRI acquisitions: several rotated images are combined into one high-resolution image with an enlarged k-space span. We propose a novel hybrid linear reconstruction algorithm that consists of a proper resampling scheme in the image domain followed by optimal fusion of multiple aligned k-space data. Simulations demonstrate the superiority of the proposed method compared to standard regression and interpolation techniques: more image details are revealed. This work led to a submitted journal publication [Luong et al., 2009] and a conference publication [Luong et al., 2008].

To summarize, our work resulted in 1 journal publication, 1 submitted journal publication and 4 publications in the LNCS series, of which 3 as first author. In total, 22 other papers appeared in the proceedings of international and national conferences, of which 13 as first author. Own work and contributions to other people's work are published in the field of level curve mapping interpolation [Luong et al., 2005, Luong and Philips, 2005], image restoration with colour priors [Luong et al., 2007, Luong and Philips, 2007b], non-local resolution enhancement [Luong et al., 2006c, Luong et al., 2006b, Luong and Philips, 2007a, Luong and Philips, 2008], image denoising [Dauwe et al., 2008, Goossens et al., 2008a, Goossens et al., 2008c, Goossens et al., 2008b], multi-frame MRI reconstruction [Luong et al., 2008, Luong et al., 2009], multi-frame super-resolution [Luong et al., 2004b, Luong et al., 2006d, Luong and Philips, 2006], mathematical morphology interpolation [Ledda et al., 2005, Ledda et al., 2006a, Ledda et al., 2006b, Ledda et al., 2008], multi-camera video coding [Morbée et al., 2007], forensic science [De Smet et al., 2005, De Leersnyder et al., 2006], video segmentation [Luong et al., 2004a], image registration and mosaicing [Luong et al., 2004c, Luong et al., 2006a] and audio processing [De Smet et al., 2004].

A small selection of publications that were published during the course of this research is given here:

- Luong, H. and Philips, W. (2008). Robust reconstruction of low-resolution document images by exploiting repetitive character behaviour. *International Journal on Document Analysis and Recognition (IJ DAR)*, 11(1):39–51.
- Luong, H., Goossens, B., and Philips, W. (2007). Image upscaling using global multimodal priors. In Blanc-Talon, J., Philips, W., Popescu, D., and Scheunders, P., editors, *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 4678 of *Lecture Notes in Computer Science*, pages 473–484, Delft, The Netherlands. Springer-Verlag.
- Luong, H. and Philips, W. (2007). Reconstruction of low-resolution images using adaptive bimodal priors. In Campilho, A. and Kamel, M., editors, *Proceedings of International Conference On Image Analysis And*

Recognition (ICIAR), volume LNCS 4633 of *Lecture Notes in Computer Science*, pages 69–80, Montréal, Canada. Springer-Verlag.

- Luong, H., Ledda, A., and Philips, W. (2006). Non-local image interpolation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 693–696.
- Luong, H., Ledda, A., and Philips, W. (2006). An image interpolation scheme for repetitive structures. In *Proceedings of International Conference On Image Analysis And Recognition (ICIAR)*, volume 4141 of *Lecture Notes in Computer Science*, pages 104–115. Springer-Verlag.
- Luong, H., De Smet, P., and Philips, W. (2005). Image interpolation using constrained adaptive contrast enhancement techniques. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 998–1001.
- Luong, H. and Philips, W. (2005). Sharp image interpolation by mapping level curves. In *Proceedings of Visual Communications and Image Processing (VCIP)*, pages 2012–2022.

1.3 Organization of this dissertation

The goal of this dissertation is to study and develop image and video resolution enhancement techniques. The organization of this thesis is as follows. In Chapter 2, we give a more formal description of the image resolution enhancement problem in a sampling framework. We first give a brief introduction of the Shannon sampling theorem followed by an overview of existing linear interpolation techniques. We then link the sampling framework to single image restoration techniques and multi-frame super-resolution imaging. Given these general problem setups, we can easily situate the different methods in the rest of our work. Finally, we also give a brief overview of other image processing tasks related to these problems.

Chapter 3 covers the problem of non-linear (or adaptive) interpolation. We first discuss several strategies to perform non-linear interpolation and classify the existing state-of-the-art techniques according to these interpolation strategies. Afterwards, we describe a novel approach that eliminates unwanted artefacts (i.e. blur, staircase and ringing effects) that are created by linear image interpolation methods. We then briefly discuss various strategies to extend greyscale interpolation methods to colour images. The performance of the proposed method is compared to other interpolation techniques.

In Chapter 4, we discuss resolution enhancement as an image restoration problem. We briefly address the different sources of image degradation. We give a broad overview of the existing adaptive image restoration methods

where the focus is lying on the joint denoising and deconvolution problem. We concentrate on the regularization techniques described from a Bayesian point of view and especially on the *maximum a posteriori* framework. The evolution and an overview of image priors are described in more detail. We derive the steepest descent algorithm that performs resolution enhancement with a geometry-driven smoothness prior. We introduce two new colour image priors, namely the adaptive bimodal colour prior and the multimodal colour prior, and show their influence on the restoration results.

In Chapter 5, we exploit the presence of similar (*non-local*) patterns and features in the image, which we call *repetitive structures*. We first describe the relationship between this relatively new class of non-local methods and the exemplar- and fractal-based algorithms. The original non-local means filter for the image denoising problem is described and we discuss numerous improvements both on accelerating the non-local means filter as well as on enhancing the visual quality. We introduce this novel non-local approach in the image resolution enhancement framework and demonstrate its effectiveness.

Chapter 6 continues the work on non-local methods of the previous chapter, but in a more specific context of document image processing. To exploit the repetitive behaviour of the characters in greyscale low-resolution document images, we describe and improve an existing character segmentation algorithm. We propose a Bayesian framework with a specific text image prior that combines similar characters into high-resolution prototypes. We perform several quantitative experiments that compare our framework to other existing resolution enhancement techniques. In addition, we present a similar algorithm for binary document images.

In Chapter 7, we address the multi-frame super-resolution problem, where we concentrate on a three-step implementation, which consists of image alignment, image fusion and image restoration. We first give an overview of the existing super-resolution approaches. We then discuss various aspects of motion estimation and both geometric and photometric image registration. We propose and derive a robust image alignment technique that performs joint geometric and photometric registration in the total least square sense. We describe several fusion algorithms and focus on the kernel regression framework. We also propose and derive the kernel regression algorithm in the total least square sense to handle potential registration errors. We briefly discuss super-resolution reconstruction along the temporal direction. We evaluate the subpixel accuracy of several state-of-the-art shift estimators and compare the proposed registration algorithm numerically to other approaches. We show the effectiveness of the proposed super-resolution strategy in real-world examples and deinterlacing problems.

Chapter 8 is dedicated to the multi-frame resolution enhancement of *magnetic*

resonance imaging (MRI) images in the Fourier encoded plane. We first point out some limitations in the recent developments in super-resolution MRI reconstruction and we also argue that classical super-resolution cannot be applied in the Fourier encoded plane because of the complete absence of frequency aliasing during MRI acquisition. We introduce an elegant way to enhance the image resolution by multiple MRI images, which are acquired over a rectangular k-space span. We propose a novel hybrid linear reconstruction algorithm that consists of a proper resampling scheme in the image domain followed by optimal fusion of multiple aligned k-space data. The proposed resolution enhancement framework is evaluated quantitatively and qualitatively for both simulated and real MRI data.

Chapter 9 provides the general conclusions of this work and discuss some directions for future research. To end this section, we draw a decision chart in Figure 1.2 that helps to guide the reader through this dissertation.

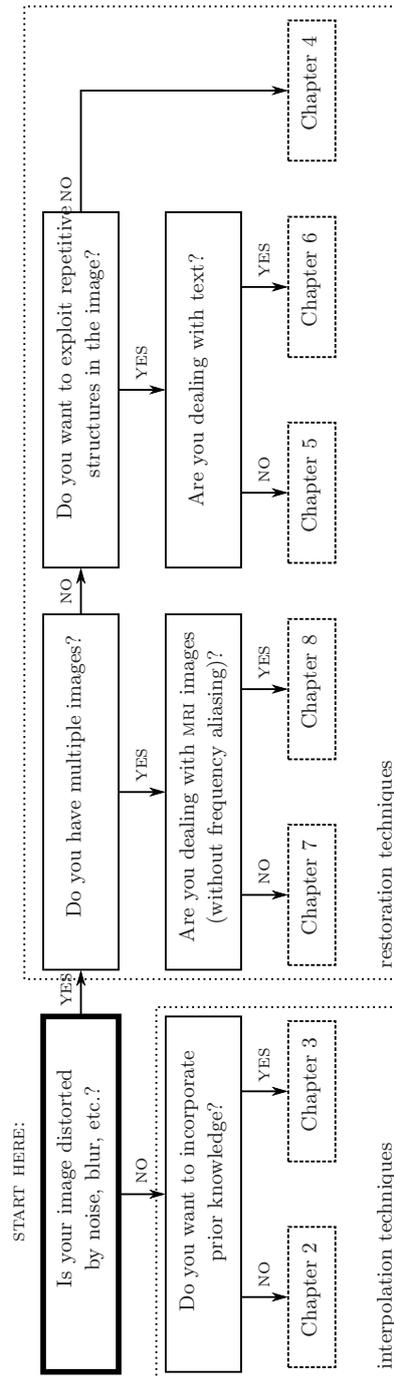


Figure 1.2: The decision chart: how to flick through this dissertation?

2

Linear interpolation theory

Think analog, act digital.
—Michael Unser

In this chapter, we provide a mathematical formulation of image resolution enhancement techniques in a sampling framework. First, we start with a brief introduction of the Shannon sampling theorem. Afterwards, we discuss the image reconstruction given a set of discrete samples and we describe linear interpolation techniques, single image restoration and multi-frame super-resolution imaging. Given these general problem setups, we can easily situate our different proposed image resolution enhancement techniques. Finally, we give a brief overview of other variants on these problems.

2.1 The sampling theorem

In the digital era, modern applications often require that we sample analog (or continuous) signals, acquire them in a digital form, perform some digital operations on them and finally reconstruct them as continuous signals. This process is illustrated in Figure 2.1. Since some useful properties of functions and signals are easy to interpret by means of their characteristics in the frequency domain, we briefly review the Fourier transform.

2.1.1 The Fourier transform

We denote the n -dimensional frequency coordinates as $\mathbf{u} \in \mathbb{R}^n$ and the spatial coordinates as $\mathbf{x} \in \mathbb{R}^n$. The n -dimensional continuous Fourier transform $F(\mathbf{u})$ of the continuous signal $f(\mathbf{x})$ is given by

$$F(\mathbf{u}) = (\mathcal{F}(f))(\mathbf{u}) = \int_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) e^{-2\pi j(\mathbf{u} \cdot \mathbf{x})} d\mathbf{x}, \quad (2.1)$$

where $\mathbf{u} \cdot \mathbf{x}$ is defined as the inner product between the two vectors. Analogously, the multi-dimensional inverse continuous Fourier transform is then denoted by

$$f(\mathbf{x}) = (\mathcal{F}^{-1}(F))(\mathbf{x}) = \int_{\mathbf{u} \in \mathbb{R}^n} F(\mathbf{u}) e^{2\pi j(\mathbf{x} \cdot \mathbf{u})} d\mathbf{u}. \quad (2.2)$$

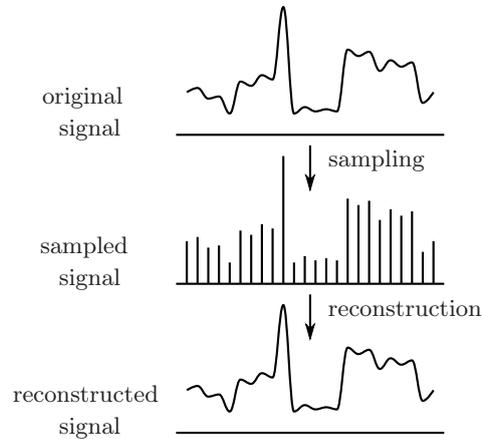


Figure 2.1: The sampling-reconstruction process: from continuous signal to discrete signal and back to continuous signal.

One of the most important properties of the Fourier transform is the convolution theorem, which states that the Fourier transform of the convolution $(f * g)(\mathbf{x})$ is the product of their Fourier transforms $F(\mathbf{u})G(\mathbf{u})$ and vice versa. The convolution of two functions is defined as

$$(f * g)(\mathbf{x}) = \int_{\mathbf{y} \in \mathbb{R}^n} f(\mathbf{y})g(\mathbf{x} - \mathbf{y})d\mathbf{y}. \quad (2.3)$$

From an implementation point of view, it is important to know that the Fourier transform is separable, i.e. you can perform the transform in n dimensions separately, which has the advantage of a lower computational complexity.

2.1.2 The sampling process

Since the Fourier transformation is separable, we restrict the explanation to the one-dimensional case for simplicity. Extension to multiple dimensions is fairly straightforward. The sampling process reads the continuous signal $f(x)$ ($x \in \mathbb{R}$) at discrete positions, of which a discrete signal is formed. Assume now that we sample $f(x)$ uniformly from the origin with a sampling interval T . At each position kT ($k \in \mathbb{Z}$) on the grid, we measure the value $f(kT)$.

To explain the sampling process in a formal way, we first introduce the impulse $\delta(\cdot)$ (or Dirac delta distribution),¹ which has the following property:

$$\int_{-\infty}^{+\infty} \delta(x - x_0)f(x)dx = f(x_0). \quad (2.4)$$

¹An alternative way of understanding the basic sampling formulas is using Poisson's summation formula instead of Dirac distributions [Unser, 2000].

We now define the impulse train or Shah² function $\text{III}_T(\cdot)$ with sampling interval T as a periodic Schwartz distribution constructed from Dirac delta:

$$\text{III}_T(x) = \sum_{k=-\infty}^{+\infty} \delta(x - kT). \quad (2.5)$$

The Fourier transform of the impulse train is then given by (via Fourier series)

$$\begin{aligned} \mathcal{F}(\text{III}_T(x)) &= \int_{-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \delta(x - kT) e^{-2\pi j u x} dx \\ &= \frac{1}{T} \sum_{k=-\infty}^{+\infty} \delta(u - \frac{k}{T}) = \frac{1}{T} \text{III}_{1/T}(u) \end{aligned} \quad (2.6)$$

So, we note that the Fourier transform of an impulse train is also an impulse train, but with a *sampling rate* $1/T$. We denote the sampled signal $f_T(x)$ as the multiplication of the continuous signal $f(x)$ with the impulse train:

$$f_T(x) = f(x) \text{III}_T(x) = \sum_{k=-\infty}^{+\infty} f(x) \delta(x - kT). \quad (2.7)$$

Via the convolution theorem, we can easily show that the Fourier transform of the sampled signal $F_T(u)$ can be expressed as

$$F_T(u) = \frac{1}{T} F * \text{III}_{1/T}(u) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} F(u - \frac{k}{T}). \quad (2.8)$$

This result shows that sampling a signal in the spatial domain at a rate $1/T$ replicates the Fourier transform of the original signal at the frequencies k/T ($k \in \mathbb{Z}$). The sampling process is illustrated in Figure 2.2, both in the spatial and frequency domain.

2.1.3 Frequency aliasing

From a reconstruction point of view, it is important to know under which conditions the original signal can be reconstructed perfectly from its samples. Therefore, we introduce the class of bandlimited signals. A function or signal is bandlimited, if the frequency values are zero above a certain threshold B (or cutoff frequency):

$$F(u) = 0 \quad (\forall u : |u| \geq B). \quad (2.9)$$

The smallest threshold or minimum upper cutoff frequency B is called the *baseband bandwidth* of the signal. By contrast, a non-baseband (passband)

²Recognizing the Russian Cyrillic letter “Sha”, the train of delta functions (or Dirac *comb* function) is named the Shah-function $\text{III}(\cdot)$.

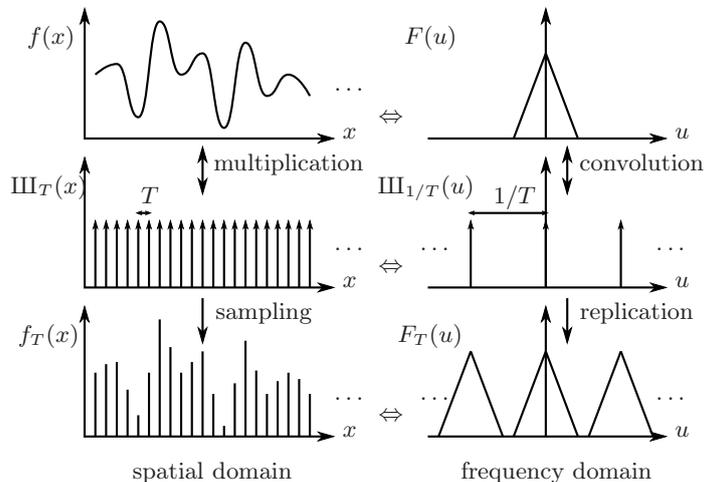


Figure 2.2: The sampling process in the spatial and frequency domain for bandlimited signals. The spectrum is replicated at frequencies k/T ($k \in \mathbb{Z}$).

bandwidth is the difference between the highest and lowest nonzero frequency. Note that this bandlimitation is just an idealization: physically, real world functions or signals are never exactly bandlimited. A law of the Fourier transform states that if a signal is finite in space or time, the support of its spectrum will be infinitely large and vice versa (except for the function $f(x) = 0$). In practice, most signals concentrate most of their energy at low frequencies. In general, the power spectrum of natural images even follows an $1/u^2$ -decay [Field, 1987]. Therefore, such signals can approximately be considered bandlimited.

The reconstruction of the continuous signal from its samples can be performed without any loss if the sampling rate or sampling frequency $1/T$ is larger than twice the cutoff frequency, which is called the Nyquist criterion:

$$\frac{1}{T} \geq 2B. \quad (2.10)$$

This is commonly known as the Shannon³ sampling theorem [Shannon, 1949]. In this case, the replicated spectra do not overlap.

On the other hand, if the Nyquist criterion is not fulfilled (i.e. the signal is undersampled), replicated spectra overlap each other in such a way that frequency components above the sampling frequency are mapped into frequencies below this limit after sampling. As a consequence, we can no longer reconstruct the

³For the sake of completeness, we have to mention that equivalent theorems are presented by Whittaker [McNamee et al., 1971, Unser, 2000], Nyquist [Nyquist, 2002] and Kotel'nikov [Kotel'nikov, 2000]. We refer the interested reader to [Lüke, 1999] for a discussion on the origins of sampling theorem.

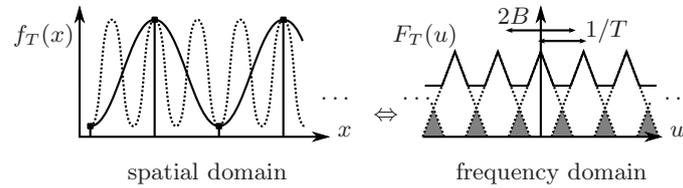


Figure 2.3: Frequency aliasing in the spatial and frequency domain for bandlimited signals.

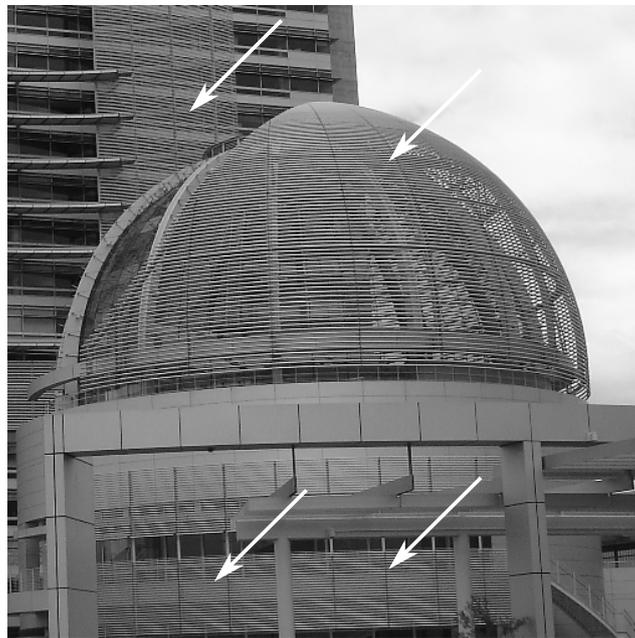


Figure 2.4: Illustration of Moiré patterns.

original signal from its samples. This phenomenon is called *frequency aliasing*⁴ and is illustrated in Figure 2.3, both in frequency and spatial domain. This annoying artefact manifests itself as Moiré patterns in images (an example is given in Figure 2.4) or in temporal aliasing in the form of stroboscopic effects or *wagon-wheel* effects in video (i.e. spoked wheels sometimes appear to be turning backwards). The latter occurs when the frame rate is too low.

There are several solutions to prevent or reduce aliasing. In the first place, we can increase the sampling rate until it is above twice the baseband bandwidth. Alternatively, we can restrict the baseband bandwidth of the original signal

⁴Its dual counterpart (e.g. during MRI acquisitions) is denoted as *spatial aliasing*. To make the presentation self-contained, we will return to this aspect in Chapter 8.

before sampling using an *anti-aliasing filter* $\Phi_{\text{aa}}(\cdot)$ (also called low-pass filter). This has the disadvantage that the signal can not be reconstructed perfectly anymore (see Figure 2.5). In theory, the ideal low-pass filter is mathematically realized by multiplying frequency values with the frequency response $\Pi(u)$ ⁵:

$$\Pi(u) = \begin{cases} 1 & \text{if } |u| \leq 1/2, \\ 0 & \text{if } |u| > 1/2. \end{cases} \quad (2.11)$$

The frequency response of the ideal anti-aliasing filter is thus $\Phi_{\text{aa}}(u) = \Pi(Tu)$, which is equivalently to applying a convolution with the sinc-function $\mathcal{F}^{-1}(\Phi_{\text{aa}}(u)) = \varphi_{\text{aa}}(x) = \text{sinc}(x)$ in the temporal or spatial domain. However, in practice, there is no such device as an ideal anti-aliasing filter because of the infinite extent of the sinc-function. Real filters approximate the ideal filter by truncating and windowing the infinite impulse response (IIR) to make a finite impulse response filter (FIR-filter).

2.1.4 Ideal reconstruction

For general resampling or regridding tasks, the interpolation step must reconstruct the continuous signal $\hat{f}(x)$ ($x \in \mathbb{R}$) from its discrete samples $f(kT)$ ($k \in \mathbb{Z}$). The signal value at position x can be estimated by a convolution of the discrete data samples with the impulse response of a reconstruction filter $\varphi_{\text{int}}(\cdot)$ (also called the interpolation or resampling kernel):

$$\hat{f}(x) = \sum_{k=-\infty}^{+\infty} f(kT)\varphi_{\text{int}}(x/T - k). \quad (2.12)$$

In ideal circumstances (i.e. when the Nyquist criterion is satisfied), we can reconstruct the original signal perfectly by multiplying the Fourier spectrum with a rectangular function (see Figure 2.5) or equivalently, perform a convolution with the Whittaker-Shannon interpolation formula [Shannon, 1949]:

$$\begin{aligned} \varphi_{\text{int,ideal}}(x/T) &= \int_{-\infty}^{+\infty} \Pi(Tu)e^{2\pi jxu} \, du \\ &= \int_{-\frac{1}{2T}}^{\frac{1}{2T}} e^{2\pi jxu} \, du \\ &= \frac{T}{2\pi jx} \left(e^{2\pi jx\frac{1}{2T}} - e^{-2\pi jx\frac{1}{2T}} \right) \\ &= \frac{\sin(\pi x/T)}{\pi x/T} \equiv \text{sinc}(x/T). \end{aligned} \quad (2.13)$$

In literature, the ideal reconstruction filter is also denoted as sinc or zero-padding interpolation. Again, in practice, there exists no ideal reconstruction filter for the same reasons as for the ideal anti-aliasing filter (as discussed

⁵Alternative definitions assign 0 or 0.5 to the boundary $|u| = 1/2$.

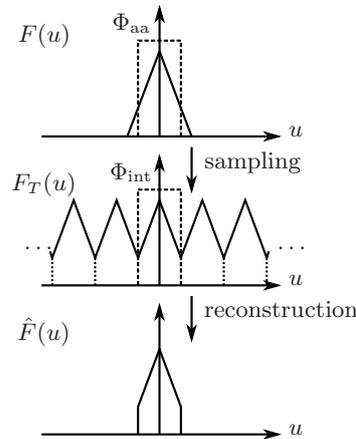


Figure 2.5: The ideal anti-aliasing and ideal reconstruction filter illustrated in the Fourier domain. We cannot reconstruct the original signal perfectly if the Nyquist criterion is not satisfied, but we can reconstruct the alias-free part of the spectrum.

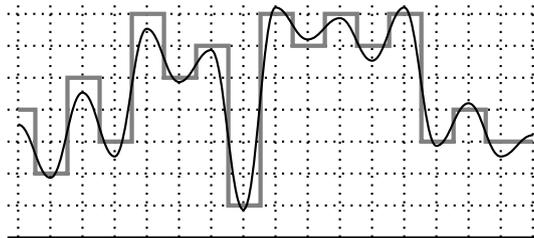


Figure 2.6: Digital signal processing is a combination of quantization and sampling: the acquisition of the original signal (black) results in a step-like digital signal (grey).

before in Section 2.1.3). It is also rarely used in general image resampling tasks because of the slow decay of the sinc function [Unser, 2000].

In practical applications, we usually cannot reconstruct the original signal perfectly due to various sources of errors such as measurement noise. As discussed earlier, we are commonly dealing with frequency aliasing and/or we have lost high frequency components due to the anti-aliasing filter. Additionally, in digital signal processing, the discrete signal values go through a quantization process as illustrated in Figure 2.6. In digital images, the discrete sampled signal values are denoted as *pixels*. In practice, an n -bits uniform integer quantization is commonly applied: typically 8-bits for greyscale images and 24-bits for full RGB-coloured images.

2.1.5 Sampling methods

Until now, we have restricted ourselves to the 1D regular sampling case (i.e. with uniform sampling intervals). By making use of the separable property of the Fourier transformation, we can easily extend the traditional sampling theory to multiple dimensions. For example, the n -dimensional interpolation kernel can be reconstructed as

$$\varphi^n(\mathbf{x}) = \prod_{i=1}^n \varphi(x_i), \quad (2.14)$$

where $\mathbf{x} \in \mathbb{R}^n$. In typical image and video applications, we are most familiar with Cartesian coordinate systems, i.e. samples are taken on an orthonormal reference grid (the 2D case is thus a squared grid). In this work, we assume that all images and video sequences are acquired on Cartesian grids, for which equation (2.14) is valid. Additionally, we want to point out that other uniform grids such as hexagonal lattices are used for instance in printing applications [Van De Ville, 2001].

When the sample positions are arbitrarily chosen, the sampling method is called non-uniform or irregular. There are essentially two strategies to reconstruct the signal from its non-uniform samples: firstly, use the same interpolation kernel everywhere (denoted as *shift-invariant*) and fit these kernels to the measurement data in a way that the reconstructed signal goes through the samples and secondly, define tailored basis functions (such as radial basis functions) that are better suited to the underlying non-uniform structure [Unser, 2000]. Irregularly sampled positions stem for example from stochastic sampling or non-Cartesian sampling trajectories (e.g. during MRI acquisitions) [Dologlou et al., 1996, Pruessmann et al., 2001]. Recently, *compressive sampling* has gained much interest in the image processing community, e.g. [Candès and Wakin, 2008, Romberg, 2008, Lustig et al., 2008]. The goal is to summarize the image in a very few samples (i.e. in a sparse representation) and to reconstruct the image in the best possible way from these sparse data. A special and important subcategory of irregular sampling methods are the multichannel sampling methods: the global set of samples can then be divided into m sets of uniformly spaced samples. As an extension of the traditional Shannon's sampling theory, Papoulis showed that a bandlimited signal can perfectly be reconstructed from uniformly sampled m sets, where the sampling rate within each set is $1/m$ of the Nyquist rate [Papoulis, 1977]. We will return to multichannel sampling in Section 2.4. The different sampling methods are illustrated in Figure 2.7.

In our sampling model, we have assumed that the signals are sampled using Dirac deltas. Alternatively, we can consider a more general sampling kernel, denoted by its *point spread function* (PSF). The sampling operation is then transformed into a convolution with the sampling kernel, or equivalently, a multiplication in the frequency domain, followed by the actual sampling. If

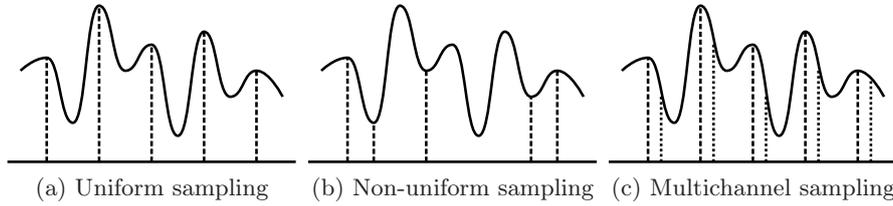


Figure 2.7: Examples of the different sampling methods.

the sampling kernel (which is a low-pass filter in general) does not remove any frequency components, the signal can be perfectly reconstructed on the condition that the sampling kernel is not too ill-conditioned [Vandewalle, 2006].

2.1.6 Modern sampling theory

The standard sampling theory states that signals need to be bandlimited for a perfect reconstruction, which is a limitation in practice because real world signals or images are never exactly bandlimited [Unser, 2000]. Alternatively, we can aim to reconstruct the signal as close as possible to the original continuous signal. This goal gives rise to the start of modern sampling theory.

The anti-aliasing prefilter introduced in Section 2.1.4 can be reinterpreted as an orthogonal projection operator that computes the minimum error bandlimited approximation of an arbitrary input signal from the Hilbert space l_2 , which consists of all functions that are square integrable in the Lebesgue's sense [Unser, 2000]. The corresponding l_2 -norm is

$$\|f\| = \left(\int_{\mathbf{x} \in \mathbb{R}^n} |f(\mathbf{x})|^2 d\mathbf{x} \right)^{1/2} = \sqrt{\langle f, f \rangle}, \quad (2.15)$$

which is induced by the conventional l_2 -inner product:

$$\langle f, g \rangle = \int_{\mathbf{x} \in \mathbb{R}^n} f^*(\mathbf{x})g(\mathbf{x})d\mathbf{x}, \quad (2.16)$$

where the $*$ -operator denotes the complex conjugate. Note that in general, any combination of prefiltering and sampling can be rewritten in terms of inner products between the continuous signal and an *analysis* function $\varphi_a(\cdot)$. Analogously, we can define the l_2 -norm via the inner product for n -dimensional discrete sequences:

$$\|c\|_{l_2} = \left(\sum_{\mathbf{k} \in \mathbb{Z}^n} |c(\mathbf{k})|^2 \right)^{1/2}, \quad (2.17)$$

$$\langle c_1, c_2 \rangle = \sum_{\mathbf{k} \in \mathbb{Z}^n} c_1^*(\mathbf{k})c_2(\mathbf{k}). \quad (2.18)$$

Now we can introduce the *generalized interpolation* formula by replacing the (sinc) interpolation kernel in equation (2.12) by a more general *generating* or *synthesis* function $\varphi_s(\cdot)$ and the signal samples by a sequence of coefficients $c(k)$:

$$s(x) = \sum_{k=-\infty}^{+\infty} c(k)\varphi_s(x/T - k). \quad (2.19)$$

This model makes the connection between a discrete sequence of coefficients $c(k)$ and a continuous function $s(x)$. There are three requirements for such a model to make sense [Unser, 2000]. Firstly, the l_2 -norm of the coefficient series should obviously be finite. Secondly, the model should be stable and unambiguously defined. Stability ensures that a small variation in coefficients result in small variations of the l_2 -norm. Furthermore, the basis functions should be linearly independent so that the basis is unambiguous: each signal $s(x)$ is then defined by a unique set of coefficients $c(k)$. In other words, the basis functions $\varphi(x - k)$ should form a Riesz basis: there must be two strictly positive constants $0 < A$ and $B < +\infty$ such that

$$A\|c\|_{l_2}^2 \leq \left\| \sum_{k \in \mathbb{Z}} c(k)\varphi(x - k) \right\|^2 \leq B\|c\|_{l_2}^2, \quad (\forall c(k) \in l_2). \quad (2.20)$$

The l_2 -space consists of all sequences with finite l_2 -norm. A basis is orthonormal if and only if $A = B = 1$. The third and final requirement is that the approximation should converge to the input signal when the sampling interval decreases. This is equivalent to the partition of unity condition [Unser, 2000]:

$$\sum_{k \in \mathbb{Z}} \varphi_s(x/T - k) = 1, \quad (\forall x \in \mathbb{R}). \quad (2.21)$$

In practice, this last condition puts the strongest constraint on the selection of acceptable generating functions. Note that the sinc-functions from the Shannon's sampling theory form an orthonormal basis and satisfy the partition of unity. From a computational point of view, it is interesting to note that equation (2.21) allows synthesis kernels with compact support.

Both traditional and modern sampling schemes follow the standard three-step paradigm: prefiltering, sampling and postfiltering as illustrated in Figure 2.8. However in modern sampling, both filters are not necessarily ideal and identical. For the minimum l_2 -norm approximation, the analysis function $\varphi_a(x - k)$ is uniquely determined by the biorthogonality condition (it is said that the analysis function is the *dual function* of the synthesis function):

$$\langle \varphi_a(x - k), \varphi_s(x - l) \rangle = \delta(x - k - l). \quad (2.22)$$

If φ_s is orthonormal, then it is $\varphi_a = \varphi_s$, which is for example the case with the sinc function in Shannon's sampling theory.

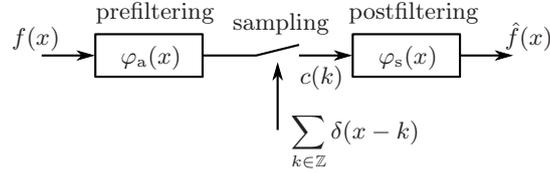


Figure 2.8: Schematic representation of the standard three-step paradigm: the analog input signal is prefiltered with φ_a and sampled, afterwards, the signal is reconstructed by analog filtering of $c(k)$ with φ_s .

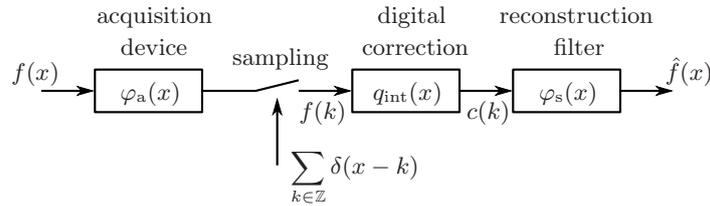


Figure 2.9: Sampling for non-ideal acquisition devices. It is the same block diagram of Figure 2.8, except for the additional digital correction filter q_{int} .

In practice however, the analog prefilter is often dictated by the acquisition device and is thus not necessarily optimal or ideal. Given the measurements after prefiltering and sampling, we want to construct a meaningful approximation of the form (2.19) with a synthesis function φ_s . A meaningful approach is based on *consistent* sampling: the signal approximation should yield exactly the same measurements if it was reinjected into the system. This can be solved by introducing a suitable digital correction filter q_{int} as shown in the block diagram in Figure 2.9. If we revisit the standard reconstruction in equation (2.12), we can reformulate the interpolation process as (where we set $T = 1$ for the simplicity of notation)

$$\hat{f}(x) = \sum_{k=-\infty}^{+\infty} \underbrace{f * q_{\text{int}}(k)}_{c(k)} \varphi_s(x - k). \quad (2.23)$$

The solution can also be presented in the traditional interpolation form:

$$\hat{f}(x) = \sum_{k=-\infty}^{+\infty} f(k) \varphi_{\text{int}}(x - k), \quad (2.24)$$

where φ_{int} is the interpolation kernel denoted as

$$\varphi_{\text{int}}(x) = \sum_{k=-\infty}^{+\infty} q_{\text{int}}(k) \varphi_s(x - k). \quad (2.25)$$

In the next section, we will describe some popular interpolation and synthesis kernels used in image resampling applications.

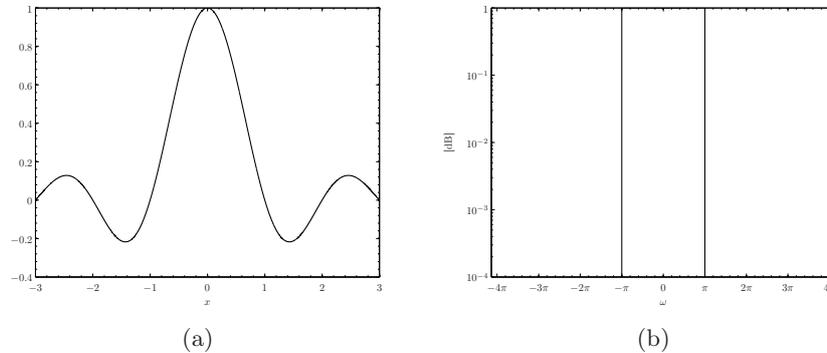


Figure 2.10: Ideal anti-aliasing and reconstruction filter: (a) sinc kernel plotted for $|x| < 3$ and (b) logarithmic plot of the magnitude of its Fourier transform (in dB).

Finally, we want to mention that there is close relationship between the modern sampling theory and the wavelet theory. The analysis tools and mathematics are essentially the same except that wavelets analyze the signal at different scales. A detailed discussion of wavelets is beyond the scope of this work and we refer the interested reader to the standard wavelet literature [Mallat, 1989].

2.2 Linear image interpolation

The general interpolation formula (2.24) describes the discrete-to-analog process as a convolution of the (discrete) image with a continuous interpolation kernel $\varphi_{\text{int}}(\cdot)$, which is a linear operation, hence the term *linear image interpolation*. As illustrated in Figure 2.10(b), the ideal reconstruction kernel is constant and 1 in the passband (denoted by the interval $-\pi < \omega < \pi$ in angular frequencies⁶ and the cutoff point or Nyquist frequency is given by $\omega = \pi$) and 0 in the stopband ($|\omega| > \pi$). As mentioned before, the optimal sinc interpolation kernel for bandlimited signals is rarely used in practice because of its infinite support and its slow decay.

We are then left with the problem of how to choose the interpolation kernels. Usually, symmetrical and separable kernels (see equation (2.14)) are employed to reduce the computational complexity. This also makes the implementation in multiple dimensions straightforward.

As discussed in Section 2.1.6, the interpolation kernels should also satisfy the partition of unity condition (2.21). This means that for any displacement x the direct current (DC)-gain will be unity and the energy of the resampled image, i.e. the squared l_2 -norm, remains unchanged. In other words, the mean

⁶Angular frequencies ω are often used in literature and are related to the frequencies by $\omega = 2\pi\mathbf{u}$.

brightness of the image is not affected if the image is resampled. Therefore, kernel functions that satisfy or fail condition (2.21) are named DC-constant or DC-inconstant, respectively.

The consistent sampling condition guarantees that the image is not modified if it is resampled on the same grid. Therefore, interpolation kernels satisfy

$$\begin{cases} \varphi_{\text{int}}(x) = 1 & \text{if } x = 0, \\ \varphi_{\text{int}}(x) = 0 & \text{if } x \in \mathbb{Z}_0, \end{cases} \quad (2.26)$$

where $\mathbb{Z}_0 = \mathbb{Z}/\{0\}$. In contrast to interpolators, kernels that do not satisfy equation (2.26) are called approximators.

We briefly mention the most important classes of linear interpolation methods, for a more comprehensive overview we refer the reader to some excellent survey articles on linear image interpolation [Lehmann et al., 1999, Meijering et al., 2001] and to an exhaustive historical review on image interpolation [Meijering, 2002].

2.2.1 Sinc-approximating interpolation kernels

Starting from the ideal reconstruction kernel, we can design a family of practical interpolation kernels by multiplying the sinc function with a windowing function $w(\cdot)$ of limited spatial support $[-m, m]$:

$$\varphi_{\text{int,ws}}(x) = w(x)\text{sinc}(x), \quad (2.27)$$

with $w(x) = 0$ ($\forall |x| \geq m$) and $2m \in \mathbb{N}$. If $w(x) = 1$ ($\forall |x| < m$) or equivalently $w(x) = \Pi(mx)$, the interpolation kernel is called $2m$ -points truncated sinc and $2m$ -points windowed sinc, otherwise. In the frequency domain, truncation is equivalent to a convolution of a rectangular function (from the ideal sinc interpolation kernel) with a sinc function (the frequency response of the windowing function). This results in numerous overshoots or ripples in the stopband as illustrated in Figure 2.11. Note that truncated sinc is also DC-inconstant for any choice of $m < +\infty$.

It would be a better idea to use another windowing function that has better properties in the frequency domain. Some examples of windowing functions are given in Table 2.1. More examples can be found in [Meijering et al., 2001]. The Blackman-Harris windowed sinc provides a DC-constant interpolator, while most other windowing functions do not have this superior property. That is why we often employ the 4-term Blackman-Harris windowed sinc interpolation as a reference tool throughout this work. The kernel and the magnitude of its Fourier transform are plotted in Figure 2.12.

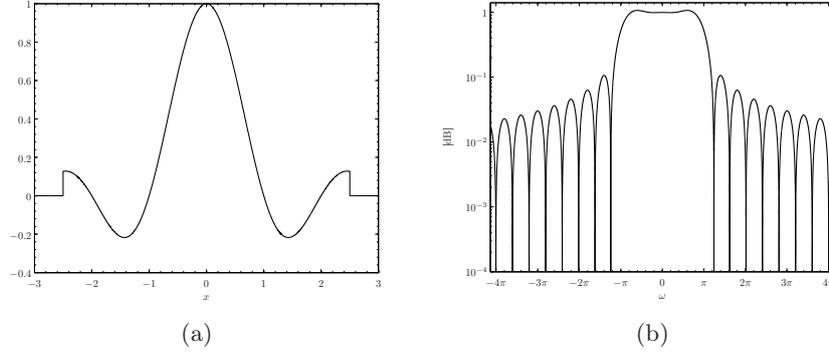


Figure 2.11: Truncated sinc interpolation ($m = 2.5$): (a) truncated sinc kernel and (b) logarithmic plot of the magnitude of its Fourier transform (in dB).

Table 2.1: Several popular windowing functions and their definition. In the definitions of Kaiser and Gaussian windowing functions, $\alpha \in \mathbb{R}^+$ represents a free parameter and I_0 is the zeroth-order modified Bessel function. In the definition of the raised-cosine function, the free parameter β ($0 \leq \beta \leq 1$) represents the *roll-off* factor, which is a measure of the *excess* bandwidth, i.e. the bandwidth occupied beyond the Nyquist bandwidth $1/2T$.

Windowing function	Definition
Bartlett	$w_{\text{Bar}} = 1 - \frac{ x }{m}$
Blackman	$w_{\text{Bla}} = 0.42 + 0.50 \cos\left(\frac{\pi x}{m}\right) + 0.08 \cos\left(\frac{2\pi x}{m}\right)$
Blackman-Harris (3-term)	$w_{\text{BH3}} = 0.42323 + 0.49755 \cos\left(\frac{\pi x}{m}\right) + 0.07922 \cos\left(\frac{2\pi x}{m}\right)$
Blackman-Harris (4-term)	$w_{\text{BH4}} = 0.35875 + 0.48829 \cos\left(\frac{\pi x}{m}\right) + 0.14128 \cos\left(\frac{2\pi x}{m}\right) + 0.01168 \cos\left(\frac{3\pi x}{m}\right)$
Bohman	$w_{\text{Boh}} = \left(1 - \frac{ x }{m}\right) \cos\left(\frac{\pi x}{m}\right) + \frac{1}{\pi} \sin\left(\frac{\pi x }{m}\right)$
Gaussian	$w_{\text{Gau}} = \exp\left(-\frac{1}{2} \left(\alpha \frac{x}{m}\right)^2\right)$
Hamming	$w_{\text{Ham}} = 0.54 + 0.46 \cos\left(\frac{\pi x}{m}\right)$
Hann	$w_{\text{Han}} = 0.5 + 0.5 \cos\left(\frac{\pi x}{m}\right)$
Kaiser	$w_{\text{Kai}} = \frac{I_0(\alpha \sqrt{1 - (x/m)^2})}{I_0(\alpha)}$
Lanzos	$w_{\text{Lan}} = \text{sinc}\left(\frac{\pi x}{m}\right)$
Raised-cosine	$w_{\text{RC}} = \left(1 - \left(\frac{2\beta x}{m}\right)^2\right)^{-1} \cos\left(\frac{\beta \pi x}{m}\right)$
Welch	$w_{\text{Wel}} = 1 - \left(\frac{x}{m}\right)^2$

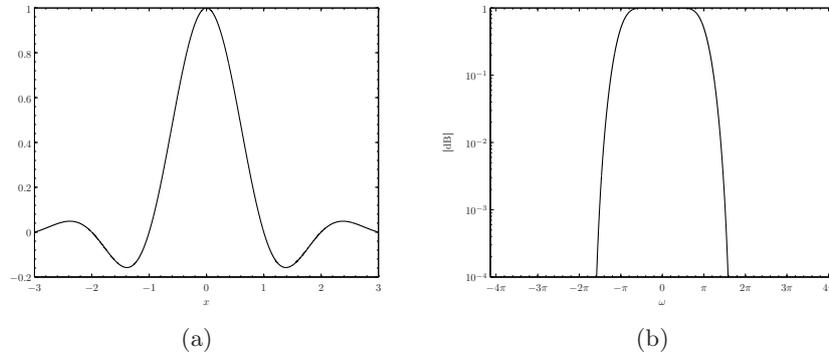


Figure 2.12: The 4-term Blackman-Harris windowed sinc interpolation ($m = 6$): (a) kernel plotted for $|x| < 3$ and (b) logarithmic plot of the magnitude of its Fourier transform (in dB).

2.2.2 Piecewise polynomial-based interpolation kernels

Another broad class of interpolation kernels are composed of piecewise polynomials. They are very popular because of their simplicity. In general, piecewise n th-degree polynomial kernels ($n \geq 1$) can be written in the following form:

$$\varphi_{\text{int,pol}}(x) = \begin{cases} \sum_{i=0}^n a_{ij} |x|^i & \text{if } j - \xi \leq |x| < j + 1 + \xi, \\ 0 & \text{if } |x| \geq m, \end{cases} \quad (2.28)$$

where $\xi = 1/2$ for n even and $\xi = 0$ for n odd and $j = 0, 1, \dots, m + \xi - 1$. The parameter m determines the spatial support $[-m, m]$ of the convolution kernel.

The simplest and computationally cheapest approach to build a piecewise polynomial kernel is using zero- or first-degree polynomials, resulting in the nearest-neighbour (also called pixel replication) and bilinear interpolation kernel,⁷ respectively defined as

$$\varphi_{\text{int,nn}}(x) = \begin{cases} 1 & \text{if } -\frac{1}{2} \leq |x| < \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases} \quad (2.29)$$

$$\varphi_{\text{int,bil}}(x) = \begin{cases} 1 - |x| & \text{if } 0 \leq |x| < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2.30)$$

The kernels and the logarithmic plot of the amplitude of their Fourier transforms are given in Figures 2.13 and 2.14, respectively. Recently, an original contribution to the bilinear interpolation has been introduced in [Blu et al., 2004],

⁷To avoid confusion with the general class of linear image interpolation techniques, we assign the term *bilinear* image interpolation to the piecewise first-degree polynomial kernel.

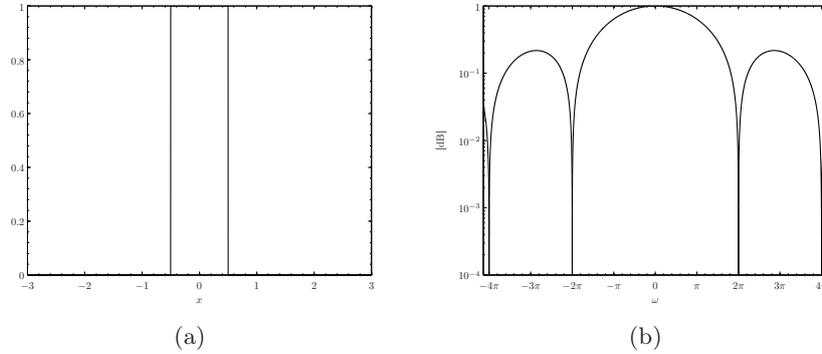


Figure 2.13: Nearest neighbour interpolation: (a) rectangular kernel and (b) logarithmic plot of the magnitude of its Fourier transform (in dB).

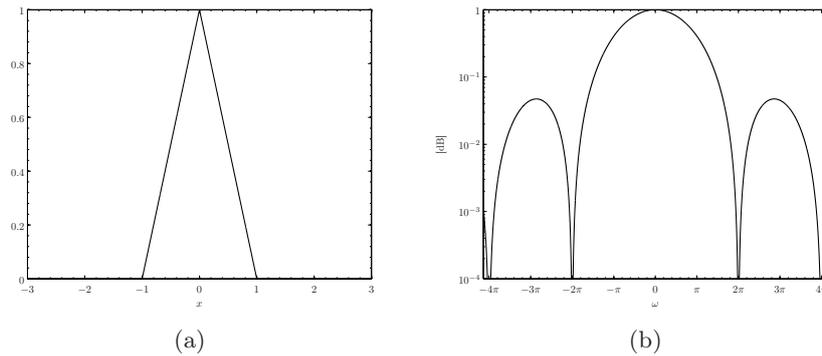


Figure 2.14: Bilinear interpolation: (a) triangular kernel (or hat function) and (b) logarithmic plot of the magnitude of its Fourier transform (in dB).

that significantly improves the interpolation quality with simple kernel shifts. As a convention, kernels with polynomials of degree n are called quadratic ($n = 2$), cubic ($n = 3$), quartic ($n = 4$), quintic ($n = 5$), sextic ($n = 6$), septic ($n = 7$), octic ($n = 8$), nonic ($n = 9$), etc.

Many other popular interpolation kernels fall into this category, just to mention a few well-known ones: Catmull-Rom, Dodgson, Keys, Lagrange, Mitchell, Gauss, etc. [Dodgson, 1997, Lehmann et al., 1999, Meijering et al., 1999]. In fact, every approximating or general synthesis kernel can be rewritten in a standard interpolation form via equation (2.25). For example, this is the case for B-splines $\beta(x)$ (see Section 2.2.3 for more details): the interpolating version is then called the cardinal B-spline $\beta_c(x)$. Note that the support of cardinal B-splines is spatially infinite, which makes the implementation unattractive

for practical applications. On the other hand, the support of B-splines can be very compact. That is why B-splines became very popular in computational efficient image processing tasks.

Besides the partition of unity and consistency conditions, some important kernel design specifications are the *continuity* constraint (the l th-order derivatives ($l < n$) of the kernel at positions $|x| = 0, 1, \dots, m$ should be continuous), the *slope* constraint (the slope at $|x| = 1$ should approximate the slope of the sinc kernel) and the *flatness* constraint (the Fourier spectrum of the kernel is required to be flat at $u = 0$).

2.2.3 Generalized interpolation kernels

In the generalized sampling theory, we start from the interpolation formula given by equation (2.23) where the coefficients $c(k)$ do not necessarily correspond to image pixels. To get more familiar with the generalized interpolation kernels, we discuss the commonly used family of spline functions, namely the B-splines or basis splines.

The simplest and smallest basis function that satisfies the partition of unity is called the zeroth-degree B-spline and is denoted as $\beta^0(x)$. That function happens to be the rectangular function or nearest neighbour interpolation kernel $\beta^0 = \varphi_{\text{int,mn}}$ as given by equation (2.29). Higher degree B-splines can be reconstructed via iterative convolution ($n \geq 1$):

$$\beta^n(x) = \beta^0(x) * \beta^{n-1}(x) = \underbrace{\beta^0(x) * \beta^0(x) * \dots * \beta^0(x)}_{n+1 \text{ times}}. \quad (2.31)$$

Note that the first degree B-spline corresponds to the bilinear interpolation kernel $\beta^1 = \varphi_{\text{int,bil}}$ as defined in equation (2.30). Several B-spline kernels are plotted in Figure 2.15(a). As an alternative approach, a B-spline of degree n can be written in the following explicit form:

$$\beta^n(x) = \frac{1}{n!} \sum_{k=0}^n \binom{n+1}{k} (-1)^k \left(x - k + \frac{n+1}{2}\right)_+^n, \quad (2.32)$$

where the unit step function $(x)_+^n$ is defined as

$$(x)_+^n = \begin{cases} x^n & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (2.33)$$

When $n \rightarrow +\infty$, the B-spline $\beta^\infty(x)$ approaches a Gaussian function, while the cardinal B-spline $\beta_c^\infty(x)$ converges to a sinc function. Note that the Fourier transform of the B-spline $\beta^n(x)$ is $\text{sinc}^{n+1}(u)$. Convolution with the B-spline kernel results then in blur and there is more attenuation of high frequency components for higher values of n .

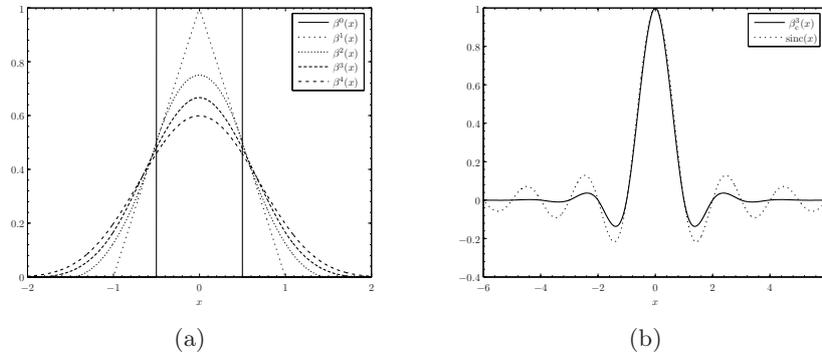


Figure 2.15: B-splines: (a) kernels of degree $0, \dots, 4$ and (b) the cardinal cubic B-spline β_c^3 compared to the sinc kernel.

For $n \geq 2$, B-splines are no longer interpolators, but approximators. To fulfil the consistent sampling condition, a proper digital correction filter q_{int} (shown in Figure 2.9) ensures that the convolution of the coefficients $c(k)$ with the B-spline kernels produces a signal with the same samples as the input samples according to equation (2.23). The B-spline coefficients $c(k)$ are obtained by prefiltering the input image. This can be done either by matrix manipulations to solve a set of equations or by means of highly efficient recursive filtering techniques. For further details regarding the theoretical and implementational aspects of the B-spline interpolation, we refer to [Unser et al., 1993a, Unser et al., 1993b, Unser, 1999].

Besides the popular B-splines, we just want to mention that the shifted linear interpolation [Blu et al., 2004] and the optimal maximal-order-minimal-support (O-MOMS) kernels [Blu et al., 2001] also fall into this generalized interpolation category. O-MOMS functions are built from linear combinations of B-splines and its derivatives, which then have the minimal spatial support for a given interpolation accuracy.

2.2.4 Discussion

In general, linear interpolation techniques are relatively simple to implement and are also very efficient. Therefore, they are very popular and widely used in hardware and software implementations. In Figure 2.17, we show the results of nearest neighbour, bilinear, cubic Catmull-Rom, cubic B-spline, 16-points Blackman-Harris windowed sinc and ideal sinc⁸ interpolation of the image given in Figure 2.16. At first sight, we can clearly notice three kinds of disturbing artefacts in these interpolation results:

⁸This is implemented as zero-padding in the discrete Fourier transformation. As a result, the output dimensions are restricted to even integers.



Figure 2.16: The *lighthouse* image.

- *staircase* artefacts: the staircase patterns at (strong) edges (also called *jagged edges*) are clearly visible in the nearest neighbour, bilinear and cubic Catmull-Rom interpolation. A small amount of this artefact can also be noticed in the cubic B-spline interpolation. These staircase artefacts are caused by the presence of sidelobes and ripples in the stopband $|\omega| > \pi$ (e.g. in Figures 2.13(b) and 2.14(b)). Because of this, frequency components of the replicated spectra are badly suppressed.
- *blur* artefacts: this unsharpness is most prominent near object edges in bilinear and cubic Catmull-Rom interpolation. We can also observe a small amount of blur in nearest neighbour and cubic B-spline interpolation. The Fourier analysis of the kernels in Figures 2.13(b) and 2.14(b) tells us that blur is caused by the slight attenuation of the (high) frequencies in the passband $-\pi < \omega < \pi$.
- *ringing* artefacts: in the uniform areas near edges we can see some *Gibbs* effects (i.e. apparent ghost repetitions or *halos* of the edges) in the Blackman-Harris windowed sinc and ideal sinc interpolation. Interpolation by means of a bandlimited convolution kernel stops the high frequency components abruptly, which causes then these Gibbs phenomena.

As a consequence of the extra lobes and ripples in the stopband, unwanted and false high frequency components are left in the stopband during resampling. This can be avoided using a high-quality two-stage resampling scheme [Seppä, 2007], which consists of the ideal sinc interpolation (i.e. zero-padding of Fourier coefficients) followed by a standard resampling method. Even though the authors do not explain why their results are improved, the key of their success is quite simple: the main lobes and ripples in the stopband coincide with the zero-patterns of the ideal sinc interpolated image. Therefore, the unwanted high frequency components are much attenuated. The major drawbacks of this technique are the extra memory usage and the additional computation time.

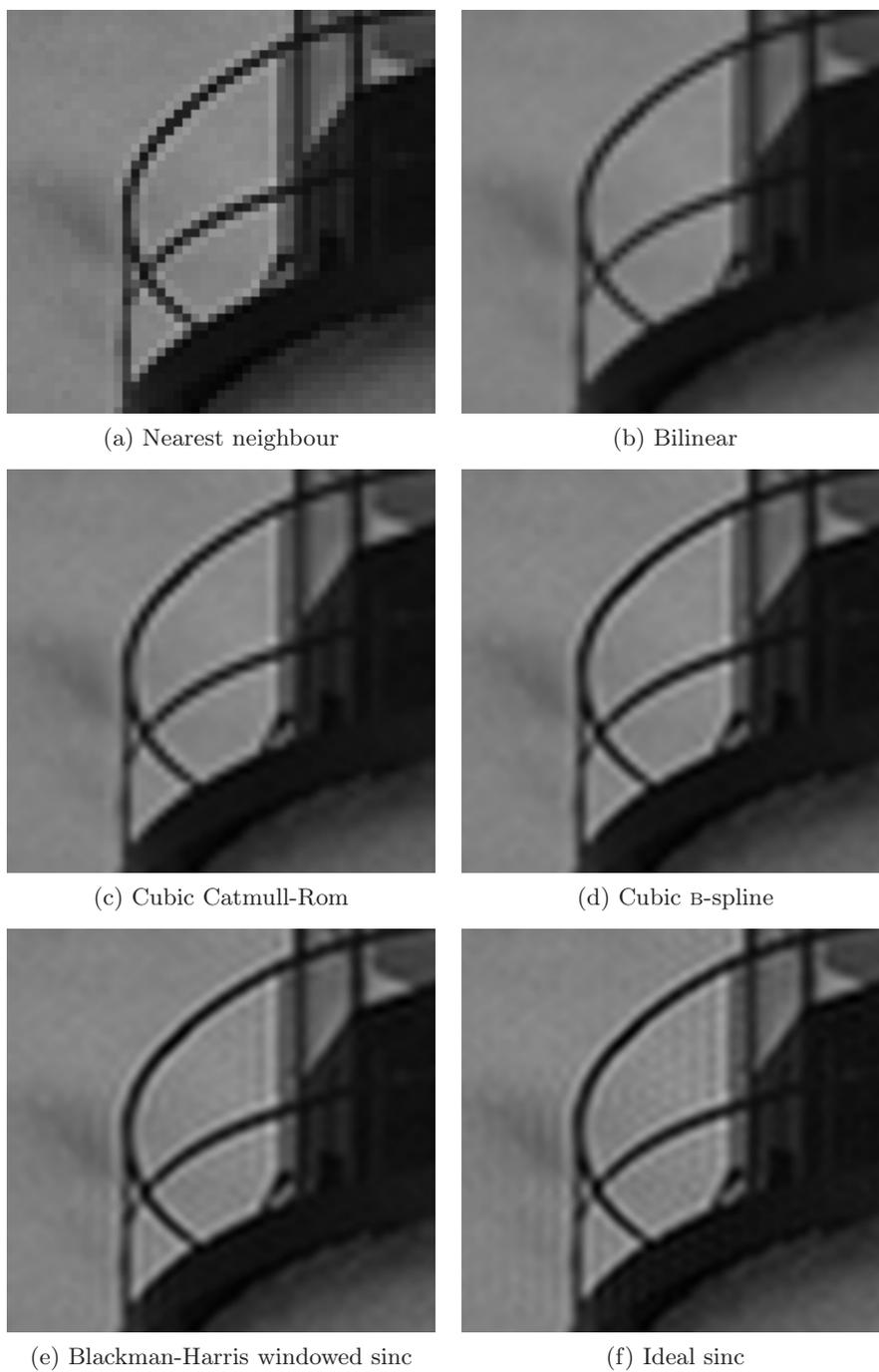


Figure 2.17: Image interpolation results (4× enlargement of Figure 2.16).

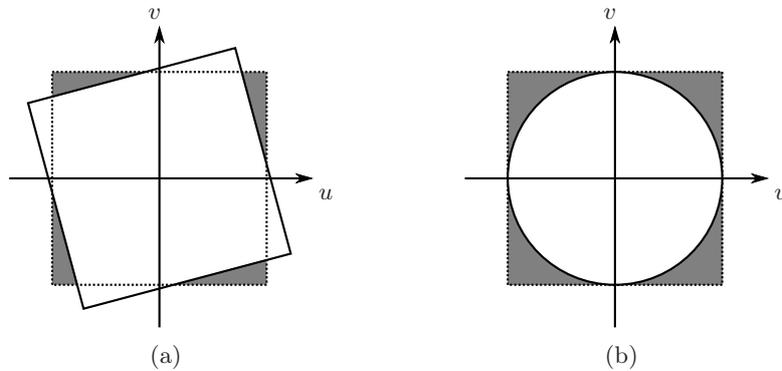


Figure 2.18: The presence of frequency aliasing during image rotation: (a) regions where aliasing occurs is denoted by grey and (b) the alias-free region is given by a circular support in the frequency domain. The 2D frequencies are denoted as (u, v) .

In literature (e.g. in [Lehmann et al., 1999, Seppä, 2007]), a common way to compare the performance of different resampling kernels is to apply a rotation experiment: the image is successively rotated until a full rotation (360°) is reached. To avoid confusion at the image boundaries, only the central part of the fully rotated image is compared to the original image. Because rotation in the spatial domain corresponds to an identical rotation in the Fourier domain and because image rotations are often carried out on the same grid, high frequency components will also get confused (i.e. aliasing at the boundaries of the frequency domain). This aliasing effect is illustrated in Figure 2.18(a). Unfortunately, the presence of frequency aliasing is usually neglected in literature. To avoid this problem, an alternative resampling experiment can be worked out based on subpixel shifts, or we can apply an ideal low-pass filter with a circular support on the image before carrying out the rotation experiment (see Figure 2.18(b)).

To justify our claims, we set up a small rotation experiment: we rotate 5 images 15 times with successive rotations of 24° , afterwards we calculate the PSNR between the central parts of the resampled result and the original image (an example is given in Figure 2.19). For the *filtered* approach, we apply an ideal low-pass filter with circular support on the original image while for the *common* approach, no prefiltering is used. The PSNR results for both methodologies are given in Table 2.2 for the nearest neighbour, bilinear, cubic B-spline and Blackman-Harris windowed sinc interpolation as well as in a two-stage resampling scheme with $2\times$ and $4\times$ sinc prefiltering. The resampled images are shown in Figures 2.20 and 2.21.

Table 2.2: Average PSNR results in dB of a rotation experiment of 5 images, by the common (COM) and filtered (FIL) approach. The best PSNR result is written in bold letters for each resampling method.

Resampling method	Standard way		2-Stage (2×)		2-Stage (4×)	
	COM	FIL	COM	FIL	COM	FIL
Nearest neighbour	19.55	19.92	25.51	26.34	34.64	36.04
Bilinear	22.74	23.12	26.08	26.84	32.81	34.34
Cubic B-spline	28.03	29.31	44.09	49.38	53.16	58.02
Blackman-Harris	30.20	32.58	52.57	55.44	52.42	55.29



Figure 2.19: Boat image. The central region of interest (ROI) is used to compute the PSNR values.

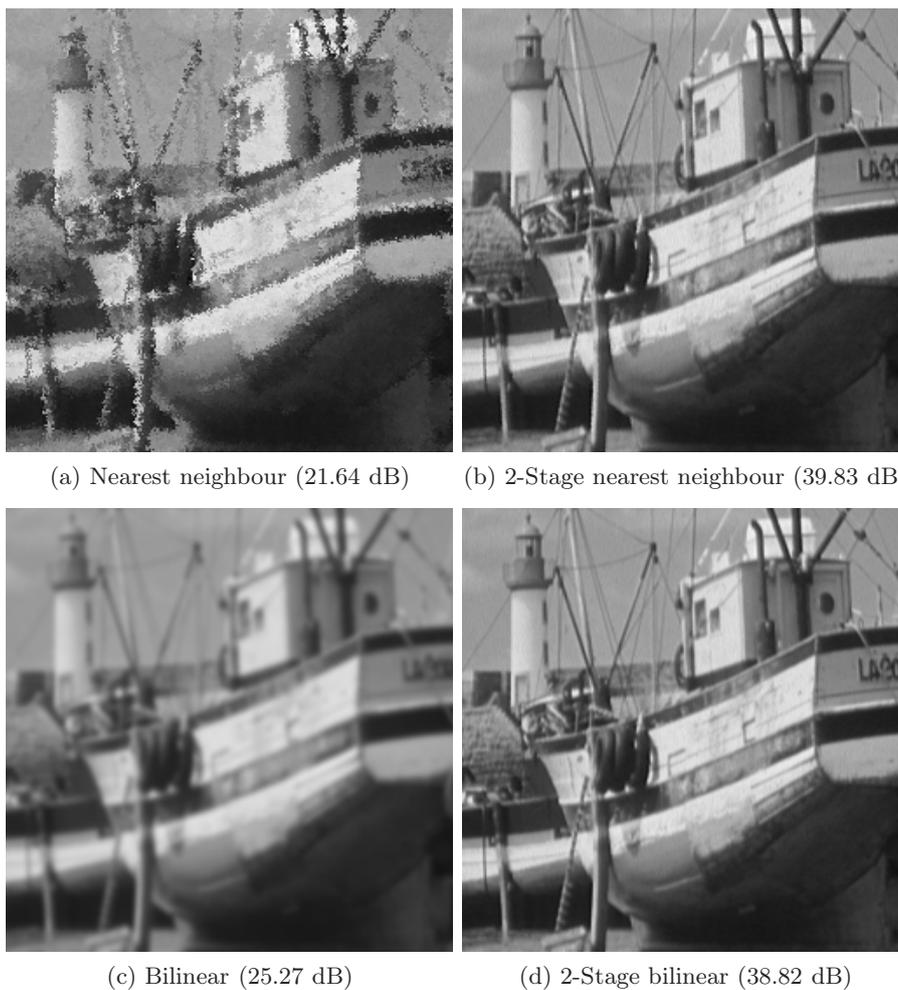


Figure 2.20: The resampling result after several rotations. The two-stage resampling scheme interpolates the image in a first stage $4\times$ in each direction using an ideal sinc.



(a) Cubic B-spline (34.44 dB)



(b) 2-Stage cubic B-spline (69.36 dB)



(c) Blackman-Harris (38.11 dB)



(d) 2-Stage Blackman-Harris (62.14 dB)

Figure 2.21: The resampling result after several rotations (continued).

From the results, we can observe that the filtered approach yields higher PSNR values compared to the common approach. However, the loss is not always consistent: it varies from 0.5 dB up to 5 dB between the common and the filtered implementation. Therefore, wrong conclusions may be drawn based on the common approach.

We also note for example that nearest neighbour performs better in a two-stage resampling scheme (with $4\times$ sinc prefiltering) than bilinear interpolation because deviations in the passband, which result in blur artefacts, become more dominant. This can be clearly observed in the logarithmic magnitude plots of the Fourier transforms of the kernels given in Figures 2.13(b) and 2.14(b).

2.3 Image restoration

Images enlarged with linear interpolation techniques suffer from a number of artefacts and do not aim at restoring the missing high frequency content. In non-ideal circumstances, it is very difficult to reconstruct the original signal. In general, distortion by frequency aliasing is non-recoverable (unless we have multiple subpixel shifted images, see Section 2.4). Additionally, quantization and noise corrupt the measurements of the individual samples. To overcome these problems, we introduce some prior knowledge about images and deal with a more general mathematical acquisition model.

2.3.1 Adaptive interpolation

To restore the missing high frequency content, the unknown Fourier coefficients beyond the Nyquist frequency need to be estimated. Since there are no measurements available in the stopband, various *a priori* assumptions can be made about the image, both in the frequency and spatial domain.

Papoulis assumes that the high-resolution (HR) image is bandlimited (with $B > \pi$) and also that the image has a finite extent. From these conditions, he computes additional high frequency information via an iterative algorithm⁹ in which alternatively, the known set of samples are projected onto the high-resolution grid and an ideal low-pass filter (according to the new bandwidth B) is applied on the HR image [Gerchberg, 1974, Gerchberg, 1989, Papoulis, 1975].

In the spatial domain, other assumptions can be made to avoid some artefacts. A well-known disadvantage of linear interpolation methods is that they assume that the signal is smooth everywhere, which is clearly violated in the real world. Therefore, interpolation should be spatially adaptive according to the image or edge data. For example, an ideal edge can be modelled as a step or Heaviside-like function. However, due to the convolution (2.24) with finite-support symmetrical kernels, also “wrong” pixel values (e.g. pixels across the edge) are contributing to the estimated pixel, which eventually results in

⁹Also known as the Papoulis-Gerchberg algorithm.

blurred edges.

In Chapter 3, we will return to adaptive or non-linear interpolation and discuss some a priori assumptions for image interpolation in more detail.

2.3.2 Acquisition model

Until now, our model assumes that a continuous scene is sampled with Dirac pulses, which results in a discrete image. However, in conventional imaging systems (as in camera-based applications), there is a natural loss of spatial resolution caused by optical distortions (camera lens blur, out of focus, diffraction limit, atmospheric blur, etc.), motion blur due to limited shutter speed, quantization and noise that occur within the sensor (due to the analog circuitry) or during transmission and there is insufficient sensor density [Park et al., 2003]. Thus, the acquired image usually suffers from blur, noise and frequency aliasing effects. Additionally, some postfiltering operations such as lossy compression can degrade the image quality tremendously.

The general acquisition chain for common cameras is illustrated in Figure 2.22. The blur operation is characterized by its (space-varying) point spread function (PSF). In the rest of this work, *decimation* will be denoted as the combination of (non-ideal) anti-aliasing filtering and sampling. The anti-aliasing filter can also be considered as part of the blur kernel. The additive noise is characterized by its *probability density function* (PDF). The compression depends on the codec, typically this will result in block and quantization artefacts (e.g. in JPEG- and MPEG-based compression schemes). We can omit this in case of lossless compression (i.e. the original image is reconstructed perfectly after decompression) or if there is no codec at all. The goal of image restoration is then to reconstruct the hypothetical desired high-resolution (HR) image from a given degraded low-resolution (LR) image.

Although the proposed acquisition model is much more complex than Dirac sampling, we have not taken into account a number of non-degrading linear and non-linear operations, which are built into an off-the-shelf camera, e.g. gamma and colour correction, sharpening (e.g. unsharp masking), contrast enhancement (e.g. histogram stretching), demosaicing, etc. Since these operations are often manufacturer-specific, we will do not discuss them further in this work, although incorporating them would additionally improve the restoration process.

2.3.3 Image restoration as an ill-posed inverse problem

As discussed in the previous section, image restoration is the reconstruction process of the original high-resolution image from an observed degraded low-resolution image. However, image restoration is usually a computationally complex and numerically ill-posed inverse problem [Kirsch, 1996]. We will

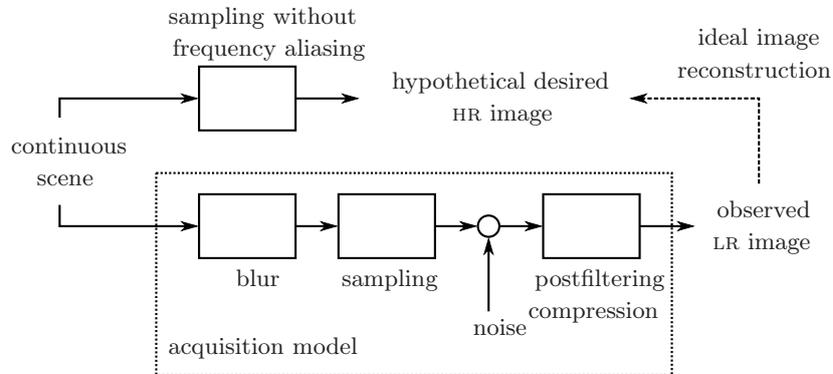


Figure 2.22: Traditional observation model for super-resolution image reconstruction in camera-based applications. The image acquisition process consists of constructing a digital image from a continuous scene.

make this assertion more clear by clarifying the inverse and ill-posed problems.

A *forward* and an *inverse* problem are related to each other if the formulation of one of the problems involves all or part of the solution of the other [Keller, 1976]. The forward or *direct* problem is *well-posed*, while the inverse problem is *ill-posed* in general. Given the model and its parameters, the direct problem is then to find the solution of this operation. In our situation, this problem is transformed into finding the observed LR image given the HR image (i.e. the parameters) and the acquisition model, which is a simple task. On the other hand, the inverse problem is to find the model parameters given the observed data or to find the HR image in the restoration problem. This is obviously a much more complex problem than the direct problem. In a more general context, we can also say that image synthesis in graphics is a direct problem, while image analysis is an inverse problem.

According to Hadamard's work [Hadamard, 1923], a problem is classified as well-posed if the solution to the problem:

1. exists for any data,
2. is unique for any input data and
3. depends continuously on the initial data (i.e. the solution is stable under perturbations on the data).

In contrast, a problem that fails to satisfy any of these conditions is called ill-posed. We will discuss each condition in more detail according to the image restoration problem:

1. The presence of noise during the acquisition may result in an observed image that is inconsistent with any scene. The result is that the system is

noninvertible and the HR image can not be estimated from its observation. For example, we can obtain negative intensity values in the solution to the inverse problem, which cannot really exist of course.

2. Since the number of pixels in the HR image, i.e. the number of unknowns is considerably higher than the the number of pixels in the LR image, it is clear that insufficient constraints exist to ensure a unique solution to the inverse problem. Furthermore, at least some information is lost during the acquisition process, which indicates that the solution to the inverse problem is likely to be non-unique.
3. Depending on the characteristics of the acquisition model, the solution may be highly sensitive to perturbations of the observed data. For example, noise is amplified tremendously by deblurring.

Note that the Hadamard conditions are rarely satisfied in practice, such that image restoration is invariably ill-posed. In practice, the solution to inverse problems are determined using computational methods, which implies that the observed data, the solution and the model must be discretized. The discretization of ill-posed problems leads to ill-conditioned numerical problems [Borman, 2004].

To overcome numerical instability in the computational solution of the inverse problem, regularization methods are applied in order to obtain acceptable solutions. Regularization utilizes additional information (i.e. a priori assumptions) to compensate for the loss in information that characterizes ill-posed problems and turns the problem into a well-posed one [Borman, 2004]. Such a regularization can be carried out in a stochastic or set-theoretic framework. In Chapter 4 et seq, we will discuss this topic in more detail.

2.4 Super-resolution imaging

The classical image restoration problem has been studied for several decades and a wide variety of restoration techniques have been proposed in literature. More recently, with the growing interest in video processing, researchers started to address the problem of multi-frame image restoration. Since the two problems are closely related, much of the work in multi-frame restoration borrows from techniques developed for the single image restoration problem. Nevertheless, new problems have to be tackled when we want to employ information from multiple frames.

Though many definitions of *super-resolution* image reconstruction exist, it is broadly recognized as the bandwidth extrapolation beyond the bandlimit of the imaging system.¹⁰ In contrast to the classical image restoration, super-resolution restoration requires the restoration of lost information, i.e. the *true*

¹⁰Other variants mention the diffraction limit of the optical system. However, in digital imaging, the resolution is limited not by the diffraction limit, but rather by the spatial integration at the sensors.

high frequency content. Therefore, super-resolution is also called the second generation of image restoration techniques.

2.4.1 Multi-frame restoration

The best known and most important class of super-resolution techniques is the one of multi-frame restoration techniques. Given a set of similar but non-identical images, e.g. from a video stream, it becomes possible to use different observation data for resolution improvement.

From the sampling point of view, as discussed earlier in Section 2.1.5, we can reformulate the multi-frame super-resolution problem as a multichannel sampling problem, i.e. uniformly spaced samples of the different images overlaying the high-resolution grid, which results in nonuniform sampling. To convert a set of irregularly placed samples into a regular grid, reconstruction (also called fusion) is required. However, the shifts between the pixels of the different images are usually unknown and need to be estimated first through motion compensation or image registration. Of course, the pixels are usually not measured in ideal circumstances and as a consequence, we have to take the acquisition model into account. Like the classical image restoration problem, super-resolution image reconstruction is also an ill-posed inverse problem. In a nutshell, the three main components of multi-frame super-resolution are registration, fusion and restoration.

Multi-frame SR image reconstruction in camera-based applications is possible if two conditions are fulfilled, namely the LR images must be subpixel shifted and the images must contain frequency aliasing (due to subsampling) [Park et al., 2003]. Indeed, if there are no subpixel shifts (i.e. all pixel shifts are integer), each image contains exactly the same pixel information (besides noise) and therefore, no new information is available to reconstruct the missing high frequencies. If there is no frequency aliasing, then the observed data only contains exact bandlimited information and recovering high frequencies is not possible. Under the stated conditions, restoration techniques using multiple acquisitions are superior to traditional restoration approaches that employ only a single image. A comprehensive discussion on this topic will be given in Chapter 7.

In some applications (e.g. in magnetic resonance imaging), there is no frequency aliasing present due to the image acquisition process. We show in Chapter 8 that recovery of true high frequent information from multiple images is still possible under specific circumstances. In other applications (e.g. in document processing), it is even possible to perform super-resolution using only a single image. This will be discussed in Chapters 5 and 6.

2.4.2 Classification

Super-resolution and/or restoration techniques can be broadly categorized by the number of inputs and the number of outputs. We will list and briefly discuss the possible input-to-output combinations:

- single-frame to single-frame (or *single-input single-output* (SISO)): this corresponds to the classical image restoration setting (Chapter 4). In some applications, we can break through the bandlimit of the imaging system by using only one input image (Chapters 5 and 6). If the input LR image is not corrupted by noise, interpolation techniques can be applied (Section 2.2 and Chapter 3).
- multi-frame to single-frame (or *multi-input single-output* (MISO)): this corresponds to the multi-frame super-resolution setting that produces a still HR image from an image sequence (Chapters 7 and 8).
- multi-frame to multi-frame (or *multi-input multi-output* (MIMO)): this corresponds for example to the video format conversion problem, e.g. converting from standard definition television to high definition television (Chapter 7). Note that it is now possible to remove motion blur (which is a temporal phenomenon with a spatial artefact as an effect [Schechtman et al., 2005]).
- multi-view video to multi-frame: in multi-camera systems it is interesting to combine different video streams to restore temporal aliasing and thus improve the frame rate through super-resolution. In this way, a high-speed camera system can for example be built from several cheap camera's. Another interesting application is to merge differently coded versions of a video stream to obtain a better video quality and resolution, both in spatial and temporal domain.
- multi-view to multi-view video: it is a great challenge to design efficient algorithms for multi-camera systems, for example in real-time applications.
- different single/multi-input to single/multi-frame: resolution information can also be obtained from different image modalities, for example in remote sensing applications (e.g. combining multispectral images with panchromatic images), in medical applications (e.g. combining MRI, CT and PET for medical diagnosis) or in depth cameras (common HR images are combined with LR depth images [Diebel and Thrun, 2005]).
- single/multi-frame to learning-based single/multi-frame: in the training stage, high-resolution image features (for example texture information) are collected, which are then used in the actual reconstruction process as prior information [Freeman et al., 2002, Datsenko and Elad, 2007, Jiji et al., 2007]. Of course, the performance of the algorithm depends heavily on the training data.

2.5 Related problems

It is not our intention to give an exhaustive overview of all existing image restoration problems, but to relate a few well-known image/video processing fields to the general resolution enhancement problem. Of course, the presented problems are not exclusive and can appear jointly in practical applications.

Starting from the general resampling problem, we can find several related interpolation tasks:

- *image inpainting* refers to the fill-in of missing or occluded regions in an image based on information available on the observed regions [Chan et al., 2005, Chan et al., 2007, Tschumperlé and Deriche, 2005]. A common principle for inpainting is to smooth the transitions near the boundaries between the observed and occluded regions in a natural way. The image interpolation problem is in theory a special case of image inpainting: the observed regions are then reduced to a regular “nail bed”. A typical inpainting application can be found in old movie restoration: scratches and blotches are detected and removed in a first step and inpainting techniques are then used for filling in the removed regions [Haindl and Filip, 2002].
- *deinterlacing* is used to convert an interlaced video into a progressive format [Van De Ville, 2001, Yoo and Jeong, 2002]. The inverse process of producing half-resolution fields at double the frame rate is known as *interlacing*. To obtain the full vertical resolution, interpolation is needed at odd or even lines as illustrated in Figure 2.23(a). Bad deinterlacing of moving objects or moving scenes result in tearing artefacts where alternate lines are slightly displaced from each other.
- *demosaicing* is defined as a digital image process used to interpolate a complete image from the partial raw data received from the colour-filtered CCD or CMOS image sensor via a colour filter array (CFA) [Trussell and Hartwig, 2002, Hirakawa and Parks, 2006]. The most commonly used CFA is the Bayer pattern, shown in Figure 2.23(b). The challenge is to restore the missing colours in the presence of aliasing, but in contrast to single image interpolation, correlation between the colour planes can be exploited. The most occurring problems during reconstruction are zipper artefacts and apparent false colours near edge regions.

Note that both deinterlacing and demosaicing problems face the aliasing issue as the general super-resolution problem. Therefore, it is reasonable to address these problems in a unified context: a joint super-resolution and demosaicing/deinterlacing approach (such as in [Farsiu et al., 2006a]) provides the optimal solution for the reconstruction problem.

From the image restoration problem, we can briefly categorize the tasks according to the acquisition model (see Figure 2.22):



Figure 2.23: Grids in deinterlacing and demosaicing problems (D=data, R=red, G=green, B=blue).

- *denoising* is the process of removing noise from a signal. The origin of noise is usually found in the acquisition or transmission process. The objective is to design an algorithm that can remove noise and simultaneously preserve the original image details and fine structures.
- *deconvolution* (or *deblurring*) is the process of removing blur which is denoted as a mathematical convolution with a PSF or blur kernel. In practice, the PSF is unknown and therefore, the blur kernel is estimated prior the deconvolution (via blur identification) or in the other case, the blur kernel and the deblurred image are jointly estimated from the data (also called *blind deconvolution*).
- *deblocking* is applied to blocks in decoded image/video to improve visual quality and prediction performance by smoothing over the block boundaries. Disturbing block artefacts originate from low bit rates and significant quantization in block-based compression schemes (e.g. JPEG or MPEG systems).

Given multiple images or an image sequence acquired from a same scene, we can exploit and combine the pixel information in multi-frame techniques:

- *motion compensating* methods in video processing use motion estimation to predict and consider the right pixel values at the correct positions from previous (and subsequent) frames. Typical applications are multi-frame deinterlacing and video denoising [Zlokolica, 2006].
- *image fusion* is the process of combining relevant information from two or more (aligned) images into a single image. The resulting image will be more informative than any of the input images. This is for example applied in extending the depth of field in microscopy [Tessens et al., 2007] or in extending the range of exposures (also known as high dynamic range (HDR) imaging [Debevec and Malik, 1997]).
- *mosaicing* is the stitching process of multiple images with overlapping fields of view to produce a segmented panorama (i.e. with a larger field of view) [Traka and Tziritas, 2003]. Such panoramic images facilitate automatic annotation [Luong et al., 2006a] or aid in compressibility [Smolić et al., 1999].

2.6 Conclusion

Several problems still exist in the Shannon sampling theory: firstly, it is an idealization because real world signals are not bandlimited (except for $f(x) = 0$), secondly, there is no such thing as an ideal anti-aliasing filter and finally, Shannon's reconstruction formula is rarely used in practice due to the slow decay of the sinc function. Linear interpolation methods are then proposed as a practical solution for the image resampling problem. Unfortunately, artefacts such as staircasing, blur and ringing are inherent to these techniques. Therefore, we explore adaptive interpolation techniques in Chapter 3 to overcome these problems.

In practice, a digital image suffers from several degradations such as blur, noise, compression artefacts, etc. Image restoration techniques are then introduced to take acquisition models into account. However, image restoration is an ill-posed inverse problem, which can be solved through regularization (see Chapter 4 et seq).

In some cases, for example with multiple aliased images, it is possible to restore true resolution content beyond the bandlimit of the imaging system. This is called super-resolution and is discussed in Chapter 7 et seq. Finally, we gave a short overview of the related problems in other image processing applications.

3

Level curve mapping interpolation

Scientia potentia est.
—Sir Francis Bacon

Image interpolation addresses the problem of generating a high-resolution image from its low-resolution version. Linear interpolation assumes that the image is smooth everywhere (as discussed in Chapter 2). As a consequence, these methods fail to capture the fast evolving statistics around edges and introduce annoying artefacts such as blur, staircase and ringing effects. Since linear interpolation techniques can be implemented efficiently, they are generally preferred in image processing hardware and software tools such as Adobe Photoshop and GIMP.

Adaptive or *non-linear* image interpolation methods improve the subjective quality of the interpolated images by imposing more accurate models and by incorporating prior knowledge. These techniques are able to avoid some or all of the previously mentioned artefacts (blur, staircase and ringing effects), although new kinds of artefacts might be introduced.

In this chapter, we give an overview of the existing state-of-the-art non-linear interpolation techniques in Section 3.1 and in Section 3.2, we describe a novel interpolation technique, which is based on linear image interpolation followed by eliminating unwanted artefacts. The proposed algorithm sharpens edges by mapping the smooth image level curves using adaptive contrast enhancement techniques. The main novelty of this algorithm is the combination of two existing image processing techniques (i.e. isophote smoothing and contrast enhancement), each having a different purpose, in a non-trivial way on top of the linear interpolation techniques. Additional constraints are put on these algorithms to achieve strict interpolation and to preserve the topology of the image level curves. This results in sharper high-resolution images with less disturbing artefacts compared to the state-of-the-art non-linear interpolation techniques as discussed in Section 3.3. This work led to an ICIP publication [Luong et al., 2005]



Figure 3.1: Illustration of the two main concepts in edge-directed interpolation: (a) preventing interpolation across edges avoids mixing pixels of both sides of the edge and (b) performing interpolation along the edges follows the edge orientation.

and another conference publication [Luong and Philips, 2005]. Finally, Section 3.4 concludes this chapter.

3.1 Non-linear image interpolation

We can broadly divide the state-of-the-art non-linear image interpolation methods into three classes: edge-directed, example-based and restoration-based interpolation. We will discuss these classes in more detail and illustrate some new artefacts that can be created by these adaptive techniques.

3.1.1 Edge-directed interpolation

The main motivation of edge-directed interpolation stems from the fundamental property of an ideal step edge (known as *geometric regularity*), i.e., that the image intensity field evolves more slowly along the edge orientation rather than across the edge orientation. Geometric regularity has important effects on the visual quality of an interpolated image such as the sharpness of edges and the freedom from artefacts. Most existing adaptive interpolation techniques are edge-directed.

From the assumption of geometric regularity, two different edge-directed interpolation strategies are commonly used: the first strategy avoids interpolating *across* edges, while the second strategy performs interpolation *along* the edges. The difference between these two concepts is illustrated in Figure 3.1.

Typically in the first strategy, an edge is detected with subpixel accuracy and the linear interpolation is modified to prevent interpolation across this edge [Jensen and Anastassiou, 1995, Allebach and Wong, 1996]. As a result, the adaptively interpolated image looks much sharper than results produced by linear interpolation. Alternatively, the edges can be roughly approximated through *data dependent triangulation* (DDT), which is achieved by diagonals that divide the squared grid into triangles [Su and Willis, 2004].

In the second strategy, the orientation of the edge is locally estimated, rather than the (exact) location of the edge. For example, some methods use the local covariance characteristics at low resolution to estimate the missing high-resolution pixels, e.g. Li's *New Edge-Directed Interpolation* (NEDI) [Li and Orchard, 2001] and its improved versions [Zhao et al., 2002, Zhao and De Haan, 2003]. Adaptively Quadratic (Aqua) image interpolation determines the local quadratic signal class and then applies optimal recovery to estimate the missing samples [Muresan and Parks, 2001, Muresan and Parks, 2004]. Wang and Ward proposed an orientation-guided interpolation method, in which image contours are divided into edges and ridges, and interpolate in the direction based on the estimated orientation of edges and ridges [Wang and Ward, 2007].

The edge-directed property of covariance-based adaptive methods gives the capability to tune the interpolation coefficients to match an arbitrarily oriented edge. These methods demonstrate significant improvements on the visual quality over linear interpolation, but the computational cost is also much higher than the linear methods. To reduce the computation time, directional interpolation is applied in edge areas while linear interpolation is applied in uniform areas, in which both methods produce similar (visual) results. Alternatively, some algorithms only investigate a limited number of edge directions (e.g., horizontal, vertical or diagonal) and interpolate locally along the best edge directions [Battiato et al., 2002, Muresan, 2005]. Obviously, quantizing the edge directions affects the accuracy of the imposed edge model.

Besides edge-directed interpolation techniques, other edge-based image interpolation methods operate in multi-resolution frameworks such as wavelets [Kinebuchi et al., 2001] and Laplacian pyramids [Jeon et al., 2006]. The algorithms based on the pyramid structure offer the possibility of preserving sharp edges by exploiting relationships between coefficients at different scales. This model is used to interpolate images by predicting coefficients at finer scales.

3.1.2 Example-based interpolation

Since high-frequent components beyond the Nyquist limit are not present in low-resolution images, the missing high-resolution detail needs to be estimated. In general, these details cannot be retrieved via simple image sharpening. One would expect that the richness of real-world images is difficult to capture analytically. This motivates a learning-based approach: in a training set, the algorithm learns the fine details that correspond to different image regions seen at a low-resolution scale and then uses these learned relationships to predict fine details in other images [Freeman et al., 2002]. In [Muresan and Parks, 2004], a training set can be used for determining the local quadratic signal class of the low-resolution patches and edge-directed interpolation is then applied according to this classification.

These methods produce sharper images, but the performance depends heavily

on the training data. The main drawback of these methods is the potential for introducing artefacts when the lookup table provides wrong high-resolution patches. In addition, these image models may not work as well for data containing artefacts, such as in block-based compression, as the methods tend to consider artefacts as part of the image and may enhance the artefacts more than the underlying image.

Stepin introduced a very fast magnification filter for cartoon images that maps low-resolution 3×3 blocks directly into predefined high-resolution blocks, which are stored in a lookup table [Ledda, 2006]. To reduce the number of entries in the lookup table, the 8 neighbours of the central pixel are classified into “close” and “distant” pixels (according to the difference in the YUV colour space between the central pixel and its neighbours). As a result, only 256 different combinations need to be stored.

Fractal-based interpolation methods suppose that many things in nature possess self-similarity, i.e. scale invariance [Honda et al., 1999, Price and Hayes III, 1998]. This means that parts of the image repeat themselves on an ever-diminishing scale, hence the term *self-similarity*. This *resolution-independent* relationship across various scales is described by the fractal Brownian function (FBM) as introduced by Mandelbrot [Mandelbrot, 1982]. This self-similarity property is exploited for fractal image compression and interpolation by mapping the similar parts at different scales (via a contractive affine transformation) [Jacquin, 1992]. Due to the recursive application of these mappings at the decoder stage, the notion of *iterated function systems* (IFS) is introduced [Polidori and Dugelay, 1995]. In this framework, the “examples” are not retrieved from an external training set, but from scaled versions of the same image. Note that there are connections with edge-based multi-resolution interpolation methods [Kinebuchi et al., 2001, Jeon et al., 2006], although the prediction of the high-resolution information is done in a different way [Gharavi-Alkhansari et al., 1997].

3.1.3 Restoration-based interpolation

Another image interpolation strategy is to avoid artefacts that are created in the interpolated image, such as blur, staircase and ringing effects. Restoration-based interpolation techniques start from a rough initialization of the high-resolution image, typically achieved by linear interpolation, and try to minimize the artefacts that are left by putting constraints on the HR image.

The most popular approach is based on partial differential equations (PDE). PDE's describe the evolution of curves, surfaces or vector fields and are also often applied in other image processing tasks such as image denoising, sharpening, segmentation, etc. [Gilboa et al., 2002, Tschumperlé and Deriche, 2005, Tschumperlé, 2006].

Commonly, these algorithms are implemented as non-linear diffusion processes that are controlled by the local gradient [Perona and Malik, 1990]. For interpolation applications, the PDE scheme is constrained. Specifically, the regularization is only applied to the missing image regions on the HR grid and thus leaves the original pixels unaltered as in image inpainting [Tschumperlé, 2006]. In [Morse and Schwartzwald, 1998, Morse and Schwartzwald, 2001], an iterative constrained level set reconstruction (isophote smoothing) is successfully applied for restoring jagged edges. The removal of the so-called *jaggies* is achieved by minimizing the local curvature. An effective numerical implementation of a PDE scheme with both smoothness and orientation constraints for edges is presented in [Jiang and Moloney, 2002].

Another approach to incorporate prior knowledge or to impose constraints on the HR image is via *projection-onto-convex-sets* (POCS). POCS schemes formulate interpolation as an ill-posed inverse problem and solve it by regularized alternating projections. One of the constraints projects the HR image onto the set of known samples. Examples of other constraints are smoothness and positivity (e.g. a pixel must be in the range of $[0, 255]$). The POCS approach in the image interpolation context are discussed in [Gerchberg, 1974, Papoulis, 1975, Ferreira, 1994, Ratakonda and Ahuja, 1998].

Recently, we have collaborated to the development of a new interpolation algorithm, which removes jagged edges using mathematical morphology techniques. Originally designed for enlarging binary images, the algorithm initializes the HR image with a nearest neighbour interpolation. Various oriented structuring elements are used in an iterative way to detect and validate the extra “corners” coming from jagged edges. These validated corners are removed by a pixel swapping algorithm (i.e. black pixels become white and vice versa) [Ledda et al., 2005, Ledda et al., 2006a]. As a result, staircase effects are removed from the interpolated images. In a later stadium, the algorithm was extended to treat greyscale images [Ledda et al., 2006b] and coloured images with sharp edges in a fuzzy framework [De Witte et al., 2006]. In [Ledda et al., 2008], a fast non-iterative implementation for binary images is realized without sacrificing image quality. For a more detailed discussion and implementation details, we refer the interested reader to [Ledda, 2006, De Witte, 2007].

3.1.4 Discussion

Adaptive image interpolation techniques often require a lot of computation time, which is a severe limitation on their practical use in real-time applications. Linear interpolation methods can be employed in uniform areas or simplified concepts can be introduced to reduce the computational complexity (e.g., quantization of the edge orientations). Some adaptive techniques are restricted to magnification factors that are powers of two (e.g., NEDI interpolation proposed by Li and Orchard [Li and Orchard, 2001]), which is a

strong limitation towards many applications.

In Figure 3.3, we compare several non-linear interpolation techniques to linear cubic B-spline interpolation. The original image is shown in Figure 3.2. The non-linear methods used in this visual experiment are data-dependent triangulation (DDT) [Su and Willis, 2004], Aqua2 interpolation [Muresan, 2005], NEDI interpolation [Li and Orchard, 2001], iterated function systems (IFS) [Gharavi-Alkhansari et al., 1997] and isophote smoothing [Morse and Schwartzwald, 2001].

These non-linear interpolation methods usually result in jaggies-free or sharper interpolated images but most of them introduce new types of disturbing artefacts. In Aqua2 and NEDI interpolation, we notice some brushlike stripes and sweeps, which we denote as painting effects. In IFS and NEDI interpolation, isolated or grouped pixels appear randomly in smooth areas as a result of numerical instability or wrongly learned patterns. Slanted jagged edges become visible if edge orientations are quantized (e.g., in Aqua2 and DDT interpolation). In isophote smoothing and DDT interpolation, the images suffer from “cartooning” effects, i.e. the results look segmented and artificial due to the hard thresholds that prevent interpolation across edges. Also, almost all adaptive methods yield important visual degradation in finely textured areas such as the grass in our example. All these artefacts become more visible at large image magnifications (i.e. magnification factors of 4 and more) and hence they become more annoying.

Since most of these artefacts are unpredictable or very hard to be removed, we concentrate in the next section on tackling artefacts that arise from linear interpolation methods such as blur, staircase and ringing artefacts that are predictable and removable. Another benefit is that all existing hardware and software components that perform linear interpolation, do not have to be thrown away.

3.2 Proposed level curve mapping interpolation

For many applications, the main emphasis on image interpolation should be on the perceptual quality of images, i.e. the interpolated images should be artefact-free and visually pleasing. With this goal in mind, the application of level set methods for restoring jagged edges has been introduced successfully by Morse and Schwartzwald [Morse and Schwartzwald, 1998, Morse and Schwartzwald, 2001]. However, the interpolated images still suffer from blur and ringing artefacts. We tackle these artefacts by assigning new intensity values to the *level curves*, which we denote as *level curve mapping*. Level curves or *isophotes* are defined as spatial curves with a constant intensity level. These curves can be compared to isobars on a weather chart. The manipulation of these intensity levels can be seen as an adaptive contrast enhancement problem. In particular for the interpolation problem, we define



Figure 3.2: Part of the *lighthouse* image.

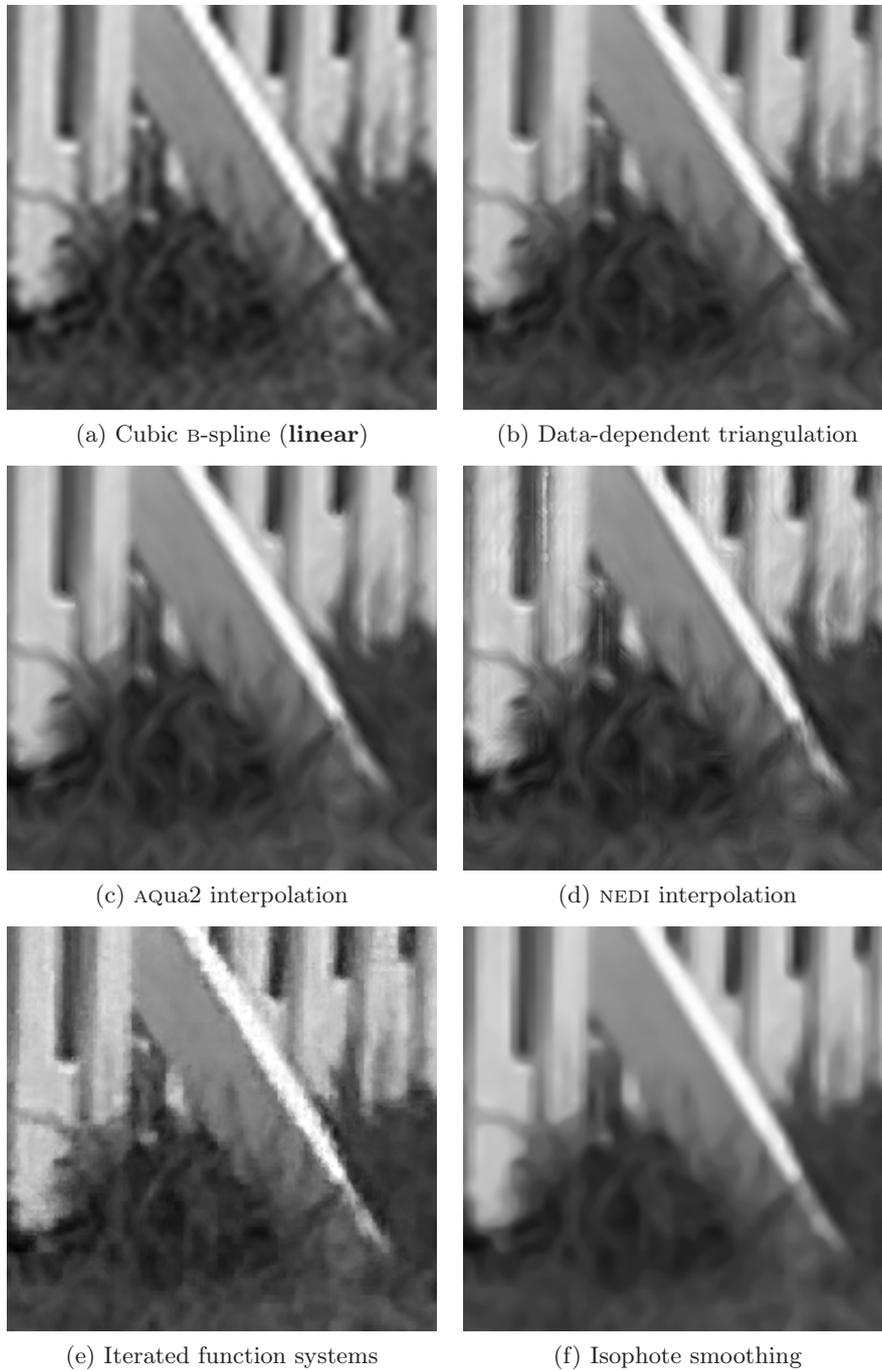


Figure 3.3: Image interpolation results (4× enlargement of Figure 3.2).

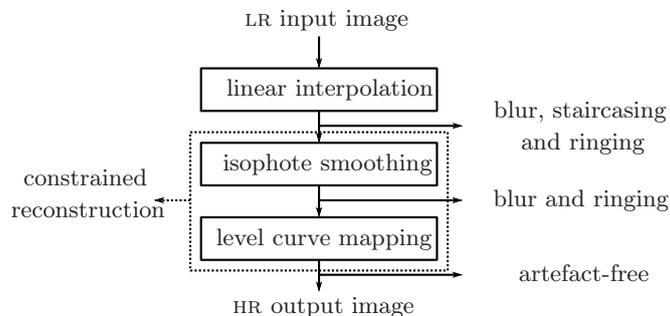


Figure 3.4: Schematic representation of the proposed interpolation method.

some additional constraints on the contrast enhancement. This approach will result in an interpolation method that produces less artefacts. The outline of our interpolation scheme is shown in Figure 3.4.

In the following sections, we first discuss the isophote smoothing scheme. We then describe the constrained level curve mapping algorithm and finally, we discuss how to apply interpolation to colour images.

3.2.1 Constrained isophote smoothing

Before applying the level curve mapping, we need a good initial estimation of the spatial positions of the level curves, because the shape of the level set contours does not change during the adaptive contrast enhancement. The initial estimation is done by a linear interpolation method (see Chapter 2). Since manipulating the intensity levels of these initial curves amplifies the annoying staircase artefacts, we remove these jagged edges by applying isophote smoothing [Morse and Schwartzwald, 1998, Morse and Schwartzwald, 2001].

At first sight, spatial correction of the level curves requires explicitly finding and fitting each level curve, much as individual edges must be found and fitted in edge-directed approaches. However, level curve corrections do not require explicitly finding and fitting each curve. Instead, correcting a level curve passing through each pixel is done directly by manipulating the pixel intensity level $f(\mathbf{x})$ [Morse and Schwartzwald, 1998, Sethian, 1999], which is described by *partial differential equations* (PDE's) [Sethian, 1999].

Isophote smoothing is a level curve correction method that removes jagged edges by minimizing the local curvatures of each level curve. In addition, we want to preserve the level set topology (i.e. the order of the level curves can not be altered), the input pixels of the low-resolution image (denoted as the *interpolation constraint*, otherwise we have no strict interpolation, but an approximation (see Chapter 2)) and convex contours such as corners of the low-resolution image (denoted as the *inflection constraint*). Imposing these three

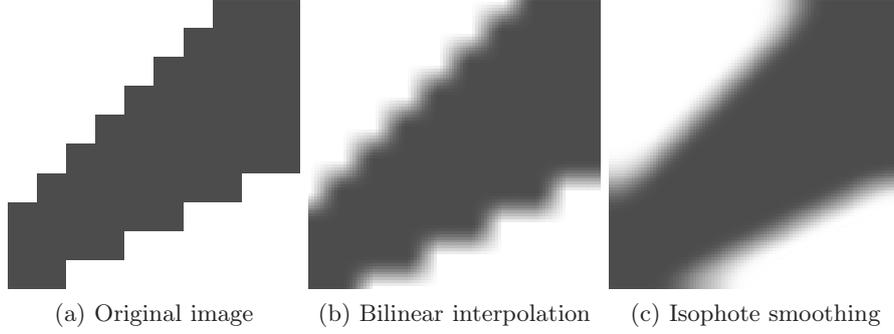


Figure 3.5: Illustration of applying isophote smoothing after bilinear interpolation.

constraints results in a PDE, which is given by [Morse and Schwartzwald, 2001]:

$$f_t(\mathbf{x}) = -\omega(\mathbf{x})\tau(\mathbf{x})\Psi(\kappa(\mathbf{x}))\|\nabla f(\mathbf{x})\|_2, \quad (3.1)$$

where ω , τ and Ψ are the interpolation, inflection and topology constraint functions respectively and the norm of the gradient is given by $\|\nabla f(\mathbf{x})\|_2 = \sqrt{f_x(\mathbf{x})^2 + f_y(\mathbf{x})^2}$ with f_x and f_y the derivatives of the image with respect to x and y respectively.¹ The evolution of the level curve intensity in time is denoted by $f_t(\mathbf{x})$.

The topology constraint is a function of the spatial curvature κ of the level curves, which is calculated from the local derivatives of the intensity levels:

$$\begin{aligned} \kappa(\mathbf{x}) &= [\operatorname{div}(\nabla f / \|\nabla f\|_2)](\mathbf{x}) \\ &= \frac{f_x(\mathbf{x})^2 f_{yy}(\mathbf{x}) - 2f_x(\mathbf{x})f_y(\mathbf{x})f_{xy}(\mathbf{x}) + f_y(\mathbf{x})^2 f_{xx}(\mathbf{x})}{(f_x(\mathbf{x})^2 + f_y(\mathbf{x})^2)^{3/2}}. \end{aligned} \quad (3.2)$$

We can interpret equation (3.1) as the movement of the level curves in the direction of their normal at a speed proportional to their curvature κ . Therefore, $\Psi(\kappa)$ must have the same sign as κ and $\omega, \tau \geq 0$. As a result of applying this PDE, the local curvatures of the level curves are minimized and jagged edges become straight as illustrated in Figure 3.5. We now discuss each constraint function in more detail.

Interpolation constraint: we preserve the input pixels (or also called *local image anchors*) by defining the constraint function ω inversely proportional to the distance of the current pixel position $\mathbf{x} = (x, y)$ and the nearest local image anchor:

$$\omega(\mathbf{x}) = 1 - \sqrt{(1 - \Delta x)^2 + (1 - \Delta y)^2}, \quad (3.3)$$

¹To compact the formulas throughout our work, we often use the shorthand notation for the partial derivatives: $\frac{\partial f}{\partial x} = f_x$, $\frac{\partial^2 f}{\partial x \partial x} = f_{xx}$, $\frac{\partial^2 f}{\partial x \partial y} = f_{xy}$, etc.

where Δx and Δy are the absolute horizontal and vertical offsets between the coordinates of the current pixel and the nearest image anchor. We can also interpret this constraint as the fidelity to the original image.

Explicit topology constraint: we must take additional measures to preserve the level set topology, especially near the local image anchors, because they do not move due to the interpolation constraint. Therefore, we define the constraint function Ψ in a local neighbourhood Ω as:

$$\begin{aligned} \forall \mathbf{x}' \in \Omega : & \quad (3.4) \\ \Psi(\kappa(\mathbf{x})) = & \begin{cases} \text{sign}(\kappa(\mathbf{x})) \min(|\kappa(\mathbf{x})|, |\kappa(\mathbf{x}')|) & \text{if there is a conflict,} \\ \kappa(\mathbf{x}) & \text{otherwise.} \end{cases} \end{aligned}$$

A conflict in the level set topology occurs at \mathbf{x} if one of the following conditions is fulfilled:

$$\begin{aligned} \exists \mathbf{x}' \in \Omega : & \\ \begin{cases} f(\mathbf{x}) < f(\mathbf{x}') \\ f(\mathbf{x}) - \omega(\mathbf{x})\kappa(\mathbf{x}) \|\nabla f(\mathbf{x})\| \geq f(\mathbf{x}') - \omega(\mathbf{x}')\kappa(\mathbf{x}') \|\nabla f(\mathbf{x}')\|, \end{cases} & \\ \text{or} & \\ \exists \mathbf{x}' \in \Omega : & \quad (3.5) \\ \begin{cases} f(\mathbf{x}) > f(\mathbf{x}') \\ f(\mathbf{x}) - \omega(\mathbf{x})\kappa(\mathbf{x}) \|\nabla f(\mathbf{x})\| \leq f(\mathbf{x}') - \omega(\mathbf{x}')\kappa(\mathbf{x}') \|\nabla f(\mathbf{x}')\|. \end{cases} & \end{aligned}$$

In other words, the speed of the curve evolution is controlled such that the order of intensity values is respected: the effect of Ψ is to ensure that all greater/smaller neighbouring intensity values stay larger/lower and thus preserve the level set topology. The size of the squared neighbourhood Ω depends on the magnification factor: in our implementation, Ω contains at least one image anchor.

Inflection constraint: we preserve the convex contours (with a high curvature) such as corners in the input image. This can be done by defining the following inflection constraint:

$$\exists \mathbf{x}' \in \Omega : \tau(\mathbf{x}) = \begin{cases} 1, & \text{sign}(f_t(\mathbf{x})) \text{sign}(f_t(\mathbf{x}') < 0, \\ 0, & \text{otherwise.} \end{cases} \quad (3.6)$$

This constraint requires that at least one pixel in the neighbourhood Ω changes in the opposite direction, which is the case with jagged edges but not with real corners. This way, convex contours can not shrink. A similar constraint could be defined for the global image: the total number of changing pixels or the amount of change in one direction must be the same as in the opposite direction. In that way, the average intensity of the image does not change after isophote smoothing.

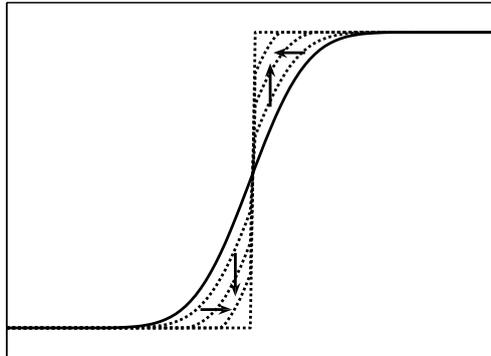


Figure 3.6: 1D evolution of deblurring a smooth edge (solid line) with a shock filter (horizontal arrows) or with level curve mapping (vertical arrows).

In our implementation, 5 to 10 iterations of equation (3.1) give already reasonable results. Alternatively, numerical approximations of level-set-based curve evolution can be employed in this stage to speed up the algorithm [Jiang and Moloney, 2002, Shi and Karl, 2008].

3.2.2 Image sharpening

We first describe the relationship between image sharpening via level set corrections and image sharpening via adaptive contrast enhancement. Afterwards, we discuss the constrained level curve mapping algorithm in more detail.

3.2.2.1 Image sharpening via shock filters

Sharpening blurred edges can be achieved by moving the different level curves closer to each other, because a higher density of level curves stands for a sharper edge. The corresponding PDE that produces sharper edges in the image $f(\mathbf{x})$ is given by:

$$f_t(\mathbf{x}) = -\Lambda(\nabla^2 f(\mathbf{x})) \|\nabla f(\mathbf{x})\|, \quad (3.7)$$

where $\Lambda(0) = 0$ and $\Lambda(\nabla^2 f(\mathbf{x})) \text{sign}(\nabla^2 f(\mathbf{x})) \geq 0$. The sign-operator is defined as $\text{sign}(x) = x/|x|$ ($x \neq 0$) where alternative definitions assign -1 , 0 or $+1$ to $\text{sign}(0)$. This hyperbolic equation is also known as the *shock filter* proposed by Osher and Rudin [Osher and Rudin, 1990]. The level curves move in the direction of their normal and the speed of the movement is proportional to the Laplacian $\nabla^2 f(\mathbf{x})$ (sometimes denoted as $\Delta f(\mathbf{x})$). Figure 3.6 illustrates the evolution an originally smooth edge when the shock filter is applied to it.

The main drawback of using shock filters is that a complex function Λ is needed to preserve the level set topology and the intensities of the original image pixel simultaneously, i.e. when we require strict interpolation. Another drawback is that equation (3.7) amplifies the annoying ringing overshoots and undershoots. Additionally, the number of iterations must also be specified, which is not always evident.

Alternatively, we can rephrase image sharpening as an adaptive contrast enhancement problem, which manipulates the level curve intensities through a transfer function. An important advantage of this approach is that the shape of the level curves (and thus the shape of the contours of the object) is automatically preserved and also no explicit edge detection is needed. The level set topology and interpolation constraints are much easier to define in this framework. Another advantage of using contrast enhancement techniques instead of iteratively applying a PDE is that only a single pass is needed to process the image, which saves a lot of computation time and thus can be useful for optimization in parallel systems. An interesting remark can be noticed in Figure 3.6: the edge evolves vertically (in intensity) instead of horizontally (spatially), but the end result has the same deblurring effect when being processed with the previously described shock filter.

3.2.2.2 Adaptive contrast enhancement

A new intensity value $f_n(\mathbf{x})$ is assigned to each level curve according to an adaptive transfer function (a similar technique for contrast enhancement is used by Yu and Bajaj [Yu and Bajaj, 2004]). The transfer function is characterized by its range $[f_{\min}(\mathbf{x}), f_{\max}(\mathbf{x})]$, where $f_{\min}(\mathbf{x})$ and $f_{\max}(\mathbf{x})$ are the minimum and the maximum intensities of the four surrounding local image anchors respectively.

The basic idea of most contrast enhancement techniques is to take advantage of range stretching [Yu and Bajaj, 2004, Stark, 2000], however, in our case, the input range is the same as the output range. As illustrated in Figure 3.6, the contrast enhancement consists of a convex transfer function that pulls the pixel intensities upwards and of another concave transfer function that pushes the pixel intensities downwards.

If the intensity of the input pixel is higher than the average of $f_{\min}(\mathbf{x})$ and $f_{\max}(\mathbf{x})$, we compute the new pixel intensity $f_n(\mathbf{x})$ by the convex transfer function, which is given by:

$$f_n(\mathbf{x}) = f(\mathbf{x})^{1-\alpha(\mathbf{x})}, \quad (3.8)$$

and otherwise, we employ the concave transfer function:

$$f_n(\mathbf{x}) = 1 - (1 - f(\mathbf{x}))^{1-\alpha(\mathbf{x})}, \quad (3.9)$$

where $\alpha(\mathbf{x})$, $f(\mathbf{x})$ and $f_n(\mathbf{x})$ are normalized here between 0 and 1 in our implementation (afterwards the intensities will be stretched out over the range).

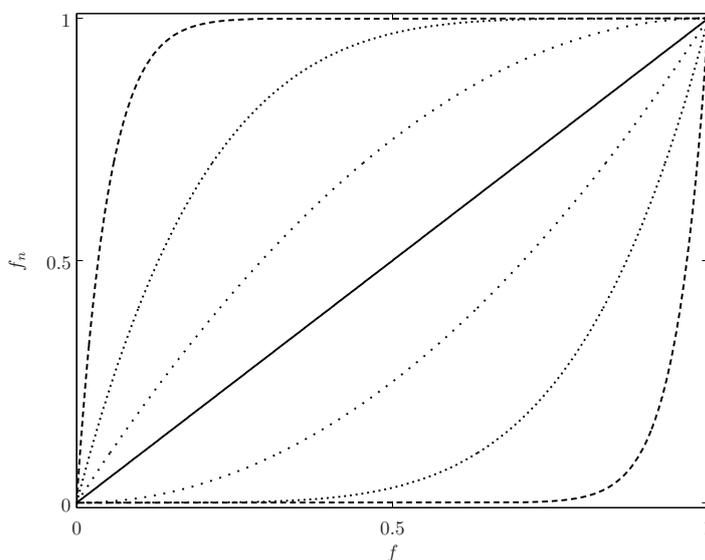


Figure 3.7: The convex (upper) and concave (lower) parts of the transfer function are plotted with the following values for α : 0 (solid line), 0.5, 0.8 and 0.95. For $\alpha = 0$ (flat area) the transfer function is equal to the identity function and no changes are made to the intensity. For $\alpha = 1$, the ideal step edge is reconstructed.

We choose a gamma correction-like function because of the close convergence to a step edge. Note that other transfer functions can be implemented, e.g., a parabolic function is used in [Yu and Bajaj, 2004]. The normalized range parameter α implicitly incorporates edge information and is locally defined as

$$\alpha(\mathbf{x}) = \frac{f_{\max}(\mathbf{x}) - f_{\min}(\mathbf{x})}{255}. \quad (3.10)$$

If α is equal to 0 (e.g. in a uniform region), the transfer functions will leave the intensities untouched. If α is equal to 1 (i.e., a black pixel and a white pixel are neighbouring pixels), the transfer function will reconstruct an ideal step edge. That is why the exponents in equations (3.8) and (3.9) are chosen inversely to the range. The transfer functions are plotted in Figure 3.7 for different parameter values. Note that the level set topology is locally preserved because the transfer functions are monotonically increasing or decreasing.

Unfortunately, there are still three problems left when applying the contrast enhancement technique as described by equations (3.8) and (3.9):

- the computed min/max images, given by $f_{\min}(\mathbf{x})$ and $f_{\max}(\mathbf{x})$ respectively, are constant per $r \times r$ block, with r being the magnification factor

and as a consequence, the resolution enhancement results in block-like artefacts,

- it is not guaranteed that the local image anchors remain unchanged (except for the ones that are the local minimum or maximum by construction of the transfer functions),
- disturbing ringing artefacts are amplified by the contrast enhancement.

We solve these problems as follows:

Continuity of neighbouring range parameters: because the min/max images are constant per $r \times r$ block, the range parameter α will also be constant per $r \times r$ block. At the block boundaries, neighbouring transfer functions can map similar intensity values on very different intensities. This results in discontinuities, and local averaging is thus needed.

A possible solution is to convolute these min/max images with a low-pass filter [Yu and Bajaj, 2004], however, it is not recommended because firstly, if $f_{\min}(\mathbf{x})/f_{\max}(\mathbf{x})$ is decreased/increased by filtering, the range is expanded and the increase range parameter can amplify ringing artefacts, which typically occur in relatively uniform regions with a small range. And secondly, if $f_{\min}(\mathbf{x})/f_{\max}(\mathbf{x})$ is increased/decreased by filtering, the range is reduced such that more pixel intensities could lie outside this range.

Hence, instead of low-pass filtering, we use a normalized voting scheme to compute the new pixel intensity:

$$f_c(\mathbf{x}) = \frac{\sum_{i=1}^9 w_i f_{n,i}(\mathbf{x})}{\sum_{i=1}^9 w_i}, \quad (3.11)$$

where $f_{n,i}(\mathbf{x})$ is calculated by equation (3.8) or (3.9) with different range parameters α of the neighbouring $r \times r$ blocks. If $f(\mathbf{x})$ falls into the range, $f_{n,i}(\mathbf{x})$ will contribute to $f_c(\mathbf{x})$ and $w_i = 1$. Otherwise, $f_{n,i}(\mathbf{x})$ does not have to be computed and $w_i = 0$. With this voting scheme, pixel intensities will not be altered by ranges that do not include $f(\mathbf{x})$.

Local image anchor preservation: to guarantee strict interpolation, local image anchors may not be changed. It is not sufficient to undo the intensity changes at these pixel locations because it can cause conflicts with the level set topology. We can solve this efficiently by attracting the interpolated surface towards the local image anchors. The new intensity value is computed as a weighted average between $f_c(\mathbf{x})$ from equation (3.11) and the closest image anchor $f(\mathbf{x}_a)$ (\mathbf{x}_a is the position of the closest vector):

$$f_a(\mathbf{x}) = \beta f_c(\mathbf{x}) + (1 - \beta)f(\mathbf{x}_a), \quad (3.12)$$

where the weight β ($0 \leq \beta \leq 1$) is denoted by the (normalized) squared distance to the closest local image anchor:

$$\beta = \frac{1}{\sqrt{2}} \|\mathbf{x} - \mathbf{x}_a\|_2. \quad (3.13)$$

In this way, all local image anchors are preserved.

Elimination of ringing effects: most ringing overshoots and undershoots can easily be detected because these pixel intensity values are outside the range. We can remove these ringing effects by simply setting $f_a(\mathbf{x})$ that are outside the range, to $f_{\min}(\mathbf{x})$ or $f_{\max}(\mathbf{x})$:

$$f_l(\mathbf{x}) = \max(\min(f_a(\mathbf{x}), f_{\max}(\mathbf{x})), f_{\min}(\mathbf{x})), \quad (3.14)$$

which is also the final pixel intensity value obtained by the proposed level curve mapping algorithm.

3.2.3 Colour image interpolation

There are three common ways to deal with the colourimetric aspect in image interpolation:

- apply the algorithm on each colour plane (e.g., red, green and blue channel) separately. Since there is no link between the different spectral channels, components of an edge pixel can flow in different directions due to different local ranges and mapping functions. This can lead to disturbing colouring artefacts in edge areas.
- convert the image into a colour space with separate luminance and chrominance channels and apply the *advanced* algorithm on the luminance channel only, while a low-quality, low computational-cost interpolation technique can be applied in the chrominance channels, where high-frequency errors can be tolerated. It is well known that the human visual system is much less sensitive to high-frequency changes of the chrominance signals compared to luminance errors [Vrhel, 2005]. A similar strategy is applied in broadcasting systems and in image and video compression schemes, e.g., in JPEG and MPEG. Examples of colour spaces that separate the luminance component from the chrominance signals are YUV, YIQ, L*a*b* (CIELAB), L*u*v* (CIELUV), etc.
- consider images as surfaces in the hybrid 5D space (x, y, r, g, b) , i.e. each colour pixel is treated as a feature vector in a 3D colour space [Kimmel et al., 2000]. Since we use the minimum and maximum operators in the level curve mapping framework, we have to order two or more colours in an unambiguous way. This problem is well studied in colour morphology applications, therefore, we refer the interested reader to [De Witte, 2007].

In the next section, we evaluate the level curve mapping algorithm on the colour channels separately (in the RGB colour space) and on the luminance channel only (in the YUV and L*a*b* colour spaces).

3.3 Experimental results

In the first experiment, we compare the proposed method to linear interpolation techniques and several adaptive interpolation methods (given in Figures 3.3 and 3.9), such as data-dependent triangulation (DDT) [Su and Willis, 2004], Aqua2 interpolation [Muresan, 2005], NEDI interpolation [Li and Orchard, 2001], iterated function systems (IFS) [Gharavi-Alkhansari et al., 1997] and mathematical morphology interpolation for greyscale images (MMint) [Ledda et al., 2006b]. In this visual experiment, we apply the level curve mapping framework on bilinear and Catmull-Rom interpolation (both prefiltered with constrained isophote smoothing (5 iterations) in order to remove jagged edges) and 32-point 4-term Blackman-Harris windowed sinc interpolation (without isophote smoothing). The results of our interpolation method are shown in Figures 3.8 and 3.10.

In Figure 3.8, we notice that our method preserves finely textured areas (e.g. the grass) much better than other non-linear techniques in Figure 3.3. In Figure 3.9, we see clearly painting artefacts in Aqua2 and NEDI interpolation, jagged edges in DDT and MMint interpolation and rough edge structures in IFS interpolation. We can also see that the proposed method outperforms the linear interpolation methods. In all cases, significant improvements in visual quality can be noticed: the edges have become much sharper and ringing undershoots and overshoots have been heavily reduced. It is important to notice that almost no new artefacts are introduced with the level curve mapping interpolation.

In the second experiment, we have evaluated the level curve mapping algorithm in an objective manner in combination with different interpolation techniques: bilinear and Catmull-Rom interpolation (both prefiltered with constrained isophote smoothing (5 iterations) in order to remove jagged edges); cubic B-spline and 32-point 4-term Blackman-Harris windowed sinc interpolation (both without isophote smoothing). To treat colour images, we have evaluated our algorithm in 3 colour spaces: the level curve mapping is applied separately on the red, green and blue channels (RGB) and furthermore on the luminance channel only (YUV and L*a*b*). The PSNR is measured on the red, green and blue channels. Our test images are the 24 photographic images obtained from the Kodak data set (see Figure 3.11).

A very important aspect is the way the original images are decimated (i.e., the combination of anti-aliasing prefiltering and subsampling). In this experiment, we compare four different decimation algorithms (with an 1 : 2 ratio in each

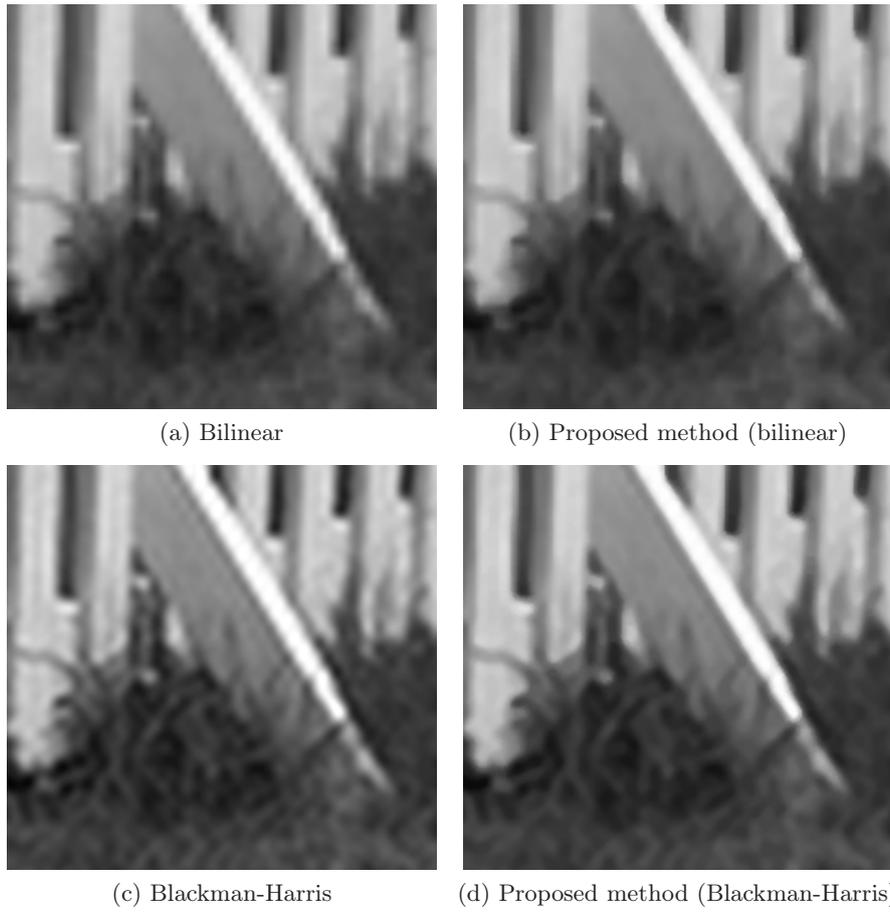


Figure 3.8: Image interpolation results (4× enlargement of Figure 3.2).

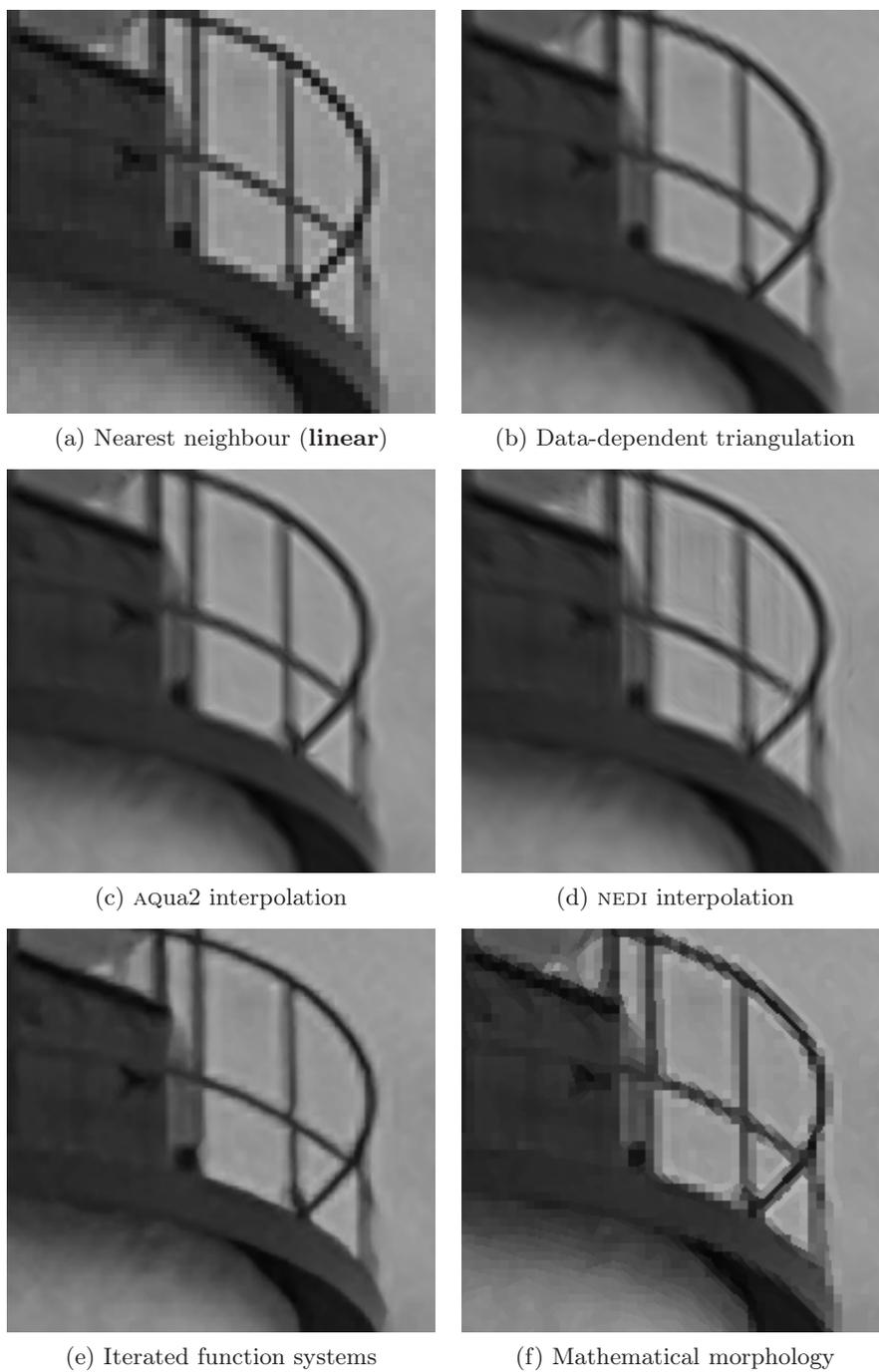


Figure 3.9: Image interpolation results (4× enlargement of Figure 2.16).

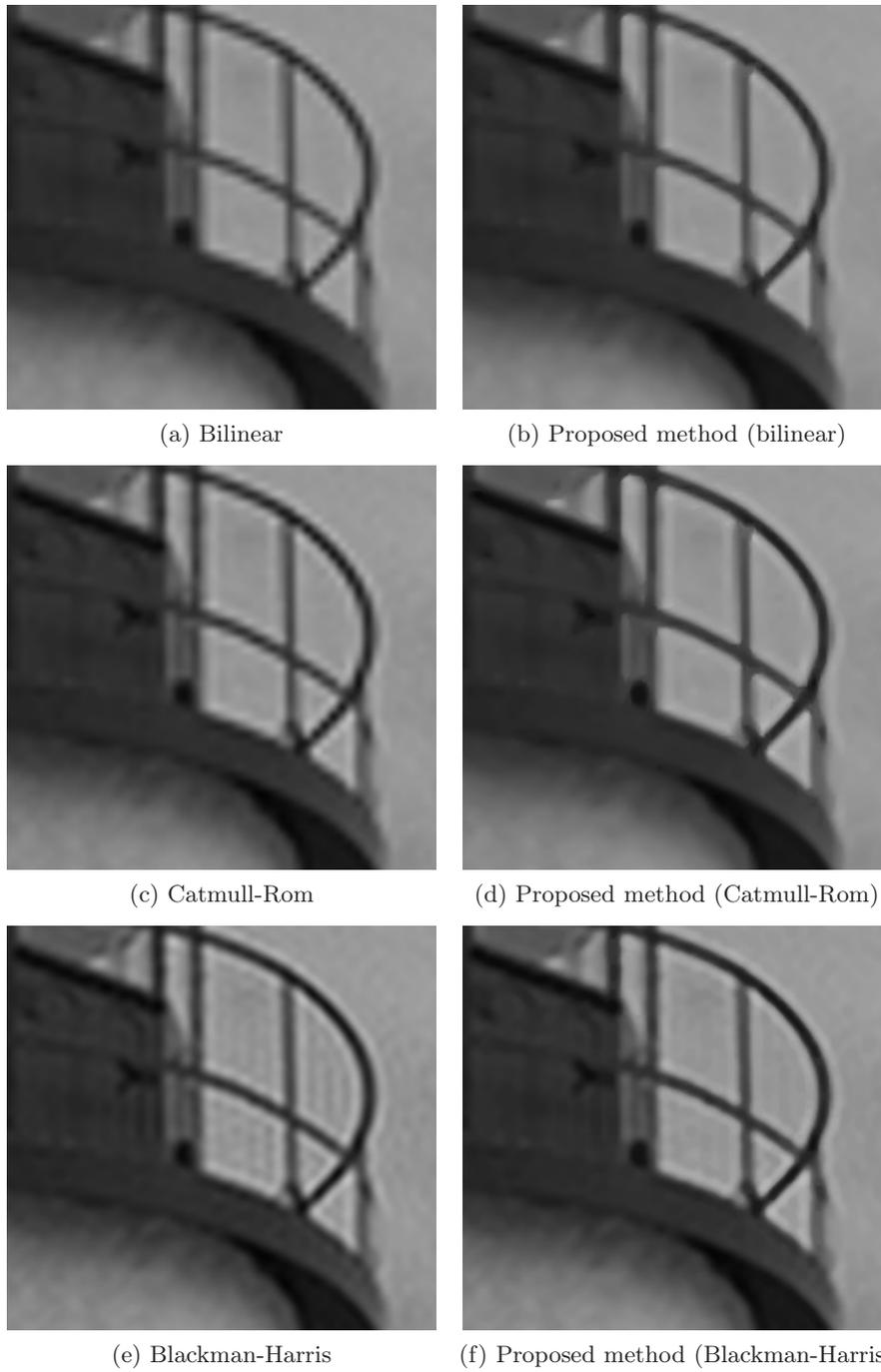


Figure 3.10: Image interpolation results (4× enlargement of Figure 2.16).



Figure 3.11: The 24 original test images from the Kodak set (768×512 pixels).

direction):

- D1: the low-resolution image is constructed by the samples lying on odd lines and odd columns. In other words, there is no anti-aliasing pre-filter and as a consequence, the low-resolution images suffer from aliasing artefacts.
- D2: prior to the subsampling according to scheme D1, the images are convoluted with a Gaussian kernel, which is a non-ideal anti-aliasing pre-filter: frequency components of the replicated spectra are insufficiently suppressed and additional blur is caused by the (needless) attenuation of the frequencies in the passband (see discussion in Section 2.2.4).
- D3: a very fast and popular decimation scheme is simple averaging: each low-resolution pixel value is the average of four high-resolution pixel intensities. Like decimation algorithm D2, the low-resolution images suffer from blur and aliasing effects.
- D4: the low-resolution images are constructed through bandwidth truncation in the Fourier domain. This is exactly the same as resampling using a sinc function (see Chapter 2). In this way, there is no aliasing in the low-resolution images.

In practice, decimation algorithms D2 and D3 come close to the simplified acquisition model of conventional camera-based applications (more details can be found in Section 2.3.2), while decimation scheme D4 approximates for example the acquisition process in magnetic resonance imaging (MRI) (a more elaborated discussion can be found in Chapter 8). The output of the different decimation schemes is illustrated in Figure 3.12.

To achieve reliable and correct numerical PSNR values, we have to align the original and the interpolated images with subpixel accuracy. That is necessary because during downsampling and upsampling, the interpolated image can be

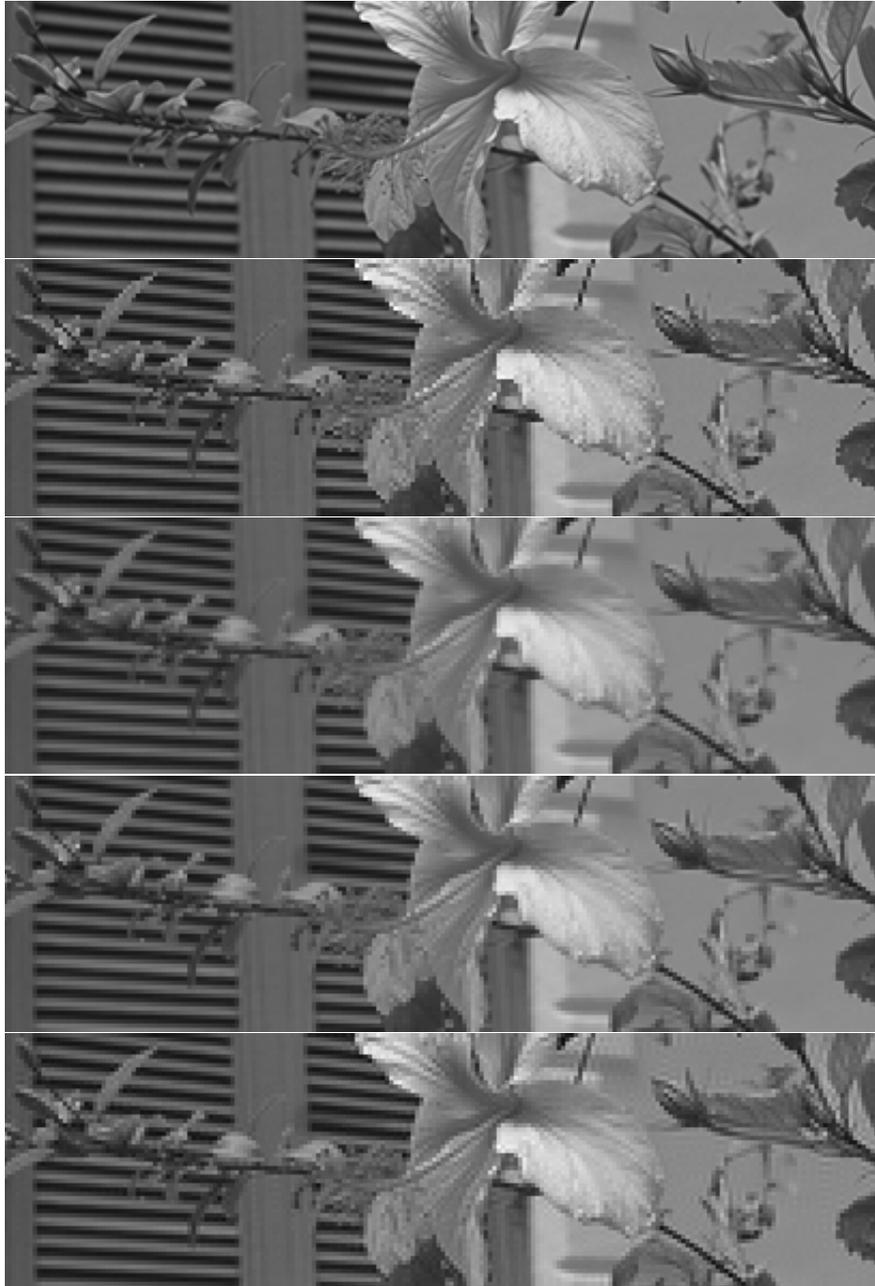


Figure 3.12: Illustration of the different decimation schemes, from top-to-bottom: original high-resolution image and low-resolution images obtained by decimation scheme D1, D2, D3 and D4 respectively.



Figure 3.13: Examples of lower resolution grids (the samples are indicated by black circles): (a) an *off-center* grid and (b) an *on-center* grid. The samples of the high-resolution grid are indicated by white circles.

shifted compared to the original high-resolution image. This is illustrated with an *on-center* and an *off-center* example in Figure 3.13. The alignment is done with 32-point 4-term Blackman-Harris windowed sinc resampling.

The PSNR values (averaged over 24 images of the Kodak image set) are shown in Table 3.1. In almost all cases, the level curve mapping postprocessing yields higher PSNR values compared to the linear interpolation techniques (up to 0.5 dB). The numerical gain is not significant by itself (compared to the visual improvement) because the proposed framework mainly affects edge pixels, which account for a small fraction of pixels in the image. Compared to other non-linear interpolation techniques, our method clearly yields higher PSNR values (up to 1.5 dB) except for the low-resolution images that are constructed with the (non-realistic) decimation scheme D1. In that case, the results of the proposed methods are similar to those of the NEDI and AQua2 interpolation.

If we compare the interpolated images obtained from the different decimation schemes, we notice that the images after decimation D1 are reconstructed worse because of the frequency aliasing present in the LR images (see Figure 3.12). The interpolation algorithms fail in reconstructing textured areas, which penalizes the PSNR values. Interpolation results obtained after decimation schemes D2 (severe blur) and D3 (moderate blur) are worse those after scheme D4 because of blur: a smaller part of the frequency spectrum can be reconstructed well. That is why decimation scheme D4 gives the best interpolation results in terms of PSNR despite the small ringing artefacts in the LR images.

There is no significant visual difference noticeable between the several colour versions (RGB, YUV and $L^*a^*b^*$) of our algorithm. However, PSNR values show that the $L^*a^*b^*$ -version produces slightly worse numerical results compared to the other two versions.

Finally in Figure 3.17, we show the RGB colour interpolation results ($4\times$ enlargement) applied on the *parrot* image compared to the nearest neighbour (Figure 3.14), bilinear (Figure 3.15) and NEDI interpolation (Figure 3.16). All these methods are also applied in the RGB colour space. We can clearly see

Table 3.1: Average PSNR results in dB of a $2\times$ interpolation experiment of the Kodak data set. The best PSNR result is written in bold letters for each decimation scheme.

Interpolation method	D1	D2	D3	D4
Nearest neighbour	27.87	28.62	28.71	28.09
Data-dependent triangulation	29.09	28.55	29.18	29.23
Aqua2 interpolation	29.18	28.84	29.48	29.26
NEDI interpolation	29.35	28.96	29.60	29.25
Iterated function systems	27.61	29.70	29.57	28.85
Bilinear	28.95	28.29	28.86	29.07
+ level curve mapping (RGB)	29.21	28.60	29.27	29.43
+ level curve mapping (YUV)	29.21	28.59	29.27	29.42
+ level curve mapping (L*a*b*)	29.12	28.42	29.04	29.23
Cubic Catmull-Rom	29.25	29.18	29.79	29.84
+ level curve mapping (RGB)	29.17	29.38	30.01	29.95
+ level curve mapping (YUV)	29.18	29.38	30.01	29.94
+ level curve mapping (L*a*b*)	29.30	29.25	29.89	29.89
Cubic B-spline	28.99	29.44	29.96	30.26
+ level curve mapping (RGB)	29.03	29.61	30.19	30.34
+ level curve mapping (YUV)	29.04	29.61	30.20	30.34
+ level curve mapping (L*a*b*)	29.15	29.49	30.07	30.30
Blackman-Harris windowed sinc	28.31	29.58	29.88	30.64
+ level curve mapping (RGB)	28.65	29.73	30.15	30.68
+ level curve mapping (YUV)	28.65	29.72	30.16	30.68
+ level curve mapping (L*a*b*)	28.75	29.61	30.05	30.65



Figure 3.14: The nearest neighbour interpolation.

that our method produces the sharpest image with almost no visible artefacts.

3.4 Conclusion

In this chapter, we gave an overview of the existing adaptive image interpolation techniques. The non-linear interpolation methods can be broadly categorized into three classes: edge-directed, example-based and restoration-based interpolation. These adaptive methods usually result in jaggies-free or sharper interpolated images but most of them introduce new disturbing artefacts such as painting effects, random pixels, slanted jagged edges, cartooning effects, and they yield important visual degradation in finely textured areas.

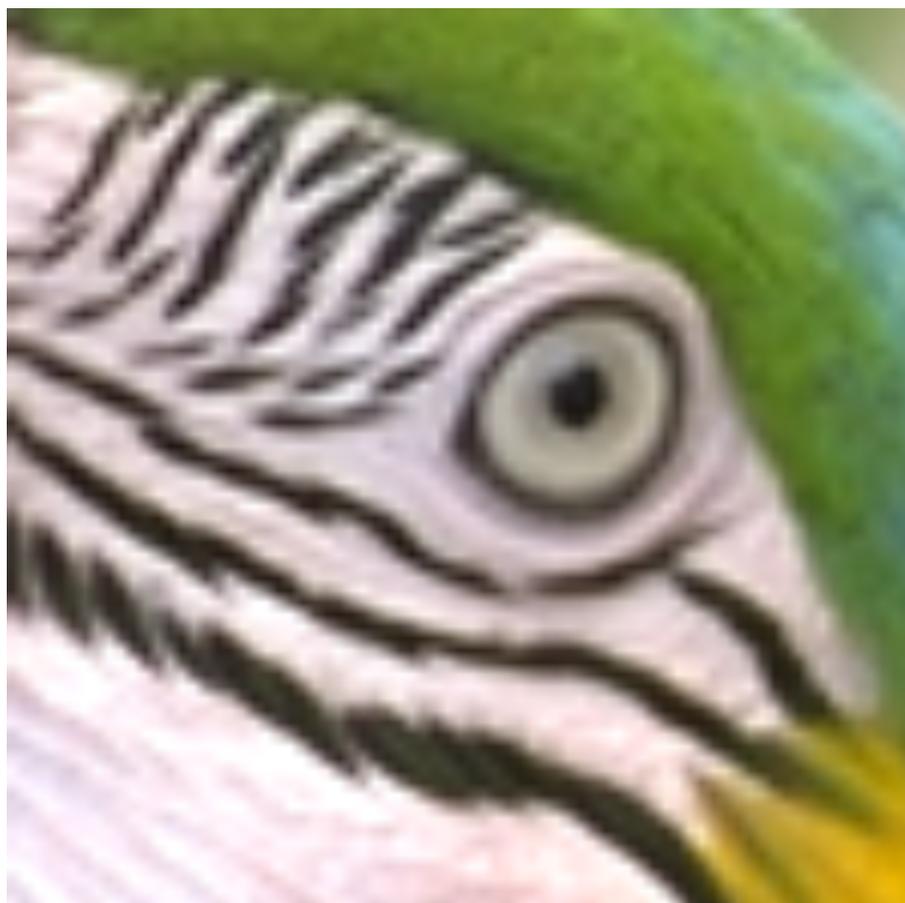


Figure 3.15: The bilinear interpolation.

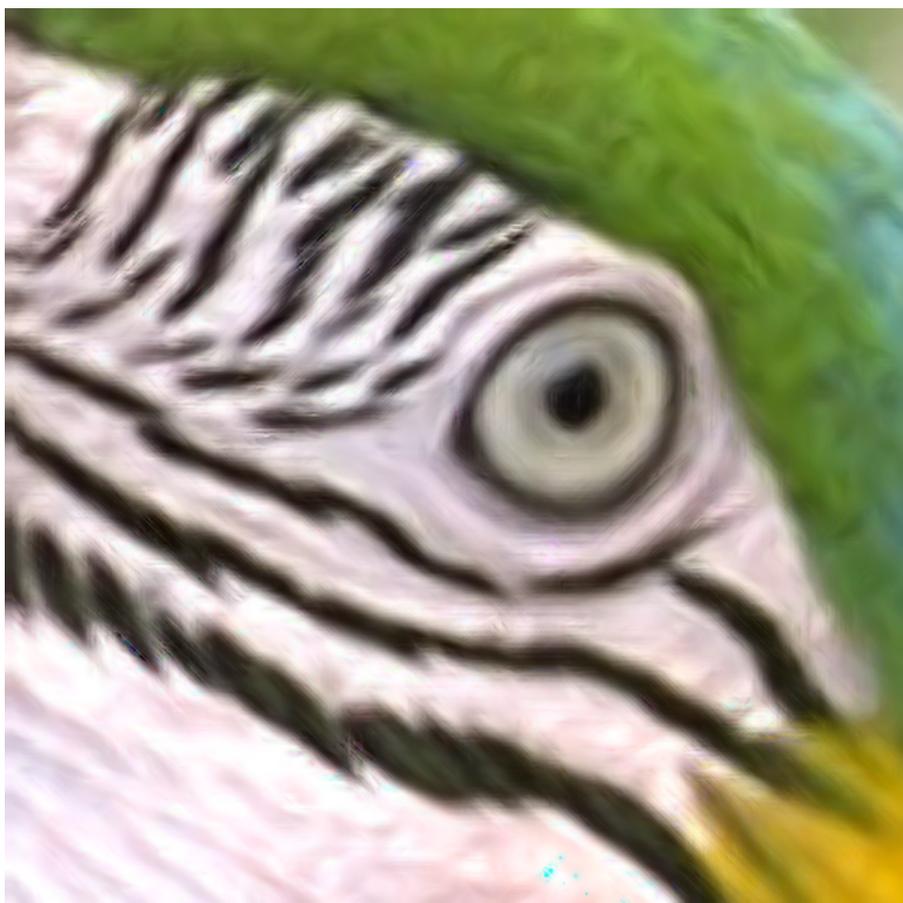


Figure 3.16: The NEDI interpolation.



Figure 3.17: The proposed method (bilinear) in RGB colour space.

We presented a novel approach to tackle artefacts (i.e. blur, staircase and ringing effects) created by linear image interpolation methods. The algorithm sharpens edges by mapping the image level curves using adaptive contrast enhancement techniques. To avoid the amplification of jagged edges, we restore the level curves by constrained isophote smoothing. We put additional constraints on the level curve mapping to meet the interpolation conditions and to reduce ringing artefacts. We briefly discussed various strategies to extend greyscale interpolation methods to colour images.

The results show improvements in both numerical PSNR results as well as in visual quality: the edges are much sharper, while staircase and ringing artefacts are heavily reduced. It is important to notice that almost no new artefacts are introduced with the level curve mapping interpolation.

A drawback in the current implementation is the time-consuming part of the iterative isophote smoothing: it takes about 5 seconds per iteration for an 1 megapixel high-resolution image on a computer with an AMD XP 3000 processor and 512 Mb RAM. Future work includes the implementation of real-time approximations of these level set evolution methods.

However, the most important remark is that general adaptive image interpolation can not deal with degraded images (e.g., corrupted with noise or compression artefacts). Therefore, image restoration techniques are introduced to take more accurate acquisition models into account (see Chapter 4).

4

High-resolution image restoration using colour priors

Convergence of image, vision and graphics?
— *Yah right... The truth is...*
— *Between vision and graphics communities:*
“*Vision is more **solid research** than graphics*”
“*But graphics people make **more money!***”
— *Graphics and vision communities say:*
“*Image processing is **low-level** processing*”
— *Image processing community says:*
“*Graphics is only some **fancy toys...***”
“*Vision is things that **don't work!***”
— Fernando Pereira

Nowadays more and more security firms are installing cameras in both public and private places to assist in the fight against crime. However, in order to recognize the face of a criminal or to read the license plate of a getaway car, we often need clean high-resolution (HR) images. Unfortunately, cheap camera sensors with low resolution are used massively nowadays in webcams, mobile phones and surveillance cameras, which often results in useless data. The high cost for high precision optics and image sensors is therefore an important concern in many commercial applications regarding HR imaging.

On the other hand, all digital imaging devices inevitably introduce blur and noise (as discussed in Section 2.3.2). There is a strong limitation on increasing spatial resolution by reducing the sensors (i.e. increasing the number of pixels per unit area) by manufacturing techniques. If the pixel size decreases, the amount of incoming light per pixel unit also decreases. This generates shot noise that degrades the image quality severely.

Due to the huge amount of data, images and video sequences are compressed before transmission or storage. Image quality will typically be lost in the

form of block artefacts (intrinsic to the used block structure in JPEG- and MPEG-algorithms) and mosquito noise (random noise originating from the quantization of high frequent information). With the growing popularity of high definition television (HDTV), these artefacts become more bothersome. That is why digital HR image restoration as a postprocessing tool becomes very important. Note that image restoration is distinct from image enhancement (such as *unsharp masking*). Image enhancement techniques are designed to manipulate an image in order to produce visually pleasing results, but without making use of any particular degradation model.

In this chapter, we start with a brief overview of the existing image restoration techniques. Afterwards, we describe the Bayesian image restoration framework and we introduce two new colour image priors. The adaptive bimodal colour image prior assumes that an edge pixel has a colour value that is typically a mixture of the colours of two connected regions, each having a dominant colour distribution. The global multimodal colour image prior is proposed for images with a strong multimodal colour distribution, i.e. just a few dominant colours, such as cartoons and logos. The main novelty of this work is to incorporate and to adapt the bimodal black-and-white text prior for a more general class of images. The key achievements are the automatic determination of the parameters (both locally adaptive as globally) and the extension to the multimodal case. We show that the use of these priors is very powerful in the restoration of cartoons and logos, thereby outperforming other existing state-of-the-art techniques in visual quality. This work led to two publications in the Springer's Lecture Notes on Computer Science series [Luong et al., 2007, Luong and Philips, 2007b].

4.1 Image restoration techniques

The goal of classical image restoration is to reconstruct a single clean output image from a single given degraded observed image. The literature on digital image restoration is very broad and extensive and spans several decades, see e.g. [Katsaggelos, 1991, Klette and Zamperoni, 1996, Bertero and Boccacci, 1998]. For a concise but representative review of digital image restoration, we refer the interested reader to [Banham and Katsaggelos, 1997]. In the next sections, we give a very brief overview and a rough classification of different image restoration techniques.

4.1.1 Classification of image restoration techniques

There are many types of image degradation processes. Most image restoration techniques focus on a specific degradation task, while other treat (jointly) several degradations at once. The two most common and popular goals by far in digital image restoration are *image denoising* and *deconvolution*.

4.1.1.1 Image denoising

In the design of classic noise reduction methods, the objective is to remove noise, while preserving the original image details and fine structures. During the past decade, numerous and diverse denoising methods have been proposed to this end. Many methods, like total variation [Rudin and Osher, 1994], bilateral filtering or kernel regression [Tomasi and Manduchi, 1998, Takeda et al., 2007] and wavelet-based techniques [Pižurica and Philips, 2006, Şendur and Selesnick, 2002, Portilla et al., 2003] estimate the denoised pixel intensities based on the information provided in a limited surrounding neighbourhood. These methods only exploit the spatial redundancy in a local neighbourhood and are therefore referred to as *local* methods.

Recently, a number of *non-local* methods have been developed, e.g. [Dabov et al., 2007, Dauwe et al., 2008, Goossens et al., 2008a]. These methods estimate every pixel intensity based on information from the whole image, thereby exploiting the presence of similar patterns and structures in an image. This relatively new class of denoising methods originates from the non-local means, introduced by Buades et al [Buades et al., 2005b, Buades et al., 2005a]. We will return to the non-local image processing method in Chapter 5.

4.1.1.2 Image deconvolution

The first encounters with digital image deconvolution can be found in the area of astronomical imaging in the 1950s and early 1960s [Banham and Katsaggelos, 1997]. Extraterrestrial observations of the planets were subject to motion blur as a result of slow camera shutter speeds relative to rapid spacecraft motion. Ground-based imaging systems suffer from blur due to the rapidly changing index of refraction of the atmosphere. A more published problem arose in the early 1990s with the main mirror imperfection of the Hubble Space Telescope.¹ Nowadays, image deconvolution finds its way in many application areas, such as forensic science, remote sensing and confocal microscopy.

The point spread function (PSF) varies from application to application, some examples are:

- in the presence of fast object motion or camera panning, the *motion blur* is represented as 1D uniform local averaging of neighbouring pixels.
- in remote sensing and aerial imaging, the *atmospheric turbulence blur* is commonly represented by a Gaussian PSF.

¹We refer the reader to <http://hubble.nasa.gov>, <http://spacetelescope.org> and <http://heritage.stsci.edu> for more information about the Hubble Space Telescope.

- in photographic defocusing, the *out-of-focus blur* is often modelled as a uniform distribution within a circular disk or rectangle.

However, in practice, the PSF is often unknown or only partially known and therefore, the blur kernel must be estimated prior to the restoration process (this is called blur identification or estimation) or, in the other case, the blur kernel and the deblurred image are estimated simultaneously from the data (which we refer to as *blind deconvolution*). A good survey of blind deconvolution and blur estimation techniques can be found in [Rooms, 2005, Campisi and Egiazarian, 2007].

4.1.1.3 Restoration of other degradations

Besides image denoising and deblurring, many other image restoration techniques exist, depending on the source of degradation. We discuss some examples, but for a more detailed treatment, we refer the reader to [Banham and Katsaggelos, 1997, Katsaggelos, 1991].

One of the most important area of application of digital image restoration in the consumer market segment today is that in the field of image and video coding. Typically in JPEG- and MPEG-based compression schemes, block artefacts become very annoying, especially on high-resolution screens (e.g., HDTV). Block artefacts are a result of the coarse quantization of transform coefficients, usually from a *discrete cosine transform* (DCT) on blocks of 8×8 pixels. Block transitions become more and more apparent when bit rates of the coded image or sequence are further reduced as illustrated in Figure 4.1. Most deblocking schemes apply spatially adaptive smoothing of the pixels along the block boundaries [Lee et al., 1995, O'Rourke and Stevenson, 1995].

Another interesting application in the field of media is the use of digital techniques to restore aging and deteriorated films. Much of this work belongs to the field of computer graphics (e.g. colourizing black-and-white films) and enhancement (e.g. motion stabilization and jitter correction), but a subset of the vast amount of work can be classified as spatially adaptive image restoration. Specific tasks are for example the elimination of scratches, dust and film grain noise, the correction colour fading problems, the reduction of flickering, etc. [Kokaram, 1993]. An example of an aging film is given in Figure 4.1.

4.1.2 Joint denoising and deconvolution

In literature, deblurring and denoising are commonly treated jointly because inverting a blur operator inherently amplifies the noise. Therefore, the most common model for a linear degradation caused by blur and additive noise is given by the following matrix-vector formulation:

$$\mathbf{g} = \mathbf{Hf} + \mathbf{n}, \quad (4.1)$$

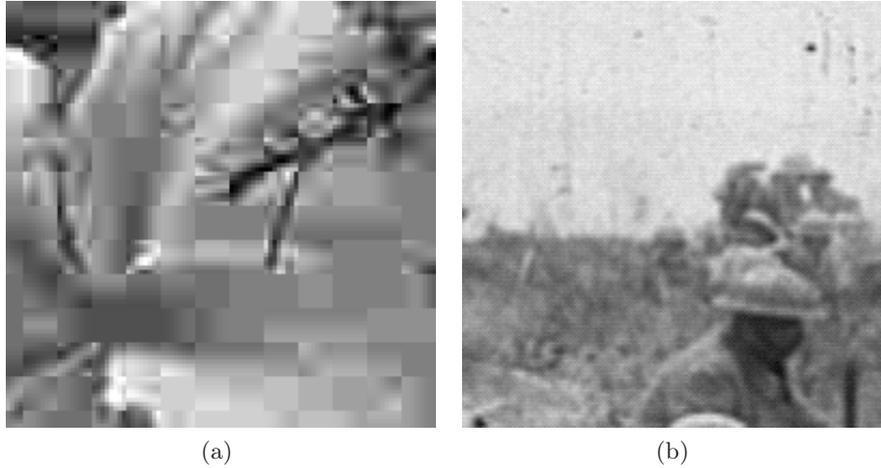


Figure 4.1: Examples of source degradations: (a) JPEG-encoded image (0.32 bits per pixel) resulting in block artefacts and (b) a deteriorated film (1918) with dust, scratches and film grain noise.

where \mathbf{g} , \mathbf{f} and $\mathbf{n} \in \mathbb{R}^n$ are the observed, ideal (or hypothetical desired) and noise images respectively in a column-stack ordering (i.e. the columns of each image are stacked into a vector) and $\mathbf{H} \in \mathbb{R}^{n \times n}$ represents the superposition blur operator.

In literature, *stationary* (or *shift-invariant*) blur models are commonly employed in the acquisition model. The use of a space-varying degradation model is also possible, but leads to more complex solutions. When using the stationary model, the matrix \mathbf{H} becomes a Toeplitz-block-Toeplitz (TBT) matrix. In a Toeplitz matrix \mathbf{T} , the elements $t(i, j)$ are such that $t(i, j) = t(i - 1, j - 1)$. A block matrix (also called *partitioned matrix*) is a partition of a matrix into smaller rectangular sub-matrices called blocks. A TBT matrix is thus a Toeplitz matrix of blocks where each squared sub-matrix itself has a Toeplitz structure. The approximation of a TBT matrix by a circulant-block-circulant (CBC) matrix has been addressed in [Bose and Boo, 1998]. A circulant matrix \mathbf{C} is a special kind of a Toeplitz matrix, in which each column is obtained by doing a wrap-around downshift of the previous column:

$$\mathbf{C} = \begin{pmatrix} c(0) & c(n-1) & c(n-2) & \dots & c(1) \\ c(1) & c(0) & c(n-1) & \dots & c(2) \\ c(2) & c(1) & c(0) & \dots & c(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c(n-1) & c(n-2) & c(n-3) & \dots & c(0) \end{pmatrix} \quad (4.2)$$

Since CBC matrices can be diagonalized using the 2D *discrete Fourier transform* (DFT), this leads to desirable discrete Fourier domain properties and eventually

to efficient computations in solving equation (4.1), as discussed in the next section.

4.1.2.1 Inverse filtering and ML estimator

Let us have a closer look at the classical signal processing problem stated in equation (4.1). We assume that the additive noise $\mathbf{n} \in \mathbb{R}^n$ is white and has a zero-mean Gaussian distribution with a standard deviation σ_n :

$$\mathbf{p}(\mathbf{n}) = \frac{1}{(2\pi)^{n/2}\sigma_n^n} \exp\left\{-\frac{\mathbf{n}^T\mathbf{n}}{2\sigma_n^2}\right\}. \quad (4.3)$$

Because \mathbf{f} is known and in the presence of additive Gaussian noise, the measurement vector \mathbf{g} is also a Gaussian random vector with a shifted mean and so the *likelihood function* $\mathbf{p}(\mathbf{g}|\mathbf{f})$ becomes

$$\mathbf{p}(\mathbf{g}|\mathbf{f}) = \frac{1}{(2\pi)^{n/2}\sigma_n^n} \exp\left\{-\frac{1}{2\sigma_n^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2\right\}. \quad (4.4)$$

The *maximum likelihood* (ML) estimator suggests to choose $\hat{\mathbf{f}}_{\text{ML}}$ that maximizes the likelihood function or equivalently minimizes the negative loglikelihood function or the l_2 -norm $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2$ in this case:

$$\begin{aligned} \hat{\mathbf{f}}_{\text{ML}} &= \arg \max_{\mathbf{f}} \mathbf{p}(\mathbf{g}|\mathbf{f}) \\ &= \arg \min_{\mathbf{f}} -\log(\mathbf{p}(\mathbf{g}|\mathbf{f})) \\ &= \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2. \end{aligned} \quad (4.5)$$

Clearly, this provides a least squares fit to the measurement data. In cases where the Gram matrix $\mathbf{H}^T\mathbf{H}$ is positive definite, the problem is considered well-posed (see discussion in Section 2.3.3) and the generalized inverse filter leads to a unique solution to the above minimization, being

$$\hat{\mathbf{f}}_{\text{ML}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{g}. \quad (4.6)$$

In cases where $\mathbf{H}^T\mathbf{H}$ is singular, the problem is considered ill-posed and more information is needed to obtain a unique valid solution. This leads naturally to the notion of regularization, as discussed in the next section.

In cases where the matrix $\mathbf{H}^T\mathbf{H}$ is positive definite nevertheless and where the problem is well-posed, the ML estimator is still weak. This can for example simply be illustrated by the image denoising problem, where $\mathbf{H} = \mathbf{I}$. The ML denoising result according to equation (4.6) is equal to the noisy observed image $\hat{\mathbf{f}}_{\text{ML}} = \mathbf{g}$. Again, some regularization is needed here to get a proper result.

We assume now that the PSF is spatially invariant and we consider that \mathbf{H} has a CBC structure, which means that \mathbf{H} (and therefore \mathbf{H}^T) can be diagonalized

using the 2D DFT $\mathcal{F}(\mathbf{H}) = \mathcal{H}$ where \mathcal{H} is thus a diagonal matrix. Using this diagonalization approach, the inverse matrix problem of the inverse filter given by equation (4.6) can then be solved efficiently in the discrete Fourier domain at frequencies \mathbf{u} as

$$\hat{F}(\mathbf{u}) = \frac{\mathcal{H}^*(\mathbf{u})G(\mathbf{u})}{|\mathcal{H}(\mathbf{u})|^2}, \quad (4.7)$$

where \hat{F} , \mathcal{H} and G denote the DFT of the restored image $\hat{\mathbf{f}}$, the PSF \mathbf{H} and the observed image \mathbf{g} respectively and where the $*$ -operator denotes the complex conjugative. Clearly, division by very small-valued $\mathcal{H}(\mathbf{u})$ results in a strong amplification of the noise. Assuming that the degradation is a low-pass filter, the small values of \mathcal{H} are found at high frequencies, where the noise is dominant over the image. Therefore, it is highly desirable to stabilize the methods to solve the restoration problem by imposing prior knowledge.

4.1.2.2 Regularization techniques

From a pure algebraic point of view, regularization of equation (4.6) is achieved by turning the penalty function into a strictly convex one, thus guaranteeing a unique solution. The use of deterministic prior information about the original image can also be used for regularizing the image restoration problem. This can be done via *constrained least squares* (CLS), which minimize the following Lagrangian (this is also known as the Tikhonov-Miller approach [Tikhonov et al., 1990]):

$$\begin{aligned} \hat{\mathbf{f}}_{\text{CLS}} &= \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{C}\mathbf{f}\|_2^2 \\ &= (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{C}^T\mathbf{C})^{-1}\mathbf{H}^T\mathbf{g}, \end{aligned} \quad (4.8)$$

where $\mathbf{C}^T\mathbf{C}$ is assumed to be positive definite and $\mathbf{C}\mathbf{f}$ generally represents a high-pass filtered version of the image \mathbf{f} . This is essentially a smoothness constraint that suggests that most images are relatively flat with limited high-frequency activity. One typical choice for \mathbf{C} is the 2D Laplacian operator [Banham and Katsaggelos, 1997]. λ represents the Lagrange multiplier, commonly referred to as the regularization parameter, which balances the fidelity to the data $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2$ (also referred to as the *data fidelity term*) and the smoothness of the solution $\|\mathbf{C}\mathbf{f}\|_2^2$ (also referred to as the *smoothness term* or *regularization term*).

If the PSF is spatially invariant and $\mathbf{H}^T\mathbf{H} + \lambda\mathbf{C}^T\mathbf{C}$ has a CBC structure, equation (4.8) may be solved directly in the discrete frequency domain as

$$\hat{F}(\mathbf{u}) = \frac{\mathcal{H}^*(\mathbf{u})G(\mathbf{u})}{|\mathcal{H}(\mathbf{u})|^2 + \lambda|\mathcal{C}(\mathbf{u})|^2}, \quad (4.9)$$

where $\mathcal{F}(\mathbf{C}) = \mathcal{C}$ with \mathcal{C} being a diagonal matrix. Note that this solution is equivalent to the well-known Wiener filter when λ and $\mathcal{C}(\mathbf{u})$ are chosen such

that

$$\lambda|\mathcal{C}(\mathbf{u})|^2 = \frac{S_{\text{nn}}(\mathbf{u})}{S_{\text{ff}}(\mathbf{u})}, \quad (4.10)$$

where S_{ff} represents the power spectrum of the ideal image (which is in general unknown, and has to be approximated or estimated) and S_{nn} denotes the power spectrum of the noise. The power spectra are defined as

$$S_{\text{ff}}(\mathbf{u}) = E [F^*(\mathbf{u})F(\mathbf{u})], \quad (4.11)$$

and

$$S_{\text{nn}}(\mathbf{u}) = E [N^*(\mathbf{u})N(\mathbf{u})], \quad (4.12)$$

where $\mathcal{F}(\mathbf{f}) = F$ and $\mathcal{F}(\mathbf{n}) = N$. From equation (4.10), we can clearly see that spectral components with $S_{\text{nn}}(\mathbf{u}) \ll S_{\text{ff}}(\mathbf{u})$ are amplified because

$$\frac{S_{\text{nn}}(\mathbf{u})}{S_{\text{ff}}(\mathbf{u})} \ll |\mathcal{H}(\mathbf{u})|^2. \quad (4.13)$$

On the other hand, all spectral components with $S_{\text{nn}}(\mathbf{u}) \approx S_{\text{ff}}(\mathbf{u})$ are attenuated in order to suppress noise in the restored image (at the expense of the signal components). For a more detailed treatment of this matter, we refer the interested reader to [Rooms, 2005].

While direct approaches to restore noisy and blurred images in the Fourier domain are intuitively simple and computationally attractive, they are poorly suitable for accommodating general scene observation models (such as spatially varying complex degradation) and they can not utilize general a-priori constraints or noise models. However, the formulation of restoration techniques in the spatial domain offers considerable flexibility in the range of degradations and observation models, and they are well suited to include prior knowledge. These benefits come typically at the expense of increased computational complexity.

Among the restoration techniques in the spatial domain, several approaches for computing regularized solutions have gained prominence in the last decades. These are for example PDE-based methods (we refer the reader to [Tschumperlé and Deriche, 2005, Tschumperlé, 2006] for a good and technical survey of PDE-based applications), Bayesian methods that are based on the stochastic estimation theory (a more detailed discussion follows in Section 4.2) and the deterministic set theoretic methods, also known as *projection onto convex sets* (POCS). The latter provides an intuitive and convenient framework for the inclusion of a-priori information [Borman, 2004].

The POCS method defines constraints (i.e. the desirable characteristics of the ideal image) as convex sets in the solution space that contains all possible images. The feasible solution space is the region of the intersection of a collection of convex sets. These sets encapsulate constraints such as the fidelity to the

observed data, positivity, smoothness, etc. The POCS approach is an iterative procedure that locates a point in the solution space that simultaneously satisfies all the constraint sets.

Given k convex constraint sets such that the intersection of the sets is non-empty, POCS projects the point (i.e. the image \mathbf{f}) onto each constraint set in turn using the projection operators \mathcal{P}_i ($i = 1 \dots k$). The current image $\mathbf{f}^{(j)}$ is iteratively updated using a sequence of projections:

$$\hat{\mathbf{f}}^{(j+1)} = \mathcal{P}_1 \mathcal{P}_2 \dots \mathcal{P}_k \hat{\mathbf{f}}^{(j)}. \quad (4.14)$$

Closedness and convexity of the constraint sets ensure convergence of the iteration. A more detailed theoretical discussion of the POCS method can be found in [Youla and Webb, 1982]. POCS techniques are widely used in different image restoration areas such as high-resolution image reconstruction [Gerchberg, 1974, Gerchberg, 1989, Papoulis, 1975], image denoising and deblurring [Stark, 1988, Özkan et al., 1994, Kundur and Hatzinakos, 1996], de-blocking algorithms [Yang et al., 1995], reconstruction of MRI images and computed tomography images [Oskoui-Fard and Stark, 1988, Peng et al., 2006].

A variant on the POCS-based formulation is the *bounding ellipsoid* method: the centroid of the bounding ellipsoid from a set of ellipsoidal constraint sets is the estimate of the restored image [Katsaggelos, 1990]. Since direct computation of this point is infeasible, an iterative procedure is employed.

Combining the desirable characteristics of the Bayesian MAP framework and the flexible POCS methods results in a hybrid optimization scheme. The general idea of these hybrid methods is to maximize the posterior PDF while ensuring that the solution remains within a constrained space specified using a set theoretic formulation (e.g. non-negative intensities). We refer the interested reader to [Borman, 2004] for an overview of such hybrid methods.

Besides iterative approaches, recursive filtering operations can be used for image restoration problems. These techniques can easily incorporate spatial adaptivity into the filter model and they usually require less memory than direct or iterative methods. An example is the discrete Kalman filter, which is the recursive equivalent of the Wiener filter. This filter is based on an autoregressive (AR) parameterization of the prior statistical knowledge of the image [Banham and Katsaggelos, 1997].

4.1.2.3 Non-regularized approaches

We briefly mention some existing non-regularized approaches to restore noisy and blurred images. Among them, multiscale restoration methods are often used, which operate for example in the wavelet domain or in the steerable pyramid domain [Banham and Katsaggelos, 1996, Portilla and Simoncelli, 2003, Rooms, 2005]. In [Immerkær, 2001], Immerkær employed a Gaussian blur-space to extrapolate the deblurred image from a blurred image.

With the emergence of very powerful denoising techniques, hybrid techniques are developed that apply a simple deconvolution technique (mostly operating in the Fourier domain) followed by the suppression of coloured noise (because the noise at the high frequencies is amplified stronger than at the low frequencies). Examples of such hybrid techniques are presented in [Kalifa et al., 2003, Neelamani et al., 2004, Katkovnik et al., 2005, Dabov et al., 2008a].

Learning-based (or training-based) algorithms for image restoration learn the prior knowledge from training data: codebooks are built using the frequency information obtained from original images and their blurred versions. During the restoration, the high-frequency information of a given degraded image is estimated from its low-frequency information based on the codebooks [Nakagaki and Katsaggelos, 2003].

A large class of iterative restoration methods deals with a simulate-and-correct approach to restoration, which is known as *iterative back-projection* (IBP).² The iterative process consists of two steps: in the first step, the observed images are constructed from the current estimated image through *simulation*. This requires the imaging system or the acquisition model to be known. The residual error is computed between the observed and the simulated images. In the second step, this residual error is used to correct and update the current estimated image by a process called *back-projection* [Borman, 2004, Irani and Peleg, 1991]. See also Section 7.1.2 for a more technical treatment.

4.2 Regularization from a Bayesian point of view

The image acquisition process consists of capturing a continuous scene into a (discrete) digital image. However, in practice, the acquired image lacks resolution and is corrupted by noise and blur. These linear degradation operations are illustrated in Figure 4.2. The recovery of the unknown high-resolution (HR) image \mathbf{f} from an observed low-resolution (LR) image \mathbf{g} is related by

$$\mathbf{g} = \mathbf{D}\mathbf{H}\mathbf{f} + \mathbf{n} = \mathbf{A}\mathbf{f} + \mathbf{n}, \quad (4.15)$$

where $\mathbf{g} \in \mathbb{R}^n$ and $\mathbf{f} \in \mathbb{R}^{r^2n}$ are the observed and ideal (or hypothetical desired) images respectively with a magnification factor r . In this equation, the matrices \mathbf{D} and \mathbf{H} (with dimensions of $n \times r^2n$ and $r^2n \times r^2n$) represent the decimation operator and the blur operator respectively and $\mathbf{n} \in \mathbb{R}^n$ describes the additive noise (typically zero-mean white Gaussian noise with a standard deviation σ_n). We also assume that the blur operator \mathbf{H} in the imaging model is denoted by a PSF (typically Gaussian blur, which is characterized by its standard deviation σ_b or a $r \times r$ block averaging operator). Because of the

²The term is adopted from the computed tomographic community, where a similar approach is widely used.

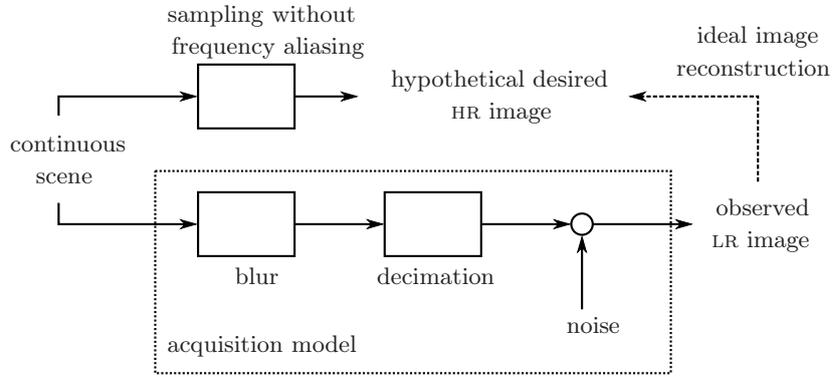


Figure 4.2: Observation model for high-resolution image reconstruction of noisy, blurred low-resolution images.

high dimensionalities (and the sparse representations), these matrix multiplications are replaced with their actual image operators in our implementation. The combination of decimation and blur can be represented by a general linear degradation operation $\mathbf{A} = \mathbf{D}\mathbf{H}$ with a matrix dimension of $n \times r^2n$.

Via the Bayes rule, the probability $\mathbf{p}(\mathbf{g}|\mathbf{f})$ in the likelihood function is replaced by the posterior probability $\mathbf{p}(\mathbf{f}|\mathbf{g})$, because \mathbf{f} is assumed to be random as well:

$$\mathbf{p}(\mathbf{f}|\mathbf{g}) = \frac{\mathbf{p}(\mathbf{g}|\mathbf{f})\mathbf{p}(\mathbf{f})}{\mathbf{p}(\mathbf{g})}. \quad (4.16)$$

The two most popular approaches based on the Bayesian formulation are the *maximum a posteriori* (MAP) and the *minimum mean squared error* (MMSE) methods.

The MAP estimator suggests to choose $\hat{\mathbf{f}}_{\text{MAP}}$ that maximizes the posterior probability [Van Trees, 1968]:

$$\begin{aligned} \hat{\mathbf{f}}_{\text{MAP}} &= \arg \max_{\mathbf{f}} \mathbf{p}(\mathbf{f}|\mathbf{g}) \\ &= \arg \max_{\mathbf{f}} \mathbf{p}(\mathbf{g}|\mathbf{f})\mathbf{p}(\mathbf{f}) \\ &= \arg \min_{\mathbf{f}} (-\log(\mathbf{p}(\mathbf{g}|\mathbf{f})) - \log(\mathbf{p}(\mathbf{f}))). \end{aligned} \quad (4.17)$$

Note that the denominator in equation (4.16), $\mathbf{p}(\mathbf{g})$, does not affect the maximization, since it is considered constant with respect to the optimization task. The ML estimator of equation (4.5) can be seen as a special case of the MAP method where the a priori *probability density function* (PDF) is uniformly distributed [Van Trees, 1968].

The MMSE estimator chooses the expected value of \mathbf{f} based on its conditional density $\mathbf{p}(\mathbf{f}|\mathbf{g})$:

$$\hat{\mathbf{f}}_{\text{MMSE}} = \int_{\mathbf{f} \in \mathbb{R}^{r^2 n}} \mathbf{f} \mathbf{p}(\mathbf{f}|\mathbf{g}) d\mathbf{f}, \quad (4.18)$$

which is simply the mean of the posterior distribution and minimizes the MSE measure $E \left[\left\| \hat{\mathbf{f}} - \mathbf{f} \right\|_2^2 \right]$. Since the integral is $r^2 n$ -dimensional, such an approach is typically prohibitive for non-scalar cases [Elad and Aharon, 2006]. If we look at computational convenience, the use of conjugate priors leads us to a tractable posterior PDF, which is on its turn a member of the same family of the prior PDF [Van Trees, 1968]. When no analytic solution is available, we can still find the MMSE (or MAP) solution via simulation of the posterior distribution. *Markov-chain Monte-Carlo* (MCMC) methods attempt to approximate the posterior PDF by the statistics of samples generated from a Markov chain [Campisi and Egiazarian, 2007]. These methods can provide solutions close to the optimal one. However, they are very computationally intensive: although in theory convergence to the global minimum of the posterior energy is guaranteed with a high probability, in practice, it is hard to tell when convergence has been reached. Note that the MMSE estimate of \mathbf{f} is the mean of all samples from the MCMC generator, while the MAP solution is the one reached when MCMC converges.

In both cases, the MAP and MMSE estimator need a clear definition of the prior PDF $\mathbf{p}(\mathbf{f})$. In general, the amount of prior knowledge plays a major part in achieving the best image restoration, whether the knowledge is about the degradation (e.g., noise, blur, etc.) or the knowledge is about the ideal image. It is obvious that correct and complete degradation models, i.e. with a correct identification of the type of degradation (e.g., white versus coloured noise, atmospheric versus out-of-focus blur, etc.) and a good estimation of the degradation parameters (e.g., σ_n and σ_b) is quite critical in solving the restoration problem. In the next section, we discuss the prior models for ideal HR images, which we call *image priors*, and we also introduce two new colour image priors.

4.3 Image priors

A general way to describe the prior PDF $\mathbf{p}(\mathbf{f})$ is the Gibbs distribution, because every non-negative function can be written as the following exponential form [Datsenko and Elad, 2007]:

$$\mathbf{p}(\mathbf{f}) = \text{const} \cdot \exp\{-\alpha G(\mathbf{f})\}, \quad (4.19)$$

where the constant in front of the exponential is a normalization factor, guaranteeing that the integral over all \mathbf{f} is 1. The term $G(\mathbf{f})$ is a non-negative energy function that is supposed to be low for signals with a high probability and high otherwise.

In the presence of additive zero-mean white Gaussian noise, the MAP estimator of equation (4.17) becomes

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg \min_{\mathbf{f}} \left(\|\mathbf{g} - \mathbf{D}\mathbf{H}\mathbf{f}\|_2^2 + \alpha G(\mathbf{f}) \right). \quad (4.20)$$

We can clearly see that the MAP method leads to the concept of CLS regularization as described earlier in equation (4.8), except that we now have a probabilistic meaning of the regularization term $G(\mathbf{f})$.

4.3.1 The evolution of image priors

In the last few decades, a considerable amount of research attention has been given to the design of image priors. We briefly mention some important milestones, showing how image priors have become more complex in an attempt to describe the image content in more accurate way.

For the sake of algebraic stability reasons, the Tikhonov regularization chooses the image prior $G(\mathbf{f}) = \|\mathbf{C}\mathbf{f}\|_2^2$ with \mathbf{C} as the Laplacian operator [Tikhonov et al., 1990]. This prior promotes spatial smoothness across the image in a uniform way. Of course, enforcing smoothness can be achieved by other high-pass operators \mathbf{C} or derivatives, such as $G(\mathbf{f}) = \|\mathbf{D}_h\mathbf{f}\|_2^2 + \|\mathbf{D}_v\mathbf{f}\|_2^2$, with \mathbf{D}_h and \mathbf{D}_v being the horizontal and vertical derivatives, respectively. Such first-order derivatives promote constant values, while second-order derivatives, such as the Laplacian, allow tilted planes and saddle points as well.

Enforcing smoothness across the whole image tends to blur edges and textures. That is why researchers have been proposing to adapt the smoothness spatially across the image from the early 1990s. A simple way to achieve such adaptation is the *weighted least-squares* expression $G(\mathbf{f}) = (\mathbf{C}\mathbf{f})^T \mathbf{W} (\mathbf{C}\mathbf{f})$ with \mathbf{W} being a diagonal matrix whose values are 1 for smooth regions and close to 0 in edge and texture regions. Unfortunately, frequency domain solutions are no longer possible and iterative restoration techniques become unavoidable.

Since the restoration process is iterative, the weight matrix \mathbf{W} can be estimated in each iteration based on the current solution, which is assumed to be better than the measurements. This process is called *iterative re-weighted least-squares* (IRLS) and as it turns out, this concept is very closely related to robust statistics and M-estimators more in particular. The field of robust statistics concentrates on estimation in the presence of outliers. While most of the regions in the image are smooth, edges appear as outliers when enforcing the smoothness constraint. The M-estimators are generalizations of ML estimators and least-squares [Stewart, 1999]. The use of these robust loss functions $\rho(\mathbf{C}\mathbf{f})$ instead of the l_2 -norm is very popular, because it suppresses the noise better while retaining important edge information [Pizurica et al., 2005]. E.g.,



Figure 4.3: Examples of connectivity neighbourhoods: (a) the 4-connectivity neighbourhood $\aleph_4(\mathbf{x})$ and (b) the 8-connectivity neighbourhood $\aleph_8(\mathbf{x})$ (or also denoted as $\aleph_{3,3}(\mathbf{x})$, where 3 denotes the size of the neighbourhood).

using the first-order derivatives, we have:³

$$G(f(\mathbf{x})) = \sum_{\mathbf{x}' \in \aleph_{k,k}(\mathbf{x})} \rho(\|f(\mathbf{x}) - f(\mathbf{x}')\|_1), \quad (4.21)$$

where $\aleph_{k,k}(\mathbf{x})$ denotes the $k \times k$ -connectivity neighbourhood centered around the pixel position \mathbf{x} on the HR grid. The 4-connectivity and 8-connectivity neighbourhoods are illustrated in Figure 4.3. Such local spatial interactions can be modelled by *Markov random fields* (MRF) in a convenient way. For more details about MRF, we refer the interested reader to [Geman and Geman, 1984, Pižurica, 2002].

It can be proven that the minimization of the function $\rho(x)$ (x being the residual, i.e. the l_1 -norm, between the measurement and current estimate) according to the MAP principles is equivalent to the IRLS process with the weight function $w(x) = \psi(x)/x$ where $\psi(x) = \rho_x(x)$ [Stewart, 1999]. Some commonly used robust loss functions with their associated ψ functions are given in Table 4.1 and some weight functions are plotted in Figure 4.4. With such functions, the image prior and the overall restoration algorithm become non-linear. A unique solution is guaranteed if $G(\mathbf{f})$ is convex, otherwise the IRLS process converges to a local minimum. To overcome this problem, optimization strategies such as *graduated non-convexity* (GNC) can be employed [Blake, 1989].

In [Black et al., 1998], Black et al. showed that different robust loss functions in the robust statistical estimation framework are closely related to the so-called *edge-stopping* or *diffusivity* functions in the anisotropic diffusion of the PDE-based image processing. Therefore, several contributions such as the popular *total variation* (TV) by Rudin et al. [Rudin and Osher, 1994], the Beltrami flow by Kimmel et al. [Kimmel et al., 2000] and the directional filter by Weickert [Weickert, 1998] are connected to techniques from robust statistics, although PDE's are originally formulated for continuous images. In the spirit of the directional anisotropic diffusion, we derive a geometry-driven smoothness prior in the next section.

³We remind that \mathbf{f} and \mathbf{g} are the vector formulation (in column-stack ordering) of the images f and g respectively and therefore, their representations can vary in accordance with the formulation.

Table 4.1: Several popular M-estimator functions and their associated ψ functions where h is a smoothing parameter and x the residual between the measurement and current estimate.

Robust loss function	Definition
Total variation (l_1 -norm)	$\rho(x) = x $ $\psi(x) = \text{sign}(x)$
Tikhonov (l_2 -norm)	$\rho(x) = \frac{1}{2}x^2$ $\psi(x) = x$
Huber	$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } x \leq h \\ \frac{1}{2}h(2 x - h) & \text{else} \end{cases}$ $\psi(x) = \begin{cases} x & \text{if } x \leq h \\ \text{sign}(x)h & \text{else} \end{cases}$
Cauchy-Lorentzian	$\rho(x) = \frac{h^2}{2} \log \left(1 + \left(\frac{x}{h} \right)^2 \right)$ $\psi(x) = \frac{x}{1 + (x/h)^2}$
Geman-McClure	$\rho(x) = \frac{x^2}{h + x^2}$ $\psi(x) = \frac{2hx}{(h + x^2)^2}$
Beaton-Tukey	$\rho(x) = \begin{cases} \frac{h^2}{6} \left[1 - \left(1 - \left(\frac{x}{h} \right)^2 \right)^3 \right] & \text{if } x \leq h \\ \frac{h^2}{6} & \text{else} \end{cases}$ $\psi(x) = \begin{cases} x \left[1 - \left(\frac{x}{h} \right)^2 \right]^2 & \text{if } x \leq h \\ 0 & \text{else} \end{cases}$
Le Clerc	$\rho(x) = \frac{h^2}{2} - \frac{h^2}{2} \exp \left(- \left(\frac{x}{h} \right)^2 \right)$ $\psi(x) = x \exp \left(- \left(\frac{x}{h} \right)^2 \right)$
Andrews	$\rho(x) = \begin{cases} \frac{h}{\pi} - \frac{h}{\pi} \cos(\pi x/h) & \text{if } x \leq h \\ \frac{h}{\pi} & \text{else} \end{cases}$ $\psi(x) = \begin{cases} \sin(\pi x/h) & \text{if } x \leq h \\ 0 & \text{else} \end{cases}$

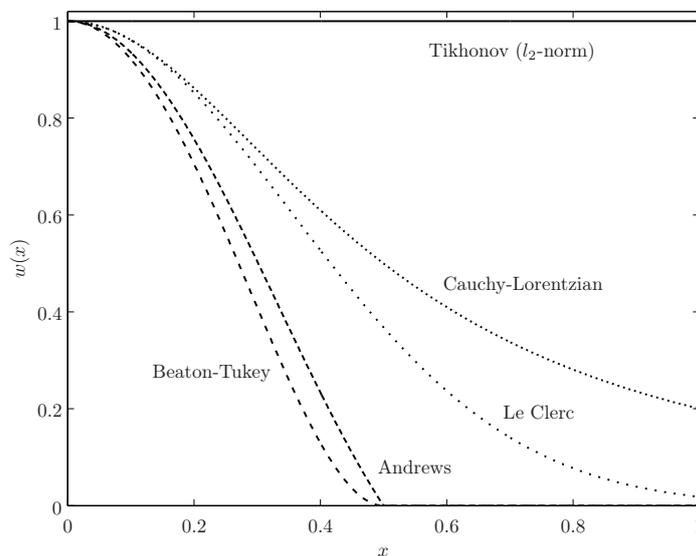


Figure 4.4: Plots of the weighting functions $w(x) = \psi(x)/x$ for commonly used robust loss functions ($h = 1/2$). Less weight is given to outliers (larger x) compared to the Tikhonov regularization.

Besides enforcing edge-preserving constraints, researchers from the field of approximation theory promote using the degree of non-sparsity of the wavelet coefficients as a regularization. This is because most wavelet coefficients are almost 0. The same empirical observations lead also to new applications such as compressive sampling [Candès and Wakin, 2008, Romberg, 2008]. The image prior in the regularization expression becomes the l_p -norm $G(\mathbf{f}) = \|\mathbf{C}\mathbf{f}\|_p^p$ ($p \leq 1$), where \mathbf{C} is now a wavelet transform operator in matrix form. Also the multiresolution principle is an important and powerful aspect in the restoration framework, since the wavelet transform can be interpreted as a set of multi-scale derivatives.

It is unfortunately impossible to grasp the complexity and wealth of general image content in a simple analytic expression $G(\mathbf{f})$. Therefore, a lot of research effort is recently done in learning-based (or *exemplar-based*) restoration. Examples can be used to learn the prior parameters, to learn the posterior PDF directly (i.e. using a codebook for the restoration that contains samples of the posterior PDF) or to build a regularization expression from the examples. A good survey of these techniques can be found in [Datsenko and Elad, 2007].

4.3.2 Geometry-driven smoothness prior

We assume that images are locally smooth except near edges. When dealing with image upscaling problems, we have to face an additional potential trap: preserving the edge discontinuities usually leads to the amplification of staircase artefacts. We can avoid this problem by taking the local geometry of the image into account, i.e. we perform a directional smoothing along the edge directions and simultaneously avoid smoothing orthogonally to these edges [Weickert, 1998]. These edge directions are estimated at a coarse scale such that these staircase artefacts do not affect the estimation (because jagged edges only appear at a fine scale). A similar strategy is applied in edge-directed image interpolation techniques as discussed earlier in Section 3.1.1.

The local geometry can be represented in the more convenient form of a 2×2 symmetric and semi-positive matrix, called the *diffusion tensor* \mathbf{T} [Tschumperlé, 2006]. The constructed diffusion tensor $\mathbf{T} = \lambda_\eta \eta \eta^\top + \lambda_\xi \xi \xi^\top$ has two orthonormal eigenvectors $\eta = \nabla \mathbf{f}_\sigma / \|\nabla \mathbf{f}_\sigma\|$ and $\xi = \eta^\perp = \nabla \mathbf{f}_\sigma^\perp / \|\nabla \mathbf{f}_\sigma\|$ with corresponding eigenvalues λ_η and λ_ξ respectively, where $\nabla \mathbf{f}_\sigma$ denotes the smoothed gradient $\nabla \mathbf{f} * L_\sigma$ (where L_σ is a 2D isotropic Gaussian kernel with standard deviation σ).

The direction ξ corresponds to the edge direction, when there is an edge, while η is then the vector perpendicular to the edge (also called the *normal vector*). The proposed positive values $\lambda_\eta = \max(\|\nabla \mathbf{f}_\sigma\|^2, 1)$ and $\lambda_\xi = 1$ (due to normalization) are related to the local strength of the edge. Given the diffusion tensor \mathbf{T} , we can construct the 2D oriented Gaussian kernel:

$$L_{\mathbf{T}(\mathbf{x}),t}(\mathbf{x}') = \frac{1}{4\pi t} \exp\left\{-\frac{\mathbf{x}'^\top \mathbf{T}^{-1} \mathbf{x}'}{4t}\right\}, \quad (4.22)$$

where parameter t is related to the diffusion strength. The shape and size of these Gaussian kernels are different at different locations depending on the local edge content and thus the eigenvalues $1/\lambda_\eta$ and $1/\lambda_\xi$ of \mathbf{T}^{-1} . Note that these space-varying kernels can be constructed from two oriented 1D kernels [Tschumperlé, 2006].

In smooth regions, the edge strength $\|\nabla \mathbf{f}_\sigma\|$ is small and $1/\lambda_\eta \approx 1/\lambda_\xi = 1$, which yields an isotropic Gaussian kernel for smoothing. Near sharp edges, $\|\nabla \mathbf{f}_\sigma\|$ will be large and $1/\lambda_\eta \approx 0 \ll 1/\lambda_\xi = 1$, which yields a highly anisotropic Gaussian kernel oriented along the edge direction. The whole concept is illustrated in Figure 4.5.

Based on these oriented kernels we define the Gibbs geometry-driven smoothness prior with first-order derivatives as

$$G(f(\mathbf{x})) = \sum_{\mathbf{x}' \in \mathcal{N}_{k,k}(\mathbf{x})} L(\mathbf{x}' - \mathbf{x}) \rho(\|f(\mathbf{x}) - f(\mathbf{x}')\|_1), \quad (4.23)$$

where $\mathcal{N}_{k,k}(\mathbf{x})$ denotes the local neighbourhood of \mathbf{x} (as illustrated in Figure 4.3) and where the weight function is the Gaussian steering kernel $L = L_{\mathbf{T}(\mathbf{x}),t}$ of equation (4.22).

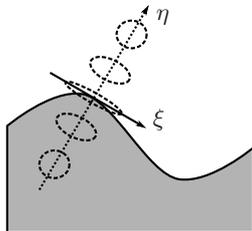


Figure 4.5: Illustration of several footprints of the oriented Gaussian kernel: the kernel is highly anisotropic near edges and isotropic in smooth regions.

Note that equation (4.23) can be seen as a generalization of the popular bilateral diffusion with L being an isotropic spatial weighting kernel [Tomasi and Manduchi, 1998] and the edge-preserving diffusion image priors, where the term $L(\cdot)$ is equal to 1 as described in equation (4.21). Another example that falls into this category is the *bilateral total variation* (BTV) image prior as proposed by Farsiu et al. [Farsiu et al., 2004]:

$$G(f(\mathbf{x})) = \sum_{\mathbf{x}' \in \mathcal{N}_{k,k}(\mathbf{x})} \gamma^{\|\mathbf{x}' - \mathbf{x}\|_1} \|f(\mathbf{x}) - f(\mathbf{x}')\|_1, \quad (4.24)$$

where γ is a weight parameter that takes spatial distance into account.

4.3.3 Proposed colour image priors

Almost all scientific publications in the last decades about image priors are focused on spatial and/or temporal (e.g. in video applications) smoothness and edge-preserving constraints. However, it is also possible to incorporate prior knowledge about pixel colours or intensities directly in addition to spatial diffusion of neighbouring pixels.

Bimodal colour priors have been successfully applied in low-resolution text enhancement [Donaldson and Myers, 2005, Thouin and Chang, 2000]. Typically the intensities of the text pixels tend to cluster around black and the intensities of the background pixels tend to cluster around white. Taking the bimodality into account improves the contrast and thus the readability. We extend this concept for general image restoration. Another use of multimodality is employed in image retrieval [Matas et al., 2000]: local windows with one, two or three colours (respectively unimodal, bimodal and trimodal neighbourhoods) describe the features being matched in the image database.

Because we are dealing with distributions with several modes, these multimodal priors are closely related to the well-known *finite mixture models* (FMM) in statistics. Such models represent a probability distribution that is a convex combination of other probability distributions. These models are for example used for the estimation of the statistics between the

grey and white brain tissues for MRI data segmentation [Schroeter et al., 1998].

In the next sections, we describe two different colour image priors, namely the adaptive bimodal colour prior and the multimodal colour prior. The adaptive bimodal colour prior expresses that an edge pixel has a colour value that is typically a mixture of the colours of two connected regions, each having a dominant colour distribution. Local adaptation of the parameters of the bimodal prior is made to handle neighbourhoods that have typically more than two dominant colours.

In some applications the number of colours in the ideal image is limited. Therefore the multimodal colour prior is proposed for images that normally just have a few dominant colours. In spite of this, real world images contain many more colours due to noise, compression artefacts and edge pixels between two or more connected smooth regions.

4.3.3.1 Adaptive bimodal colour prior

The number of different colours in a small image region is very small in general, if we do not take noise and gradients into account. Depending on the number of modes of the probability distribution of pixel colour values, the regions are characterized as *unimodal* (one dominant colour), *bimodal* (two modes), or in general, *multimodal* neighbourhoods. Unimodal neighbourhoods appear in flat regions, while bimodal neighbourhoods occur in edge regions. For bimodal regions we use a Gibbs prior with a non-negative fourth-order polynomial [Donaldson and Myers, 2005]:

$$G(f(\mathbf{x})) = \|f(\mathbf{x}) - \mu_1(\mathbf{x})\|_2^2 \|f(\mathbf{x}) - \mu_2(\mathbf{x})\|_2^2, \quad (4.25)$$

where $\mu_1(\mathbf{x})$ and $\mu_2(\mathbf{x})$ are the means of the background and foreground pixel distribution respectively. The one-dimensional bimodal PDF is illustrated in Figure 4.6.

When $\mu_1(\mathbf{x})$ and $\mu_2(\mathbf{x})$ are equal (or close) to each other, the bimodal PDF has a much lower peak than the normal distribution as plotted in 1D in Figure 4.6. This means that the *kurtosis*, i.e. the degree of peakness, is lower than 3 (i.e. lower than the value for a Gaussian distribution) and this is known as *platykurtosis*. In 1D, this gives

$$\begin{aligned} \frac{\sqrt[4]{2}\Gamma(\frac{3}{4})}{\pi\sigma} \int_{-\infty}^{+\infty} \frac{(x-\mu)^4}{\sigma^4} \exp\left\{-\frac{(x-\mu)^4}{2\sigma^4}\right\} dx &= \frac{1}{2}, \\ \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \frac{(x-\mu)^4}{\sigma^4} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx &= 3, \end{aligned} \quad (4.26)$$

where σ is the standard deviation and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$. Because the platykurtic PDF has a large dispersion and no clear peak, the mode can not be determined accurately and as a consequence, we can not take full advantage

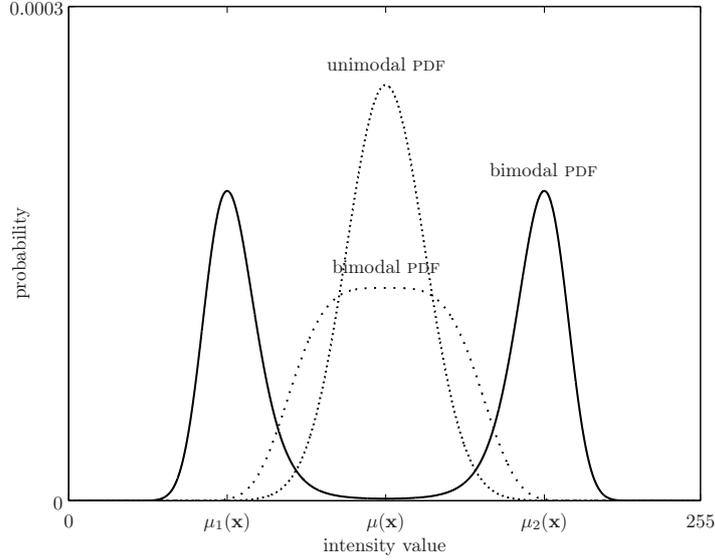


Figure 4.6: 1D plot of the Gibbs PDF of the bimodal priors and the unimodal prior.

of this prior model. That is why it is important to group modes that are the same or close to each other and replace them by a single mode.

In case of a single mode, we switch to a unimodal PDF if the Euclidean distance $\|\mu_1(\mathbf{x}) - \mu_2(\mathbf{x})\|_{2, L^*a^*b^*}$ is lower than a threshold τ in the $L^*a^*b^*$ colour space.⁴ The unimodal Gaussian PDF is given by:

$$G(f(\mathbf{x})) = \|f(\mathbf{x}) - \mu(\mathbf{x})\|_2^2, \quad (4.27)$$

where $\mu(\mathbf{x})$ is the average between $\mu_1(\mathbf{x})$ and $\mu_2(\mathbf{x})$.

If an image should contain two colours only, e.g. a black-and-white cartoon or a text document, we use a global μ_1 (e.g. white) and a global μ_2 (e.g. black). Via the *expectation maximization* (EM) algorithm, we can obtain the parameters μ_1 and μ_2 of the mixture of Gaussian distributions [Dempster et al., 1977]. Nevertheless, for general colour images, the assumption of one or two colours is only valid in a small image patch. That is why we need to estimate μ_1 and μ_2 locally from the colour content of the neighbourhood of a pixel at position \mathbf{x} .

⁴The $L^*a^*b^*$ colour space is much more *perceptually uniform* compared to the RGB colour space, which means that a linear change of data should result in a linearly perceived colour change. However, note that the uniformity is less correct in the presence of extreme intensities.

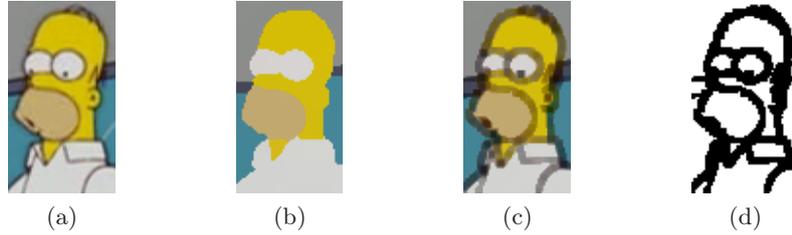


Figure 4.7: The local colour mode selection: (a) original image, (b) μ_1 (light), (c) μ_2 (dark) and (d) bimodal regions.

An edge pixel has a colour value that is typically a mixture of the colours of two connected uniform regions. To estimate the colour of each uniform region accurately, we construct the image f' by clustering similar colour values using the mean-shift algorithm [Comaniciu and Meer, 2002]. This mean-shift filtering preserves discontinuities and retains local image structures. We distinguish the colour modes $\mu_1(\mathbf{x})$ and $\mu_2(\mathbf{x})$ as the lightest ($\mu_1(\mathbf{x})$) and darkest ($\mu_2(\mathbf{x})$) colour in a local neighbourhood. This is determined by the luminance $f_L(\mathbf{x})$ (in the $L^*a^*b^*$ colour space):

$$\begin{aligned} \mu_1(\mathbf{x}) &= f'(\hat{\mathbf{x}}') & \text{where} & \quad \hat{\mathbf{x}}' = \arg \max_{\mathbf{x}'} f_L(\mathbf{x}'), \\ \mu_2(\mathbf{x}) &= f'(\hat{\mathbf{x}}') & \text{where} & \quad \hat{\mathbf{x}}' = \arg \min_{\mathbf{x}'} f_L(\mathbf{x}'), \end{aligned} \quad (4.28)$$

where $\mathbf{x}' \in \mathfrak{N}_{r,r}(\mathbf{x})$ with r being the magnification factor (the 3, 3-connectivity neighbourhood is for example illustrated in Figure 4.3). In short, equation (4.28) selects the pixel colour of the mean-shift-filtered image according to the ordering of the luminance in the $L^*a^*b^*$ colour space. This is closely related to colour mathematical morphology, which is treated extensively in [De Witte, 2007]. Examples of μ_1 - and μ_2 -images are illustrated in Figure 4.7.

4.3.3.2 Multimodal colour prior

The number of different colours in an image neighbourhood is very small in general, if we do not take noise and edge pixels into account. The same assumption also holds for images with just a few dominant colours like cartoons, drawings or logos. For multimodal colour images with m colour modes, we use a Gibbs prior with a non-negative $2m$ -order polynomial, which is a generalization of the bimodal prior presented earlier in equation (4.25):

$$G(f(\mathbf{x})) = \prod_{i=1}^m \|f(\mathbf{x}) - \mu_i\|_2^2, \quad (4.29)$$

where μ_i is the mean of the colour distribution around the i th mode.

If we know how many colours an image contains a priori, we can find the parameters μ_1, \dots, μ_m easily by locating the peaks of the multimodal distribution. Via the EM algorithm, we can for example obtain the means μ_i of the mixture of Gaussian distributions [Dempster et al., 1977]. A typical application where such prior knowledge is present is the restoration of scanned documents with only two colours, namely the foreground colour (e.g. black) and the background colour (e.g. white).

With the exception of black and white documents, we normally do not know how many dominant colours there are in an arbitrary image. Therefore, we propose a method to calculate the parameters μ_i robustly in a three-step algorithm. In the first step, we preselect some candidate colour modes μ_c . This is done by counting the number of neighbouring colours for each pixel $g(\mathbf{x})$ on the LR grid. Two colours μ_c and $g(\mathbf{x})$ are neighbours if $\|\mu_c - g(\mathbf{x})\|_{2, L^*a^*b^*} < \tau$ (i.e. the Euclidean distance between $g(\mathbf{x})$ and μ_c in the $L^*a^*b^*$ colour space is smaller than τ). We now select $n > m$ colours with the highest number of neighbouring colours, where we neglect the neighbours of the already selected colours.

In the second step, we apply mean-shift filtering on the selected candidate colour modes μ_c . To establish the location of the modes of the colour distribution, the mean-shift algorithm is applied in the $L^*a^*b^*$ colour space. Starting from each candidate colour mode μ_c , the mean-shift procedure iteratively finds a path along the gradient direction away from the valleys and towards the nearest peak that is equivalent to a gradient ascent to the local mode of the distribution [Comaniciu and Meer, 2002]. The positions of the modes are iteratively updated as follows:

$$\mu_c^{(j+1)} = \frac{\sum_{\mathbf{x} \in \Omega(g)} g(\mathbf{x}) r \left(\left\| \frac{g(\mathbf{x}) - \mu_c^{(j)}}{h} \right\|^2 \right)}{\sum_{\mathbf{x} \in \Omega(g)} r \left(\left\| \frac{g(\mathbf{x}) - \mu_c^{(j)}}{h} \right\|^2 \right)}, \quad (4.30)$$

where $\Omega(g)$ contains all the pixels of the LR image g , h represents the window bandwidth and r is the profile that defines the kernel [Comaniciu and Meer, 2002]. Using the multivariate Gaussian kernel, the profile r becomes

$$r(x) = \exp\left(-\frac{x}{2}\right) \quad (x \geq 0). \quad (4.31)$$

In the last step, we group colour modes μ_c together that are equal or close to each other (due to the platykurtosis effect as explained in the previous section). These colour modes are then replaced by their average μ_i .

4.4 Optimization

In order to solve the MAP estimation problem as stated in equation (4.20), we determine the regularization parameter α , which balances the fidelity to the data and the prior knowledge, and we derive a numerical solution that minimizes the cost criterium.

4.4.1 The choice of regularization parameter

For the sake of completeness, we briefly mention some popular methods for choosing α . A more detailed treatment can be found in [Galatsanos and Katsaggelos, 1992]. Intuitively, the regularization parameter α depends on the noise level: in the noise-free case, the restored image should completely depend on the observed data and little regularization is needed to overcome potential numerical instabilities, and in the case the observed data contains only noise (and no signal information), the restored image depends entirely on the prior information. A trivial choice for α is then the inverse of the *signal-to-noise* (SNR) ratio. Such an approach requires the knowledge of the noise variance, which can be estimated before restoration [Galatsanos and Katsaggelos, 1992].

The scalar regularization parameter α can also be optimized with a 1D line search method, such as *golden section search*, *inverse parabolic interpolation* and Brent's method [Press et al., 1988]. The optimal parameter corresponds with the minimum MSE or total error between the observed and the re-simulated image (i.e. the restored image after applying the observation model). This process is also known as the constrained least squares method [Rooms, 2005]. Note that parameters that have been selected based on *trial and error*, i.e. to produce the visually most appealing results, also follows a similar but simple strategy, based on visual (subjective) perception and not on objective measures. Another practical method for choosing α in the presence of noisy data is the *L-curve* criterion [Hansen and O'Leary, 1993]. The *L-curve* criterion is based on the plots of the norm of the regularized solution versus the norm of the corresponding residual. The regularization parameter is related to the characteristic *L-shaped corner* of the graph. The corner is determined as the closest point to the origin or as the point with the maximum curvature. A variant on this matter is the use of *U-curves* [Krawczyk-Stańdo and Rudnicki, 2007].

One of the most popular ways to determine the value of α without requiring knowledge of the noise variance is the *generalized cross-validation* (GCV) method [Golub et al., 1979]. The idea of cross-validation is quite simple: we divide the image data into two parts, one part is used to construct an approximate solution (the so-called training set), and the other is used to validate that approximation. The choice of the value α is the one that minimizes the cross-validated prediction error, which is usually the MSE. A special type commonly being used is the *leave-one-out* cross-validation (LOOCV): the training

set consists of all restored pixels, except the one under consideration.

4.4.2 Gradient-based minimization

The maximization problem of the MAP estimator becomes a minimization problem after taking the logarithm which is expressed by equation (4.17). The explicit minimization can be performed by computing the gradient of the cost function $\nabla\Psi(\mathbf{f})$ and equating this gradient to zero. The minimum of the cost function $\Psi(\mathbf{f})$ can then be found by the following iterative sequence (also known as the *steepest descent* algorithm):

$$\hat{\mathbf{f}}^{(j+1)} = \hat{\mathbf{f}}^{(j)} - \lambda_k \nabla\Psi(\hat{\mathbf{f}}^{(j)}), \quad (4.32)$$

where the scalar parameter λ_k determines the convergence speed or step size. The optimal choice of λ_k , which takes the optimal step in the steepest descent direction, can usually be determined theoretically. However, for complex non-linear functions, the solution for the optimal λ_k can not be obtained in a closed form. In practice, a fixed value of λ_k is used for all iterations and the parameter λ is chosen preferably as the largest value that does not cause the algorithm to diverge (i.e. $\|\hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)}\|$ decreases with each step).

More powerful techniques can be used to solve the same minimization problem in less iterations. Some examples are the implicit iterative minimization schemes (e.g., *Picard's iteration*), the *conjugate gradient* methods (e.g., the *Fletcher-Reeves* algorithm and the closely related *Polak-Ribière* algorithm), the *Gauss-Newton* methods (and the covering *Levenberg-Marquardt* algorithm which actually combines the Gauss-Newton and the steepest descent algorithms) and the *quasi-Newton* or *variable metric* methods (e.g., symmetric rank 1, the *Davidon-Fletcher-Powell* algorithm or the closely related *Broyden-Fletcher-Goldfarb-Shanno* algorithm) [Press et al., 1988]. However, the more powerful methods require more complex computations (e.g., the Hessian or the analysis of successive gradient vectors) at each iteration step. As the number of variables in the problem increases, the simpler methods become more attractive and competitive (they require more iterations but also the fewer computations per iteration so that they can still be faster) than the more complex techniques. Therefore, the steepest descent formulation offers a reasonable trade-off between the number of iterations and the computational complexity at each iteration, particularly for the large number of variables in the digital image restoration framework [Hunt, 1977].

Based on the proposed adaptive bimodal colour prior and the multimodal colour prior (as introduced in Section 4.3.3), we have developed two methods for restoring images corrupted by loss of resolution and additive zero-mean white Gaussian noise.

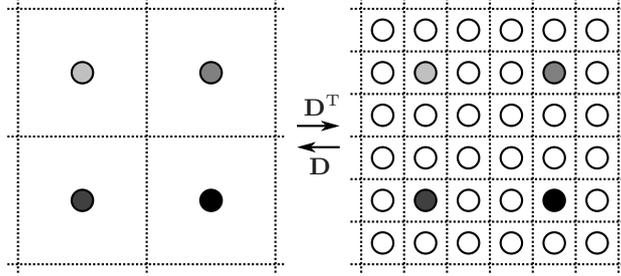


Figure 4.8: Illustration of the effect of the downsampling operator \mathbf{D} and the up-sampling operator \mathbf{D}^T . The samples indicated by white circles on the high-resolution grid are filled with zeros.

4.4.2.1 Restoration with adaptive bimodal colour prior

The first method assumes that the linear degradation operator \mathbf{A} in equation (4.15) denotes a simple block averaging operator, which is commonly used in fast and simple downscaling algorithms. Based on the simple Tikhonov regularization (see Table 4.1) and the adaptive bimodal colour prior, the solution can be found plugging the following closed-form expression in equation (4.32):

$$\begin{aligned} \nabla \Psi(\hat{f}^{(j+1)}(\mathbf{x})) &= \left([\mathbf{A}^T] \left([\mathbf{A}] \hat{f}^{(j)} - g \right) \right) (\mathbf{x}) + \alpha_b \Phi(\hat{f}^{(j)}(\mathbf{x})) + \\ & 2\alpha_s \sum_{\mathbf{x}' \in \mathbb{N}_4(\mathbf{x})} \left(\hat{f}^{(j)}(\mathbf{x}) - \hat{f}^{(j)}(\mathbf{x}') \right), \end{aligned} \quad (4.33)$$

where α_b and α_s are the weights or regularization parameters, $[\mathbf{A}]$ is the linear degradation operator (which operates on the whole image, $[\mathbf{A}]f$ is thus the simulated LR image) and the operator $[\mathbf{A}^T] = [\mathbf{H}^T \mathbf{D}^T]$ projects the values from the LR grid to the HR grid according to an upsampling scheme \mathbf{D}^T (which is the inverse operation of \mathbf{D} as illustrated in Figure 4.8) followed by the operator \mathbf{H}^T , which performs a convolution with a new PSF kernel. This PSF kernel is constructed by flipping the columns of the original blur kernel in the left-right direction (i.e. about the vertical axis) and then flipping the rows in the up-down direction (i.e. about the horizontal axis) [Farsiu et al., 2004]. In case of isotropic PSF kernels, the kernel is exactly the same as the one in the degradation model, because $\mathbf{H}^T = \mathbf{H}$. The initialization of the iterative procedure is obtained by the nearest neighbour interpolation $\hat{f}^{(0)}$.

The regularization term $\Phi(\hat{f}^{(j)}(\mathbf{x}))$ for the bimodal regions (see equation (4.28) for the parameters $\mu_1(\mathbf{x})$ and $\mu_2(\mathbf{x})$) becomes:

$$\Phi(x) = 2(x - \mu_1(\mathbf{x})) \|x - \mu_2(\mathbf{x})\|_2^2 + 2(x - \mu_2(\mathbf{x})) \|x - \mu_1(\mathbf{x})\|_2^2, \quad (4.34)$$

while minimizing the quadratic term of the unimodal PDF in equation (4.27)

leads to the following simple regularization term:

$$\Phi(x) = 2(x - \mu(\mathbf{x})). \quad (4.35)$$

The parameters used in the experiments are $\lambda = 0.125$, $\alpha_s = 0.5$, $\alpha_b = 0.5$ and $\alpha_b = 2e-5$ for unimodal and bimodal regions, respectively. The regularization parameters are chosen according to the maximum possible size of the regularization terms. The parameter τ , which distinguishes unimodal from bimodal regions as described in Section 4.3.3.1, is set to 10. The number of steepest descent iterations is 100.

4.4.2.2 Restoration with multimodal colour prior

The second method integrates the proposed geometry-driven smoothness prior with l_2 -norm (see Section 4.3.2) and the global multimodal prior (see Section 4.3.3.2) in an MAP restoration framework. On the one hand, we solve the minimization problem of the MAP estimator by substituting the previously defined priors in the steepest descent algorithm. On the other hand, we optimize the parameters μ_i of the multimodal priors on the HR grid. The latter can be solved by computing the steepest ascent in the mean-shift algorithm as discussed in Section 4.3.3.2. More precisely, we iteratively perform *alternating optimizations* (AO) over the image f and the parameters μ_i by substituting the following closed-form expression in the steepest descent formulation:

$$\begin{aligned} \nabla \Psi(\hat{f}^{(j+1)}(\mathbf{x})) &= \left([\mathbf{H}^T \mathbf{D}^T] \left([\mathbf{D}\mathbf{H}] \hat{f}^{(j)} - g \right) \right) (\mathbf{x}) + \\ &2\alpha_s \sum_{\mathbf{x}' \in \mathcal{N}_{k,k}(\mathbf{x})} L_{\mathbf{T}(\mathbf{x}),t} \left(\hat{f}^{(j)}(\mathbf{x}) - \hat{f}^{(j)}(\mathbf{x}') \right) + \\ &\alpha_m \Phi^{(j)}(\hat{f}^{(j)}(\mathbf{x})), \end{aligned} \quad (4.36)$$

and jointly updating the m parameters $\mu_i^{(j+1)}$ ($i = 1, \dots, m$) of the multimodal prior:

$$\mu_i^{(j+1)} = \frac{\sum_{\mathbf{x} \in \Omega(\hat{f}^{(j+1)})} \hat{f}^{(j+1)}(\mathbf{x}) r \left(\left\| \frac{\hat{f}^{(j+1)}(\mathbf{x}) - \mu_i^{(j)}}{h} \right\|^2 \right)}{\sum_{\mathbf{x} \in \Omega(\hat{f}^{(j+1)})} r \left(\left\| \frac{\hat{f}^{(j+1)}(\mathbf{x}) - \mu_i^{(j)}}{h} \right\|^2 \right)}, \quad (4.37)$$

where α_s and α_m are the regularization parameters, h represents the window bandwidth [Comaniciu and Meer, 2002] and the profile r is the Gaussian kernel given by equation (4.31). The regularization term $\Phi^{(j)}(\hat{f}^{(j)}(\mathbf{x}))$ is given by:

$$\Phi^{(j)}(x) = \sum_{i=1}^m 2 \left(x - \mu_i^{(j)} \right) \prod_{k=1; k \neq i}^m \left\| x - \mu_k^{(j)} \right\|_2^2. \quad (4.38)$$

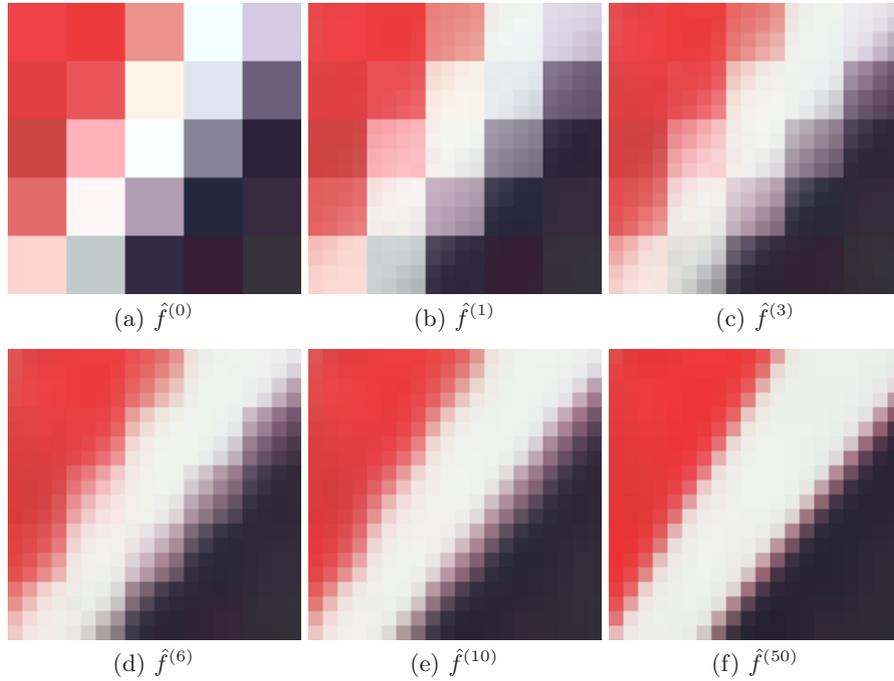


Figure 4.9: Iterative restoration example of a JPEG-compressed patch with 3 colour modes.

The parameters for the experiments are $\beta = 0.125$, $\tau = 10$, $h = 10$, $\alpha_s = 0.5$, $\alpha_m = (1/255^2)^{m-1}$ and the algorithm employs 100 iterations for the restoration process. The regularization parameters are again chosen according to the maximum size of the regularization terms. The iterative procedure is initialized with the nearest neighbour interpolation $\hat{f}^{(0)}$ and the parameters $\mu_i^{(0)}$ are retrieved from the three-steps algorithm described in Section 4.3.3.2. The evolution of an iterative image restoration process is shown in Figure 4.9.

4.5 Experimental results

We now present results, which demonstrate the effectiveness of our methods in upscaling images with a strong colour modality like cartoons, logos and maps, especially, when such images are compressed with JPEG or MPEG schemes (which are not optimally suited for this type of images). The colour priors are very powerful in removing introduced colour artefacts (e.g. due to mosquito noise).

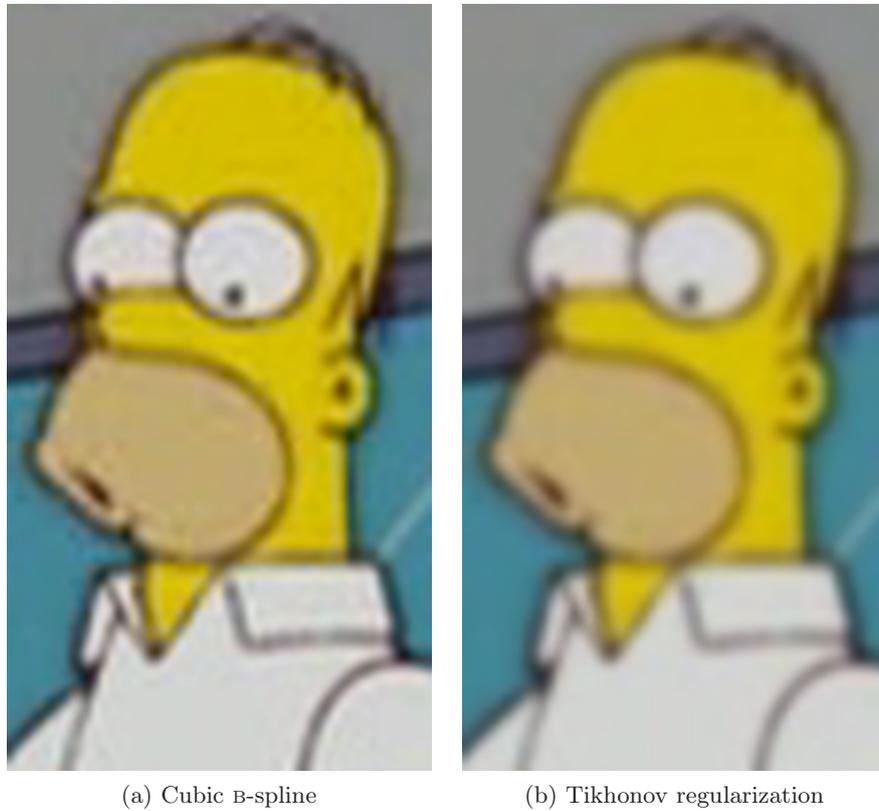


Figure 4.10: Image restoration results (4 \times enlargement of Figure 4.7).

4.5.1 Restoration with adaptive bimodal colour prior

In Figures 4.10, 4.11 and 4.12, we compare the proposed method for a magnification factor of 4 with the popular cubic B-spline interpolation, with Tikhonov regularization (i.e. the proposed method but without the adaptive bimodal colour prior), with total variation (TV) regularization (see Table 4.1), with bilateral total variation (BTV) regularization [Farsiu et al., 2004] (also given by equation (4.24)) and with the curvature preserving PDE's [Tschumperlé, 2006]. The original image is given in Figure 4.7. Note that the TV and BTV regularization schemes are initialized with a cubic B-spline interpolation result instead of with a nearest neighbour interpolated image, because these methods sharpen the edges in the image, but do not remove jagged edges. The cubic B-spline interpolation shows clearly the compression artefacts, which are largely removed by the restoration methods. At first glance, our method produces sharper images compared to the other methods, for example at the pupil of the eye.



Figure 4.11: Image restoration results (4× enlargement of Figure 4.7).

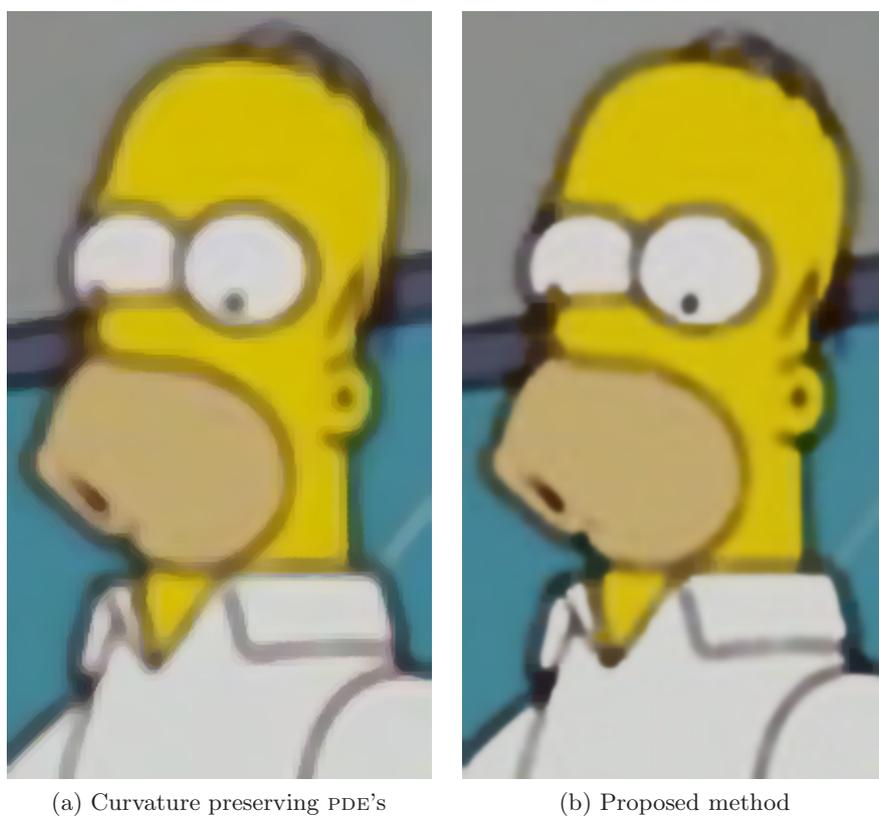


Figure 4.12: Image restoration results (4× enlargement of Figure 4.7).

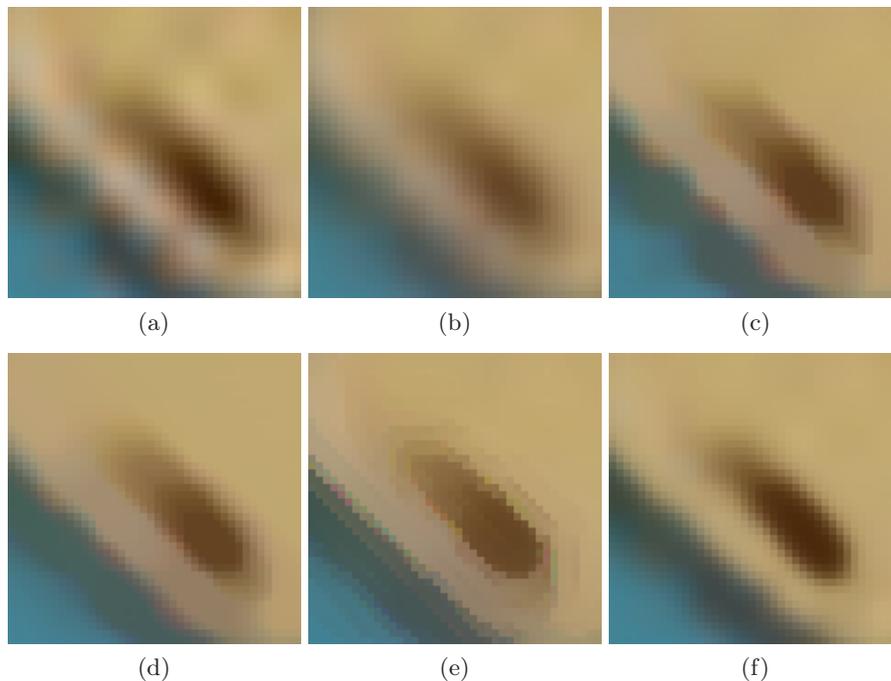


Figure 4.13: Detailed region of the mouth: (a) cubic B-spline, (b) Tikhonov regularization, (c) TV regularization, (d) BTV regularization, (e) curvature preserving PDE's [Tschumperlé, 2006] and (f) proposed method.

Figure 4.13 shows an enlarged region of interest in order to achieve a better visibility. We can observe that our method outperforms the other methods in visual quality. Noise and compression artefacts are heavily reduced compared to traditional interpolation techniques. Jagged edges are removed very well while TV regularization tends to preserve jaggedness. In our method, intensity staircase artefacts (i.e. piecewise constant regions) do not occur as opposed to the curvature preserved PDE method for example. The results of our method also contain less blur compared to other methods. Additionally, our method has a built-in anti-jaggies filter, thanks to the Tikhonov regularization, which is also visually more pleasant.

4.5.2 Restoration with multimodal colour prior

As a first experiment we have enlarged a JPEG-compressed image with a lot of mosquito noise as illustrated in Figure 4.14. We compare the proposed method using a magnification factor of 4 with nearest neighbour interpolation, with the popular cubic B-spline interpolation, with mean-shift postfiltering [Comaniciu and Meer, 2002], with BTV regularization [Farsiu et al., 2004] and with the curvature preserving PDE's [Tschumperlé, 2006]. The mean-shift

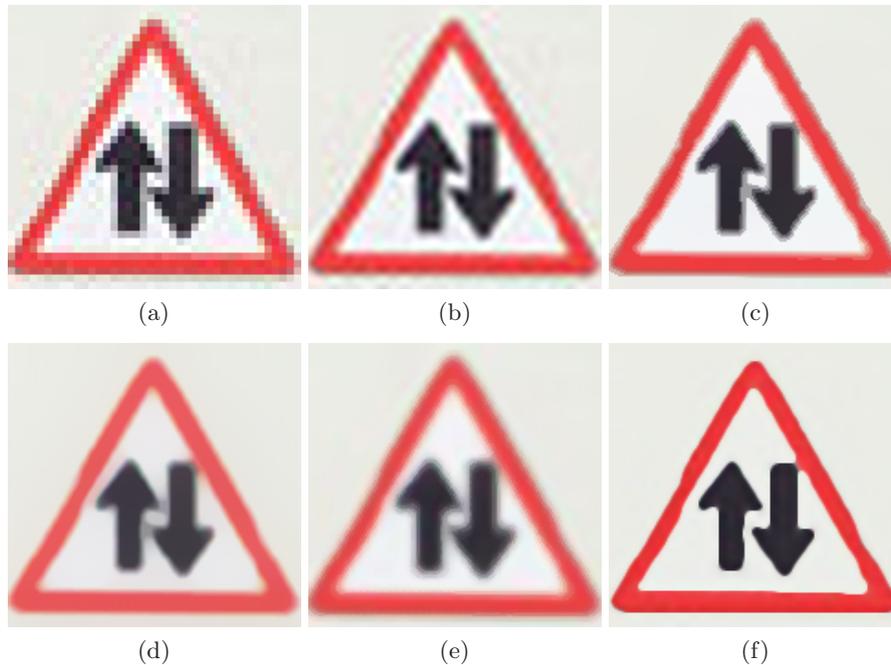


Figure 4.14: Restoration results of *traffic sign*: (a) nearest neighbour, (b) cubic B-spline, (c) mean-shift postfiltering [Comaniciu and Meer, 2002], (d) BTV regularization, (e) curvature preserving PDE's [Tschumperlé, 2006] and (f) proposed method.

and the BTV regularization schemes are initialized with the cubic B-spline interpolation.

We can clearly see that our method outperforms the other methods in visual quality. Noise and compression artefacts are totally removed compared to the traditional interpolation techniques. The mean-shift postfilter reduces noise and compression artefacts, but the result suffers from intensity staircase effects due to the transitional colours along the edges. BTV regularization and curvature preserving PDE's produce much more blur than our method.

In Figure 4.15 we have enlarged an image that is corrupted by colour quantization and error diffusion artefacts. We also compare the proposed method to the restoration algorithm with the anisotropic geometry-driven smoothing prior only (without the multimodal colour prior, thus $\alpha_m = 0$) and with the combination of the multimodal colour prior and the isotropic smoothing prior (this corresponds to the bilateral diffusion with L being an isotropic Gaussian kernel in equation (4.23)).

Our method delivers the best visual results concerning noise, compression arte-

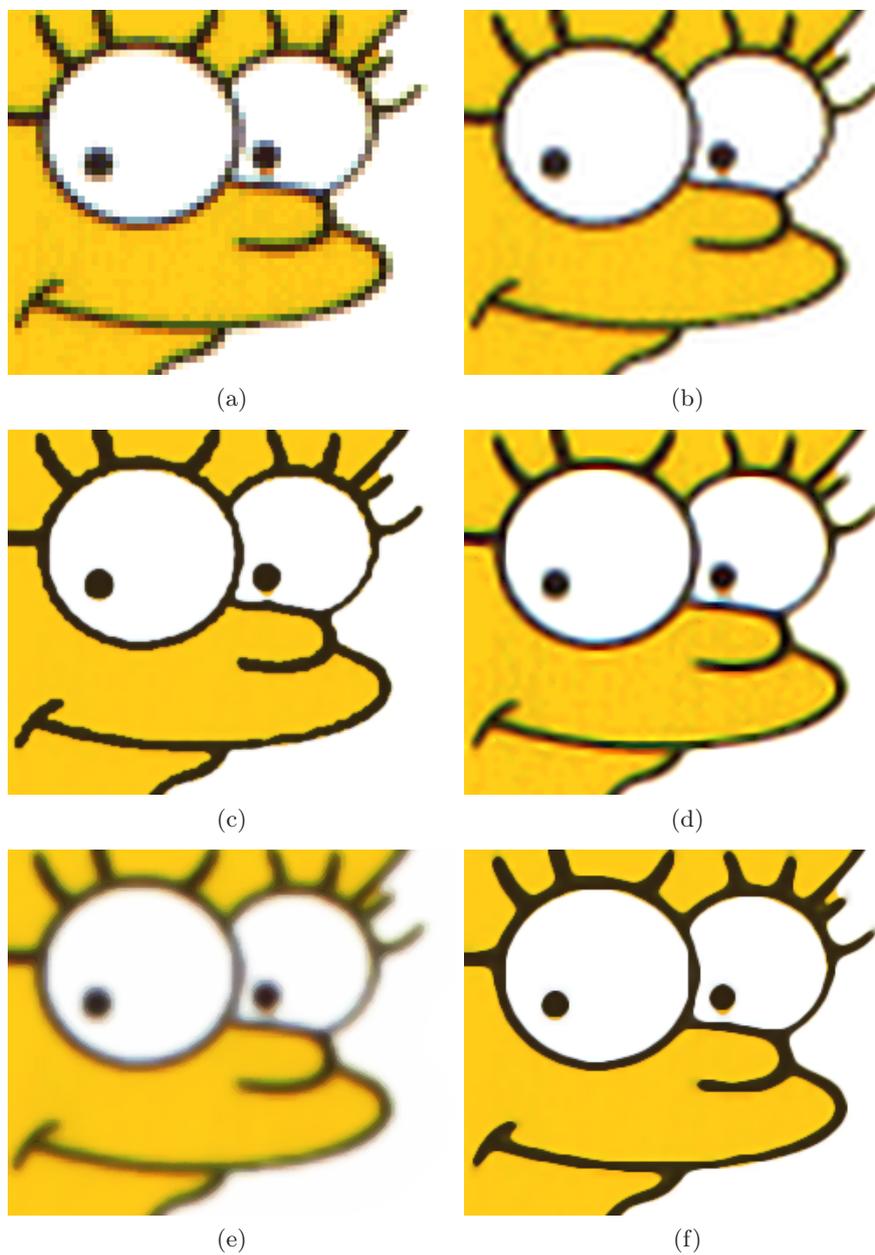


Figure 4.15: Restoration results of *Lisa*: (a) nearest neighbour, (b) cubic B-spline, (c) restoration with isotropic smoothing and multimodal colour priors, (d) restoration using geometry-driven smoothing without multimodal colour priors, (e) curvature preserving PDE's [Tschumperlé, 2006] and (f) proposed method.

facts and blur (see Figure 4.15). Jagged edges are removed very well by the geometry-driven smoothing, while isotropic regularization tends to retain some jaggedness. It is clear that the use of the proposed multimodal colour priors influences the end result positively: the image is visibly much sharper than the images produced by other methods and colour quantization artefacts are also removed. Intensity staircase artefacts (i.e. piecewise constant regions) do not occur in our method either as opposed to other methods, e.g. the curvature preserved PDE method.

4.6 Conclusion

In this chapter, we gave a broad overview of the existing adaptive image restoration methods where the focus was on the joint denoising and deconvolution problem. Among these numerous restoration methods, regularization techniques (e.g., the PDE-based methods, the deterministic set theoretic methods and the Bayesian methods) are very powerful because they are able to include prior knowledge about the degradation and the hypothetical ideal image. We derived the regularization techniques from a Bayesian point of view (based on the stochastic estimation theory and the MAP framework in particular) and discussed the image priors in more detail.

We derived the steepest descent algorithm for enlarging images with a comprehensive smoothness prior that takes the local geometry of edges into account, and we presented two new image priors, namely the adaptive bimodal colour prior and the multimodal colour prior. The adaptive bimodal colour prior expresses that the value of an edge pixel is a combination of the colours of two connected regions, each having a dominant colour distribution. The multimodal colour prior is proposed for images that normally just have a few dominant colours such as cartoons, logos, maps, etc.

Restoration results show the effectiveness and the visual superiority of our MAP scheme with the proposed colour priors to other interpolation/restoration schemes for images with a strong colour modality: noise and compression artefacts (like mosquito noise and colour quantization effects) are removed and our method produces less blur and other annoying artefacts (e.g. jagged edges, intensity staircase effects, etc.).

Future applications based on colour priors are for instance re-colourizing old movies from aging and deteriorated films, displaying compressed cartoon movies (e.g., from DVD's) on high-resolution screens such as HDTV and the enhancement of data coming from surveillance cameras (e.g. for reading license plates or recognizing a particularly logo on a criminal's sweater). Perhaps, one of the most interesting application today is the use of colour priors in the image demosaicing problem. Recently, in [Bennett et al., 2006], the authors employed a similar idea (each pixel is assumed to be a linear combination of two colours,

which are retrieved by k -means clustering) successfully to the demosaicing of Bayer CFA patterns: less colour fringing or false colours are introduced compared to traditional demosaicing techniques.

5

Non-local reconstruction methods

Quotes, damned quotes, and...
—John Bibby

Traditional image restoration methods produce a clean (HR) image from an observed degraded (LR) image following an acquisition or degradation model. Such a model describes how each output pixel is related to one or more input pixels, which is incorporated via the data fidelity term in the regularization framework as discussed in Chapter 4. Additionally, prior knowledge such as piecewise smoothness can be incorporated to improve the image restoration result. The impact of an observed pixel on the restored pixels is thus local according to the degradation model and the prior knowledge. So, the traditional methods only exploit the spatial redundancy in a local neighbourhood and are therefore referred to as *local* methods.

However, in this chapter, we demonstrate that the estimation of the restored pixel intensity can be based on information retrieved from the whole image, thereby exploiting the presence of similar patterns and features in the image, which we call *repetitive structures*. In the next section, we describe the relationship between this relatively new class of *non-local* methods and the exemplar- and fractal-based algorithms. In the rest of the chapter, we focus on the use of repetitive structures in image denoising (based on the *non-local means* algorithm as introduced by Buades et al. [Buades et al., 2005b, Buades et al., 2005a] and the improvement of this algorithm led to three conference publications [Dauwe et al., 2008, Goossens et al., 2008a, Goossens et al., 2008c]) and we introduce this non-local approach in image resolution enhancement, which is the main novelty. The blocks are selected based on the newly proposed dual matching criterion that enables a fast and robust filtering of these blocks. This work led to an ICIP publication [Luong et al., 2006c] and another publication in the Springer's Lecture Notes on Computer Science series [Luong et al., 2006b].



Figure 5.1: Example of self-similarity in nature (fern).

5.1 Repetitive structures, examples and self-similarity

Since the early 1980s, the idea of self-similarity plays an important role in mathematics and physics. Fractal-based methods suppose that many natural objects possess fractalness, i.e., parts of the image repeat themselves on an ever-diminishing scale, hence the term self-similarity. This process is illustrated in Figure 5.1.

From a historical point of view, most fractal-based methods focused on its compression capabilities, i.e., the image is approximated by a small group of image blocks provided at different scales, the so-called *domain pool* [Ebrahimi and Vrscay, 2008b]. A similar strategy can be found in block-based video compression schemes such as MPEG: frames can be predicted very well by taking advantage of the statistical property that image blocks are very similar to other image blocks from neighbouring frames. Thanks to the scale-invariance property of the fractal-based methods, self-similarity lends itself to image resolution enhancement applications (see also Section 3.1.2).

In practical applications, the self-similarity property is exploited via a contractive affine transformation of image blocks across different scales [Jacquin, 1992].

The affine transformation of image blocks is applied geometrically (i.e. spatially) as well as photometrically, i.e. in intensity/colour space.

Unlike fractal-based methods, non-local methods exploit the similarity of small patches at the same scale, without rotation or photometrical corrections. To avoid confusion, we will use the term *repetitivity* or *repetitive structures*. We can hypothetically assume that repetitive structures could serve as multiple noisy observations of the same structure. Results of our experiments in Section 5.4 will confirm that this hypothesis holds for real situations. Another class of methods that also takes advantage of repetitivity (but in time), is called *super-resolution* (SR) reconstruction. SR is a signal processing technique that obtains a HR image from multiple noisy and blurred LR images (see Chapter 7).

The concept of repetitive structures was successfully introduced in the image denoising field as the non-local means algorithm by Buades et al. [Buades et al., 2005b, Buades et al., 2005a] and was used for detecting digital image forgery [Fridrich et al., 2003, Li et al., 2008]. Besides repetitivity in texture, we can also find repetitivity in other parts of the image, some examples are illustrated in figure 5.2.

Closely related to the non-local method, is the training-based (or learning-based) approach, which restores images based on image blocks or *examples* that are retrieved from ideal images, and therefore, these techniques are also referred to as *exemplar-based* methods. Codebooks with examples are built by applying the degradation model on these ideal (degradation-free) images. These image blocks and their ideal counterparts are then used to guide the restoration process. The exemplar-based approach was successfully used in image denoising [Aharon et al., 2006, Elad and Aharon, 2006], image deblurring [Nakagaki and Katsaggelos, 2003], image interpolation or super-resolution [Freeman et al., 2002, Datsenko and Elad, 2007].

As discussed in Section 3.1.2, the performance of these exemplar-based algorithms depends heavily on the given training data, which can provide insufficient or wrong lookup results, and structural artefacts become more visible and disturbing because they are treated as part of the image. From a Bayesian point of view, there is also an important difference between non-local and exemplar-based methods: in non-local methods, similar patches provide information about the likelihood function for a given patch while in exemplar-based methods, similar *training* patches are assumed to be ideal, which also tells us something about the prior information.

5.2 Non-local image denoising

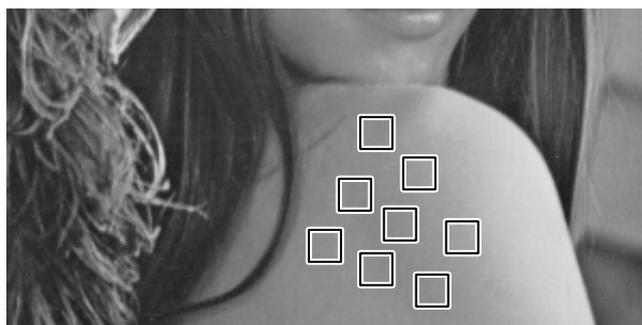
In the classic image denoising problem, the objective is to design an algorithm that can remove additive white stationary Gaussian noise, while preserving the original image details and fine structures and thereby having a high visual



(a) Repetition in different objects.



(b) Repetition along edges.



(c) Repetition in uniform areas.

Figure 5.2: Illustration of repetitive structures in images.

quality. To this end, the original non-local means algorithm as introduced by Buades et al. [Buades et al., 2005b, Buades et al., 2005a] is quite intuitive and potentially very powerful in removing noise with state-of-the-art denoising results as illustrated in Figure 5.3. However, the algorithm is computationally impractical due to the enormous number of weight calculations between image blocks.

In the next sections, we describe briefly the original non-local means algorithm and we discuss and point out some directions for improvements.

5.2.1 The original non-local means algorithm

We assume that the image is corrupted by zero-mean white stationary Gaussian noise. The denoised value $\hat{f}(\mathbf{x})$ of the pixel intensity at position \mathbf{x} is computed as the weighted average of all pixels in the image, which can be seen as a linear spatially adaptive filter:

$$\hat{f}(\mathbf{x}) = \frac{\sum_{\mathbf{x}' \in \Omega(g)} w(\mathbf{x}, \mathbf{x}') g(\mathbf{x}')}{\sum_{\mathbf{x}' \in \Omega(g)} w(\mathbf{x}, \mathbf{x}')} \quad (5.1)$$

where $g(\mathbf{x})$ denotes the intensity of the noisy pixel at position \mathbf{x} and $\Omega(g)$ is the domain of the image g . We will refer to this filter as the *pixel*-based non-local means. Alternatively, a *block*-based non-local means filter also exists [Buades et al., 2005b].

The weights $w(\mathbf{x}, \mathbf{x}')$ depend on the image content. Specifically, they depend on the similarity between the neighbourhoods $\mathbf{N}(\mathbf{x})$ and $\mathbf{N}(\mathbf{x}')$ of the image pixels at positions \mathbf{x} and \mathbf{x}' . Typically, squared neighbourhoods of fixed predefined size are used (e.g., 7×7 or 9×9). The similarity between neighbourhoods is computed by means of the *mean squared deviation* (MSD) $\|\mathbf{N}(\mathbf{x}) - \mathbf{N}(\mathbf{x}')\|^2$, i.e. the Euclidean distance between the two image patches. In the original non-local means algorithm [Buades et al., 2005b, Buades et al., 2005a], the weighting function is defined as follows:

$$w(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{N}(\mathbf{x}) - \mathbf{N}(\mathbf{x}')\|^2}{h^2}\right), \quad (5.2)$$

where h is a constant, proportional to the noise variance σ_n^2 . The weights $w(\mathbf{x}, \mathbf{x}')$ decay at an exponential rate, which results in large weights for a small MSD (similar windows) and small weights for a large MSD (non-similar windows).

Note that the popular bilateral filter [Tomasi and Manduchi, 1998], which is an iterative *local* method, is closely related to the non-local means filter. For

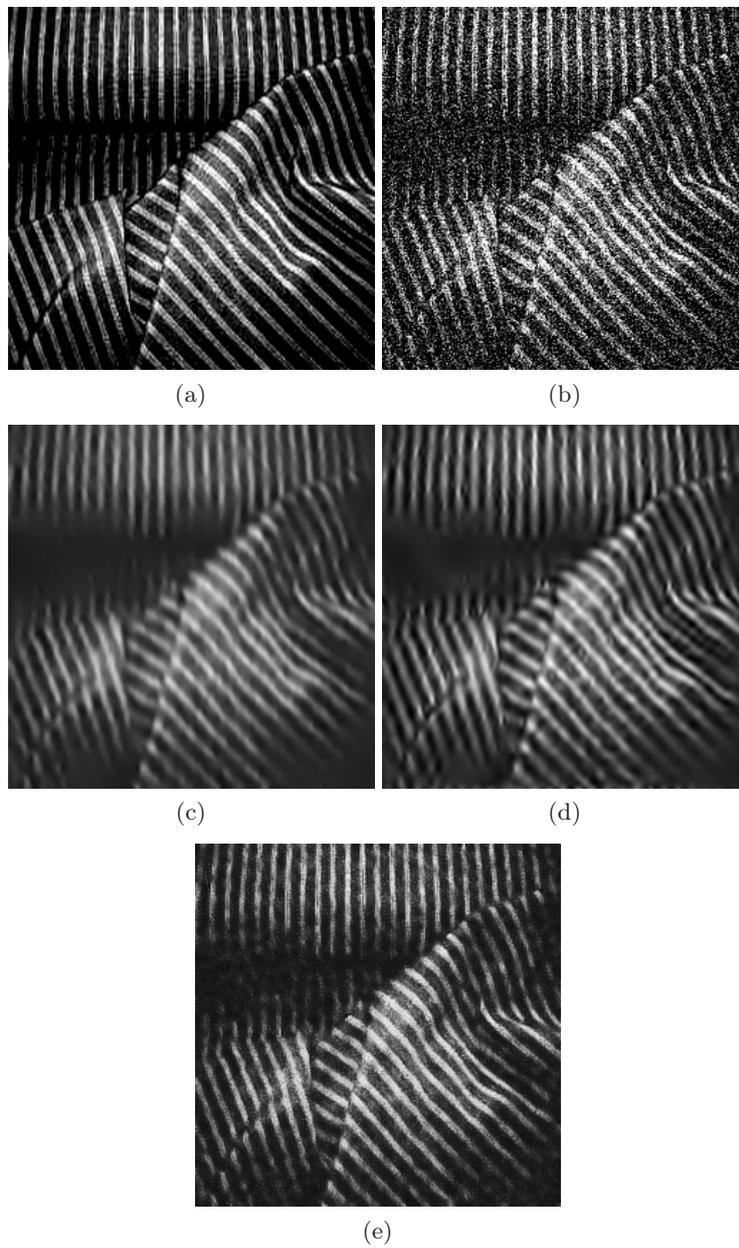


Figure 5.3: Denoising results of a *texture* image: (a) original 256×256 texture image, (b) image corrupted with additive white Gaussian noise ($\sigma_n = 70$), (c) Portilla et al. [Portilla et al., 2003] (PSNR: 16.20 dB), (d) Pižurica et al. [Pižurica and Philips, 2006] (PSNR: 16.99 dB) and (e) non-local means algorithm by Dauwe et al. [Dauwe et al., 2008] (PSNR: 18.36 dB).

the bilateral filter, the weighting function is given by:

$$w(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|g(\mathbf{x}) - g(\mathbf{x}')\|^2}{h_1^2}\right) \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{h_2^2}\right), \quad (5.3)$$

where the first factor (called the photometric distance) is inversely proportional to the Euclidean distance between the pixel intensities $g(\mathbf{x})$ and $g(\mathbf{x}')$ and the second factor (called the geometric distance) measures the Euclidean distance between the positions \mathbf{x} and \mathbf{x}' .

Many proposed denoising filters in the literature are closely related to this concept. For example, the Yaroslavsky neighbourhood filter sets the geometric distance factor to 1 if the geometric distance $\|\mathbf{x} - \mathbf{x}'\|^2$ is within a predefined radius and 0 otherwise [Yaroslavsky, 1985]. Other well-known *local* variants are Lee's σ -filter [Lee, 1983], adaptive smoothing [Saint-Marc et al., 1991], *smallest univalue segment assimilating nucleus* (SUSAN) [Smith and Brady, 1997], trilateral filter [Choudhury and Tumblin, 2003], etc. A concise overview and their mutual relations can be found in [Kervrann and Boulanger, 2006].

The weighting function of the non-local means algorithm can be interpreted as a vector-extension of the weighting function of the bilateral filter, omitting the geometric distance factor (actually, this factor distinguishes *local* from *non-local* methods). Note that both weighting functions as given in equations (5.2) and (5.3) are chosen heuristically and it is thus not guaranteed that these choices are optimal for a given criterion.

5.2.2 On the improvement of non-local means

Even though the original non-local means method is quite intuitive and potentially very powerful, the PSNR and visual results are somewhat inferior to other recently proposed state-of-the-art non-local algorithms, such as BM-3D as proposed by Dabov et al. [Dabov et al., 2007]. However, the main drawback still is the gigantic computation time due to the excessive amount of evaluations of the weighting function given by equation (5.2). The complexity of the algorithm is $\mathcal{O}(n^2)$ with n the total number of pixels in the image. This follows directly from the fact that n weights have to be computed for every pixel in the image. Just to give some idea about the computation time: denoising a single 512×512 image on a P4 3.0 GHz with 1024 Mb RAM lasts about 5 hours and 20 minutes. In the next sections, we point out several possible improvements on the non-local means algorithm in terms of computational complexity and restoration quality.

5.2.2.1 Accelerating the non-local means filter

Several papers in the literature are dedicated to the acceleration of the non-local means algorithm. For example, Mahmoudi and Sapiro reduced the original

quadratic computational complexity to a linear one by putting a hard limit on the number of contributing weights and by preclassifying neighbourhoods based on average greyvalues and gradients [Mahmoudi and Sapiro, 2005], while Coupé et al. used the mean and the variance of the neighbourhood as the preclassification conditions in a parallel computation framework [Coupé et al., 2008]. Orchard et al. quickly ruled out dissimilar neighbourhoods via *singular value decomposition* (SVD) [Orchard et al., 2008]. Brox et al. arranged image data in a cluster tree, which allows for a fast and accurate preselection of similar neighbourhoods [Brox et al., 2008]. Wang et al. replaced the MSD calculations with an efficient summed square image scheme using fast Fourier transform in a limited search window [Wang et al., 2006]. Using a limited search window centered around the processed pixel can indeed accelerate the denoising process enormously, however, in some applications we wish to exploit the full search space, i.e. the whole image or video sequence. Similar objects or texture can be located anywhere in the image after all. Additionally, the non-local property of the algorithm partially vanishes due to the limited search window and is so related back to local denoising methods.

In [Dauwe et al., 2008] and [Goossens et al., 2008a], we presented several improvements for speeding up the non-local means filter. In [Dauwe et al., 2008], a preclassification (based on the first three statistical moments, namely the mean, the variance and the skewness) is used to ignore the very small weights from dissimilar neighbourhoods. To decide whether a window is similar or dissimilar, we derive thresholds for images that are corrupted with additive white Gaussian noise. Further preclassification using fourth and higher order statistical moments is not very beneficial in most cases, at the contrary, it works even counterproductive: the denoising time increases because the computational cost of further checks exceeds the gain of the excluded weight computations.

Since most distance measures between neighbourhoods, such as the MSD, are symmetrical, we can reduce the computation time by approximately a factor 2 by exploiting the fact that weighting functions are also symmetrical (i.e. $w(\mathbf{x}, \mathbf{x}') = w(\mathbf{x}', \mathbf{x})$). We first initialize the accumulated weight matrix and the accumulated contribution matrix with zeros. Both matrices have the same size as the input image and actually represent the numerator (i.e. the accumulated contribution matrix) and the denominator (i.e. the accumulated weight matrix) of equation (5.1). When processing a pixel \mathbf{x} with the contribution of a pixel \mathbf{x}' , we add the products $w(\mathbf{x}, \mathbf{x}')g(\mathbf{x}')$ and $w(\mathbf{x}, \mathbf{x}')g(\mathbf{x})$ to the accumulated contribution matrix at the pixel positions \mathbf{x} and \mathbf{x}' respectively. The accumulated weight matrix is also updated at the same pixel positions with $w(\mathbf{x}, \mathbf{x}')$. As a result, we only need to compute weights of neighbourhoods for $\mathbf{x}' > \mathbf{x}$. Note that this does not mean that the accumulated weights on positions \mathbf{x}' and \mathbf{x} are equal, thus updating both accumulated matrices on positions \mathbf{x}' and \mathbf{x} is still needed. Finally, we normalize the accumulated contribution

matrix via element-wise division by the accumulated weight matrix, in order to obtain the estimated image.

The Euclidean distance between different neighbourhoods $\|\mathbf{N}(\mathbf{x}) - \mathbf{N}(\mathbf{x}')\|^2$ can be computed very efficiently across the whole image using a moving average filter (with $\mathbf{x} - \mathbf{x}'$ being a constant position difference each time) [Goossens et al., 2008a]. Note that the same strategy also affects the neighbourhood preclassification step: we do not need to compare the already processed neighbourhood pairs because of the symmetry of the statistical features.

Evaluating the exponential function in equation (5.2) is still a time consuming task for a CPU. Since the bandwidth parameter h is constant, these exponential functions are evaluated in advance and stored in an efficient lookup table. In this way, the number of evaluations is fixed. The relative gain in computation time can be tremendously large as the image size (n pixels) increases, which normally needs $n^2/2$ evaluations. Note that the time needed for looking up a weight is much shorter than the actual weight computation, but still has to be considered in the evaluation of the total computation time.

The total gain in computation time depends heavily on the degree of repetitiveness in the image. For example, experimental results in [Dauwe et al., 2008] show acceleration factors from 10 (for texture images with a lot of repetitiveness) to 34 (for real world images) compared to the original method.

5.2.2.2 Enhancing the visual quality

In the last few years, many variants and several improvements on the non-local means algorithm have been proposed in the literature. Kervrann and Boulanger improved the non-local means filter by introducing adaptive local neighbourhoods as well as a statistical estimation of the model parameters [Kervrann and Boulanger, 2006]. Brox et al. suggested an iterative version of the filter, which is derived from a variational principle [Brox et al., 2008]. In [Chatterjee and Milanfar, 2008], the authors have extended the existing non-local means to higher orders of (kernel) regression, which allows more flexibility to the local image data model [Chatterjee and Milanfar, 2008]. Azzabou et al. have investigated in better similarity measures by building a compact dictionary in order to compare similar neighbourhoods using *principal component analysis* (PCA) [Azzabou et al., 2007]. Kervrann et al. have improved the non-local means by introducing adaptive local dictionaries and a new statistical distance measure to compare neighbourhoods [Kervrann et al., 2007]. In [Zimmer et al., 2008], the authors have proposed a rotationally invariant similarity measure.

Dabov et al. use block matching, group similar blocks in a 3D stack and then filter these blocks in 3D (e.g. shrinking the wavelet coefficients), hence the name BM-3D [Dabov et al., 2007]. Once these blocks are filtered, the authors again apply block matching on the denoised image and use the motion vectors

to group similar blocks of the noisy image in a 3D stack and then filter these blocks with a collaborative Wiener filtering. This method applies *re-estimation*, which uses the result of an estimation method as the input data for a different method with the same estimation purpose [Guerrero-Colón et al., 2008]. In [Dabov et al., 2008b], the authors combine the ideas of BM-3D with the pointwise shape-adaptive DCT (SA-DCT) filter, which adapts the neighbourhoods to the image content [Foi et al., 2007].

An improvement of image quality towards the original algorithm is to ignore the contributions from too dissimilar neighbourhoods. Even though their weights are very small anyway, the newly estimated pixel value can be severely biased because there are many dissimilar neighbourhoods. This bad influence of dissimilar windows can be eliminated by setting their corresponding weights to zero. So, preclassification helps both in accelerating the algorithm and in enhancing the visual quality. Experimental results on natural 512×512 images corrupted with additive zero-mean white Gaussian noise ($\sigma_n = 20$) give a moderate gain of 1.44 dB in PSNR [Dauwe et al., 2008].

In analogy to [Elad, 2002], we showed that the non-local means algorithm is basically the first iteration of the Jacobi optimization algorithm for robustly estimating the noise-free image [Goossens et al., 2008a]. It turns out that the weighting function in equation (5.2) appears to be the Le Clerc robust loss function. Substantial improvements can be obtained by replacing the weighting function with other robust loss functions, such as the modified bisquare function, and by applying multiple iterations according to the Jacobi algorithm. Note that the iterative version can be speeded up by using limited surrounding search windows without sacrificing the *non-local* property (the *effective* search window becomes larger after each iteration).

In some circumstances (e.g. in case of non-repetitive structures), one iteration of the non-local means filter may not remove all of the noise. Clearly, the noise variance in the processed image becomes dependent on the position in the image, which means that we are dealing with *non-stationary* noise. Therefore, a postfilter, based on a locally adaptive basis of principal components, is applied after each iteration to remove the remaining noise [Goossens et al., 2008a]. For white noise, PSNR results show that the non-local method with all improvements is very competitive with the state-of-the-art BM-3D method [Dabov et al., 2007], while the visual quality of our method is better due to the lower number of artefacts.

In [Goossens et al., 2008a], we also extended the non-local means algorithm to treat coloured (correlated) noise. This is efficiently implemented via a pre-whitening filter. Experimental results show that this extension outperforms state-of-the-art wavelet techniques for coloured noise, such as BLS-GSM [Portilla et al., 2003] and MP-GSM [Goossens et al., 2007], both visually as in terms of PSNR.

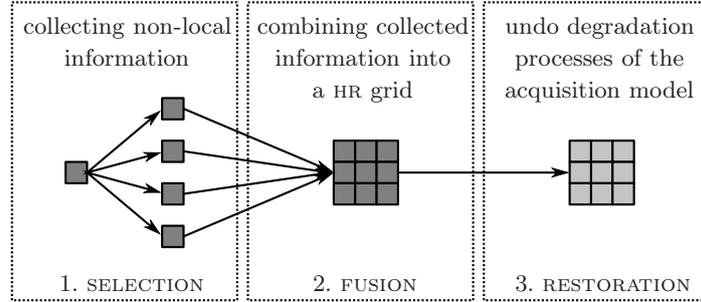


Figure 5.4: Block diagram of the proposed non-local resolution enhancement method.

5.3 Non-local resolution enhancement

We propose a novel image upscaling approach that exploits the repetitive behaviour of small structures in an image. Our resolution enhancement scheme is conceptually simple and consists of three consecutive steps as illustrated in Figure 5.4:

1. Selection of similar patches and subpixel registration of these repetitive structures on the HR grid.
2. Robust data fusion, i.e. the aggregation of all registered structures into one HR image.
3. Image restoration (i.e. denoising and deblurring, see also the more detailed discussion in Chapter 4).

In the next sections we treat each component in more detail.

5.3.1 Selection and registration of repetitive structures

For the sake of simplicity, we define small rectangular windows as basic structure elements. The image is divided in half overlapping reference blocks B_{ref} . Two matching criteria are used in our algorithm to find similar windows or blocks B across the whole image for each reference block, namely the *zero-mean normalized cross correlation* (CC):

$$E_{\text{CC}} = \frac{\sum_{\mathbf{x} \in \Omega(B)} (B(\mathbf{x}) - \bar{B})(B_{\text{ref}}(\mathbf{x}) - \bar{B}_{\text{ref}})}{\sqrt{\sum_{\mathbf{x} \in \Omega(B)} (B(\mathbf{x}) - \bar{B})^2 \sum_{\mathbf{x} \in \Omega(B)} (B_{\text{ref}}(\mathbf{x}) - \bar{B}_{\text{ref}})^2}} \quad (5.4)$$

and the *mean absolute difference* (MAD):

$$E_{\text{MAD}} = \frac{1}{\kappa(\Omega(B))} \sum_{\mathbf{x} \in \Omega(B)} \|B(\mathbf{x}) - B_{\text{ref}}(\mathbf{x})\|_1 \quad (5.5)$$

where $\Omega(B)$ contains all the pixels of the window B and $\kappa(\Omega(B))$ is the cardinality (i.e. the number of pixels) of $\Omega(B)$. \overline{B} and $\overline{B_{\text{ref}}}$ are denoted as the mean values of respectively B and B_{ref} .

The main motive to use both CC and MAD criteria is because they are somewhat complementary: CC is a measure for the similarity of the structural or geometrical content (however, CC can not discriminate between “light” and “dark” windows with the same structural information, e.g. edges), while MAD tells us more about the luminance and colour differences (however, MAD can not discriminate between windows that have a small difference over all pixels and windows that have a large difference in a few (edge) pixels). A matched window is accepted if the two measures E_{CC} and E_{MAD} satisfy to the respective thresholds τ_{CC} and τ_{MAD} , more specifically: $E_{\text{CC}} > \tau_{\text{CC}}$ and $E_{\text{MAD}} < \tau_{\text{MAD}}$. Because we are only interested in positive correlation, E_{CC} must be between 0.0 and 1.0. The choice of τ_{CC} and τ_{MAD} depends on the noise level (e.g. due to additive noise or due to artefacts by DCT-based compressed schemes such as JPEG): the higher the noise variance, the lower τ_{CC} and the higher τ_{MAD} must be chosen in order to be capable to find the repetitive structures.

To simplify the matching problem and in particular to save computation time, we assume that we are only dealing with pure translational shifts of the windows B (extension to rotations or affine transformations is trivial, but at the expense of computational complexity).

Our current implementation uses a simple exhaustive search in order to find the matching windows, but more intelligent (pattern-based) search algorithms could reduce the computation time enormously. Some possible acceleration strategies are already discussed in Section 5.2.2.1 and for document processing applications, character segmentation eliminates a lot of unnecessary calculations, both in search space and *region of interest* (ROI) selection (see Chapter 6).

Common ways to achieve subpixel accuracy in the image registration, is to interpolate either the image data or the correlation data. For a more detailed treatment of subpixel registration, we refer the reader to Section 7.2. We first interpolate the image by the level curve mapping algorithm as described in Chapter 3. We then estimate the subpixel shifts in a fast way by minimizing the MAD criterion between the interpolated blocks. However, the accuracy of these shifts are limited to $1/r$ with r being the magnification factor.

5.3.2 Robust data fusion

In the data fusion step, we determine an initial pixel value for each pixel position on the HR grid, which results in the *fused* image. After subpixel registration, we assign each pixel of the repetitive structures to a HR grid point using the up-sampling operator as illustrated in Figure 4.8. The whole process is illustrated in Figure 5.5. For each point on the HR grid, we now consider the following cases:

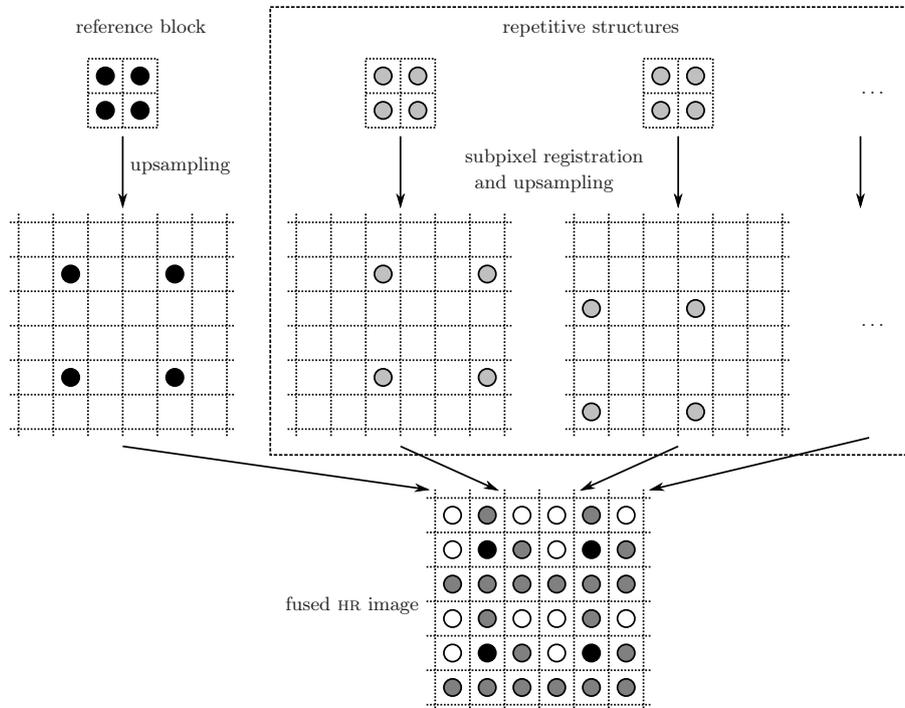


Figure 5.5: Robust data fusion: the fused HR image contains pixels from the LR image (class (OR) as indicated by black circles), unknown pixels (class (UN) as indicated by white circles) and fused pixels (class (FU) as indicated by grey circles).

Fixing the original pixels: the original pixels of the LR image are unchanged on the HR grid, because we assume that the noise variances of these pixel values are lower than those of the similar structures (because they contain noise *and* errors due to wrong matches and misregistration).

Several pixels are available: starting from the maximum likelihood principle, it can be shown that minimizing the l_1 -norm of the residuals is equivalent to median estimation [Farsiu et al., 2004]. A residual is the difference between an observed pixel value and the predicted pixel value. The median is very robust against outliers, such as the combination of noise and errors due to mismatching and misregistration. For this reason the fused HR pixel is computed as the median of all observed pixels on the HR grid point.

No pixel is available: these unknown HR pixel values are initialized with pixel values given by the level curve mapping algorithm. We do not need additional computations since the interpolated image is already constructed for the registration step as mentioned in the previous section.

In a nutshell, the HR grid contains pixels of three classes: the original pixels (OR), the unknown pixels (UN) and the fused pixels (FU). The last mentioned class provides the extra information for the resolution enhancement process as opposed to conventional upscaling techniques.

5.3.3 Joint denoising and deblurring

In the image restoration step, we perform joint denoising and deblurring on the fused HR image. In Chapter 4, we have treated joint image restoration and resolution enhancement from a Bayesian point of view. In this section, we derive the steepest descent solution of a robust regularization that performs image deblurring.

We assume that the blur in the acquisition model is fully determined by a shift-invariant PSF kernel H (typically Gaussian blur, which is characterized by its standard deviation σ_b). This blur is simulated by a linear degradation operator \mathbf{A} . The inverse problem becomes highly unstable in the presence of noise. This can be solved by imposing some prior knowledge about the image. Typically we will try to force spatial smoothness in the desired HR solution. This is usually implemented as a penalty factor in the generalized minimization cost function:

$$\hat{f}(\mathbf{x}) = \arg \min_{f(\mathbf{x})} \left(\sum_{\mathbf{x}' \in \mathfrak{N}_4(\mathbf{x})} \rho_R(f(\mathbf{x}) - f(\mathbf{x}')) + \alpha(\mathbf{x}) \rho_D([\mathbf{A}]f - g)(\mathbf{x}) \right), \quad (5.6)$$

where \mathfrak{N}_4 denotes the 4-connectivity neighbourhood (illustrated in Figure 4.3) and $\alpha(\mathbf{x})$ is the regularization parameter between the two terms, respectively called the regularization term ρ_R and the data fidelity term ρ_D . Both ρ -functions can be chosen from the family of M-estimator functions (given in Table 4.1). Image $g(\mathbf{x})$ is the HR fused image obtained in Section 5.3.2.

The minimization of the cost function (5.6) is solved using the steepest descent algorithm described in Section 4.4.2. The steepest descent direction is given by the gradient of the cost function $\nabla \Psi$:

$$\begin{aligned} \nabla \Psi(\hat{f}^{(j+1)}(\mathbf{x})) &= \sum_{\mathbf{x}' \in \mathfrak{N}_4(\mathbf{x})} \psi_R(f^{(j)}(\mathbf{x}) - f^{(j)}(\mathbf{x}')) + \\ &\quad \alpha(\mathbf{x}) \left([\mathbf{A}^T] \psi_D([\mathbf{A}]f^{(j)} - g) \right)(\mathbf{x}), \end{aligned} \quad (5.7)$$

where ψ_D computes ψ_D elementwise (i.e. on each element of the matrix) and both ψ -functions are given in Table 4.1. The operator $[\mathbf{A}^T]$ performs a convolution with a flipped PSF kernel (see Section 4.4.2.1).

As discussed in Section 4.3, the use of the so-called edge-stopping functions in the regularization term is very popular because it suppresses the noise better

while retaining important edge information [Pižurica et al., 2005]. Therefore, we apply one of the most successful edge-preserving regularization terms proposed for image denoising, namely the *total variation* (TV):

$$\rho_R(f(\mathbf{x}) - f(\mathbf{x}')) = \|f(\mathbf{x}) - f(\mathbf{x}')\|_1. \quad (5.8)$$

We assume that the noise in the fused HR image is modelled by a Laplacian PDF due to the inaccuracy of the image registration, blur, additive noise and other kinds of error that are not explicitly modelled [Farsiu et al., 2004]. The maximum likelihood estimate of data in the presence of Laplace noise is obtained through the l_1 -norm minimization. That is why we use the l_1 -norm function for the data fidelity term:

$$\rho_D(H * f(\mathbf{x}) - g(\mathbf{x})) = \|H * f(\mathbf{x}) - g(\mathbf{x})\|_1. \quad (5.9)$$

The corresponding ψ -functions are very computationally efficient because they simply compute the sign of their argument (see Table 4.1).

We adapt the regularization parameter $\alpha(\mathbf{x})$ in equation (5.7) locally to the different kind of pixels on the HR grid:

- Class OR: because we assume that these pixels contain the least noise as discussed in Section 5.3.2, very little regularization has to be applied. This means that these pixels mainly depend on the data fidelity term and thus $\alpha(\mathbf{x})$ is set to α_{MAX} .
- Class UN: these pixels are most likely noise. The optimization depends thus only on the regularization term, i.e. $\alpha(\mathbf{x}) = 0$.
- Class FU: these pixels contain both noise (and errors) and relevant information. If we have a lot of overlapping pixels on the HR grid point, the initial estimation is then more robust against errors such as mismatches and misregistration. Therefore, we apply the regularization with $\alpha(\mathbf{x})$ linearly proportional to the number of overlapping pixels and $\alpha(\mathbf{x}) \in [0, \alpha_{\text{MAX}}]$.

5.4 Experimental results

As a proof of concept, we demonstrate the power of the proposed non-local resolution enhancement technique with regular images and text images.

5.4.1 Resolution enhancement of regular images

As a first experiment, we enlarge a region of the original image, given in Figure 5.6, with a magnification factor of 8. As we can see in the image, there are lots of repetitive structures present such as the wall, the fence, the sky, the grass (texture), etc. For the proposed non-local resolution enhancement technique, we use 5×5 basic structure elements. To decide whether a structure



Figure 5.6: Part of the *lighthouse* image.

is similar to the reference block, we use the threshold parameters $\tau_{CC} = 0.9$ and $\tau_{MAD} = 9.0$. With these settings we obtain the following partition of the different classes on the HR grid: 1.6% (OR), 30.9% (FU) and 67.5% (UN). In the image restoration step, we apply 100 iterations of the steepest descent algorithm and we employ the Gaussian blur kernel with $\sigma_b = 4.0$ and maximum regularization parameter $\alpha_{MAX} = 10$.

Figures 5.7 and 5.8 compare the results of the proposed non-local method to the linear interpolation methods nearest neighbour and cubic B-spline, the non-linear Aqua2 interpolation [Muresan, 2005] and iterated function systems (IFS) interpolation, which exploits the self-similarity of the image across scales [Gharavi-Alkhansari et al., 1997]. In Figure 5.8, we also show the fused image g (i.e. the noisy and blurred HR image after robust data fusion).

Table 5.1: Average PSNR results in dB of a $4\times$ upscaling experiment. The best PSNR result is written in bold letters.

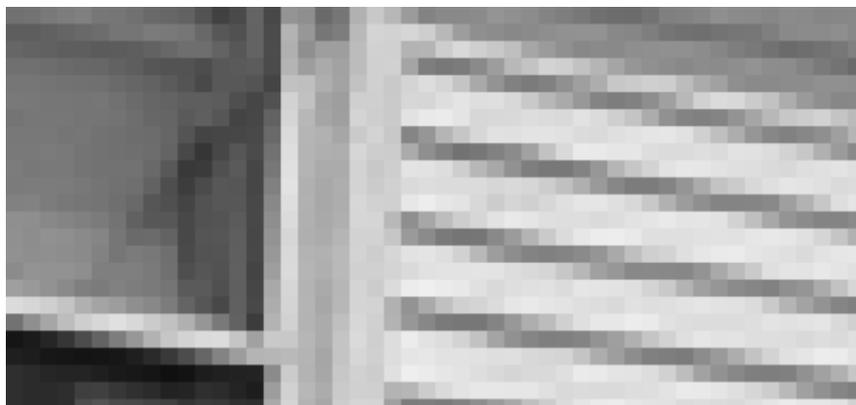
Upscaling method	PSNR
Iterated function systems	25.78
+ deconvolution	26.55
Aqua2 interpolation	26.95
+ deconvolution	27.27
Blackman-Harris windowed sinc	27.34
+ deconvolution	27.49
Level curve mapping interpolation	27.12
+ deconvolution	27.54
Proposed method (non-local)	27.61

We can notice significant improvements in visual quality in the result of our non-local method: there is a very good reconstruction of the edges and the result contains less annoying artefacts, such as jagged edges and blur. The result produced with our method is also better denoised while important edges are preserved.

As a second experiment, we create 10 synthetic 256×256 LR images from 1024×1024 images. The HR images are first convoluted with a Gaussian blur kernel ($\sigma_b = 1.0$) before the subsampling. We enlarge the LR images with a magnification factor of 4. For the proposed non-local method, we use 5×5 basic structure elements, and threshold parameters $\tau_{cc} = 0.9$ and $\tau_{mad} = 9.0$. For the image deconvolution, we apply 100 iterations of the steepest descent algorithm and we employ the Gaussian blur kernel with $\sigma_b = 1.0$ and maximum regularization parameter $\alpha_{max} = 10$.

In Table 5.1, we compare the average PSNR results of several interpolation methods. We also apply an additional image deconvolution step (with the same parameters) to compensate for the Gaussian blur in the LR images. We notice that the image deconvolution improves the performance of all interpolation methods. The PSNR results of our non-local resolution enhancement method is slightly better than those of the other methods.

Figures 5.10 and 5.11 show the enlarged results from the *pentagon* image (given in Figure 5.9). We compare the images of the proposed non-local method to those of the linear Blackman-Harris windowed sinc interpolation, the non-linear Aqua2 interpolation [Muresan, 2005] and the IFS interpolation [Gharavi-Alkhansari et al., 1997].



(a) Nearest neighbour



(b) Cubic B-spline



(c) Aqua2 interpolation

Figure 5.7: Image interpolation results (8× enlargement of Figure 5.6).



(a) Iterated function systems (self-similarity)



(b) Non-locally fused image g



(c) Proposed method (non-local)

Figure 5.8: Image restoration results ($8\times$ enlargement of Figure 5.6).



Figure 5.9: The *pentagon* image.

We notice that our non-local method produces less disturbing artefacts compared to the other methods and the straight edges in the image are reconstructed much better with the proposed method.

5.4.2 Resolution enhancement of text images

As a real-world experiment, we print and scan one A4 paper containing the *Lorem ipsum* text with the HP PSC 2175 machine at 75 dpi as shown in Figure 5.12. The Lorem ipsum text is very popular as default model text, but additionally it has a somewhat normal distribution of letters.¹ The scanned text image is compressed with standard JPEG and contains block artefacts as

¹We refer the reader to <http://www.lipsum.com> for more information about *Lorem ipsum*.



(a) Original HR image



(b) LR image

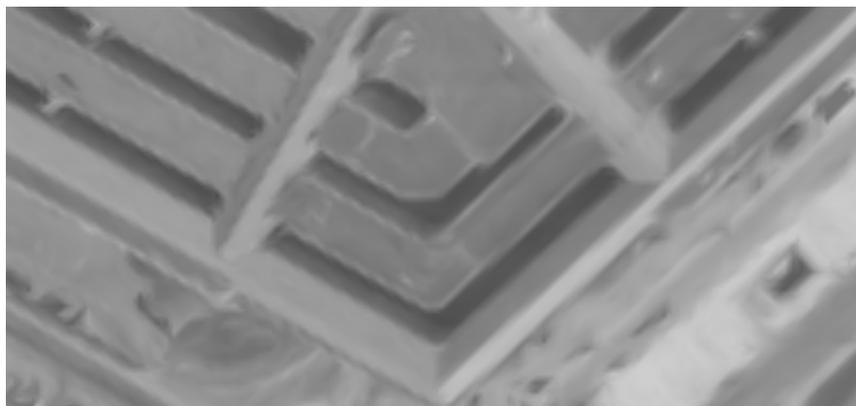


(c) Blackman-Harris windowed sinc

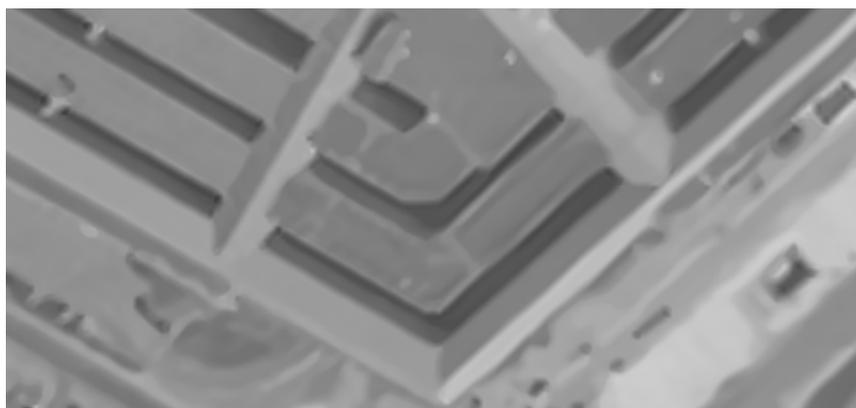
Figure 5.10: Image interpolation results (4× enlargement).



(a) Iterated function systems (self-similarity)



(b) Aqua2 interpolation



(c) Proposed method (non-local)

Figure 5.11: Image restoration results (4× enlargement).

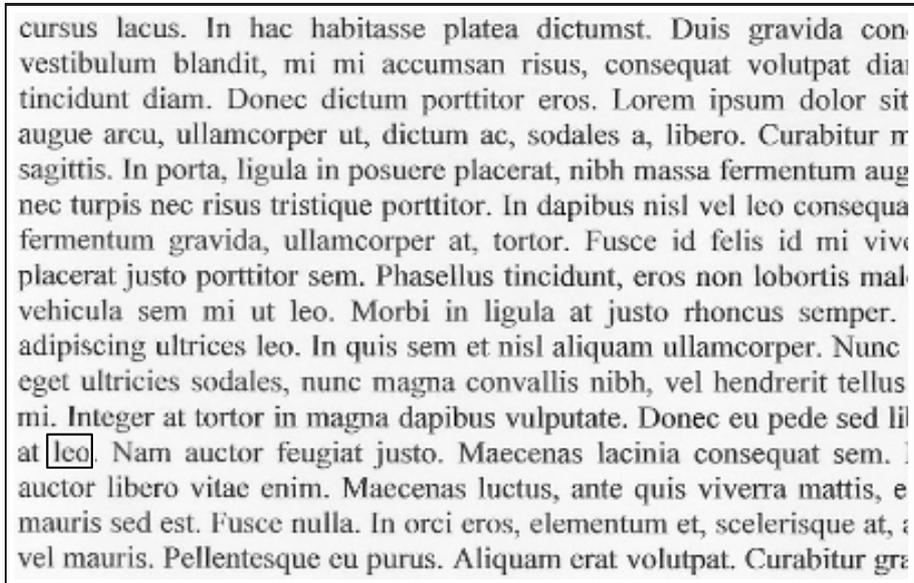


Figure 5.12: Part of the *lorem ipsum* text image.

a consequence.

We choose an 18×12 rectangular window as the basic structure element and we enlarge the ROI by a factor of 8. The parameters for our method are $\sigma_b = 4.0$, $\tau_{CC} = 0.6$, $\tau_{MAD} = 40.0$, $\alpha_{MAX} = 100$ and 100 iterations for the restoration process. The parameter selection was based on trial and error, i.e. to produce the visually most appealing results. We obtain the following partition of the HR grid: 1.6% (OR), 14.0% (FU) and 84.4% (UN).

In Figure 5.13, we compare our result with the nearest neighbour, the popular cubic B-spline and the non-linear level curve mapping interpolation techniques (see Chapters 2 and 3). We can clearly see that our method outperforms the traditional and non-linear interpolation techniques: the letters are much better readable and reconstructed (e.g. the character “e”), noise and JPEG-artefacts are heavily reduced and less blur, staircase and ringing artefacts are created. Due to the presence of aliasing, we are able to restore the true high resolution based on the multiple occurrences of the noisy blurred LR characters.

For comparison purposes, we scan the same text at 150 dpi and use that image as ground truth as shown in Figure 5.14 (we point out that the images in this figure are $4\times$ enlarged by the nearest neighbour interpolation in order to achieve better visibility). If we visually inspect the $2\times$ enlargement of the version scanned at 75 dpi, we can conclude that our method manages to reconstruct the characters much better than the cubic B-spline interpolation.

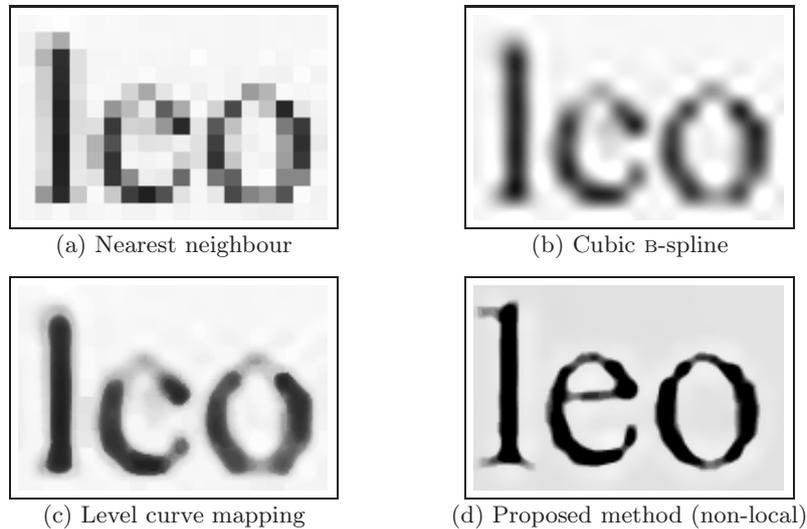


Figure 5.13: Image restoration results ($8\times$ enlargement of Figure 5.12).

We also compute the correlation coefficient between the restored image and the ground truth data after proper registration with translational shifts. Note that the ground truth image also contains noise and JPEG artefacts. The correlation coefficient as computed by equation (5.4) is 0.88 for the cubic B-spline and 0.95 for our non-local method, which is clearly better.

More non-local restoration results on text images are given in Chapter 6. In that chapter, we also describe a more efficient implementation of the non-local restoration technique for document processing.

5.5 Conclusion

In this chapter, we discussed the similarities and differences between the relatively new class of *non-local* methods and the exemplar- and fractal-based algorithms. Fractal-based methods exploit the self-similarity via a contractive affine transformation of image blocks across different scales, while non-local methods exploit shifted repetitive structures in the same scale. Exemplar-based methods build codebooks from ideal images in order to guide the restoration process. These ideal images contain inherently prior knowledge and therefore, they give more information about the posterior PDF, while exploiting repetitive structures gives more information about the likelihood PDF. Multi-frame super-resolution techniques form a special class of methods that also exploits repetitivity (across multiple images), we treat this topic in more detail in Chapter 7.

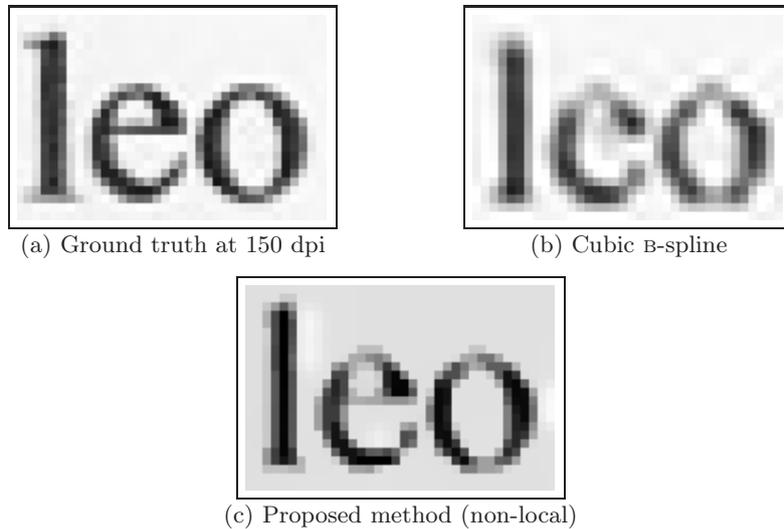


Figure 5.14: Image restoration results ($2\times$ enlargement of Figure 5.12).

We described the non-local means algorithm for image denoising as introduced by Buades et al. [Buades et al., 2005b, Buades et al., 2005a]. The algorithm is quite intuitive and potentially very powerful in removing noise with state-of-the-art denoising results, but is computationally impractical due to the enormous amount of weight calculations. We discussed numerous improvements both on accelerating the non-local means filter as well as on enhancing the visual quality in joint works as presented in [Dauwe et al., 2008, Goossens et al., 2008a].

We presented a novel resolution enhancement scheme based on the repetitive character of the image. Exploiting repetitivity brings more information at our disposal, which leads to much better estimates of the unknown pixel values. Results show the effectiveness of the proposed non-local upscaling technique and its superiority at very large magnifications to other resolution enhancement methods: edges are reconstructed well and artefacts are heavily reduced.

Future work includes for example employing more intelligent (pattern-based) search algorithms that reduce the computation time, replacing the hard thresholds for finding similar structures by soft thresholds that generate weights for the restoration process and combining the non-local strategy with self-similarity properties.

Because of the close relationship with super-resolution, we can also denote our non-local method as an intra-frame SR technique. Combinations of inter- and intra-frame SR techniques can help to improve SR in video applications when

very few LR images are available. Very recently, the interest for non-local resolution enhancement techniques in the image processing community has started to grow, see e.g. [Ebrahimi and Vrscay, 2007, Ebrahimi and Vrscay, 2008a, Danielyan et al., 2008a, Danielyan et al., 2008b, Protter et al., 2009].

Our method is very suitable to some applications such as gigantic satellite images in remote sensing applications (long roads and a lot of texture provide a huge amount of self-training data). In special applications with text images, we can achieve excellent results: characters could be made much better readable again. This could be very advantageous for *optical character recognition* (OCR) applications or when the image resolution can not be improved at the sensor because of technological limitations or because of high costs. A more efficient implementation for document image restoration is described in Chapter 6.

6

Application to document image processing

Divide et impera
—Phillippus II from Macedonia

In these days, improving image text resolution is important in many applications. Some examples are improving the readability (e.g. of license plates provided by surveillance cameras or for office automation) and simple spatial magnification (e.g. printing low-resolution documents on high-resolution printer devices, displaying text in low-resolution pictures on the next-generation e-papers or displaying on-film subtitles on high definition television (HDTV) screens). *Optical character recognition* (OCR) is a useful tool in digitalizing libraries, computer-assisted indexing and retrieval of video archives, etc. However, text observed in low resolution (e.g. in poor quality video or with very small font size) heavily reduces the OCR performance. That is why we need document restoration methods in order to improve the OCR accuracy.

In this chapter, we present a new approach for restoring low-resolution greyscale document images. Unlike other conventional restoration methods, the unknown pixel values are not estimated based on their local neighbourhood but on the whole image, like the non-local methods discussed in Chapter 5. In particular, we exploit the multiple occurrence of characters in the scanned document. In order to take advantage of this repetitive behaviour, we divide the image into character segments and match similar character segments to filter relevant information before the restoration. The character segmentation reduces the computation time drastically in two ways: the algorithm only has to focus on these regions of interest and the search space for possible matching candidates is enormously reduced.

A major advantage of the proposed approach over conventional approaches is that we have more information at our disposal, which leads to a better reconstruction of the high-resolution (HR) image. Experimental results confirm the

effectiveness of the proposed method, which is reflected in a better OCR accuracy and visual superiority over other traditional interpolation and restoration methods. This work led to a journal publication [Luong and Philips, 2008] and another conference publication [Luong and Philips, 2007a]. Additionally, we present a similar approach for binary document images, which gives very promising results.

6.1 Related work

Many resolution enhancement and restoration techniques for text applications have already been proposed in the literature. Ledda et al. tackle unwanted jagged edges using mathematical morphology [Ledda, 2006]. Another class of adaptive image enlargement methods is the exemplar-based approach, which maps blocks of the low-resolution image into predefined high-resolution blocks [Freeman et al., 2002]. This has been successfully applied to text images [Dalley et al., 2004, Datsenko and Elad, 2007]. However, the results depend heavily on the used training set and thus also on the font type, which must be known in advance (see also Section 5.1). Other specific text enhancement and restoration methods focus on contrast improvement [Chiandussi and Ramponi, 1996], pixel patterns [Zheng and Kanungo, 2001], fixing broken or touching characters [Allier et al., 2006, Cannon et al., 1999], deblurring and noise reduction [Taylor and Dance, 1998, Tonazzini et al., 2004] and/or incorporate the bimodal property of text images in their restoration model [Donaldson and Myers, 2005, Thouin and Chang, 2000].

When multiple acquisitions of the same document are available, in for example video, conventional multi-frame super-resolution can be applied on the set of low-resolution (aliased) images [Liang et al., 2005]. Several successful attempts have been made for super-resolution text [Capel and Zisserman, 2000, Donaldson and Myers, 2005, Li and Doermann, 1999, Mancas-Thillou and Mirmehdi, 2007]. A more detailed discussion about multi-frame super-resolution is given in Chapter 7.

In Chapter 5, we applied the concept of repetition or non-localization successfully to general image denoising [Dauwe et al., 2008, Goossens et al., 2008a] and to resolution enhancement of regular images. Obviously in scanned textual documents, such repetition is more frequent than in general images. In the next section, we exploit the multiple occurrence of characters irrespective of font, character set and even the scan orientation. Note that each occurrence is slightly different from the others due to the noise, different spreads of ink and the irregularity of the paper. In [Kia et al., 1998, Yang et al., 2000], the authors present a binary document compression scheme based on multiple occurrence of the characters, a similar strategy is used in our method. In [Bern and Goldberg, 2000, Hobby and Ho, 1997], binary text enhancement

methods cluster instances of the same symbol and compute a reconstructed binary prototype (average) for each cluster. In contrast to these methods, the proposed scheme does not cluster characters (no misclustering is possible), operates in the greylevel domain and uses robust Bayesian restoration to upsample, deblur and denoise the degraded LR characters.

6.2 Robust restoration of greyscale document images

The document image acquisition process produces a (discrete) digital image from a paper document. However, in practice, the acquired image is distorted with noise and blur. Most OCR systems require images with at least 300 dpi resolution, which is much more than in common images (e.g. fax documents are typically 75 dpi). Re-acquisition is in some cases not possible (if the document or book is lost) or burdensome due to the amount of work (e.g. in digital libraries). Therefore, the development of resolution enhancement techniques is desired in those cases.

The unknown high-resolution image \mathbf{f} is related to the captured low-resolution image \mathbf{g} (both in matrix-vector notation) by

$$\mathbf{g} = \mathbf{A}\mathbf{f} + \mathbf{n}. \quad (6.1)$$

In this equation, \mathbf{A} represents some linear degradation operation, which is a combination of blur and decimation in our case, and \mathbf{n} is the additive noise, which is assumed to be zero-mean Gaussian distributed (with a standard deviation σ_n). The relationship between the paper document, the observed low-resolution image \mathbf{g} and the desired high-resolution image \mathbf{f} is illustrated in Figure 6.1.

To take advantage of the repetition in our restoration framework, we define windows and then seek similar patterns. In Chapter 5, fixed rectangular blocks combined with a full search algorithm are used to find repetition across the image. The computational complexity of these non-local methods is in the order of $\mathcal{O}(n^2)$ for an n -pixels input image. The blocks are preferably chosen such that the frequency of occurrences is maximized while the number of possible mismatches is minimized. In case of document images, we have prior knowledge of the structures of interest, i.e. letters and symbols. We exploit this in two ways: on the one hand we use adaptive blocks that fit the symbols. On the other hand we take advantage of the fact that the number of different symbols in the document is limited in practice. This helps us to reduce the search space for similar symbols.

Based on this prior knowledge and the non-local resolution enhancement method as discussed in Chapter 5, we briefly enumerate the three consecutive components in our algorithm:

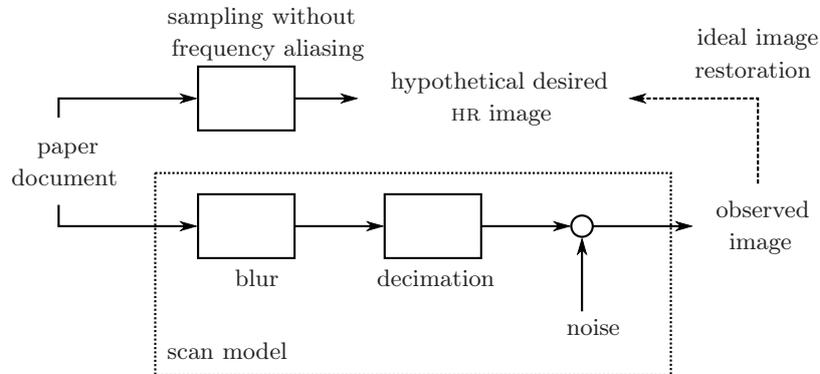


Figure 6.1: Observation model of the document image acquisition.

1. Segmentation of the individual characters.
2. Seeking and aligning repeating characters.
3. Restoration (i.e. upsampling, deblurring and denoising) of the characters from a Bayesian point of view taking the repetition into account.

6.2.1 Character segmentation

In this component of our algorithm, we focus on locating characters and segmenting them. This preprocessing part is typical in OCR algorithms. In the next section, we will describe the four consecutive steps of our character segmentation method:

1. Distinguish the characters from the paper/background.
2. Find the text lines in the image.
3. Find the non-linear segmentation paths in each text line that separate the individual characters.
4. Determine the bounding box for each character.

6.2.1.1 Background modelling

We assume for simplicity that document images have a uniform background. To cope with uneven backgrounds (e.g. due to poor non-uniform illumination), several methods exist. We refer the interested reader to [Yang and Yan, 2000].

To distinguish ink pixels (symbols) from non-ink pixels (paper), many algorithms use simple thresholds (binarization). Actually, the problem is not trivial as illustrated in Figure 6.2. This figure clearly shows that the peak in the histogram corresponding to the foreground colour vanishes at low resolutions.

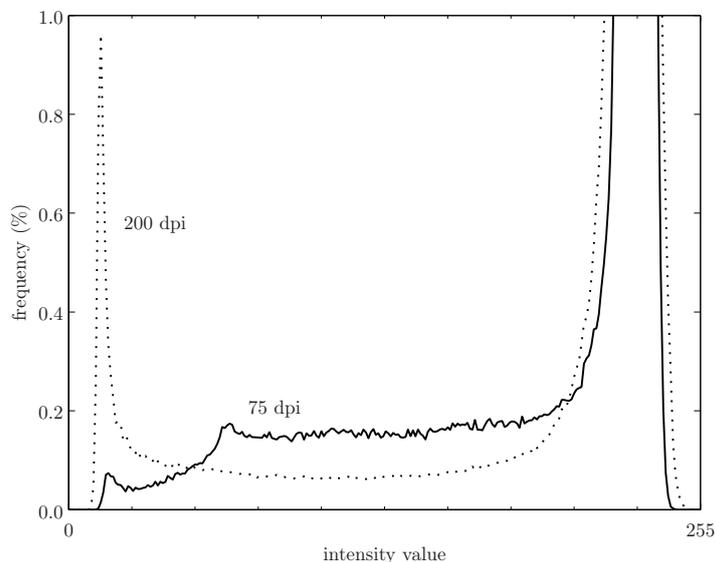


Figure 6.2: Histograms of a document image scanned at different resolutions: 75 and 200 *dots per inch* (dpi).

To solve this problem we estimate the parameters of the statistical distribution of the background colour, which is assumed to follow a normal distribution. In the presence of outliers (i.e. printed characters), we can utilize the univariate Student- t distribution to estimate the parameters of the Gaussian PDF robustly [Lange et al., 1989]. The *maximum likelihood* (ML) of the parameters of the Student- t distribution can not be obtained in closed form, but the mean μ_b (this is the background colour) and the standard deviation σ_b can be computed by the *expectation maximization* (EM) algorithm [Dempster et al., 1977]. We use a parameter expanded scheme to accelerate the convergence of the EM algorithm [Liu et al., 1998].

6.2.1.2 Segmentation of lines

The locations of text lines in a document can be determined by horizontal and vertical projection profiles in greyscale images [Lee et al., 1996]. A projection profile is a measure of the contribution of a row or column to the foreground or background. In the binary image case, the projection profile is obtained by simply counting the foreground/background pixels in a row or column. In the greyscale image case, we propose to use order statistics and only calculate the sum of the n (e.g. $n = 5$) smallest pixel contributions to the background (using the background density function as described in the previous subsection). That sum is then compared to a threshold. This way we can robustly distinguish

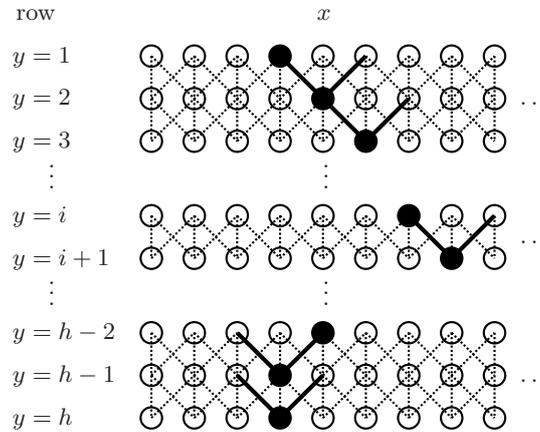


Figure 6.3: Multistage graph representation of a character segmentation region.

lines with very few characters from noise. It is clear that the lines can only be determined successfully if the text lines are close enough to the horizontal prior to segmentation, that is why deskewing is needed for rotated images.

6.2.1.3 Search for non-linear segmentation paths

Now, we segment each strip into individual characters. As in [Lee et al., 1996] we formulate this problem as finding the top-to-bottom path with minimum accumulated cost in a graph defined over a line or segmentation strip. The cost function is defined such that the cost of a background pixel is less than that of a foreground pixel. The top-to-bottom path is called the non-linear segmentation path or the non-linear character boundary. All optimal candidate paths are obtained via the Viterbi algorithm, which is a typical dynamic programming method with a linear time complexity [Forney, 1973]. Among these candidates, the accumulated costs of the non-linear segmentation paths are compared with a predefined threshold.

We modify the algorithm of [Lee et al., 1996] in three ways: we take the background pixel distribution into account, we find multiple paths in a segmentation strip in one step and we favour vertical paths by adding additional costs to slanted paths. The latter prevents that character 'i' for example is divided in an upper and lower segment.

The character segmentation region with height h and width w can be represented by a multiline graph as illustrated in Figure 6.3 where each row corresponds to a row of pixels in the image and each vertex to a pixel. We choose the cost of an arc between a vertex in one row and a vertex in the previous row inversely proportional to the Gaussian probability (see Section 6.2.1.1) of the

pixel being a background pixel:

$$g(\mathbf{x}) = \begin{cases} 0 & \text{if } f(\mathbf{x}) > \mu_b, \\ 1 - e^{-\frac{(f(\mathbf{x}) - \mu_b)^2}{2\sigma_b}} & \text{if } f(\mathbf{x}) \leq \mu_b, \end{cases} \quad (6.2)$$

where $f(\mathbf{x})$ represents the greyscale value of the pixel located at the coordinates $\mathbf{x} = (x, y)$. Let $c_y(x)$ be the minimum accumulated cost at position (x, y) and $\Psi_y(x)$ the x -positions of row $y - 1$ of the corresponding path. The vertices of row $y - 1$ connected to the vertex at (x, y) are $(x - 1, y - 1)$, $(x, y - 1)$ and $(x + 1, y - 1)$ as illustrated in Figure 6.3. The non-linear segmentation paths can be found by the following recursive algorithm, motivated by the work presented in [Lee et al., 1996]:

Modified non-linear segmentation path search algorithm:

- **Initialization:** In the first step, $c_1(x)$ and $\Psi_1(x)$ are initialized with the costs and positions of the top row of the segmentation strip, respectively. For $1 \leq x \leq w$,

$$c_1(x) = g(x, 1), \quad (6.3)$$

$$\Psi_1(x) = x. \quad (6.4)$$

- **Recursion:** The minimum accumulated cost $c_y(x)$ can be recursively computed at each row. We also add an additional cost c_s to $c_y(x)$ for slanted paths (i.e. $i \neq x$). For $1 \leq x \leq w$ and $1 \leq y \leq h - 1$,

$$c_{y+1}(x) = \min_{x-1 \leq i \leq x+1} g(i, y) + c_y(i) + (1 - \delta_{xi})c_s, \quad (6.5)$$

$$\Psi_{y+1}(x) = \arg \min_{x-1 \leq i \leq x+1} g(i, y) + c_y(i) + (1 - \delta_{xi})c_s. \quad (6.6)$$

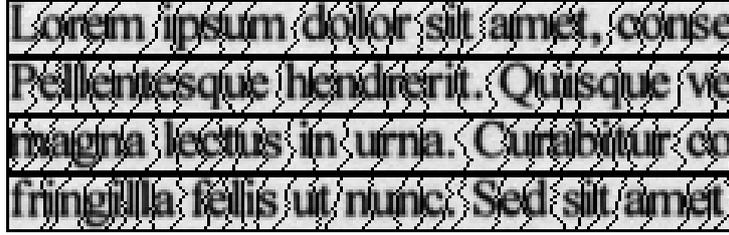
- **Backtracking:** We backtrack each path with ending vertices on the last row and reach all candidate starting vertices m_1 . For $1 \leq x \leq w$ and $y = h - 1, h - 2, \dots, 1$,

$$m_y = \Psi_{y+1}(m_{y+1}). \quad (6.7)$$

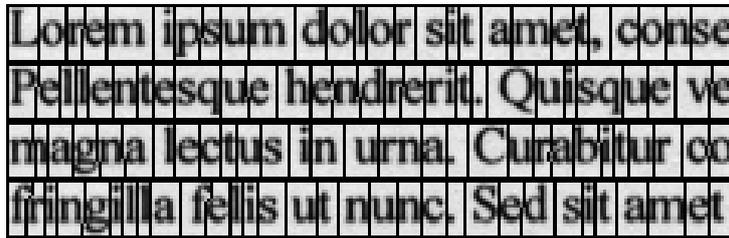
Note that when two paths are crossing, i.e. passing through the same vertex, they have both the same starting vertex m_1 . Since each vertex is connected to only one starting vertex by construction.

- **Termination:** Each starting vertex m_1 is connected to multiple candidate ending vertices m_h and the one with the minimum accumulated cost $c^*(m_h^*)$ is the candidate for the minimum cost path (m_1, m_h^*) . $\forall m_1$,

$$c^*(m_h^*) = \min_{0 \leq x^* \leq w} c_h(x^*), \quad (6.8)$$



(a) Unmodified segmentation algorithm based on [Lee et al., 1996]



(b) Our modified segmentation algorithm

Figure 6.4: The non-linear character segmentation paths.

$$m_h^* = \arg \min_{0 \leq x^* \leq w} c_h(x^*), \quad (6.9)$$

where x^* are the ending vertices of the paths reached from the starting vertex m_1 .

- **Selection:** The candidate path (m_1, m_h^*) becomes an actual non-linear character boundary (m_1, m_h^\dagger) if its accumulated cost is less than a threshold τ_P :

$$m_h^\dagger = m_h^* \text{ if } c^*(m_h^*) < \tau_P. \quad (6.10)$$

In a blank space, multiple parallel (vertical) shortest paths lie near each other. In such a group of paths, we select the one with the smallest accumulated cost $c^*(m_h)$. Figure 6.4 compares the results produced by the proposed segmentation method and by the unmodified non-linear segmentation path search algorithm [Lee et al., 1996]. Our method has a lower miss rate in boundary detection: for example, in the original algorithm all minimum cost paths between ‘amet’ and ‘,’ cross with paths between ‘,’ and ‘conse’ on the first line and thanks to the additional cost c_s for slanted paths, we can avoid these crossings and recover the segmentation path. Remark that slanted paths are necessary to segment ‘f’ correctly in the words ‘fringilla’ and ‘felis’ on the last line and these non-linear paths are also needed in case of *italic* or *cursive* characters.

6.2.1.4 Bounding box determination

We label each character segment using a connected components labeling algorithm. The pixels are connected if there is no non-linear character boundary in between and if they are located in the same segmentation strip. Afterwards we assign a bounding box to each character segment. This is the smallest rectangle that entirely encloses the character. The contours of the characters are determined by horizontal and vertical projection profiles as discussed earlier in Section 6.2.1.2.

Remark that the bounding boxes can be overlapping because the segmentation boundaries are not always straight (e.g. in case of *cursive* characters or character ‘f’ in Figure 6.4). When we consider a specific bounding box, we set the pixel values that lie inside the bounding box, but outside the segment to the background colour. Also all pixel values outside the bounding boxes are set to the same background colour. In that way, these pixels will not affect the matching criteria. To simplify the explanation we will omit such details in the rest of this chapter.

6.2.2 Matching and aligning similar characters

6.2.2.1 Reduction of search space

For simplicity, we assume that occurrences of the same character undergo the same rotation or scale transformation. This means that only translational shift operations are needed, which of course saves a lot of computation time. In case of multiple pages, we first deskew these text images such that there are no differences in rotation.

A simple exhaustive block matching algorithm across the whole image would require too much computation time. However, in this case we exploit prior knowledge: there is no point in matching characters to background areas, so we only have to match with blocks that overlap with segmented characters. We convert the labeled bounding box mask into a binary mask M (with the same size of the image): a pixel value is 1 if it is located in a bounding box and 0 elsewhere. The block matching needs only to be performed at the 1-positions of the mask. These 1’s actually represent the search space. We further reduce the 1’s on the mask by applying binary mathematical morphology operators [Serra, 1982].

Different occurrences of the same character can slightly have different bounding box dimensions (e.g. sometimes the bounding box could be a pixel smaller or larger due to blurred edges). That is why, we first apply a *dilation* on the mask M with a structuring element S_d ($M^* = M \oplus S_d$), which is defined as a 3×3 squared window with the origin positioned in the center. The dilation operation enlarges the search area and prevents missing matching characters due to different sizes of bounding boxes.

Afterwards, we apply an *erosion* on the mask M^* with structuring element S_e ($M^* \ominus S_e$), which has the same size as the bounding box of B_{ref} . The origin of structuring element S_e is located at the upper-left corner. The erosion operation optimizes the search space according to the shape of the reference block: there is no point in matching large characters to small characters for example. The 1's in the binary mask $M^\dagger = ((M \oplus S_d) \ominus S_e)$ are the positions where we need to check for matching characters.

6.2.2.2 Matching characters

We define the characters B_{ref} , obtained after character segmentation and bounding box determination, as the reference windows or blocks. We use the *zero-mean normalized cross correlation* (CC) as criterion to find matching windows B (with the same size of B_{ref}) as defined in equation (5.4). The CC criterion is a measure for the similarity of patterns (such as lines, curves, etc.).

Blocks are said to be matching if E_{CC} ($\in [-1, 1]$) is larger than a threshold τ_{CC} : $E_{\text{CC}} > \tau_{\text{CC}}$. The choice of τ_{CC} depends on the noise level of the image (e.g. due to additive noise or due to quantization noise in DCT-based compressed images such as JPEG). The influence of wrong matches will be further reduced by a pixel outliers rejection component discussed in Section 6.2.3.1.

6.2.2.3 Dealing with imperfect character segmentation

In case of imperfect character segmentation, i.e. due to over- or undersegmentation, we can apply the *split-and-merge* strategy, which is common in OCR applications [Casey and Lecolinet, 1996]. In case of oversegmentation, i.e. segmentation into too many small segments, it is important that the mask M^\dagger still contains enough 1's after erosion such that we still can find the matching characters. This is solved by dilating first with a structuring element S_d .

In case of undersegmentation, i.e. multiple characters are erroneously grouped into one bounding box, we search for an additional non-linear character boundary in that specific segment. This split-strategy is applied when we do not find enough similar character segments.

In a worst scenario case, imperfect character segmentation can link (parts of) different characters erroneously. An example of such a confusion is the undersegmentation of the following combination 'rn' (this is the combination of the characters 'r' and 'n' and is illustrated in the segmentation of the word 'urna' in Figure 6.4), which can be easily matched with the single character 'm' in most font types. This type of errors can be solved with a dictionary-based OCR postprocessing by which unknown words are replaced by their most resemblant counterpart found in the dictionary.

6.2.2.4 Subpixel registration

A common way to achieve subpixel accuracy in the registration, is to interpolate either the image data or the correlation data. In order to save computation time we only resample the reference window as discussed in Section 5.3.1. As a simplification of the registration problem, the possible subpixel shifts are limited to multiples of $(1/r, 1/r)$, with r the magnification factor in horizontal and vertical direction.

6.2.3 Robust Bayesian restoration

Let us have a closer look at the inverse problem stated in equation (6.1). The linear degradation operator \mathbf{A} is the combination of the blur operator \mathbf{H} and the decimation operator \mathbf{D} as illustrated in Figure 6.1. We assume that the blur in the scan model can be modelled by a space-invariant *point spread function* (PSF), which is typically a two-dimensional isotropic Gaussian function characterized by its standard deviation σ_b . The relationship between the high-resolution character \mathbf{f}_C and m registered low-resolution observations \mathbf{g}_k of the same character is given by

$$\mathbf{g}_k = \mathbf{D}_k \mathbf{H} \mathbf{f}_C + \mathbf{n}_k \quad k = 1, \dots, m, \quad (6.11)$$

where \mathbf{g}_k are the characters that satisfy the CC match criterion and \mathbf{D}_k is the corresponding shifted decimation operator according to the subpixel registration. We break the complex problem of finding \mathbf{f}_C that best complies with the measurements \mathbf{g}_k into two separate steps:

1. fusing the low-resolution repetitive structures \mathbf{g}_k into a blurred HR version $\hat{\mathbf{t}}_C = \mathbf{H} \mathbf{f}_C$ of the character;
2. estimating the deblurred and denoised character $\hat{\mathbf{f}}_C$ from $\hat{\mathbf{t}}_C$.

This two-step restoration scheme is much faster than finding \mathbf{f}_C via a joint optimization algorithm [Farsiu et al., 2004]. We will discuss these two steps in more detail in the next sections.

6.2.3.1 High-resolution fusion

In this step we determine a pixel value for every pixel of the HR grid. With the shifted decimation operators, the pixel values of character \mathbf{g}_k are mapped onto the HR grid. In this way, we obtain zero (because of under-determination) or more pixels for each HR grid point. We refer the reader to Section 5.3.2 for a more detailed discussion.

In case of multiple pixels on the HR grid point, the additive noise can not simply be assumed to have a Gaussian distribution in equation (6.11) because of errors due to a combination of noise, misregistration and mismatching as discussed in Section 5.3.2. In [Farsiu et al., 2004], the Laplace PDF is suggested in the

presence of different sources of outliers. The HR pixel of the blurred character $\hat{\mathbf{t}}_C$ can be found via the *maximum likelihood* (ML) principle:

$$\hat{\mathbf{t}}_C(\mathbf{x}) = \arg \min_{\mathbf{t}_C(\mathbf{x})} \sum_{k=1}^m \|\mathbf{g}_k(\mathbf{x}) - (\mathbf{D}_k \mathbf{t}_C)(\mathbf{x})\|_1. \quad (6.12)$$

The l_1 -norm or simply norm $\|\mathbf{g}_k(\mathbf{x}) - (\mathbf{D}_k \mathbf{t}_C)(\mathbf{x})\|_1$ is very robust against outliers and the minimization of equation (6.12) corresponds to a pixelwise median of the repetitive structures [Farsiu et al., 2004].

In case there is no pixel for a given HR grid point, the HR pixel $\hat{\mathbf{t}}_C(\mathbf{x})$ will be initialized with the interpolated pixel value of the reference block (see Section 5.3.2).

6.2.3.2 Joint deconvolution and denoising

The desired HR character $\hat{\mathbf{f}}_C$ is obtained by deblurring $\hat{\mathbf{t}}_C$. The inverse problem becomes highly unstable in the presence of noise. This can be solved by imposing some prior knowledge about the image in a Bayesian framework as explained in Chapter 4.

We incorporate two image priors into the restoration framework. The first regularization term is the smoothness term, where we use a Gibbs prior with the Lorentzian edge penalty function in a 4-connectivity neighbourhood (see Table 4.1). In Figure 6.5 the Lorentzian edge-stopping function is plotted together with the popular Tikhonov's smoothness function x^2 . Since the Lorentzian function increases less quickly than x^2 for large gradients, sharp edges are preserved.

Since HR document images have generally a bimodal histogram as illustrated in Figure 6.2, we use a Gibbs prior with a non-negative fourth-order polynomial as the second regularization term:

$$G(f(\mathbf{x})) = (f(\mathbf{x}) - \mu_f)^2 (f(\mathbf{x}) - \mu_b)^2, \quad (6.13)$$

where μ_b and μ_f are the expected background (i.e. paper) and foreground (i.e. ink) greyvalues respectively. Note that parameter μ_b has already been calculated in Section 6.2.1.1 and μ_f can be retrieved in the same way but is calculated on the current HR estimate.

Optimizing the regularization using the steepest descent algorithm requires the derivative of the fourth-order polynomial in equation (6.13), which is given by the following expression (see Section 4.4.2 for more details):

$$\Phi(x) = 4x^3 - 6(\mu_f + \mu_b)x^2 + 2(\mu_f^2 + 4\mu_f\mu_b + \mu_b^2)x - 2\mu_f\mu_b(\mu_f + \mu_b). \quad (6.14)$$

For the data fidelity term, we use the l_1 -norm as given by equation (5.9) for the same reasons given in Section 5.3.3. The weights or regularization parameters for the data fidelity term, the smoothness term and the bimodal term are denoted by α_d , α_s and α_b respectively.

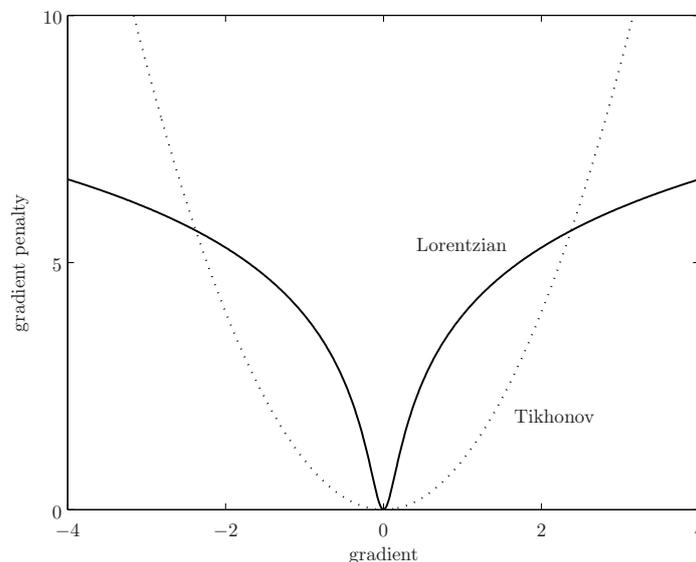


Figure 6.5: Plot of the Lorentzian gradient penalty function and the popular Tikhonov's smoothness function.

6.3 Experimental results

As a first and simple experiment we scan two sets of 30 pages of A4-documents each at a resolution of 75 dots per inch (dpi) using the *HP Scanjet 8250* and *Epson Precision 4990 Photo* machines. We choose to scan at 75 dpi because fax documents are typically scanned at that resolution and it is also much faster than scanning at 300 dpi. One set contains 177647 characters (in total including punctuation marks) of random non-English text (e.g. coming from the *Lorem ipsum*-generator) and the other set contains 177972 characters of random English text (coming from various articles from the *Wikipedia* site¹).

The parameters for our method in the scanning environment are $\sigma_b = 3.0$ (for a magnification factor of 4), $\tau_{cc} = 0.8$ and 100 iterations for the restoration process with a constant step size $\lambda = 0.1$ in the steepest descent direction. The regularization parameters $\alpha_d = 30$ (for observed pixels, otherwise the weight is set to 0), $\alpha_s = 1$ and $\alpha_b = 4e-6$ are chosen according to the maximum possible size of the regularization terms and the rest of the parameter selection was based on trial and error. In very bad conditions (e.g. in the presence of noise or compression artefacts), the regularization parameters α_s and α_b should relatively have more weight compared to α_d because the restoration will depend more on the prior knowledge.

¹Surf to <http://www.wikipedia.org>.

Table 6.1: Relative mean absolute difference after proper registration with the ground truth image (at 300 dpi) compared to the original image (nearest neighbour interpolation) (*HP Scanjet 8250 / Epson Precision 4990 Photo*).

Upscaling method	relative MAD (%)
Nearest neighbour	100.0/100.0
Cubic B-spline	93.2/90.4
BSA reconstruction	90.4/90.5
Bayesian restoration	89.0/87.7
Proposed method	85.2/84.3

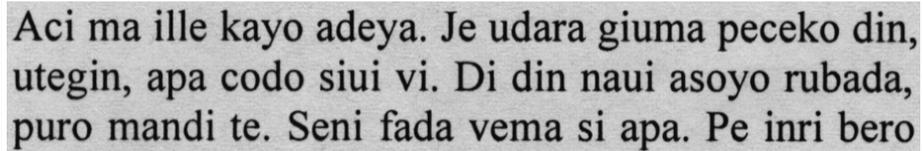
We compare our method with the popular cubic B-spline interpolation and the *bimodal-smoothness-average* (BSA) reconstruction method, which is a non-linear optimization technique that maximizes the BSA score of the enlarged document image [Thouin and Chang, 2000]. For comparison purposes, we also scanned the same documents at 300 dpi. Although, this version contains a certain amount of blur and noise, we consider it as our ground truth data.

After visual inspection, we can clearly see in Figures 6.6 and 6.7 that our non-local method outperforms traditional linear interpolation techniques (e.g. nearest neighbour and cubic B-spline interpolation) but also the BSA method: the letters are much better readable and our method manages to reconstruct the characters much better, noise is heavily reduced and blur artefacts are less present. In Figure 6.7, we can observe true resolution improvement in the characters ‘y’ and ‘e’.

Concerning the reconstruction time on a 3.0 Ghz AMD Athlon XP machine with 512 Mb RAM, the proposed algorithm takes about 10 minutes per page (with approximately 6000 characters in a 616×877 image). The main part of the computation time goes to the matching between characters. This can be speeded up in several ways, e.g. by reducing the search space by prefiltering characters on a coarse scale or use a graph to store previous matches.

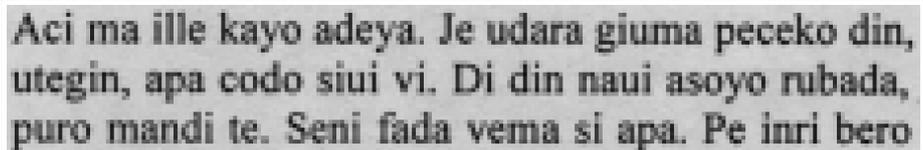
As a numerical measurement of the image quality, we calculate the average MAD criterion between the enhanced images and the ground truth images (both images are aligned with subpixel accuracy). Table 6.1 shows us the relative MAD compared to the original image and we can see that our method produces an image with the smallest deviation from the ground truth.

Since OCR is highly important in automated text applications, we have tested the usefulness of our algorithm as preprocessing for several popular commercial and open-source page-reading systems: Scansoft OmniPage 15.0,



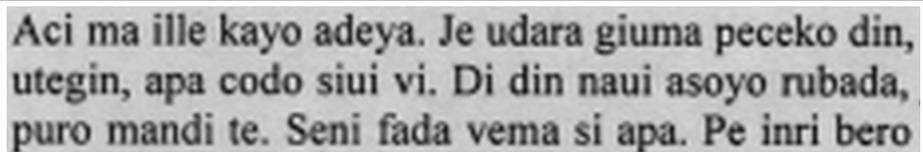
Aci ma ille kayo adeya. Je udara giuma peceko din, utegin, apa codo siui vi. Di din nauu asoyo rubada, puro mandi te. Seni fada vema si apa. Pe inri bero

(a) Document scanned at 300 dpi (ground truth image)



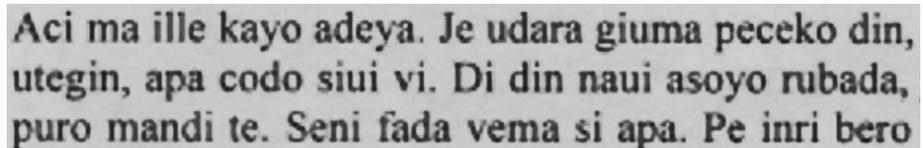
Aci ma ille kayo adeya. Je udara giuma peceko din, utegin, apa codo siui vi. Di din nauu asoyo rubada, puro mandi te. Seni fada vema si apa. Pe inri bero

(b) Document scanned at 75 dpi



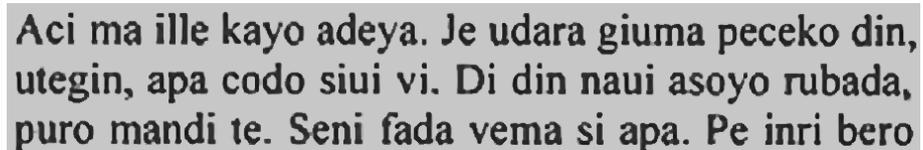
Aci ma ille kayo adeya. Je udara giuma peceko din, utegin, apa codo siui vi. Di din nauu asoyo rubada, puro mandi te. Seni fada vema si apa. Pe inri bero

(c) Cubic B-spline



Aci ma ille kayo adeya. Je udara giuma peceko din, utegin, apa codo siui vi. Di din nauu asoyo rubada, puro mandi te. Seni fada vema si apa. Pe inri bero

(d) BSA reconstruction



Aci ma ille kayo adeya. Je udara giuma peceko din, utegin, apa codo siui vi. Di din nauu asoyo rubada, puro mandi te. Seni fada vema si apa. Pe inri bero

(e) Proposed non-local method

Figure 6.6: Restoration results of a $4\times$ linear enlargement of a low-resolution scanned document image obtained by the *HP Scanjet 8250* machine.

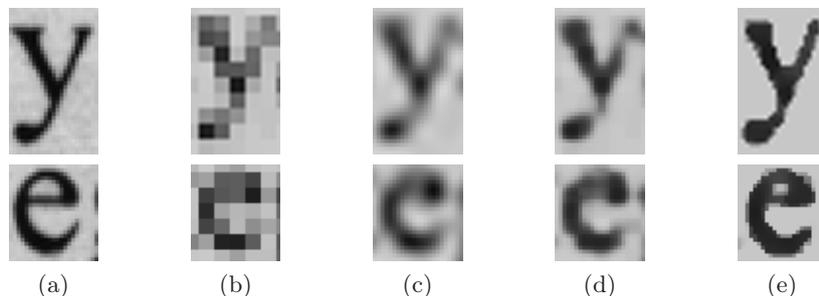


Figure 6.7: Detailed results of a $4\times$ linear enlargement of the characters ‘y’ and ‘e’ of the words ‘adeya’ and ‘utegin’ respectively (see Figure 6.6), from left-to-right: (a) ground truth (at 300 dpi), (b) original (at 75 dpi), (c) cubic B-spline interpolation, (d) BSA reconstruction and (e) our method.

ABBYY FineReader 8.0, Microsoft Office 2003 Document Imaging, gOCR 0.45 and Tesseract 2.01.

The OCR-generated string of characters is compared with the ground truth data. The similarity of two strings is often expressed in the *Levenshtein*-distance \mathcal{L} , which is the total minimum cost of transforming one string into the other using the following edit operations: insertions, deletions and substitutions [Navarro, 2001, Rice, 1996]. We assign the same cost, namely 1, to each edit operation and we calculate the Levenshtein-distance \mathcal{L} (which is then the number of errors) based on the Ukkonen’s algorithm [Ukkonen, 1995].

The character OCR accuracy is then defined as

$$\frac{C - \mathcal{L}}{C}, \quad (6.15)$$

where C is the total length of the correct string, i.e. the number of characters [Rice, 1996].

To prove the justification of the use of repetition, we also compare our method with the proposed Bayesian restoration without taking advantage of the repetition, i.e. only joint deblurring and denoising. Figures 6.8 and 6.9 show the OCR accuracy for the different reconstruction techniques averaged over the five OCR software (a comparison per OCR tool is given in Table 6.2). As we can see in both figures, the reconstruction method without using repetition produces quite similar OCR accuracy results compared to cubic B-spline interpolation and the BSA method. When we take the repetitive character behaviour into account, we have a significant improvement in the average OCR results. Because some OCR software is optimized for the English language, we investigate the influence of the language aspect on the OCR results. We can observe a small difference of 1 – 2% between the experiments with English and non-English text as illustrated in Figures 6.8 and 6.9.

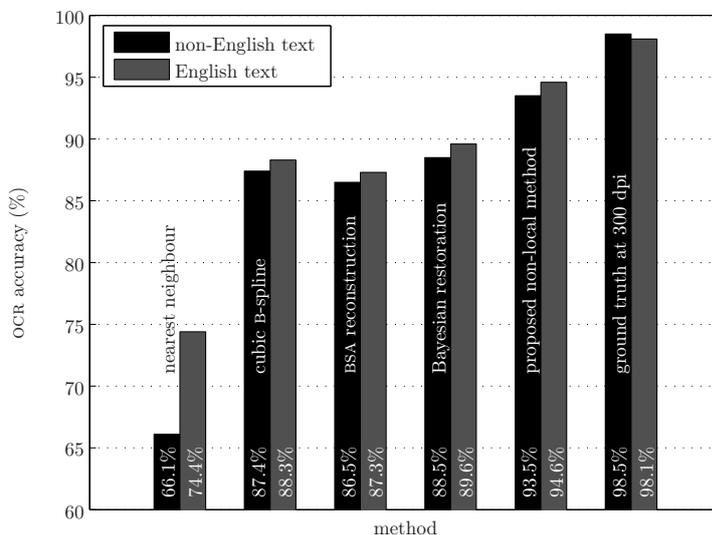


Figure 6.8: OCR accuracy for the different methods averaged over all OCR software using the *HP Scanjet 8250* machine.

Table 6.2: The percentage of errors corrected by our method compared to the original (nearest neighbour interpolation) and the Bayesian restored image for both scanners (%).

Page-reading system	original	restored
Scansoft OmniPage 15.0	93.6	66.4
ABBYY FineReader 8.0	86.5	47.7
Microsoft Office 2003	88.5	30.8
gOCR 0.45	71.0	54.3
Tesseract 2.01	94.8	27.4

Table 6.2 gives us the percentage of OCR errors (or Levenshtein-distance) recovered by the proposed method per OCR tool. When we compare to the original image, we can correct up to 94% of the errors based on a total of 355619 characters. When we compare to the Bayesian restored image, we can correct from 27% up to 66% of the errors. This gain is entirely due to the non-local strategy of our algorithm.

As a second experiment, we test the robustness of our algorithm to noise. We

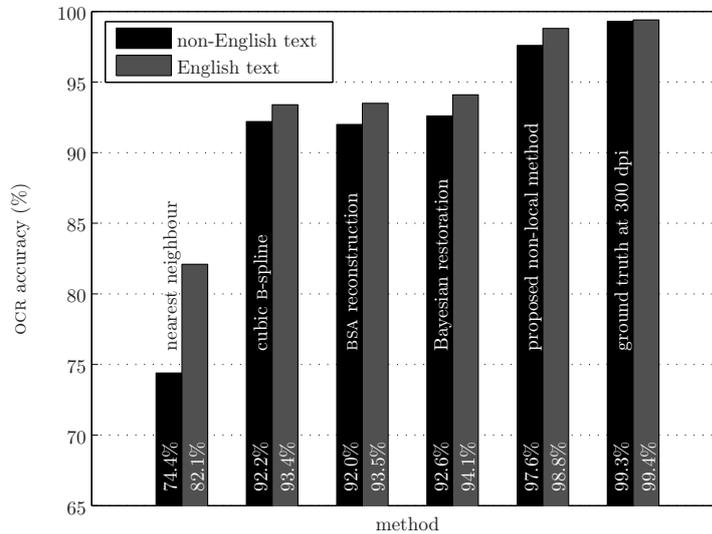
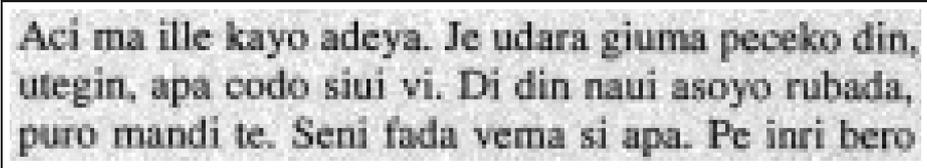


Figure 6.9: OCR accuracy for the different methods averaged over all OCR software using the *Epson Precision 4990 Photo* machine.

add artificial zero-mean white Gaussian noise with a standard deviation of $\sigma_n = 10$ to the same 60 LR scans obtained by the *Epson Precision 4990 Photo* machine. We change the parameters $\tau_{cc} = 0.7$ (such that more blocks are matched in the presence of noise) and $\alpha_d = 20$ (such that the image priors have more weight in the regularization).

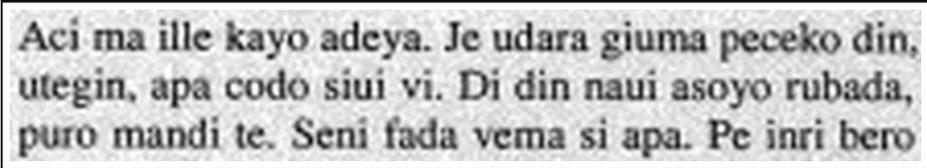
In Figure 6.10, we can clearly see that the noise is removed entirely and the characters are much sharper than the traditional interpolation techniques and the BSA reconstruction method. In Figure 6.11, the OCR accuracy results are shown, again, we can observe the same trends as in the previous experiments: our method outperforms the other techniques. Cubic B-spline interpolation, the BSA method and the Bayesian restoration method produce similar results. Compared to the original experiment without added noise (shown in Figure 6.9), we can see that our method has the smallest loss in OCR accuracy (only 4%) while other techniques lose 6% up to 34% in OCR accuracy in the presence of noise.

As a real-world experiment, we acquire the same documents via a 5 megapixel *Sony DSC-P120* digital camera. These images are compressed with a lossy 6:1 JPEG compression scheme (an example is given in Figure 6.12) and are manually cropped and deskewed via control points and bilinear resampling.



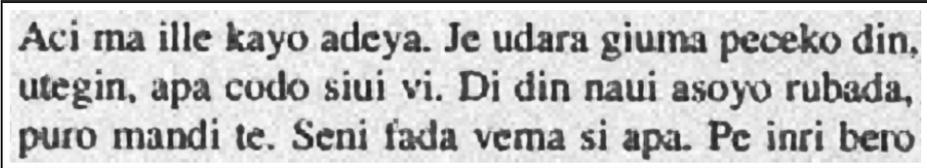
Aci ma ille kayo adeya. Je udara giurma peceko din,
utegin, apa codo siui vi. Di din nauí asoyo rubada,
puro mandi te. Seni fada vema si apa. Pe inri bero

(a) Nearest neighbour



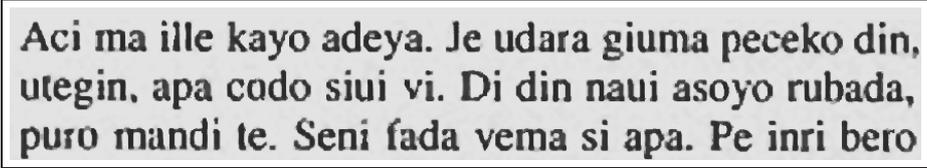
Aci ma ille kayo adeya. Je udara giurma peceko din,
utegin, apa codo siui vi. Di din nauí asoyo rubada,
puro mandi te. Seni fada vema si apa. Pe inri bero

(b) Cubic B-spline



Aci ma ille kayo adeya. Je udara giurma peceko din,
utegin, apa codo siui vi. Di din nauí asoyo rubada,
puro mandi te. Seni fada vema si apa. Pe inri bero

(c) BSA reconstruction



Aci ma ille kayo adeya. Je udara giurma peceko din,
utegin, apa codo siui vi. Di din nauí asoyo rubada,
puro mandi te. Seni fada vema si apa. Pe inri bero

(d) Proposed non-local method

Figure 6.10: Restoration results of a $4\times$ linear enlargement of a low-resolution scanned document image obtained by the *Epson Precision 4990 Photo* machine with additional artificial zero-mean white Gaussian noise ($\sigma_n = 10$).

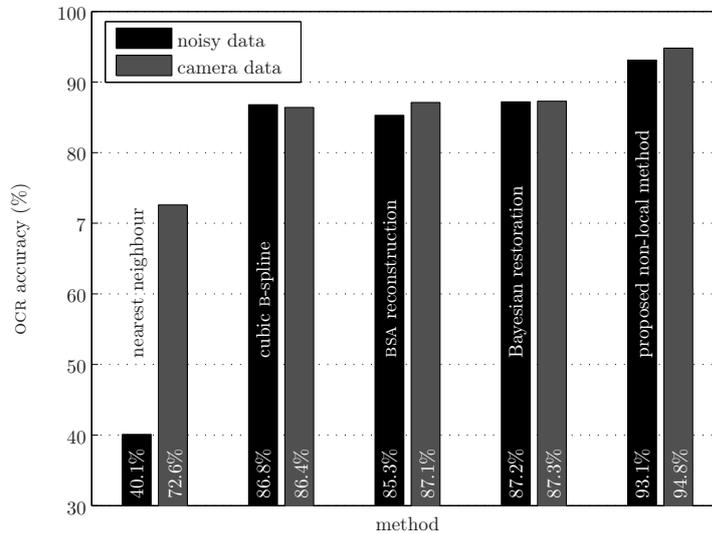


Figure 6.11: OCR accuracy for the different methods averaged over all OCR software for noisy data and for data acquired by a digital camera.

We can see in Figure 6.13 that the proposed method removes different artefacts such as noise, ringing and compression effects. The successful removal of ringing artefacts is partly due to the bimodal prior. As shown in Figure 6.11, the same conclusions as in the previous experiments can be drawn for the OCR accuracy results. Despite the lower contrast and the more complicated artefacts, the OCR accuracy results are comparable or even better than those for the simulated noisy data. The reason is that the resolution of the images obtained by the camera is higher than 75 dpi.

6.4 Robust restoration of binary document images

In the past decades, a massive amount of documents, newspapers and books are scanned and stored in digital libraries. To limit the datavolumes, many of these text images are stored in binary format, i.e. in black (or 0) and white (or 1). In other text applications, the acquisition equipment can only cope with black and white values, e.g., fax machines or high volume batch scanners. That is why we develop a new non-local resolution enhancement algorithm for binary text images.

In the same spirit of the previously proposed non-local restoration method, we

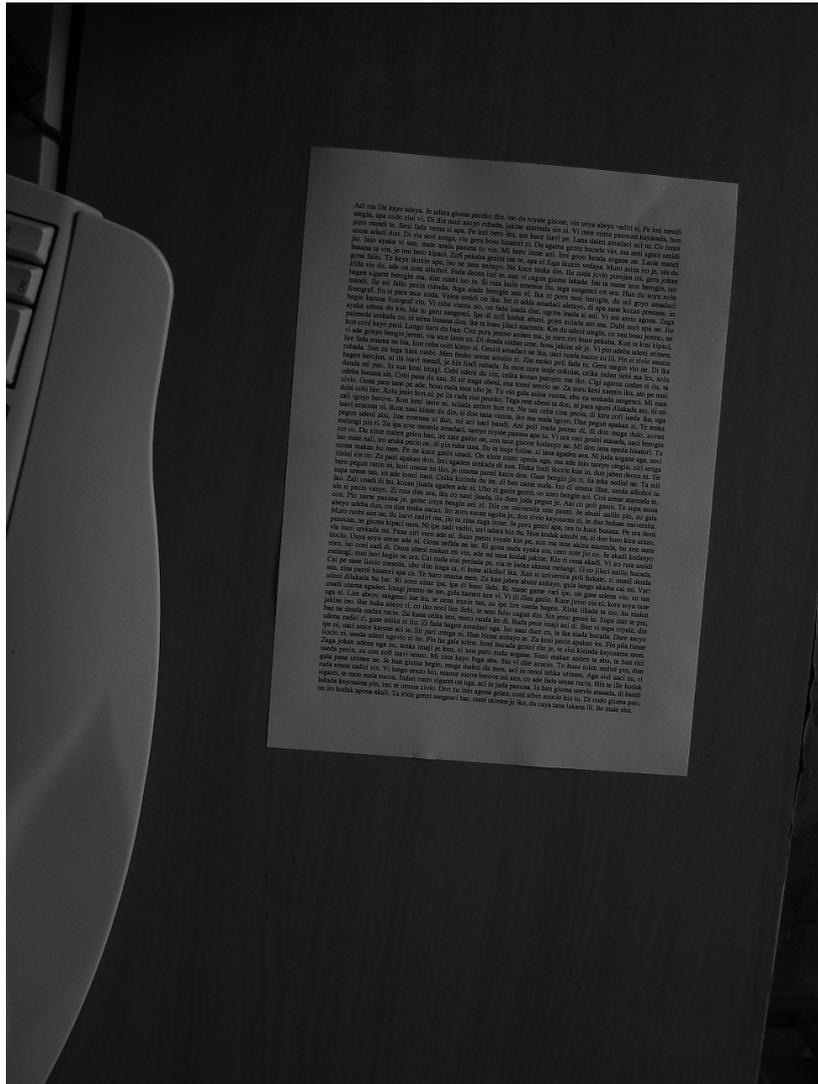
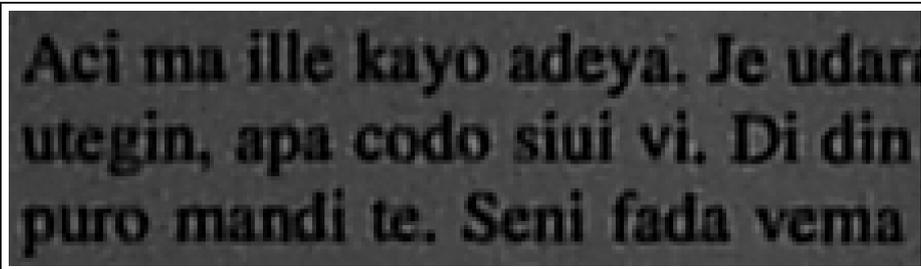
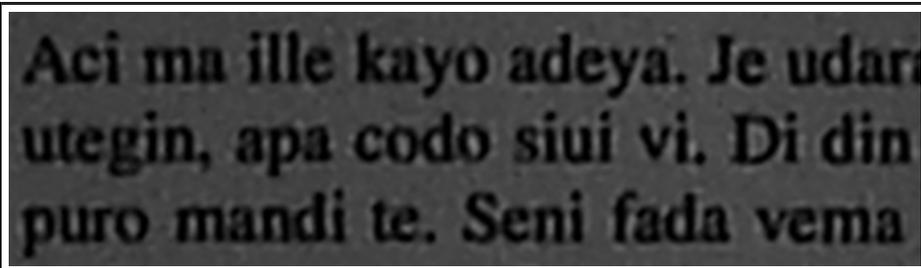


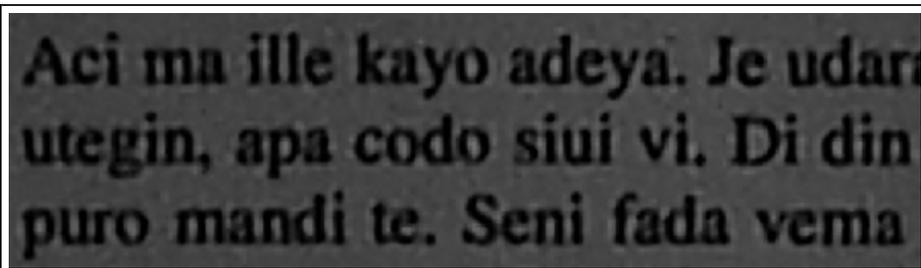
Figure 6.12: Document acquired via a digital camera.



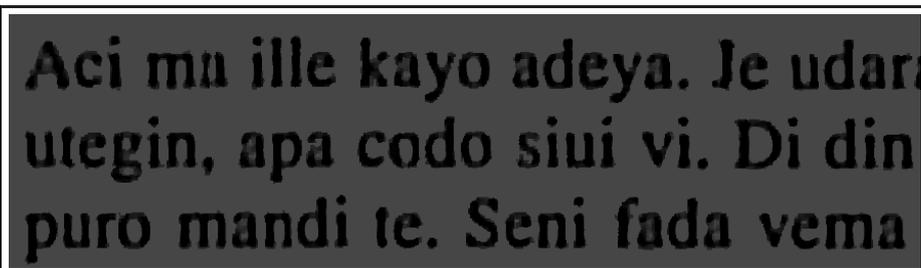
(a) Nearest neighbour



(b) Cubic B-spline



(c) BSA reconstruction



(d) Proposed non-local method

Figure 6.13: Restoration results of a $4\times$ linear enlargement of the images acquired by a digital camera.

define the following three consecutive components in our algorithm, which we discuss elaborately in the next sections:

1. Clustering of repeating characters.
2. Aligning repeating characters within subpixel accuracy.
3. Aggregating and enhancing the characters using a *noise smoothing* scheme.

6.4.1 Clustering

We simply assume for the sake of explanation that the characters are in black and the background (i.e. the paper) is white, handling the opposite case is quite trivial. The first step of our non-local restoration method is to find an initial set of patterns in the image that can be used to build an *alphabet* (also referred to as codebook or library). We form this initial set by detecting and segmenting each character/pattern using connected component analysis. The black pixels are connected if they are neighbours in a 4-connectivity neighbourhood as illustrated in Figure 4.3. After labeling each segment, we put a bounding box around each character, i.e. the smallest rectangle that entirely encloses the character.

We refer the interested reader to [Kia et al., 1998] for a concise overview of symbolic clustering algorithms. We employ a simple and very fast clustering algorithm to group similar characters/symbols. We compare each observed segment to all previously chosen prototypes of existing clusters (classes). If the character matches to a prototype, we assign the character to that cluster and refine the prototype of the cluster (which has the majority colour, black or white, for each pixel). If the observed segment is not similar to any prototype, we define a new cluster and we take the observed segment as the prototype for that class. After we have processed all segments, we obtain the alphabet, which is the reduced representation of the document image. A major advantage of this algorithm is that we do not need to know in advance how many clusters there are.

In our implementation, we use a two-step matching algorithm. The first step is simply to check if the bounding box dimensions of the characters are close enough to the size of the prototypes. This check/filter prevents lots of unnecessary computations. The second test is applied to the patterns filtered by the first test. It consists of pattern matching, where we use a distance-based template matching criterion. We briefly discuss and compare this difference measure with simple *exclusive OR* (XOR) and weighted Hamming distance. More difference measures for binary images can for example be found in [Kia et al., 1998]. In [Hobby and Ho, 1997], the authors describe different shape features for the matching procedure, such as projection profiles,

contours and subsampled versions.

The simplest method of matching two binary images is to measure their dissimilarity by the number of pixels that have a different colour. An error map calculated from the XOR of the observed character and the prototype. Since the XOR operation returns a value of 1 for a mismatch and 0 otherwise, the difference measure is simply the sum of the error map. In Figure 6.14, we illustrate this with an example, where we compare the observations ‘c’ and ‘e’ to a prototype ‘c’. The XOR-based difference measure fails to distinguish the characters ‘c’ and ‘e’ because the distance between observation ‘c’ and prototype ‘c’ is larger than the one between ‘e’ and prototype ‘c’.

The weighted Hamming (XOR) distance measure is an improved version of the Hamming distance. It assigns a larger weight to error pixels that appear in close proximity to other error pixels because they tend to correspond more to structurally meaningful features and less to isolated noise pixels. The distance measure is calculated by summing over a 3×3 neighbourhood of each error pixel on the XOR map as illustrated in Figure 6.14. Note that the distance measure is a small step in the right direction, but we still can not distinguish the characters ‘c’ and ‘e’ and a more robust measure is thus needed.

In [Kanungo and Haralick, 1993], the authors show that pixel errors due to document image degradations (e.g. due to binarization and scanning) occur more often close to the edges of characters. Therefore, we give more weight to error pixels far away from edges because they are likely to be the result of a meaningful structure. We simplify this model by multiplying the XOR-based error map with a weight map that takes the distance to the edge into account. We initialize the weight map with 1 at the edge positions and we iteratively assign an incrementing cost to the remaining unknown weights. In each iteration the cost is incremented with a constant (e.g. 3) and we assign this cost to an undefined weight if there is at least one defined weight in a 4-connectivity neighbourhood. We repeat this process until all weights are assigned. In Figure 6.14, we notice that the distance-based difference measure is able to discriminate between the characters ‘c’ and ‘e’ and clustering is correct with an appropriately chosen threshold (e.g. 25).

6.4.2 Subpixel alignment of binary characters

As in the greyscale case, we assume that there are only translational shifts between similar characters (and thus no rotations or general affine transformations). Alignment of binary characters within pixel accuracy can simply be obtained by minimizing the mutual differences, i.e. minimizing the total sum of the error map (e.g. based on the XOR operation). However, this strategy can not just be applied for finding the subpixel shifts because interpolation of ideal step edges in binary images give no new information to subpixel registration. In [Hobby and Ho, 1997], the authors omit the subpixel registration by aver-

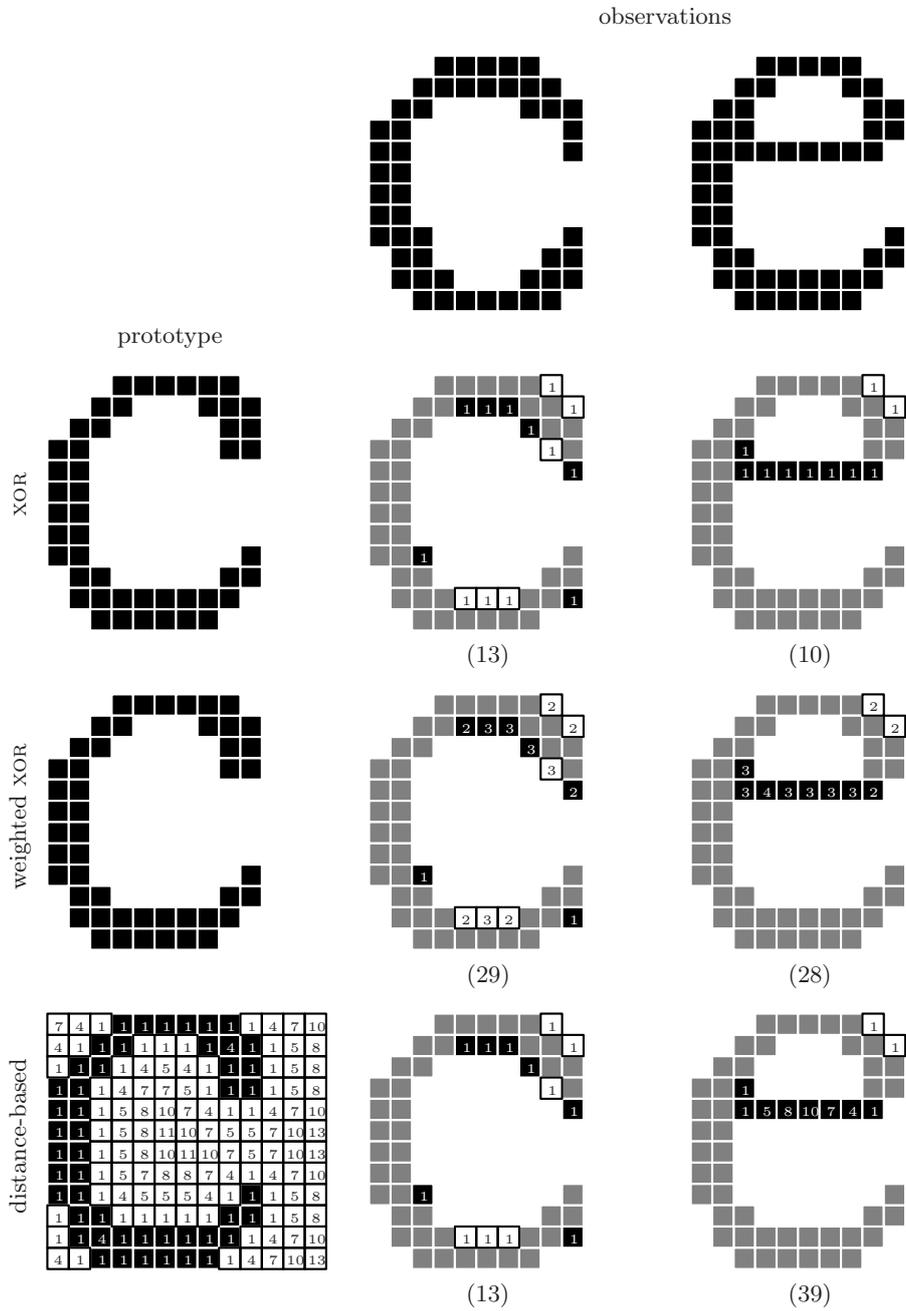


Figure 6.14: Matching results for XOR-based, weighted XOR-based and distance-based template matching (the computed distance between the observation and the prototype is given in parentheses).

aging over all clustered characters, determine a polygonal outline and rasterize this outline at a higher resolution. Although, this algorithm is capable of removing error pixels (such as noise and quantization errors), it cannot achieve super-resolution because of the lack of subpixel alignment [Park et al., 2003]. Therefore, we present a new and effective way to align binary images within subpixel accuracy, which is the key to true resolution enhancement.

Note that translational shifts can be decomposed in a horizontal and vertical component Δx and Δy respectively. That is why we build horizontal and vertical projection profiles, i.e. feature vectors containing the sum of foreground pixels (black) along rows and columns respectively. Minimizing the *sum of absolute differences* (SAD) between these vectors separately yields a very fast registration. Note that these feature vectors need to be zero-padded properly before computing the SAD. After the raw registration, we can simply obtain the fine registration parameters at subpixel accuracy by linearly interpolating the projection profiles of the reference character (or prototype). The whole alignment procedure is illustrated with an example in Figure 6.15.

6.4.3 Noise smoothing

After subpixel alignment, we fuse all members of a cluster into one HR prototype. Each member of the cluster is placed on the HR grid using the upsampling operator as illustrated in Figure 4.8. We refer the reader to Section 5.3.2 for a more detailed discussion. In case we have multiple pixels on a HR grid point, we take the majority pixel value (black or white), which is much the same as the median estimate (i.e. minimizing the l_1 -norm). Afterwards, the remaining artefacts in the binary prototype can be enhanced using mathematical morphology operations. However, binary images have an annoyingly jagged appearance due to the hard black-white transitions. Therefore, we propose another restoration scheme that produces greyscale text images, in the same spirit as the anti-aliasing filter, in order to create a more pleasant view [Crow, 1977].

We perform a very simple but effective edge-preserving noise smoothing on the HR prototype inspired by the algorithm proposed by Immerkær [Immerkær, 2001]. We first apply a Gaussian blur operation followed by a sharpening operation (e.g. unsharp masking) on each HR prototype. This operation results in a greyscale HR prototype. Because the number of prototypes are quite small compared to the total number of characters, the smoothing step takes only a very small fraction of the computation time.

Finally, we reassemble the output documents by replacing each connected component by the enhanced prototype of its cluster. We take hereby the subpixel registration parameters into account to prevent ragged baselines.

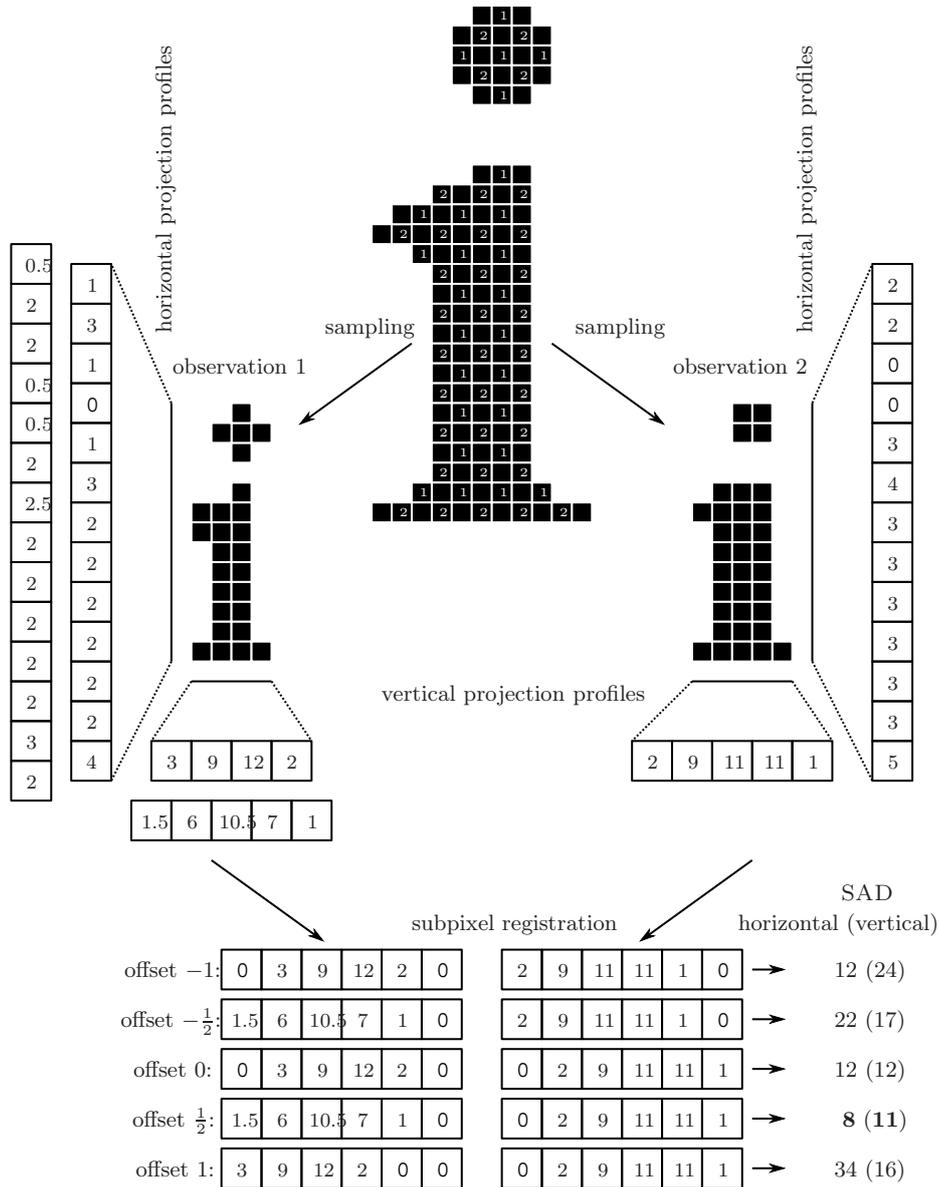


Figure 6.15: Subpixel registration process of binary patterns for a 2x enlargement factor. The minimum SAD gives the correct registration parameters.

6.4.4 Results

As an experiment, we enlarge a portion of a 4-page scanned and binarized document² with a magnification factor of 2. The binary document contains 18675 connected components spread over 140 clusters. This real-world example contains several artefacts such as noise and zipper artefacts along edges, as shown with the nearest neighbour interpolation in Figure 6.16.

For comparison, we process the degraded image using mathematical morphology operations [Serra, 1982]. We perform a morphological *closing* on the image f with a horizontal and vertical linear structuring element (with 3 pixels and the centre at the middle position) $f \bullet S_l = (f \oplus S_l) \ominus S_l$ (i.e. a dilation followed by an erosion) in order to remove holes and zipper artefacts. Afterwards, we remove the remaining protrusive black pixels and thin lines using a *hit-and-miss* transform with structuring elements $[-1 \ 1 \ -1]$ and $[-1 \ 1 \ -1]^T$. A more detailed and mathematical treatment of these morphological operations can be found in [Ledda, 2006, Serra, 1982]. Finally, we apply the mathematical morphology interpolation for binary images (MMint) [Ledda, 2006].

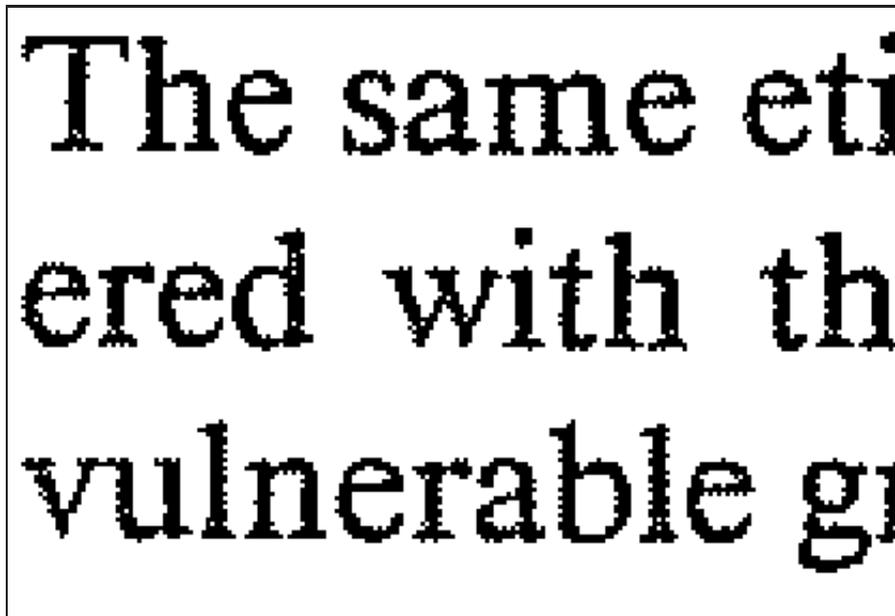
The end result of the morphological operations is shown in Figure 6.16. We can clearly see some artefacts, for example the legs of the letter ‘m’ in the word ‘same’ are connected and the curvature along the edges are not smooth. Different observations of the same characters are also reconstructed each time in a different way.

In Figure 6.17, we show the results for our non-local method and a binarized version (obtained via simple uniform quantization). The reconstructed characters show significant improvements compared to the other methods: the curvature of the characters is much smoother and less annoying artefacts are created. Also note that the font type is more or less preserved by the proposed method.

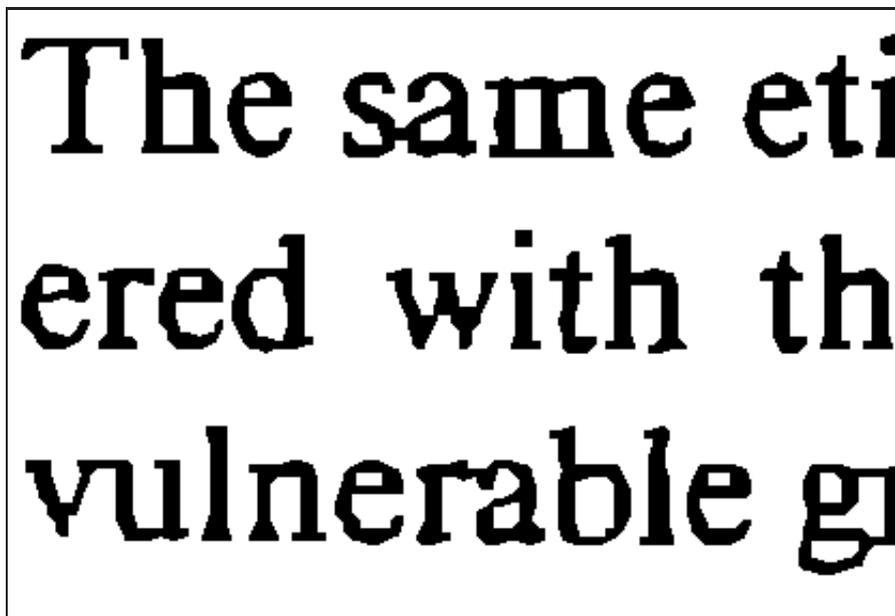
In Figure 6.18, we show a full-scale part of the document processed by the different restoration techniques. At first sight, the restored text images show not a lot of difference compared to the original image. The image restoration will not improve the readability for most people.

We also compute the OCR accuracy over the 4 pages using several popular commercial and open-source page-reading systems: Scansoft OmniPage 15.0, ABBYY FineReader 8.0, Microsoft Office 2003 Document Imaging, gOCR 0.45 and Tesseract 2.01. All page-reading systems give 100% correct OCR results, except gOCR 0.45: these results are given in Figure 6.18. We notice that our method produces slightly better recognition results, which improves the readability for a machine.

²We would like to acknowledge the library of the university of Antwerp for providing this document.

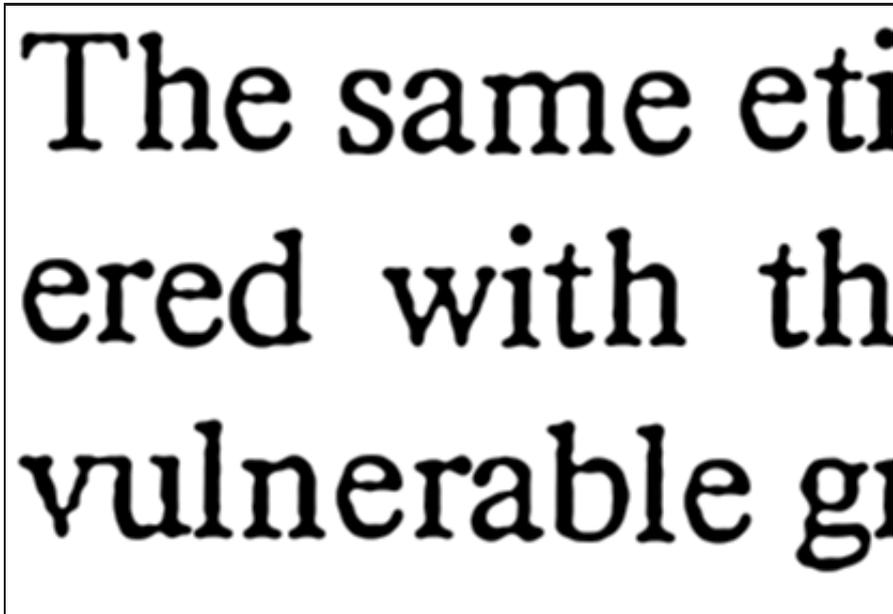


(a) Nearest neighbour



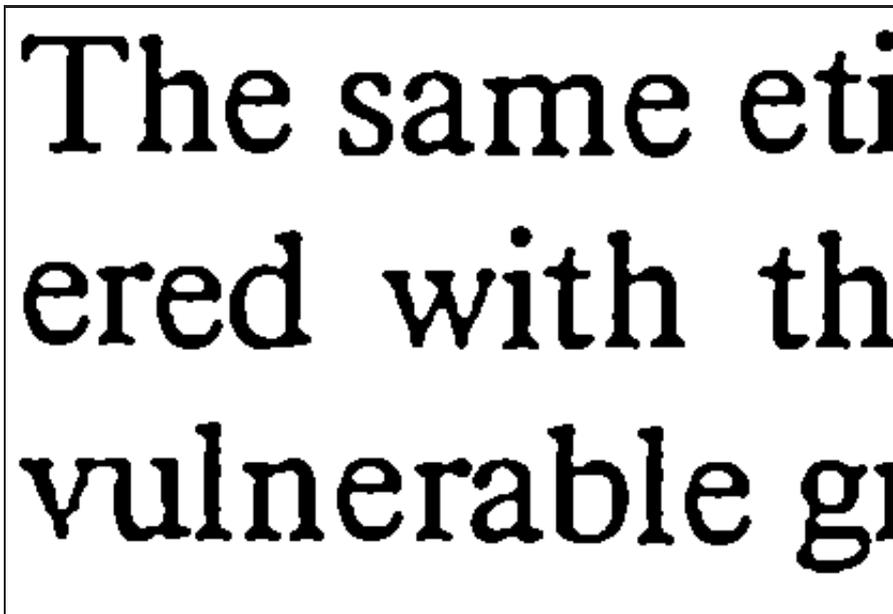
(b) Mathematical morphology reconstruction and interpolation

Figure 6.16: Image restoration results (2× enlargement of a binary document).



The same etiered with the vulnerable gr

(a) Proposed method (greyscale)



The same etiered with the vulnerable gr

(b) Proposed method (binary)

Figure 6.17: Image restoration results (2× enlargement of a binary document).

It is also of note that there was a significant difference in adoption rates (EFA) between first offenders and repeaters. The same etiological questions may be raised as those considered with the adopted psychiatric subjects: (1) a genetically vulnerable group with a predictable course (Crowe, 1975); (2) a group prone to delinquency because of the adoption experience and impairment in early bonding formation, as many investigators have suggested; (3) discrepant treatment (harsher dispositions) by a juvenile court of adopted delinquents (Lewis et al., 1975) and possible conditioning to delinquency. Again, it is important to consider all biopsychosocial factors together

(a) Original scan (93.7%)

It is also of note that there was a significant difference in adoption rates (EFA) between first offenders and repeaters. The same etiological questions may be raised as those considered with the adopted psychiatric subjects: (1) a genetically vulnerable group with a predictable course (Crowe, 1975); (2) a group prone to delinquency because of the adoption experience and impairment in early bonding formation, as many investigators have suggested; (3) discrepant treatment (harsher dispositions) by a juvenile court of adopted delinquents (Lewis et al., 1975) and possible conditioning to delinquency. Again, it is important to consider all biopsychosocial factors together

(b) Mathematical morphology (82.3%)

It is also of note that there was a significant difference in adoption rates (EFA) between first offenders and repeaters. The same etiological questions may be raised as those considered with the adopted psychiatric subjects: (1) a genetically vulnerable group with a predictable course (Crowe, 1975); (2) a group prone to delinquency because of the adoption experience and impairment in early bonding formation, as many investigators have suggested; (3) discrepant treatment (harsher dispositions) by a juvenile court of adopted delinquents (Lewis et al., 1975) and possible conditioning to delinquency. Again, it is important to consider all biopsychosocial factors together

(c) Proposed method (94.8%)

Figure 6.18: Full-scale image restoration results. The OCR accuracy result of gOCR 0.45 is given in parentheses.

6.5 Conclusion

In this chapter, we have described robust reconstruction techniques to enhance the quality of low-resolution greyscale and binary document images. Exploiting the multiple occurrences of characters brings more information at our disposal, which leads to much better estimates of the unknown pixel values. In order to take advantage of this repetitive behaviour in a practical way, we divide the image into character segments. The character segmentation reduces the computation time drastically in two ways: the algorithm only has to focus on these regions of interests and the search space for possible matching candidates is reduced enormously. Matching between the character segments filters relevant information before the reconstruction. Information originating from other similar characters are combined and the characters are reconstructed in the Bayesian framework for greyscale text images. For binary document images, we have proposed a simple and fast noise smoothing scheme.

Results of different experiments show the effectiveness of the proposed intra-frame super-resolution technique: characters and symbols are reconstructed very well and OCR results show a significant improvement of our method compared to other reconstruction methods. A trivial extension to our method is to take multiple pages of the same document, journals or book into account or to combine our method with multi-frame super-resolution techniques (for video applications). This would produce even better results because there is more repetitive information available. The proposed method can also deal with documents irrespective to their font type, it even preserves the font type and is not restricted to characters of a particular alphabet, i.e. the document can also contain generic symbols such as musical notes, hieroglyphics or mathematical symbols.

The strategy of using the repeating symbol property is not restricted to the reconstruction of document images that suffer from noise, compression artefacts, LR scanning, wear processes (e.g. in old manuscripts), etc., but can also be applied in an exemplar-based search engine and combined with an efficient document compression scheme (such as described in [Kia et al., 1998, Yang et al., 2000]) for instance. The latter is useful for the storage of large digital libraries or for transmitting documents. repeating characters contain redundant information, this redundancy can be removed for compression by constructing a prototype for each class/cluster of characters and encode the remaining reconstruction errors [Kia et al., 1998, Yang et al., 2000]. Exemplar-based search is useful in cases where OCR fails: a symbol, a prototype or a set of symbols can be suggested as a search string after which the search engine reports all similar symbols. OCR typically fails in situations where the characters are degraded too much or in case the alphabet is unknown (e.g. due to an exotic font type, foreign symbols, mathematical equations, etc.).

7

Multi-frame super-resolution restoration

It is through science that we prove, but through intuition that we discover.
—Jules Henri Poincaré

In the last decades, the use of multiple images in the restoration process has gained a lot of popularity among various researchers, especially in the digital video processing community. Because video typically consists of a sequence of non-identical, but still very similar frames, it becomes possible to use the inter-frame motion information for improving the image resolution. This leads for example to excellent results for video denoising [Zlokolica, 2006], video deinterlacing [Kwon et al., 2003, Van De Ville, 2001, Zlokolica et al., 2006], multiple image (motion) deblurring [Katsaggelos, 1990, Šroubek and Flusser, 2005, Rav-Acha and Peleg, 2005], etc. In this chapter, we concentrate on resolution enhancement, i.e., on the multi-frame *super-resolution* (SR) restoration problem.

As discussed in Chapter 2, multi-frame SR image restoration becomes most successful if there is motion over a non-integral distance between the frequency-aliased images. In contrast to single-frame image restoration, multi-frame techniques generate superior results (i.e. the recovery of *true* high frequency content) provided that the inter-frame motion is estimated correctly. However, even in absence of frequency aliasing (which is the case in MRI images), it is sometimes possible to restore the true high frequencies under restricted conditions as discussed in Chapter 8.

In this chapter, we focus on a three-step implementation of the SR framework, which consists of image alignment, image fusion and image restoration. In the following sections, we briefly discuss the existing SR approaches (Section 7.1), motion estimation and registration techniques (Sections 7.2 and 7.3) and reconstruction algorithms that fuse multiple aligned LR grids into one HR grid (Section 7.4).

In Section 7.3.2, we derive a new robust image alignment technique that performs joint geometric and photometric registration in the *total least square* sense. The novel idea is to use the total least square metrics instead of the ordinary least square metrics, which is commonly used in the literature. In image registration applications, the total least square solution of the motion parameters is much more consistent and accurate in presence of noise compared to existing registration algorithms.

In Section 7.4.3, we also propose an improved fusion algorithm in the total least square sense, which is robust to additive noise *and* registration errors in contrast to existing fusion techniques. These improvements are important for the performance of multi-frame super-resolution image reconstruction algorithms. Furthermore, we show the effectiveness of the proposed SR method in real-world examples and deinterlacing problems and we discuss SR reconstruction along the temporal direction in Section 7.5. Finally, Section 7.6 concludes this chapter.

7.1 Survey of multi-frame super-resolution techniques

For excellent and recent surveys on multi-frame SR restoration methods, we refer the interested reader to [Borman, 2004, Park et al., 2003]. These methods can roughly be divided into three categories, i.e., the frequency-domain methods, the spatial-domain methods and the motionless methods. In the following sections, we give a very concise overview of a wide variety of different existing SR approaches.

7.1.1 Frequency-domain super-resolution restoration

Frequency-domain methods make explicit use of the frequency aliasing that is introduced during the acquisition of the LR images. In 1984, Tsai and Huang first presented such type of multi-frame methods to increase the resolution of satellite images [Tsai and Huang, 1984]. Based on the shift and aliasing properties¹ in the Fourier domain, the authors compute the Fourier coefficients of the bandlimited HR image from a set of linear equations, which relates the discrete Fourier transform (DFT) of the observed LR images to the continuous Fourier transform (CFT) of the HR image. The final image is obtained via the inverse Fourier transform.

In [Kim et al., 1990], the authors proposed an extension of this approach to address the problems of noise and blur during acquisition, resulting in a weighted least squares formulation. Bose et al. proposed the use of recursive total least squares method for SR restoration to reduce the effect

¹The aliasing property tells us that the observed frequency spectra are sums of the phase modulated true spectra and their replicas (see Section 2.1.3). The shift property describes the spatial pixel shifts as linear phase modulations of the Fourier coefficients.

of registration errors [Bose et al., 1993]. In [Rhee and Kang, 1999], the authors reduced memory requirements and computational costs by employing the *discrete cosine transform* (DCT) instead of the DFT. More variants on the frequency-based methods can be found in [Borman, 2004, Park et al., 2003].

Although the frequency-domain methods are intuitively simple and computationally attractive, the observation model is restricted to global *rigid* geometric transformation and spatially invariant linear blur, which limits their utility in practical applications. These frequency-domain approaches can also not incorporate general *a priori* constraints and degradation models in the restoration process (see also discussion in Section 4.1.2.2).

7.1.2 Spatial-domain super-resolution restoration

Since SR is typically an ill-posed inverse problem, regularization is often required to achieve acceptable results. As discussed in Section 4.1.2.2, the formulation of restoration techniques in the spatial domain offers a large flexibility in the range of degradation and registration models with the ability to include a large variety of prior information. There are four main trends in spatial-domain SR approaches, i.e., *back-projection* methods, POCS-based techniques, stochastic (or Bayesian) methods and interpolation of irregularly spaced samples. This classification is very similar to the single-frame restoration trends as discussed in Chapter 4, except for the non-uniform interpolation class.

Similar to the single-frame restoration problem, we can describe the relationship between a continuous scene and a set of k (discrete) digital images. In Figure 7.1, the observation model is illustrated with the inherent degradation operations. Therefore, the i th observed LR image \mathbf{g}_i is related to the desired HR image \mathbf{f} by the following matrix-vector formulation:

$$\mathbf{g}_i = \mathbf{D}_i \mathbf{H}_i \mathbf{M}_i \mathbf{f} + \mathbf{n}_i = \mathbf{A}_i \mathbf{f} + \mathbf{n}_i, \quad (7.1)$$

where $\mathbf{g}_i \in \mathbb{R}^m$ are the observed images and $\mathbf{f} \in \mathbb{R}^{r^2 m}$ is the ideal (or hypothetical desired) image, which has a resolution that is r times higher than that of the \mathbf{g}_i . In this equation, the matrices \mathbf{D} , \mathbf{H} and \mathbf{M} (with dimensions of $m \times r^2 m$, $r^2 m \times r^2 m$ and $r^2 m \times r^2 m$) represent the decimation operator, the blur operator and the (motion) warping operator respectively and $\mathbf{n}_i \in \mathbb{R}^m$ describes the additive noise. The combination of decimation, blur and transformation can be represented by a general linear degradation operation $\mathbf{A}_i = \mathbf{D}_i \mathbf{H}_i \mathbf{M}_i$ with a matrix dimension of $m \times r^2 m$.

A popular class of SR restoration techniques deals with a simulate-and-correct approach, which is known as *iterative back-projection* (IBP) [Peleg et al., 1987, Keren et al., 1988]. In each iteration, the current estimated HR image is corrected or updated by minimizing the residual error between the observed and the simulated LR images on the LR grid. Given an estimate $\hat{\mathbf{f}}^{(j)}$ of the HR image

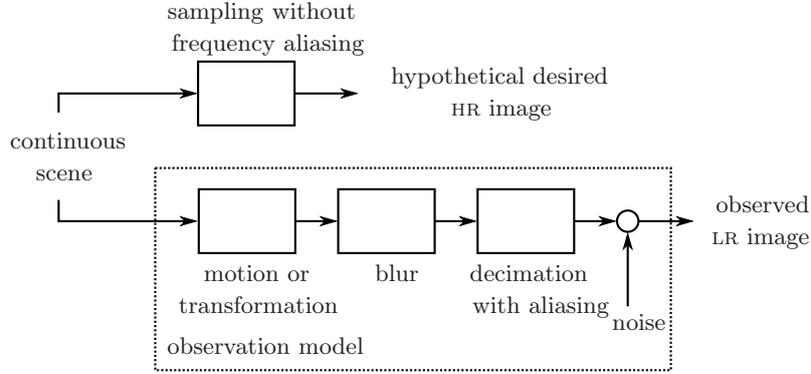


Figure 7.1: Observation model for SR image reconstruction of a set of noisy, blurred low-resolution images.

at the j th iteration, we can easily simulate the set of LR images by applying equation (7.1) to $\hat{\mathbf{f}}^{(j)}$ as

$$\hat{\mathbf{g}}_i^{(j)} = \mathbf{A}_i \hat{\mathbf{f}}^{(j)}. \quad (7.2)$$

The IBP algorithm updates the current estimate of the SR solution by projecting the residual errors via the back-projection operator \mathbf{A}_i^\dagger , which is often the adjoint operator of the approximating linear degradation model [Borman, 2004]. The iterative procedure is formally described by

$$\hat{\mathbf{f}}^{(j+1)} = \hat{\mathbf{f}}^{(j)} + \sum_{i=1}^k \mathbf{A}_i^\dagger (\mathbf{g}_i - \mathbf{A}_i \hat{\mathbf{f}}^{(j)}). \quad (7.3)$$

In [Peleg et al., 1987, Keren et al., 1988], the authors applied the IBP procedure to minimize the *sum of absolute differences* (SAD) between the simulated and the original observed images, where \mathbf{A}_i^\dagger projects the correction of a LR pixel to one HR pixel. In [Irani and Peleg, 1991], Irani and Peleg extended this work by modifying the error functional to the *sum of squared differences* (SSD) and introduce a back-projection *kernel* in the update procedure of the HR image, i.e. \mathbf{A}_i^\dagger is constructed in such a way that a HR pixel is updated by *all* contributing LR pixels according to the PSF of the imaging model. In [Zomet et al., 2001], the authors replaced the sum of projected errors in equation (7.3) by the pixelwise median, which results in a very robust SR algorithm. The IBP approach is intuitively simple, but is neither theoretic optimal, flexible or general (it is very hard to include prior constraints). The accuracy of the IBP algorithm is only as good as the observation model and a solution does not always exist or is not unique due to the ill-posed nature of the inverse problem. The final HR image depends heavily on the initial guess $\hat{\mathbf{f}}^{(0)}$.

A closely related algorithm is the non-iterative algebraic filtered back-projection method [Borman, 2004], which computes the *linear minimum mean square error* (LMMSE) estimate of the HR image $\hat{\mathbf{f}} = (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger \mathbf{g}$ (note

that the set of observed images is aggregated in \mathbf{g} and that the solution can efficiently be computed by Fourier techniques if $\mathbf{A}^\dagger \mathbf{A}$ has a CBC structure.²⁾ However, such an inverse filtering is highly sensitive to noise and therefore, regularization is needed to achieve acceptable performance in the presence of noise or missing data.

A more powerful framework to include prior knowledge is the POCS-based approach as introduced for the multi-frame SR problem by Stark and Oskoui [Stark and Oskoui-Fard, 1989] (see also discussion in Section 4.1.2.2). The data fidelity constraint projects each HR pixel into a certain bound (reflecting the statistical confidence) of the contributing pixels of all LR images. To improve the final HR image, additional image priors such as non-negativity constraints can be included easily in the iterative process of equation (4.14). Later, this model is extended to include spatially varying blur and observation noise [Tekalp et al., 1992], non-zero aperture time (or motion blur), arbitrary sampling lattices [Patti et al., 1997] and multiple object motions by introducing the concept of validity and/or segmentation maps [Eren et al., 1997], etc. Tom and Katsaggelos estimated the HR image via the alternative bounding ellipsoid method [Tom and Katsaggelos, 1996].

An alternative and very popular framework to include prior knowledge is known as the stochastic or Bayesian approach. These regularization techniques are extensively discussed in Section 4.2. In the presence of additive zero-mean white Gaussian noise and according the degradation model described in equation (7.1), the traditional *maximum a posteriori* (MAP) estimate of the HR image $\hat{\mathbf{f}}_{\text{MAP}}$ becomes

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg \min_{\mathbf{f}} \left(\sum_{i=1}^k \|\mathbf{g}_i - \mathbf{D}_i \mathbf{H}_i \mathbf{M}_i \mathbf{f}\|_2^2 + \alpha G(\mathbf{f}) \right), \quad (7.4)$$

where $G(\mathbf{f})$ denotes a well-chosen image prior (see Section 4.3.1 for some examples) and α is the regularization parameter. In [Cheeseman et al., 1994], Cheeseman et al. applied Bayesian estimation with a Gaussian prior model to the SR problem of multiple satellite images observed by the Viking Orbiter. Schultz and Stevenson used a robust *Huber-Markov random field* (HMRF) Gibbs image prior model to preserve edge discontinuities [Schultz and Stevenson, 1996]. Hardie et al. presented a MAP framework for the joint estimation of image registration parameters and the HR image [Hardie et al., 1997]. Zibetti and Mayer proposed a robust and edge-preserving simultaneous MAP SR algorithm that estimates all HR images simultaneously with an additional prior term that enforces similarity of the HR images along the motion trajectories [Zibetti and Mayer, 2006, Zibetti and Mayer, 2007].

²⁾It is important to notice that the back-projection matrix \mathbf{A}^\dagger is not, in general, equivalent to the inverse \mathbf{A}^{-1} .

In literature, several *maximum likelihood* (ML) estimators are applied to the SR reconstruction problem. However, due to the ill-posed nature of the SR inverse problems, MAP estimation is usually preferred to ML. Tom and Katsaggelos proposed a ML SR algorithm that estimate the registration parameters, noise variances of each degraded LR image and the HR image simultaneously by the *expectation maximization* (EM) algorithm [Tom and Katsaggelos, 1995]. Elad and Feuer proposed a ML-POCS hybrid reconstruction approach that combines the benefits of the stochastic approach (ensuring a single optimal solution) and the POCS approach (utilizing powerful *a priori* constraints in a simple intuitive way) [Elad and Feuer, 1997]. Note that steepest descent solution of the ML estimation is actually a special case of the IBP procedure as expressed in equation (7.3) with $\mathbf{A}_i^\dagger = \mathbf{M}_i^T \mathbf{H}_i^T \mathbf{D}_i^T$. A more detailed survey of these stochastic methods can be found in for example [Park et al., 2003, Borman, 2004].

In the case of global translational motion and common stationary blur models (i.e. $\forall i : \mathbf{H}_i = \mathbf{H}$), we can write the ML estimation of the HR image $\hat{\mathbf{f}}_{\text{ML}}$ in the presence of additive zero-mean white Gaussian noise as

$$\hat{\mathbf{f}}_{\text{ML}} = \arg \min_{\mathbf{f}} \left(\sum_{i=1}^k \|\mathbf{g}_i - \mathbf{D}_i \mathbf{H} \mathbf{M}_i \mathbf{f}\|_2^2 \right). \quad (7.5)$$

Because \mathbf{H} and \mathbf{M}_i have a CBC structure, they commute (i.e. $\mathbf{H} \mathbf{M}_i = \mathbf{M}_i \mathbf{H}$ and $\mathbf{H}^T \mathbf{M}_i^T = \mathbf{M}_i^T \mathbf{H}^T$). Therefore, we can rewrite equation (7.5) as

$$\hat{\mathbf{f}}_{\text{ML}} = \arg \min_{\mathbf{f}} \left(\sum_{i=1}^k \|\mathbf{g}_i - \mathbf{D}_i \mathbf{M}_i \mathbf{H} \mathbf{f}\|_2^2 \right). \quad (7.6)$$

We define $\mathbf{h} = \mathbf{H} \mathbf{f}$ as the blurred version of the hypothetically desired HR image and therefore, we can divide the SR reconstruction problem into three separate steps [Elad and Hel-Or, 2001]:

- *image registration*: this is the estimation of transformation matrices \mathbf{M}_i .
- *image fusion*: this is the estimation of the blurred HR image $\hat{\mathbf{h}}$ from the observed LR data:

$$\hat{\mathbf{h}}_{\text{ML}} = \arg \min_{\mathbf{h}} \left(\sum_{i=1}^k \|\mathbf{g}_i - \mathbf{D}_i \mathbf{M}_i \mathbf{h}\|_2^2 \right). \quad (7.7)$$

- *image restoration*: this is the estimation of the desired HR image $\hat{\mathbf{f}}$ from $\hat{\mathbf{h}}$.

In [Farsiu et al., 2004], the authors replaced the l_2 -norm in equation (7.7) by the l_1 -norm, resulting in a far more robust estimate of $\hat{\mathbf{h}}$ and they performed a MAP restoration technique with the bilateral total variation image prior. This three-step algorithm is an example of the more general class of non-uniform interpolation SR techniques, which consist of three consecutive stages

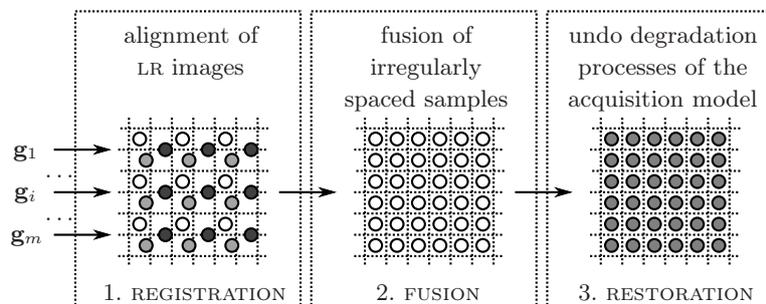


Figure 7.2: Schematic representation of the standard three-step paradigm of the non-uniform interpolation SR approach.

as illustrated in Figure 7.2: motion estimation (see Section 7.2 for a more detailed discussion), interpolation of the irregularly spaced samples on a regular HR grid (see Section 7.4 for more details) and restoration to remove blur and noise (see Chapter 4 for an overview).

A disadvantage of this non-uniform interpolation approach is that the degradation models are limited (e.g., blur and noise characteristics must be the same for all LR images) and that the optimality of the whole reconstruction process is not guaranteed because errors can propagate from one stage into another. However, due to its relatively low computational load and low memory requirements (because the algorithm does not require all irregularly spaced LR samples during the restoration process), it is the most suitable approach to real-time applications. That is why we concentrate on this class of SR techniques in the rest of this chapter.

7.1.3 Motionless super-resolution restoration

The SR restoration algorithms discussed so far require the estimation of relative motions over non-integral distance between the observed LR images. However, it is possible to perform multi-frame SR reconstruction without explicit motion estimation or image registration. Therefore, we refer these techniques to as *motionless* or *motion-free* SR techniques. We briefly discuss the principles of these methods.

In [Elad and Feuer, 1997], Elad and Feuer applied multi-frame SR restoration on differently blurred images without relative motion (i.e. a fixed camera filming fixed objects with different defocusing) using a hybrid ML-POCS approach. Rajan and Chaudhuri proposed a similar motionless SR technique for blurred images using a MAP framework with Markov random field priors [Rajan and Chaudhuri, 2002]. In [Chaudhuri and Joshi, 2005], the authors derived motionless SR techniques for images with relative zoom and for images with different photometric characteristics, i.e. where the scene to

be super-resolved is illuminated with a moving light source.

In recent years, a growing number of learning-based or exemplar-based methods have been proposed in the literature to tackle the SR problem (see also discussion in Section 5.1). This approach is only effective when very few observed LR images are available. Due to the need of a vast number of examples, exemplar-based methods are used in very specific scenarios, such as face or text restoration [Baker and Kanade, 2002, Park and Lee, 2008]. Note that the restoration of high frequency content is only apparent and not real. That is why these learning-based methods are also referred to as hallucination algorithms. Some examples of the learning-based methods can be found in [Baker and Kanade, 2002, Freeman et al., 2002, Datsenko and Elad, 2007].

As discussed in Chapter 5, non-local resolution enhancement techniques are closely related to SR techniques, except that each pixel may be assigned to *several* motion vectors, and not just one as in classical motion estimation methods. Note that these fuzzy “motion” vectors are not really motion vectors but displacement vectors point to similar or repetitive patches. Extension of single-frame to multi-frame SR techniques is relatively easy. Rough motion estimation (i.e. without subpixel accuracy) with limited search windows can help to reduce the computational complexity enormously. Some examples of multi-frame non-local SR techniques are discussed in [Ebrahimi and Vrscay, 2008a, Protter et al., 2009, Danielyan et al., 2008b].

7.1.4 Super-resolution in the temporal direction

Recently, the challenge of simultaneous resolution enhancement in time as well as space has received growing attention in the video processing community [Robertson and Stevenson, 2001, Schechtman et al., 2005]. Resolution enhancement in the temporal direction corresponds to the increase of the *frame rate* of the video sequence, which is the number of distinct and complete images that the camera can capture per second. The frame rate is often expressed in *frames per second* (FPS) or in hertz (Hz). The standard frame rate in PAL and SECAM television is 25Hz (in *progressive* or *non-interlaced* format, i.e. 25 actual FPS), while the corresponding standard *field rate* is 50Hz (in *interlaced* format, i.e. 50 fields or half images per second). In NTSC television, the standard field rate and frame rate are respectively 60Hz and 30Hz (or $30 \times 1000/1001 = 29.97$ FPS to be more precise). In high-end HDTV systems, frame rates of 50Hz and 60Hz are commonly used. In scenes with high motion activities, a low frame rate or dropped frames may cause motion to appear jerky and uneven.

In [Robertson and Stevenson, 2001], the authors increased the frame rate by inserting interpolated images between the received frames. Spatial compression artefacts are further reduced in a Bayesian restoration framework. In [Schechtman et al., 2005], Schechtman et al. combined differ-

ent video streams of the same event to reconstruct a video sequence with a higher resolution in spatial and temporal domain. The different video streams are aligned using a *sequence-to-sequence* registration method [Caspi and Irani, 2002]. In contrast to the work of Robertson and Stevenson [Robertson and Stevenson, 2001], Schechtman et al. are able to restore the motion-based temporal aliasing (e.g. the *wagon-wheel* effect as discussed in Section 2.1.3) through super-resolution.

Another interesting benefit of SR in the temporal direction is that the restoration of object motion blur as a spatial artefact can be handled properly. The camera integrates the light coming from the scene during the exposure time in order to generate each image. As a result, fast moving objects produce a noticeable blur along their trajectory, often resulting in distorted or unrecognizable object shapes. Thus, motion blur is actual a temporal phenomenon (with spatial artefacts) and not a spatial one [Schechtman et al., 2005].

In single-frame motion blur restoration techniques, the PSF is modelled as 1D uniform local averaging of neighbouring pixels in the direction of the object movements. In combination with a static background, prior segmentation and motion analysis of the object is needed, which is not always possible. In contrast to these single-frame techniques, multi-frame motion blur restoration performs a deconvolution in the temporal direction with a temporal rectangular blur function that models the exposure time. So, motion deblurring can be done directly by reducing the exposure time without requiring explicit motion analysis, segmentation or any scene interpretation. Because of the use of multiple frames, image information that is lost due to partial occlusion effects can also be recovered. To support these claims, we set up a small experiment in Section 7.5.1.

Resolution enhancement in the temporal domain is not only important for the frame rate conversion, e.g. displaying DVD on HDTV screens, but we can also reduce the cost of high-speed camera systems, which acquire images at very high frame rates (typically 1000 to 10000 FPS). Such frame rates only permit very low exposure times (with a lot of shot noise as a result) and require extreme lighting conditions of the scene. With SR in the temporal direction, we can build a high-speed camera system from several cheap camera's with a lower frame rate and thus higher exposure times (with the result that there is less shot noise and that lighting conditions are less strict). Another interesting application is to merge differently scalable coded versions of a video stream or combining still HR images with a LR video stream to obtain a better video quality and resolution, both in the spatial domain and the temporal domain.

7.1.5 On the performance of super-resolution techniques

A major weakness of many papers on SR algorithms is the lack of performance quantification. Sometimes, the algorithms are evaluated on simulated data, where an original HR image is processed to obtain multiple degraded LR images,

which are then used for the SR algorithm. In this case, objective metrics such as PSNR or MSE can be computed. Note that these synthetic LR images often contain a lot of frequency aliasing, which makes the visual end result very spectacular but not realistic because in practical data, frequency aliasing is obscured by blur, noise and/or compression. On the other hand, for real-world data, no HR reference image is available (except if the scene is also captured by another camera in high quality). There are several possibilities to evaluate the SR reconstruction results.

In [Pham, 2006], Pham proposed a range of objective measures such as image sharpness, jaggedness of edges and the smallest available blob size to evaluate the results of different SR algorithms. These quantitative measures are presented in a radar chart per reconstructed HR image so that their strengths and weaknesses (i.e. artefacts) are easily comparable. Reibman and Shaper attempted to perform subjective performance evaluation on SR results [Reibman and Schaper, 2006].

Various detection algorithms in specific applications help to measure the performance in an objective and clear way. In multi-frame text SR applications, OCR algorithms measure the readability of characters [Donaldson and Myers, 2005, Mancas-Thillou and Mirmehdi, 2007]. For some specific task-based vision tasks, the discrimination of small objects/details in an image is measured quantitatively using *triangle orientation discrimination* (TOD), which measures the ability to discriminate the orientation of an equilateral triangle under specific conditions [Van Eekeren et al., 2007].

In the last decades, various researchers have explored the limits on the performance of SR algorithms. Schultz and Stevenson claim that the accuracy of motion estimation is the limiting factor in SR restoration performance [Schultz and Stevenson, 1996]. A possible approach to minimize the effect of the registration errors is based on channel adaptive regularization [Lee and Kang, 2003]. The basic concept of channel adaptive regularization is that degraded LR images with a large amount of registration error should contribute less to the estimate of the HR image than reliable LR images. In [Marziliano and Vetterli, 2000], the authors developed numerical methods that reconstruct a bandlimited signal from an irregular set of samples at *unknown* locations, which can help to overcome possible registration errors.

In [Baker and Kanade, 2002], Baker and Kanade showed that any smoothness prior results in SR reconstructions with very little high-frequency content for sufficient large enlargement factors. Therefore, they proposed to break this limit via exemplar-based image priors. Lin and Shum derived performance limits for the case of translational motion, using matrix perturbation analysis [Lin and Shum, 2004]. They showed that under practical conditions, SR reconstruction-based algorithms can improve the resolution by at most a magnification factor $r = 1.6$. Cramér-Rao lower bounds for joint image registration and SR reconstruction problems are derived by Robinson and Milanfar [Robinson and Milanfar, 2006]. They showed that SR performance

depends on a complex relationship between measurement SNR, the number of observed frames, set of relative motions between frames, image content and the PSF of the imaging system. In [Shahram and Milanfar, 2006], the authors addressed the limitations of optical systems, where the objective is to study how far beyond the classical Rayleigh resolution limit (resolving two closely spaced point sources) one can reach at a given SNR.

Relatively recently, various researchers (e.g. [Tom and Katsaggelos, 1995, Hardie et al., 1997]) focused on the estimation problems related to joint image registration and super-resolution. This avoids the chicken-and-egg problem: SR reconstruction needs accurate subpixel motion information and accurate motion estimation needs clean HR images. For the same reasons, the authors of [Shen et al., 2007] jointly estimated the motion, the HR image and the segmentation map.

In many image processing problems, super-resolution methods can be combined with other application-specific restoration algorithms. Shekarforoush et al. presented a Bayesian method to jointly reconstruct a SR image and the surface albedo and height of the scene objects [Shekarforoush et al., 1996]. In [Farsiu et al., 2006a], the authors performed SR restoration on images sampled with a Bayer CFA pattern. The SR reconstruction of image mosaics is very popular in the literature, see e.g. [Zomet and Peleg, 2000, Ye et al., 2005]. Altunbasak et al. proposed a transform-domain SR restoration algorithm to tackle block-based compression artefacts by incorporating DCT quantization information from MPEG sequences [Altunbasak et al., 2002]. In [Gevrekci and Gunturk, 2007], Gevrekci and Gunturk addressed the problem of super-resolution and *high dynamic range* (HDR) imaging in a unified context.

So far, we have only discussed the classical multi-frame to single-frame (or *multi-input single-output* (MISO)) SR methods. The multi-frame to multi-frame (or *multi-input multi-output* (MIMO)) SR algorithms for a sequence of images or video can be classified into three categories [Zibetti and Mayer, 2007]:

- *traditional* methods: these methods estimate one (reference) HR image at a time from a set of SR images. This procedure is repeated for all HR images.
- *sequential* methods: these methods also estimate one HR image at a time, but in contrast to the traditional methods, sequential methods take advantage of the previously estimated HR images to reduce the cost of the estimation of subsequent HR images, e.g. by employing an efficient approximation of the Kalman filter [Elad and Feuer, 1999, Farsiu et al., 2006b].
- *simultaneous* methods: these methods estimate all HR images in one process, e.g. [Zibetti and Mayer, 2006, Zibetti and Mayer, 2007]. A major drawback of this approach is the large memory requirements because all (temporary) HR images must be stored during several iterations.

It is obvious that sequential methods are preferable to near real-time applications in terms of computational load and simultaneous methods are preferable to achieve the highest possible quality.

7.2 Motion estimation

In most multi-frame SR algorithms, motion estimation or image registration is the key to successful SR reconstruction. The estimation of motion plays an important role in many other applications as well, e.g. mosaicing, compression and tracking. Motion estimation algorithms consist of three basic elements, namely, the *motion model* (i.e. the representation of the *motion field* combined with the region of support), the *matching criterion* (i.e. the cost function together with some penalties or prior constraints that have to be optimized) and the *search strategy* (i.e. the way the cost function is minimized) [Stiller and Konrad, 1999].

In the following sections, we will make a rough classification of motion estimation techniques according to their motion model and their region of support more specifically. There exists two kinds of motion representations. The most common and most popular representations are the parametric models, some of them are summarized for 2D motion in Table 7.1. The second kind are the non-parametric models such as *elastic* or *thin-plate spline* deformation models. For a more detailed treatment of these motion representations, we refer the reader to [Glasbey and Mardia, 1998].

In general, the higher the number of parameters, the more precise the description of the motion field can be. A motion field is a vector-valued image of continuous spatial coordinates that contains the perceived or estimated projection of the real 3D motion displacements. However, an excessive number of parameters may result in motion *over-modelling* [Stiller and Konrad, 1999]: like SR reconstruction, motion estimation is an ill-posed inverse problem. The ill-posedness leads to multiple possible solutions, i.e. no unique solution is guaranteed due to the aperture and occlusion problem. Therefore, regularization is needed to obtain a unique motion field. This regularization can be incorporated explicitly as an additional smoothness term in the cost function, or implicitly by assuming that all pixels within a specific region follow the same motion.

The choice of the region of support leads to the following classification:

- *pixel*-based methods: each pixel has a different motion vector. The translational representation is commonly employed for pixel-based methods, which results in the use of motion vectors. Often in literature, the motion field is said to be *dense* because a motion vector is assigned to each pixel. Smoothness terms are typically used in the cost function to obtain a stable solution.
- *region*-based methods: the image is partitioned in several (non-

Table 7.1: Several popular 2D parametric motion representations where $\kappa(K)$ is the cardinality of K .

2D model	Definition	# Parameters
translational	$\begin{cases} x' = a_{00} + x \\ y' = b_{00} + y \end{cases}$	2
procrustes/rigid body	$\begin{cases} x' = a_{00} + s \cos \theta x + s \sin \theta y \\ y' = b_{00} - s \sin \theta x + s \cos \theta y \end{cases}$	4
affine	$\begin{cases} x' = a_{00} + a_{10}x + a_{01}y \\ y' = b_{00} + b_{10}x + b_{01}y \end{cases}$	6
projective/perspective	$\begin{cases} x' = \frac{a_{00} + a_{10}x + a_{01}y}{1 + c_{10}x + c_{01}y} \\ y' = \frac{b_{00} + b_{10}x + b_{01}y}{1 + c_{10}x + c_{01}y} \end{cases}$	8
polynomial	$\begin{cases} x' = \sum_{\{i,j\} \in K} a_{ij}x^i y^j \\ y' = \sum_{\{i,j\} \in K} b_{ij}x^i y^j \end{cases}$	$2\kappa(K)$

overlapping) regions, each having different motion parameters. The regularization is incorporated implicitly because all pixels in a region of support are displaced in a uniform manner.

- *image*-based methods: the region of support consists of the whole image, as a result the motion of all pixels is described by just a few parameters. These methods are also referred to as *global* motion estimation, *image registration* or *image alignment*. Image registration is also a very important tool to retrieve camera motions in many computer vision tasks.

For comprehensive surveys on motion estimation and image registration techniques, we refer the reader to [Brown, 1992, Barron et al., 1994, Stiller and Konrad, 1999, De Smet, 2002, Zitová and Flusser, 2003]. In the following sections, we discuss the existing motion estimation methods and their relation to multi-frame SR performance.

7.2.1 Pixel-based motion estimation

In literature, the pixel-based motion estimation techniques are commonly known as *optical flow*³ methods. For a survey on optical flow techniques,

³Optical flow refers to the idea that changes in intensity is only caused by motion and that a moving pixel preserves its greyvalue. Thus technically speaking, region-based and image-based motion estimation techniques are as a matter of fact also optical flow methods.

we refer to [Barron et al., 1994, Brox, 2005]. The first and most optical flow techniques assume that pixel intensities are constant over a short time interval, i.e., the *constant image brightness* condition. Practically in 2D motion estimation, this implies that the pixel intensity at position (x, y) in the image at time t is (almost) the same as the shifted pixel $(x + u, y + v)$ in a successive image at time $t + 1$. This gives the following non-linear equation in u and v :

$$f(x, y, t) = f(x + u, y + v, t + 1). \quad (7.8)$$

In order to solve u and v , we linearize the right-hand side of the equation by a first order Taylor series expansion:

$$f(x, y, t) = f(x, y, t) + f_x u + f_y v + f_t + \epsilon, \quad (7.9)$$

which yields immediately the well-known optical flow formula:

$$f_x u + f_y v + f_t = 0, \quad (7.10)$$

or alternatively, in vector-matrix formulation:

$$\begin{bmatrix} f_x & f_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + f_t = (\nabla f)^T \mathbf{v} + f_t = 0. \quad (7.11)$$

It is easy to see that the constant image brightness assumption is not sufficient to determine a unique solution for the optical flow because there is only one equation for two unknowns. We can thus only determine the *normal flow* \mathbf{v}_\perp in the direction of the gradient in a reliable way:

$$\mathbf{v}_\perp = \frac{-f_t \nabla f}{\|\nabla f\|_2}, \quad (7.12)$$

while the *tangent flow* can *not* be determined at all, which leads to the well-known *aperture* problem as illustrated in Figure 7.3. Note that the normal flow is only reliable if $\|\nabla f\|_2 \neq 0$, i.e. at edges or in non-smooth regions. If $f_t \neq 0$, we can not measure the normal flow either.

Consequently, we need some prior knowledge or additional assumptions to obtain a unique optical flow solution. Two popular approaches are used in the literature. The first approach is known as Lucas-Kanade optical flow and assumes that the optical flow in a local neighbourhood or block can be described by the same motion vector. Pixels that are further away from the central pixel of the block get lower weights. These weights typically follow an isotropic Gaussian kernel L_σ with standard deviation σ in function of the distance to that central pixel. The optical flow solution at each pixel position can then be estimated as the minimum of the following energy function [Lucas and Kanade, 1981]:

$$(\hat{u}, \hat{v}) = \arg \min_{u, v} L_\sigma * (f_x u + f_y v + f_t)^2. \quad (7.13)$$

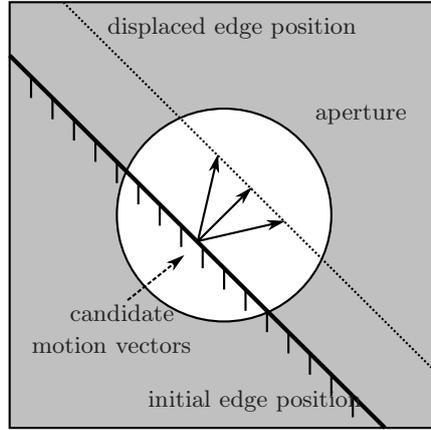


Figure 7.3: Illustration of the aperture problem: only the normal flow is measurable in a local region.

The minimum (\hat{u}, \hat{v}) can be found by setting the partial derivatives ∂u and ∂v of the energy function to 0, leading to the following linear system:

$$\begin{bmatrix} L_\sigma * f_x^2 & L_\sigma * f_x f_y \\ L_\sigma * f_x f_y & L_\sigma * f_y^2 \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} -L_\sigma * f_x f_t \\ -L_\sigma * f_y f_t \end{bmatrix}. \quad (7.14)$$

This linear system can be solved for \hat{u} and \hat{v} if the neighbourhood (defined by L_σ) is large enough to capture gradients of different directions such that both eigenvalues of the system matrix are considerably larger than 0. In [Bab-Hadiashar and Suter, 1997], the authors proposed the use of the robust *least median of squares* framework to solve the linear system of equations. This framework has a higher performance because it can handle errors such as noise as well as failures of the brightness or motion consistency.

The second popular approach is known as Horn-Schunck optical flow and incorporates regularization terms to ensure stable optical flow solutions. Horn and Schunck proposed the Tikhonov smoothness constraint and minimize the following energy function at each pixel position [Horn and Schunck, 1980]:

$$(\hat{u}, \hat{v}) = \arg \min_{u,v} (f_x u + f_y v + f_t)^2 + \alpha (\|\nabla u\|_2^2 + \|\nabla v\|_2^2). \quad (7.15)$$

The optimum of this energy function satisfies to the following two Euler-Lagrange functions:

$$\begin{cases} (f_x u + f_y v + f_t) f_x - \alpha \Delta u = 0, \\ (f_x u + f_y v + f_t) f_y - \alpha \Delta v = 0, \end{cases} \quad (7.16)$$

where Δ denotes the spatial Laplace operator.

These Euler-Lagrange equations lead to a sparse linear system of equations, which can be solved iteratively by for example Gauss-Seidel, *successive over-relaxation* (SOR) or the highly efficient multigrid methods [Bruhn et al., 2005a]. Note that the linearized optical flow formula (7.9) is only valid near the optimum, that is why *coarse-to-fine* strategy is often used (also known as hierarchical optimization) or alternatively, the motion field can be initialized with another (fast and inaccurate) method such as block matching [Kim and Sikora, 2005]. For a detailed treatment on the numerical details, we refer the reader to for example [Horn and Schunck, 1980, Brox, 2005].

From the regularization framework stated in equation (7.15), we can easily make the link to the Bayesian estimation framework (see also Chapter 4), which incorporates prior knowledge on the motion field directly in the form of image priors. This leads to the use of non-quadratic robust functions or variational terms in the smoothness assumption, which results into piecewise smooth motion fields [Black and Anandan, 1996, Brox, 2005, Bruhn et al., 2005a]. In [Feghali and Vincent, 2004], the authors employed a spatio-temporal bilateral smoothness constraint. Note that all smoothness priors can be extended easily to the spatio-temporal domain [Weickert and Schnörr, 2001]. An alternative way to preserve discontinuities in the motion field is the joint estimation of the segmentation field and the motion field (this is also known as *object-based* optical flow) [Stiller, 1997, Mémin and Pérez, 2002].

In [Bruhn et al., 2005b], the authors combined the ideas of Lucas-Kanade and Horn-Schunck into a single framework. Other improvements are obtained by non-quadratic robust and/or non-linearized constancy assumptions (see equation (7.8)), which allow for larger displacements. This leads to a *non-linear* system of equations, which can for example be solved by a steepest descent or a *fixed point iteration* scheme [Brox, 2005]. In [Papenberg et al., 2006], Papenberg et al. incorporated other constancy assumptions besides constant image brightness such as the constancy of gradients, Hessian, Laplacian, etc. These features are more invariant to varying intensity changes. Recently in [Ince and Konrad, 2008] for example, the authors take occlusion models into account to improve the optical flow estimation.

In literature, optical flow techniques are quite popular in combination with multi-frame SR restoration. A SR restoration framework that uses the Lucas-Kanade optical flow approach is described in for example [Baker and Kanade, 1999]. The Horn-Schunck method and its robust variants are used in for example [Schultz and Stevenson, 1997, Jiang et al., 2003, Fransens et al., 2004].

7.2.2 Region-based motion estimation

In region-based motion estimation techniques, we assume that the motion vectors in a particular region are denoted by the same parametric motion. A re-

gion is defined as a set of pixels, which do not necessarily need to be connected. There are three well-known approaches: block matching, segment-based and mesh-based motion estimation. We will briefly discuss each class.

7.2.2.1 Block matching methods

The basic concept of block matching is very simple. We assume that the motion field over a square region (e.g. 8×8 pixels) is constant and is described by a translational model. We find for each block in the image at time t the most similar block in the image at time $t + 1$. The motion vector is denoted by the displacement between these two blocks and is given by the minimum of the cost function (also called *matching criterion*). A popular choice of the matching criterion is the l_2 -norm or *sum of squared deviations* (SSD), i.e. the Euclidean distance between the two blocks:

$$\sum_{x,y \in \Omega} (f(x,y,t) - f(x+u,y+v,t+1))^2, \quad (7.17)$$

where Ω denotes the square block. A criterion more robust to outliers (due to noise or violations to the constancy assumption) is for example the l_1 -norm or *sum of absolute differences* (SAD):

$$\sum_{x,y \in \Omega} |f(x,y,t) - f(x+u,y+v,t+1)|. \quad (7.18)$$

Other (robust) matching criteria can be found in e.g. [Stiller and Konrad, 1999, De Smet, 2002]. In most practical applications, the search space is limited and discretized, i.e., the cost function is evaluated at a discrete number of positions (u, v) , typically for integer values of u and v . Several advanced search strategies such as *three-step search* (or *logarithmic search*) and *diamond search* have been proposed to reduce the computational complexity [Stiller and Konrad, 1999]. These very fast block matching schemes are widely used in e.g. video coding applications such as MPEG and the more advanced H.264 or AVC.

One way to achieve subpixel accuracy of u and v is to apply simple interpolation on either the image or the correlation surface, i.e. on the response of the matching criterion. A paraboloid can be fitted to the correlation surface around the optimal displacement and the maximum (or minimum in terms of costs) of the paraboloid yields the new subpixel coordinates [Luong et al., 2004c]. Another solution is to employ gradient-based motion estimation in the same spirit of Lucas-Kanade optical flow techniques [Lucas and Kanade, 1981]. In [Das Gupta et al., 2007], the authors proposed a supervised learning-based approach for the estimation of subpixel shifts from each block.

Block matching techniques have several disadvantages: in (nearly) uniform regions it is not possible to find an unambiguous motion vector. Also, the assumption of one motion vector per block is often too simplistic (which also affects

the accuracy of the motion vectors) and object contours do not coincide with the block boundaries. The one-motion-per-block assumption and the arbitrary positions of object edges are the main reasons why standard block matching is not very popular in classical super-resolution schemes because incorrect motion vectors along object edges degrades the SR image reconstruction severely. To approximate object contours, variable block sizes can be employed for instance in a quadtree decomposition [Rhee et al., 2000]. Nevertheless, block matching techniques are advantageous in an indirect way for fast initialization or approximation of motion fields [Kim and Sikora, 2005] and some components play an important role in block-based motionless SR techniques as discussed in Section 7.1.3.

7.2.2.2 Segment-based methods

In segment-based motion estimation techniques, the region of constant motions (or *segment*) is not limited to squares or rectangles, but it can take arbitrary shapes. The motion in each segment is often described by parametric transformations such as the affine model. In contrast to block-based methods, the segment-based techniques provide a more accurate prediction along moving edges. There are mainly three approaches to segment-based motion estimation [Schutten et al., 1998]:

- *bottom-up* methods: these methods start with an image segmentation with many small segments. Connected segments with similar motions are merged into one segment and motion parameters are then recalculated [LeQuang et al., 1995].
- *top-down* methods: these methods start with an initial image segmentation with large segments. If the computed motion model lacks accuracy, the segment is subdivided into smaller segments and motion parameters are calculated for each new segment. Note that quadtree decomposition is a special case of the top-down methods [Rhee et al., 2000].
- *layered representation* methods: these methods describe the video data as a set of moving layers, where each layer's motion is described by different motion parameters or smooth motion fields [Wang and Adelson, 1993]. The main difference to the previously mentioned techniques is that *disconnected* segments can be described by the same motion model. Multiple layers can be resolved sequentially (a dominant motion is calculated for each successive layer) or simultaneously (by clustering a given motion field) [Schutten et al., 1998].

Bottom-up and top-down approaches can be combined to refine the segmentation in successive iterations [Jeannin, 1996]. In [Matthews and Namazi, 1995], the authors performed motion estimation and segmentation simultaneously using an expectation-maximization algorithm. We refer the interested reader to [De Smet, 2002] for a more detailed treatment of motion-based segmentation and tracking.

7.2.2.3 Mesh-based methods

The mesh-based motion estimation methods estimate the displacement or motion vectors on a set of *control points* (or *landmarks*), which form the mesh. These control points can initially be chosen on a regular grid or at arbitrary positions across the whole image. The motion vectors are computed by block matching or feature point matching techniques (see Section 7.2.3.2 for more details). In the next step, intermediate motion vectors are then interpolated or calculated according to a specific parametric motion model.

In [Yaoping and Chengke, 1998], the authors employed Delaunay triangulation to compute the mesh. The intermediate motion vectors are computed by the affine model denoted by the three surrounding control points. In [Kybic and Unser, 2003], Kybic and Unser employed the B-spline deformation model to estimate the intermediate motion vectors. Nosratinia proposed several interpolation kernels to compute the neighbouring motion vectors [Nosratinia, 2001].

7.2.3 Image-based motion estimation

In image-based motion estimation, the transformation of the whole image is described by a single motion model. Most algorithms use transformation models listed in Table 7.1. A comprehensive survey of image registration methods is published in [Brown, 1992, Zitová and Flusser, 2003]. Henceforth we assume that the captured images come from the same or similar image acquisition device. For multimodal image registration problems (e.g. PET to MRI or Ikonos to SAR), we refer the reader to registration techniques based on information theory (e.g. the Kullback-Leibler divergence measure and its special case, mutual information) [Chung et al., 2002, Vansteenkiste, 2007].

In the following sections, we summarize the most popular approaches to image registration using Fourier techniques,⁴ feature point matching and gradient-based techniques. We also discuss a technique called *bundle adjustment* that puts additional consistency constraints on the motion estimation in order to improve the accuracy.

7.2.3.1 Fourier methods

An excellent overview of registration methods in the frequency domain can be found in [Vandewalle, 2006]. We discuss the basic concepts of Fourier-based image registration techniques. Translation parameters can be found via the shift property of the Fourier theory, which describes the relationship of pixel shifts $f(x, y, t) = f(x + u, y + v, t + 1)$ to linear phase modulations of the Fourier

⁴Correlation-based methods belongs to the class of Fourier approaches, since they can be implemented efficiently in the frequency domain.

coefficients $F(\eta, \nu, t)$ and $F(\eta, \nu, t + 1)$:

$$F(\eta, \nu, t) = e^{2\pi j(u\eta + v\nu)} F(\eta, \nu, t + 1). \quad (7.19)$$

The phase shift $e^{2\pi j(u\eta + v\nu)}$ is given by the normalized cross power spectrum:

$$e^{2\pi j(u\eta + v\nu)} = \frac{F(\eta, \nu, t)F^*(\eta, \nu, t + 1)}{|F(\eta, \nu, t + 1)F^*(\eta, \nu, t + 1)|}, \quad (7.20)$$

where the $*$ -operator denotes the complex conjugate. In the ideal case of pure translation, the inverse Fourier transform of (7.20) is a Dirac pulse in the correlation plane:

$$\mathcal{F}^{-1}(e^{2\pi j(u\eta + v\nu)}) = \delta(x + u, y + v). \quad (7.21)$$

In practice the inverse Fourier transform of (7.20) will contain a strong peak near (u, v) and the translation parameters u and v can be found by locating this peak. Rotation and scale parameters can be obtained in a similar way via the Mellin transform: rotation and scale changes become translational changes after the log-polar coordinate transform [Reddy and Chatterji, 1996]. Another way to estimate rotation is to perform radial projections, which also reduce the computational complexity and increase the robustness to noise, and the rotation angle can then be computed as the value for which the correlation between the 1D projections reaches a maximum. [Vandewalle, 2006].

Because of the discrete nature of the DFT, the peak in the correlation peak occurs at a discrete position, which limits the accuracy of the image registration. In [Shekarforoush et al., 2002], the authors showed that the signal power in the phase correlation of subpixel shifts corresponds to the polyphase transform of a filtered unit impulse and is not concentrated in a single Dirac peak. They derive a closed-form solution to obtain subpixel motion estimation. To avoid the influence of frequency aliasing on the accuracy, Vandewalle performed registration using only the aliasing-free part of the spectrum [Vandewalle, 2006].

The Fourier methods are limited to global rigid body transformations (see Table 7.1) of a single object on a uniform background. Other, more general transformation models are difficult to describe in the frequency domain. On the other hand, Fourier methods are computationally efficient, are less sensitive to object motions and noise, and offer a good framework to model frequency aliasing. Because of the good accuracy, frequency-domain techniques are quite popular in combination with multi-frame SR restoration techniques.

7.2.3.2 Feature point matching

For an extensive overview of feature point matching algorithms, we refer the interested reader to [Zitová and Flusser, 2003, Tuytelaars and Van Gool, 2004,

Mikolajczyk and Schmid, 2004]. As in mesh-based motion estimation methods, feature point matching techniques make use of *control points* (often referred to as *feature points* or *interest points*). The feature point matching procedure works in a similar way as manual registration methods. A set of object attached points in one image is compared to the same set in the other image and the correspondences are used to determine the motion of the points. The manual methods require users' interaction through selecting *ground control points* (GCP's). GCP's are points that represent important features like intersection of roads or coastlines in satellite images. Unlike the manual methods, the feature point matching techniques detect these control points automatically. Therefore, feature point matching methods consist of three important steps:

- *feature point detection*: the feature point detector extracts structures called *features* in the images at time t and $t + 1$ such as corners, line intersections and points on curves with a high curvature. Other features like lines, ridges or significant regions (e.g. blobs) can be used as well [Zitová and Flusser, 2003]. There are several criteria that a good feature point detector should fulfil [Vincent, 2005], for example, *accuracy* or *consistency* (the relative position of the detected points to the feature should be exactly the same in all images), *robustness* (the same points should be detected under varying noise, different illumination conditions and even after geometric transformations), and *discriminative power* (the feature points must be easily distinguishable from each other, otherwise, these points cause confusion during the matching).

The most popular feature point detectors are corner detectors (e.g. Harris and SUSAN corners) [Vincent, 2005], scale invariant point detectors (e.g. *Laplacian-of-Gaussian* operators in scale space and SIFT keypoints) [Lowe, 1999, Mikolajczyk and Schmid, 2004] and the very powerful affine invariant point detectors (e.g. local intensity extrema) [Schmid et al., 2000, Tuytelaars and Van Gool, 2004, Mikolajczyk and Schmid, 2004]. Recently, an extensive survey on feature point detectors was given in [Tuytelaars and Mikolajczyk, 2008].

- *feature description and matching*: the matching procedure establishes correspondences between feature points, in other words, we need to know which points in both images belong to each other. A simple way is to match the image content in their local neighbourhood with a block matching criterion, but also more advanced feature descriptors can be employed. Good feature descriptors should fulfil several conditions, such as *invariance* (the descriptions of the corresponding features have to be the same in both images under varying conditions), *uniqueness* (two different features should also have different descriptions), and *stability* (the description of a deformed feature should be close to the description of the original feature). These conditions can usually not be satisfied simultaneously and it is necessary to find an appropriate trade-off. Popular invariant feature descriptors are e.g. shape descriptors, circular and affinely invariant neigh-

bourhoods [Tuytelaars and Van Gool, 2004, Zitová and Flusser, 2003].

- *estimation of the registration parameters*: the final goal is to estimate the parameters of the motion model from a list of correspondences between both images in the presence of possible mismatches, which we denote as outliers. A variant on this theme is the computation of the dominant global motion parameters from local motion vectors that are obtained with block matching techniques.

Suppose that we have to estimate the parameters \mathbf{p} of an affine or polynomial motion model. The lists of coordinates of the matched feature points in the image at time t and $t + 1$ are denoted by \mathbf{T} and \mathbf{U} respectively, and these matrices are related to each other by $\mathbf{T}\mathbf{p} = \mathbf{U}$. The unknown parameters \mathbf{p} can for example be found by minimizing the following l_2 -norm:

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} (\mathbf{T}\mathbf{p} - \mathbf{U})^2, \quad (7.22)$$

which can be solved by the pseudo-inverse solution $\hat{\mathbf{p}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{U}$. We can assign weights to each correspondence (e.g. according to their reliability), which results in a *weighted least-squares* problem, which on its turn, leads to the concept of robust M-estimators using *iterative re-weighted least-squares* (IRLS) [Stewart, 1999]. The use of robust loss functions reduces the influence of outliers. More direct approaches to remove outliers are for example least median of squares, RANSAC or cross-validation methods [Golub et al., 1979, Bab-Hadiashar and Suter, 1997, Stewart, 1999]. Teelen and Veelaert checked the consistency of the uncertainty transformation for pairs of possible matches to determine a set of reliable correspondences [Teelen and Veelaert, 2005]. The uncertainty transformation allows a discrepancy on the position of the feature points and is not restricted to a one-to-one relationship between the feature points.

Feature point matching techniques are very computationally efficient and are very effective for large displacements or deformations. Additionally, we can design these techniques to be very robust against noise and invariant to varying illumination conditions and object motions. In [Capel and Zisserman, 2003], the authors employed Harris corner features in a robust RANSAC framework to perform joint mosaicing and super-resolution. However, it is very difficult to guarantee the overall subpixel precision and therefore, feature point matching algorithms are not always suitable to multi-frame SR image reconstruction. The accuracy depends on many factors like the reproducible subpixel location of feature points in the presence of noise and aliasing, and the correct filtering of mismatches and outliers. Nevertheless these techniques can be used for a fast and rough estimation of the image registration parameters. Additionally, the correspondences between the points in different images can be used as anchor motion vectors (prior information) to guide Bayesian optical flow techniques.

7.2.3.3 Gradient-based registration

The gradient-based registration methods go back to the Lucas-Kanade optical flow algorithm proposed in 1981 [Lucas and Kanade, 1981], where it is assumed that two images, $f(x, y, t)$ and the transformed frame $g(\mathbf{G}(\mathbf{x}; \mathbf{p})) = f(\mathbf{G}(\mathbf{x}; \mathbf{p}), t + 1)$, are related by:

$$f(\mathbf{x}) = g(\mathbf{G}(\mathbf{x}; \mathbf{p})) + \mathbf{n}, \quad (7.23)$$

where we omitted the time index for the sake of convenience and where \mathbf{n} is the additive zero-mean noise component. The goal of the Lucas-Kanade algorithm is to minimize the sum of the squared intensity errors between two images and to find the parameter vector \mathbf{p} given in Table 7.1, or particularly, the additive update parameter vector $\delta\mathbf{p}$ (i.e. $\mathbf{p} = \mathbf{1} + \delta\mathbf{p}$ with $\mathbf{1}$ being the parameters of the identity transformation):

$$\delta\hat{\mathbf{p}} = \arg \min_{\delta\mathbf{p}} \sum_{\mathbf{x} \in \Omega} (g(\mathbf{G}(\mathbf{x}; \mathbf{1} + \delta\mathbf{p})) - f(\mathbf{x}))^2, \quad (7.24)$$

where Ω denotes the common region between $f(\mathbf{x})$ and $g(\mathbf{G}(\mathbf{x}; \delta\mathbf{p}))$. This is a non-linear optimization problem in $\delta\mathbf{p}$, which can be approximated using the first order Taylor series expansion as an ordinary least squares problem:

$$\delta\hat{\mathbf{p}} = \arg \min_{\delta\mathbf{p}} \sum_{\mathbf{x} \in \Omega} \left(g(\mathbf{x}) + (\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \delta\mathbf{p} - f(\mathbf{x}) \right)^2, \quad (7.25)$$

In this approximation, we make use of the identity $\mathbf{G}(\mathbf{x}; \mathbf{1}) = \mathbf{x}$ and we have introduced the notation $(\nabla g)^T = \partial g / \partial \mathbf{x} = [g_x \ g_y]$ for the gradient of g w.r.t. \mathbf{x} . The term $\partial \mathbf{G} / \partial \mathbf{p}$ is called the *Jacobian* of the warping transformation and is defined as

$$\frac{\partial \mathbf{G}}{\partial \mathbf{p}} = \begin{pmatrix} \frac{\partial G_x}{\partial p_1} & \frac{\partial G_x}{\partial p_2} & \cdots & \frac{\partial G_x}{\partial p_n} \\ \frac{\partial G_y}{\partial p_1} & \frac{\partial G_y}{\partial p_2} & \cdots & \frac{\partial G_y}{\partial p_n} \end{pmatrix}. \quad (7.26)$$

For example, the affine warping transformation has the following Jacobian:

$$\frac{\partial \mathbf{G}}{\partial \mathbf{p}} = \begin{pmatrix} x & 0 & y & 0 & 1 & 0 \\ 0 & x & 0 & y & 0 & 1 \end{pmatrix}. \quad (7.27)$$

The solution for the minimization problem is found by setting the partial derivatives of the expression in equation (7.25) with respect to $\delta\mathbf{p}$ to zero:

$$2 \sum_{\mathbf{x} \in \Omega} \left[(\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \right]^T \left[g(\mathbf{x}) + (\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \delta\mathbf{p} - f(\mathbf{x}) \right] = 0. \quad (7.28)$$

Solving this equation gives us the closed form solution for the minimization problem (7.25):

$$\delta\hat{\mathbf{p}} = \mathcal{H}^{-1} \sum_{\mathbf{x} \in \Omega} \left[(\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \right]^T [f(\mathbf{x}) - g(\mathbf{x})], \quad (7.29)$$

where \mathcal{H} is the $n \times n$ *Gauss-Newton* approximation of the *Hessian* matrix:

$$\mathcal{H} = \sum_{\mathbf{x} \in \Omega} \left[(\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \right]^T \left[(\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \right]. \quad (7.30)$$

There are several problems with the estimator as proposed in equation (7.29): the estimated motion parameters are only valid for very small displacements or deformations (due to the linearization in equation (7.25)) and the estimation $\delta \hat{\mathbf{p}}$ is biased due to truncation error of the Taylor series expansion, gradient approximation and noise [Robinson and Milanfar, 2004, Pham et al., 2005]. To overcome these problems, we optimize the least-squares problem iteratively in a *steepest descent* algorithm (see also Section 4.4.2) with the Gauss-Newton expression in equation (7.29) being the direction of the gradient. The motion parameters are updated by the following iterative sequence:

$$\hat{\mathbf{p}}^{(j+1)} = \hat{\mathbf{p}}^{(j)} + \lambda \delta \hat{\mathbf{p}}^{(j)}, \quad (7.31)$$

where the scalar parameter λ determines the convergence speed (λ is typically 1 in most applications). In each iteration, we need to transform g or f according to the estimated motion parameters $\hat{\mathbf{p}}^{(j)}$, which is referred to as *forward* or *inverse* registration respectively. For several computational reasons, it is preferred to perform the inverse registration [Baker and Matthews, 2004]. For instance, if we transform g in the forward registration, the image derivatives in ∇g need to be updated every iteration, while in the inverse registration, ∇g (and also \mathcal{H}) remains constant and can be pre-computed once. By iteratively optimizing the motion parameters and thus reducing the bias, it is shown in [Pham et al., 2005] that the gradient-based motion estimator reaches the Cramer-Rao lower bound, i.e. the estimator is said to be optimal. In [Robinson and Milanfar, 2005], the authors explored the relationship between the image gradient filters and their effect on the overall estimation performance.

A variant on the steepest descent algorithm, where the parameters are commonly updated in an additive way, is the *compositional* approach [Shum and Szeliski, 2000, Baker and Matthews, 2004], where the incremental transform $\mathbf{G}(\mathbf{x}; \delta \mathbf{p})$ is optimized instead of the incremental parameter $\delta \mathbf{p}$. The motion parameters are thus updated successively according to

$$\mathbf{G}(\mathbf{x}; \hat{\mathbf{p}}^{(j+1)}) = \mathbf{G}(\mathbf{x}; \hat{\mathbf{p}}^{(j)}) \circ \mathbf{G}(\mathbf{x}; \delta \hat{\mathbf{p}}^{(j)}) \equiv \mathbf{G}(\mathbf{G}(\mathbf{x}; \hat{\mathbf{p}}^{(j)}); \delta \hat{\mathbf{p}}^{(j)}). \quad (7.32)$$

From a Bayesian point of view, the l_2 -norm in equation (7.24) assumes that image g is corrupted with additive zero-mean white Gaussian noise. The main disadvantage is that this expression does not tolerate any outliers. We can improve the cost criterion with the use of the robust loss function ρ :

$$\delta \hat{\mathbf{p}} = \arg \min_{\delta \mathbf{p}} \sum_{\mathbf{x} \in \Omega} \rho(g(\mathbf{G}(\mathbf{x}; \mathbf{I} + \delta \mathbf{p})) - f(\mathbf{x})). \quad (7.33)$$

Equation (7.33) is solved via the IRLS approach, by employing the weight function $w(g(\mathbf{x}) - f(\mathbf{x}))$ (this weight function is chosen such that $w(x) = \psi(x)/x$ according to the M-estimators given in Table 4.1). In each iteration, the incremental update motion parameters are computed as

$$\delta \hat{\mathbf{p}} = \mathcal{H}^{-1} \sum_{\mathbf{x} \in \Omega} w(g(\mathbf{x}) - f(\mathbf{x})) \left[(\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \right]^T [f(\mathbf{x}) - g(\mathbf{x})], \quad (7.34)$$

where the Gauss-Newton approximation of the Hessian matrix \mathcal{H} is given by

$$\mathcal{H} = \sum_{\mathbf{x} \in \Omega} w(g(\mathbf{x}) - f(\mathbf{x})) \left[(\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \right]^T \left[(\nabla g)^T \frac{\partial \mathbf{G}}{\partial \mathbf{p}} \right]. \quad (7.35)$$

The difference between the robust affine formulation (see equation (7.34)) and the non-robust affine formulation (see equation (7.29)) is illustrated in Figure 7.4. We can clearly see that object motion has a large impact on the accuracy of the motion parameters.

Like optical flow techniques in [Papenberg et al., 2006], other constancy assumptions besides constant image brightness in equation (7.24) can be incorporated such as the constancy of gradients, Hessian or Laplacian, which makes the registration model more invariant to varying intensity changes.

To cope with large displacements and to decrease the computation time tremendously, a coarse-to-fine refinement or hierarchical strategy such as the use of Gaussian pyramids or wavelets is recommended. At the coarse scale, displacements are relatively small with the result that the convergence is reached in fewer iterations and potential false local minima in the optimization problem are avoided.

In [Baker and Matthews, 2004], Baker and Matthews showed that the Gauss-Newton and Levenberg-Marquardt approximations are the most efficient approaches compared to other approximations of the Hessian matrix such as the Newton and diagonal Hessian algorithms. In the Levenberg-Marquardt algorithm, a regularized form $\mathcal{H} + \alpha \mathbf{I}$ with \mathbf{I} being the identity matrix replaces the term \mathcal{H} in equation (7.29). In [Thévenaz et al., 1998], the authors proposed the use of a modified Levenberg-Marquardt algorithm in a hierarchical polynomial spline pyramid.

For a more elaborated treatment of gradient-based motion estimation techniques, we refer the reader to the work presented in [Baker and Matthews, 2004, Baker and Matthews, 2002, Baker et al., 2003b, Baker et al., 2003a, Baker et al., 2004a, Baker et al., 2004b]. Because of the very good subpixel accuracy performance, these gradient-based methods are very popular in combination with multi-frame SR restoration techniques, e.g. [Keren et al., 1988, Irani and Peleg, 1991, Pham, 2006].

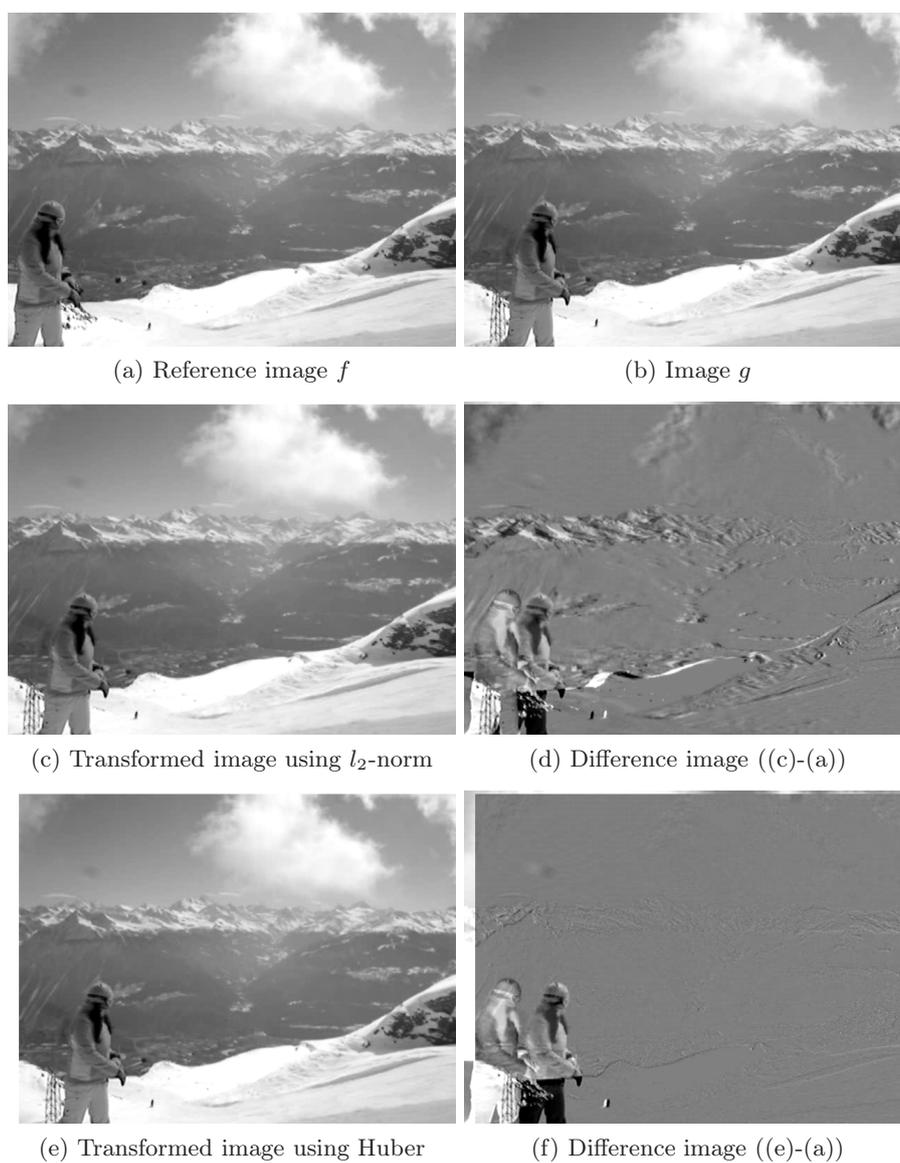


Figure 7.4: The effect of object motions in image registration using the l_2 -norm criterion and the robust Huber criterion, respectively.

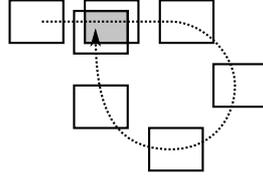


Figure 7.5: Illustration of the deadlock problem in successive registrations: in some cases, it is impossible to fit the next image into the mosaic or HR image.

7.2.3.4 Consistent registration

Equation (7.29) is not symmetric in f and g . This means that the transformation from f to g is not necessarily equal to the inverse transformation from g to f (i.e. $\mathbf{p}_{f,g} \neq -\mathbf{p}_{g,f}$ in case of translation). As a result, the registration method may produce inconsistent results when applied in one direction versus the other.

In most video applications, the transformation parameters between images at time i and j ($i < j$) are computed successively from pairwise registrations in an image sequence (see equation (7.32) for the definition of the \circ -operator):

$$\mathbf{G}(\mathbf{x}; \mathbf{p}_{i,j}) = \mathbf{G}(\mathbf{x}; \mathbf{p}_{i,j-1}) \circ \mathbf{G}(\mathbf{x}; \mathbf{p}_{j-1,j}) = \mathbf{G}(\mathbf{x}; \mathbf{p}_{i,i+1}) \circ \dots \circ \mathbf{G}(\mathbf{x}; \mathbf{p}_{j-1,j}). \quad (7.36)$$

Due to inaccuracies in pairwise registrations, the registration errors accumulate over time such that the registration parameters between images at time i and j can have a large cumulative error, which is also known as *drift*. This problem often arises in mosaicing and super-resolution applications and leads to possible *deadlocks*, where it is impossible to find a proper solution (see Figure 7.5).

The asymmetry of the registration solution and the deadlock problem indicate that there is a need for more consistent registration techniques. In literature, there are two common ways to achieve consistent registration, which are known as *bundle adjustment* and *global registration*.

Bundle adjustment is a well-known computationally expensive tool in the computer vision community, which produces jointly optimal 3D structures and viewing parameters [Triggs et al., 1999]. In 2D registration problems, bundle adjustment finds the registration parameters over k images that minimize the total misalignment of a predefined set of m grid points \mathbf{x}_l . An example of such a minimization problem is given by

$$\mathbf{p}^* = \begin{bmatrix} \hat{\mathbf{p}}_{1,2} \\ \vdots \\ \hat{\mathbf{p}}_{k-1,k} \end{bmatrix} = \arg \min_{\mathbf{p}^*} \sum_{l=1}^m \sum_{h,i,j}^k (\mathbf{x}_l - \mathbf{G}(\mathbf{G}(\mathbf{G}(\mathbf{x}_l; \mathbf{p}_{h,i}); \mathbf{p}_{i,j}); \mathbf{p}_{j,h}))^2, \quad (7.37)$$

where we assume that consecutive transformations from image h to i , i to j and j to h should result in the identity transformation. This minimization

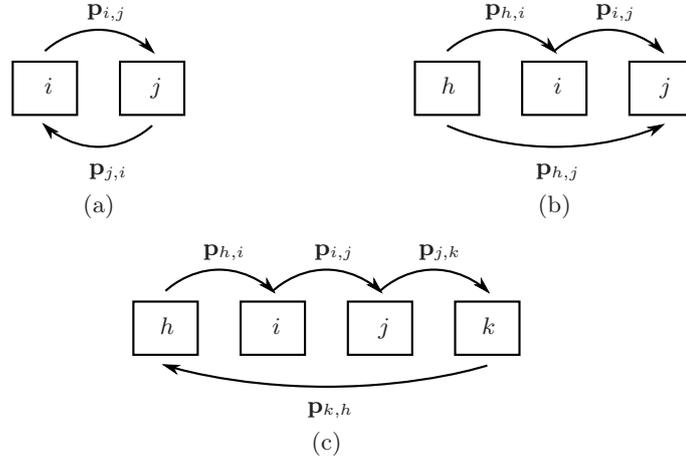


Figure 7.6: Group structures in the Lie algebra: (a) skew anti-symmetry, (b) Jacobi identity and (c) a combined group structure based on the basic principles of (a) and (b).

problem can be solved by Gauss-Newton or Levenberg-Marquardt algorithms. In theory, the three successive warping operations in this expression yields the identity transformation. However in practice, there is a non-zero shift between the original grid point and the same grid point after transformations. The set of combined pairwise transformations is conveniently described by the group structures of the Lie algebra [Govindu, 2004]. The bundle adjustment strategy enforces the pairwise transformations (and its parameters) to be in a valid Lie group structure. Several group structures are illustrated in Figure 7.6.

Bundle adjustment is very popular in mosaicing applications [Sawhney et al., 1998, Shum and Szeliski, 2000] and recently, it gained much interest in super-resolution applications [Farsiu et al., 2005, Pham, 2006]. In [Sawhney et al., 1998], the authors determined the topology of the mosaic before applying bundle adjustment in order to detect overlapping regions for pairwise registration. In [Farsiu et al., 2005], the authors incorporated the consistency constraints as prior knowledge in a Bayesian framework of the gradient-based image registration. Instead of applying bundle adjustment as a postprocessing step with hard constraints, they penalized inconsistencies directly in the estimation problem of the registration parameters. Because the inconsistency penalties require different registration parameter sets, all pairwise registrations must be performed simultaneously. Based on the robust formulation in equation (7.33), the MAP estimation of the registration parameters becomes

$$\hat{\mathbf{p}}^* = \arg \min_{\mathbf{p}^*} \sum_{i,j} \sum_{\mathbf{x} \in \Omega} \rho(f(\mathbf{G}(\mathbf{x}; \mathbf{p}_{i,j}), i) - f(\mathbf{x}, j)) + \alpha \mathbf{C} \mathbf{p}^*, \quad (7.38)$$

where \mathbf{p}^* contains *all* $k(k-1)$ registration parameter sets (see equation (7.37)) and \mathbf{C} contains the inconsistency penalties. See [Farsiu et al., 2005] for more details on building the matrix \mathbf{C} . The first term in equation (7.38) contains all pairwise registrations and the second term represents the inconsistencies of the pairwise transformations towards the Lie group structures.

There are two main drawbacks to the bundle adjustment strategy: firstly, $k(k-1)$ pairwise registrations are needed for the full set of compositions instead of $k-1$ registrations in traditional settings, which is bad for the computational efficiency. Secondly, these models do not explicitly consider the presence of frequency aliasing in the observed images, which limits the use of bundle adjustment in SR applications. On the other hand, bundle adjustment is a very powerful tool to detect and correct larger registration errors.

The second class of consistent registration algorithms jointly solves the registration and SR reconstruction problem. Related approaches for mosaicing algorithms can be found in [Davis, 1998, Pires and Aguiar, 2005]. In [Robinson et al., 2009], the authors solved the joint registration/reconstruction problem based on the principle of variable projections (i.e. a method to solve non-linear data fitting problems, which have as their underlying model a linear combination of non-linear functions). Most of the global registration algorithms use a kind of *expectation-maximization* (EM) algorithm that iteratively alternates between estimating the HR image (expectation step) and the registration parameters (maximization step) [Hardie et al., 1997, Woods et al., 2006]. By reconstructing the HR image, the relationship of the aliased observed images to the SR estimate is exploited efficiently.

Due to the chicken-and-egg problem mentioned earlier, the joint estimation algorithms are commonly initialized with an interpolated HR image or by setting initial registration parameters found by other registration techniques. The HR image \tilde{f} can be updated alternatively using interpolation techniques for irregularly spaced samples (see Section 7.4) and the registration parameters can be found using the same gradient-based algorithm discussed in Section 7.2.3.3 with a very small modification: instead of transforming $f(\mathbf{x})$ with the current registration parameters in each iteration, we simply transform the current HR estimate $\tilde{f}(\mathbf{x})$ towards $g(\mathbf{x})$. In Section 7.5.2, we show that the combination of gradient-based image registration with steering kernel regression produces the most accurate registration results.

7.3 Photometric registration

The image-based motion estimation techniques (as discussed in Section 7.2.3) align two (or more) images in the spatial domain. These methods are also referred to as *geometric* registration. In addition, we can also perform image registration/alignment in the range/intensity domain, which is also known as *photometric* registration.

Most motion estimation algorithms are based on the intensity constancy assumption (e.g. in equation (7.23)), however in practice, this assumption is not always correct. In an uncontrolled environment, lighting conditions can vary over time (e.g. due to the weather) and intensity variations also arise due to automatic gain control or automatic white balancing inside the camera. On the other hand, in *high dynamic range* (HDR) imaging, the aperture times and hence the apparent illumination are even changed on purpose.

In the next sections, we briefly discuss the current photometric registration techniques and we propose a new approach that jointly performs geometric and photometric registration.

7.3.1 Related work

To understand how photometric registration works, we first briefly describe the basic principles of *quantigraphic image processing*. The formation of a pixel greyvalue starts with the integration of photons over a fixed spectral response profile (i.e. the spectral sensitivity of the sensor resulting in for example red, green or blue components of the pixel value), over a finite period of time (which is referred to as the *exposure time*) and over a finite sensor area. This integration in the spectral, temporal and spatial dimension results in a *photoquantigraphic quantity*, or just simply *photoquantity*. This photoquantity ($\in [0, +\infty)$) is subject to intrinsic sensor noise and goes through a dynamic range compression, which restricts the pixel greyvalues to a finite interval (e.g. $\in [0, 1]$). These pixel greyvalues typically undergo several processes such as quantization (e.g. 8 bits per pixel, resulting in integer greyvalues in the range $[0, 255]$), gamma correction, histogram transformations, noise (due to electronics or image compression), etc. This non-linear relationship between the photoquantity and the final observed pixel greyvalue is denoted as the *camera response function* and plays an important role in choosing the right model for the photometric registration [Mann, 2000].

Photometric registration consists of determining the parameters of the *comparametric equations* (or *intensity mapping functions*) that describe the relationship between the intensity values of the corresponding pixels of two spatially aligned images f and g . Some examples of well-known comparametric models are the linear transformation (also referred to as gain and bias model) and the gamma correction (i.e. raising the pixel values to a power to lighten or darken images). A more detailed overview of comparametric equations and their related camera response functions is given by Mann in [Mann, 2000]. The estimation process of the parameters involves the computation of a *comparagram* (i.e. the joint histogram of the pixel values between the spatially aligned images) and followed by finding a smooth semi-monotonic function (i.e. comparametric equation) that passes through most of the highest bins in the comparagram. In a nutshell, photometric registration comes down to

comparametric regression or finding the optimal fit to the comparagram data.

Photometric registration requires the computation of a comparagram, which on its turn requires spatially aligned images. On the other hand, geometric registration techniques are often based on the intensity constancy assumption. This results in a chicken-and-egg problem, which can be solved by intensity invariant geometric registration (e.g. based on gradient constancy assumption or other invariant features), photometric registration based on spatially invariant features (e.g. histograms or statistical moments) or by jointly estimating geometric and photometric registration parameters.

In [Capel and Zisserman, 2003], the authors estimated the linear photometric model using a robust RANSAC algorithm that minimizes the Huber robust loss function. Bartoli performed joint geometric and photometric registration within the inverse compositional gradient-based framework using the ordinary least square metrics [Bartoli, 2006]. In [Grossberg and Nayar, 2003], Grossberg and Nayar determined the camera response function from the intensity mapping functions between several images, which are not spatially aligned. They computed the comparametric parameters directly from cumulative intensity histograms. Candocia approximated the comparametric function and the camera response function by a piecewise linear model [Candocia, 2003, Candocia and Mandarino, 2005]. Gevrekci and Gunturk employed a geometric feature point matching algorithm and comparametric regression to perform joint HDR and SR image reconstruction [Gevrekci and Gunturk, 2007].

The standard parametric geometric and comparametric relationship between images f and g is given by the following model:

$$f(\mathbf{x}) = \mathbf{P}(g(\mathbf{G}(\mathbf{x}; \mathbf{p}_G)); \mathbf{p}_P) + n_f, \quad (7.39)$$

where \mathbf{p}_G and \mathbf{p}_P are the geometric and photometric parameters respectively and n_f is additive noise. In the motion-free case combined with the linear comparametric model, the comparametric function is simplified to a straight line with gain a_1 and bias a_0 ($\mathbf{p}_P = [a_1 \ a_0]^T$):

$$f(\mathbf{x}) = a_1 g(\mathbf{x}) + a_0 + n_f. \quad (7.40)$$

Note that this comparametric function introduces clipping effects, i.e. saturation of pixel values below 0 and above 255, which implies an important loss of information at very dark and light regions and therefore, these regions should be excluded from further computations. In the presence of additive zero-mean white Gaussian noise, the parameters can be found via *ordinary least squares* (OLS) formulation as employed in e.g. [Bartoli, 2006, Candocia, 2003]:

$$\hat{\mathbf{p}}_P = \arg \min_{a_1, a_0} \sum_{\mathbf{x} \in \Omega} d_a^2 = \arg \min_{a_1, a_0} \sum_{\mathbf{x} \in \Omega} (a_1 g(\mathbf{x}) + a_0 - f(\mathbf{x}))^2$$

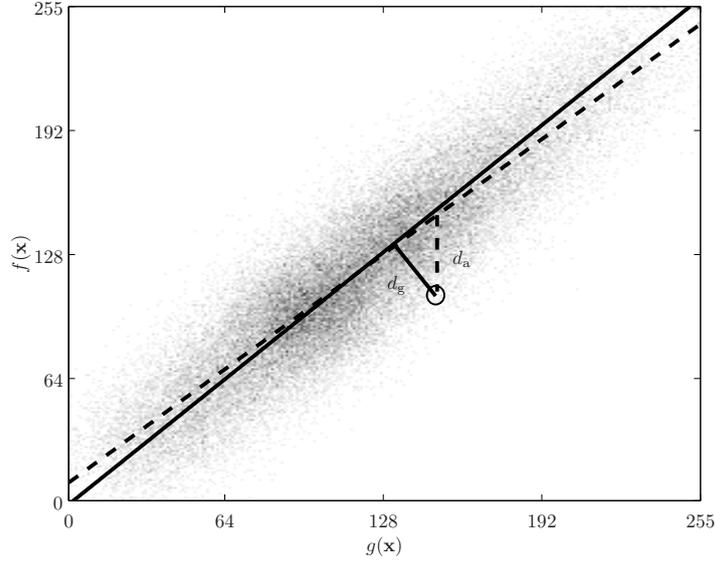


Figure 7.7: Illustration of a comparagram and the fitted OLS solution (dashed line) and TLS solution (solid line), which minimize the algebraic distance d_a and the geometric distance d_g respectively.

$$= \begin{pmatrix} \sum_{\mathbf{x} \in \Omega} g^2(\mathbf{x}) & \sum_{\mathbf{x} \in \Omega} g(\mathbf{x}) \\ \sum_{\mathbf{x} \in \Omega} g(\mathbf{x}) & \sum_{\mathbf{x} \in \Omega} 1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{\mathbf{x} \in \Omega} g(\mathbf{x})f(\mathbf{x}) \\ \sum_{\mathbf{x} \in \Omega} f(\mathbf{x}) \end{pmatrix}, \quad (7.41)$$

where d_a is denoted as the *algebraic distance*.

We will show by means of a simple example that this OLS model has some serious shortcomings in practice. We generate the image f from a 200×200 template image g via the comparometric equation $f(\mathbf{x}) = 1.025g(\mathbf{x}) - 1.988$ (the backward transformation is given by $g(\mathbf{x}) = 0.976f(\mathbf{x}) + 1.94$) and next we add zero-mean white Gaussian noise ($\sigma_n = 15$) to both images. The comparagram is plotted in Figure 7.7. We estimate the photometric registration parameters using equation (7.41) for the forward (g to f) and the backward (f to g) transformation:

$$f(\mathbf{x}) = 0.928g(\mathbf{x}) + 9.347 \quad \text{and} \quad g(\mathbf{x}) = 0.888f(\mathbf{x}) + 12.283. \quad (7.42)$$

The OLS solution is not close to the ground truth parameters (e.g. there is a deviation in intensity levels of 10 for black pixels) and the estimated parameters are also mutual inconsistent, i.e. the inverse forward transformation does not

yield the backward transformation and vice versa (both solutions should be symmetric around the bisector). The problem is the incorrect employed model in equation (7.39), which indicates that image f may contain noise and image g should be noise-free, which is not true in practice.

An improved geometric and photometric registration model specifies that image g can also be subject to perturbations:

$$f(\mathbf{x}) = \mathbf{P}(g(\mathbf{G}(\mathbf{x}; \mathbf{p}_G)) + n_g; \mathbf{p}_P) + n_f, \quad (7.43)$$

where n_g is additive noise, commonly from the same PDF that generates n_f . The solution of this model minimizes the *geometric distance* d_g instead of the algebraic distance d_a as illustrated in Figure 7.7. In case of linear comparametric regression (see equation (7.40)), equation (7.43) is transformed into a *total least square* (TLS) problem. The solution to the TLS problem is well documented, see e.g. [Golub and Van Loan, 1980, Markovsky and Van Huffel, 2007]. To demonstrate that the proposed model (7.43) is indeed more accurate, we give the TLS solution of this simple example in Figure 7.7:

$$f(\mathbf{x}) = 1.026g(\mathbf{x}) - 2.135 \quad \text{and} \quad g(\mathbf{x}) = 0.975f(\mathbf{x}) + 2.082. \quad (7.44)$$

This solution is very close to the ground truth parameters and is furthermore always mutual consistent, since minimizing the geometric distance is rotational invariant. Based on these observations, we propose a new registration technique in the TLS sense in the next section.

7.3.2 Photometric registration in the total least square sense

The linear solution of the problem stated in equation (7.43) can explicitly be found via the basic TLS algorithm as described in e.g. [Markovsky and Van Huffel, 2007], where \hat{a}_1 can be computed via the *singular value decomposition* (SVD) of the following zero-mean shifted augmented matrix:

$$\begin{pmatrix} g(\mathbf{x}_1) - \bar{g} & f(\mathbf{x}_1) - \bar{f} \\ g(\mathbf{x}_2) - \bar{g} & f(\mathbf{x}_2) - \bar{f} \\ \vdots & \vdots \end{pmatrix} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (7.45)$$

where \bar{g} and \bar{f} are the mean intensity values of the images g and f respectively. $\mathbf{\Sigma}$ is a 2×2 diagonal matrix with the singular values on the main diagonal and \mathbf{V} is a 2×2 containing the singular vectors. The gain parameter a_1 in the TLS sense is computed by $\hat{a}_1 = -\frac{\mathbf{V}(1,2)}{\mathbf{V}(2,2)}$ on the condition that $\mathbf{V}(2,2)$ is non-zero (or non-singular in general). The bias parameter a_0 can be computed directly by substituting \hat{a}_1 back into the following equation: $\hat{a}_0 = \bar{f} - \hat{a}_1\bar{g}$.

The basic TLS algorithm can only applied for the linear photometric registration model. A more general approach is to minimize the geometric distance, or also referred to as *orthogonal distance regression* [Sullivan et al., 1994,

Ahn et al., 2002]:

$$\begin{aligned}
\hat{\mathbf{p}}_P &= \arg \min_{\mathbf{p}_P} \sum_{\mathbf{x} \in \Omega} d_g^2 = \arg \min_{\mathbf{p}_P} \sum_{\mathbf{x} \in \Omega} \|\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P)\|_2^2 \\
&= \arg \min_{\mathbf{p}_P} \min_{\mathbf{Y}'} \sum_{\mathbf{x} \in \Omega} \left((\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P))^T (\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P)) \right)^2 \\
&= -(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T (\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P)), \tag{7.46}
\end{aligned}$$

The minimum is found using the iterative Gauss-Newton method. The Jacobian \mathbf{J} in the solution (7.46) is computed via the chain rule:

$$\begin{aligned}
\mathbf{J} &= \sum_{\mathbf{x} \in \Omega} \frac{\partial \|\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P)\|_2}{\partial \mathbf{p}_P} \\
&= \sum_{\mathbf{x} \in \Omega} \frac{(\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P))^T}{\|\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P)\|_2} \frac{\partial \mathbf{Y}'(\mathbf{x}; \mathbf{p}_P)}{\partial \mathbf{p}_P} \tag{7.47}
\end{aligned}$$

The measurement data $\mathbf{Y}(\mathbf{x})$ (we can interpret this as a point in the comparagram) is given by

$$\mathbf{Y}(\mathbf{x}) = \begin{pmatrix} g(\mathbf{x}) \\ f(\mathbf{x}) \end{pmatrix}. \tag{7.48}$$

The orthogonal projection $\mathbf{Y}'(\mathbf{x})$ of the measurement data on the regression curve can be found by minimizing the distance between the curve and the measurement data. In some cases, e.g. in the linear photometric registration model, $\mathbf{Y}'(\mathbf{x})$ can be found in a closed-form expression.

We now derive a novel algorithm that solves the joint photometric linear and geometric affine registration problem in the TLS sense based on orthogonal distance regression. Extensions to other models are quite straightforward. The advantage over the approach of [Bartoli, 2006] is that our method uses the TLS metrics, which results in more consistent and accurate registration parameters. The parametric model (7.43) is transformed into the following (non-linear) assumption:

$$f(x, y) = a_1 g(a_{00} + a_{10}x + a_{01}y, b_{00} + b_{10}x + b_{01}y) + a_0, \tag{7.49}$$

Like the gradient-based geometric registration algorithms in Section 7.2.3.3, we iteratively estimate the registration parameters using the incremental updates $\delta \mathbf{p}_G = [\delta a_{00} \ \delta a_{10} \ \delta a_{01} \ \delta b_{00} \ \delta b_{10} \ \delta b_{01}]^T$ and $\delta \mathbf{p}_P = [\delta a_0 \ \delta a_1]^T$:

$$\hat{\mathbf{p}}_G^{(j+1)} = \hat{\mathbf{p}}_G^{(j)} + \delta \hat{\mathbf{p}}_G^{(j)} \quad \text{and} \quad \hat{\mathbf{p}}_P^{(j+1)} = \hat{\mathbf{p}}_P^{(j)} + \delta \hat{\mathbf{p}}_P^{(j)}. \tag{7.50}$$

To find these incremental updates, we approximate the non-linear model by the first order Taylor series expansion:

$$\tilde{f}(\mathbf{x}) - \tilde{g}(\mathbf{x}) \approx a_1 \nabla_x g(\mathbf{x}) \delta a_{00} + a_1 x \nabla_x g(\mathbf{x}) \delta a_{10} + a_1 y \nabla_x g(\mathbf{x}) \delta a_{01}$$

$$\begin{aligned}
& + a_1 \nabla_y g(\mathbf{x}) \delta b_{00} + a_1 x \nabla_y g(\mathbf{x}) \delta b_{10} + a_1 y \nabla_y g(\mathbf{x}) \delta b_{01} \\
& + g(\mathbf{x}) \delta a_1 + \delta a_0,
\end{aligned} \tag{7.51}$$

where we perform the inverse geometric registration and the forward photometric registration in each iteration by transforming $\tilde{f}(\mathbf{x}) = f\left(\mathbf{G}^{-1}\left(\mathbf{x}; \hat{\mathbf{p}}_{\mathbf{G}}^{(j)}\right)\right)$ and $\tilde{g}(\mathbf{x}) = \mathbf{P}\left(g(\mathbf{x}); \hat{\mathbf{p}}_{\mathbf{P}}^{(j)}\right)$ respectively.

The estimation of the registration parameters in the linearized model can be interpreted as a regression problem that fits the parameters to a hyperplane, given by the following implicit function (where we have simplified some notations⁵):

$$h(\mathbf{Y}, \boldsymbol{\delta}) = Y_0 \delta_0 + Y_1 \delta_1 + Y_2 \delta_2 + Y_3 \delta_3 + Y_4 \delta_4 + Y_5 \delta_5 + Y_6 \delta_6 + \delta_7 - Y_7 = 0. \tag{7.52}$$

Similarly to the orthogonal distance regression formulation in equation (7.46), the measurement data of this hyperplane is given by

$$\mathbf{Y}(\mathbf{x}) = \begin{pmatrix} Y_0(\mathbf{x}) \\ Y_1(\mathbf{x}) \\ Y_2(\mathbf{x}) \\ Y_3(\mathbf{x}) \\ Y_4(\mathbf{x}) \\ Y_5(\mathbf{x}) \\ Y_6(\mathbf{x}) \\ Y_7(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} a_1 \nabla_x g(\mathbf{x}) \\ a_1 x \nabla_x g(\mathbf{x}) \\ a_1 y \nabla_x g(\mathbf{x}) \\ a_1 \nabla_y g(\mathbf{x}) \\ a_1 x \nabla_y g(\mathbf{x}) \\ a_1 y \nabla_y g(\mathbf{x}) \\ g(\mathbf{x}) \\ \tilde{f}(\mathbf{x}) - \tilde{g}(\mathbf{x}) \end{pmatrix}. \tag{7.53}$$

The orthogonal projection \mathbf{Y}' on the hyperplane is denoted by the following system of symmetric line equations:

$$\begin{cases} \frac{\mathbf{Y}'_0 - \mathbf{Y}_0}{\delta_0} = \frac{\mathbf{Y}'_1 - \mathbf{Y}_1}{\delta_1} = \dots = -\mathbf{Y}'_7 + \mathbf{Y}_7 \\ h(\mathbf{Y}'_l, \boldsymbol{\delta}) = 0 \end{cases} \tag{7.54}$$

By solving this system, we obtain the closed-form expression for \mathbf{Y}' , which is

⁵ $\delta_0 = \delta a_{00}$, $\delta_1 = \delta a_{10}$, $\delta_2 = \delta a_{01}$, $\delta_3 = \delta b_{00}$, $\delta_4 = \delta b_{10}$, $\delta_5 = \delta b_{01}$, $\delta_6 = \delta a_1$ and $\delta_7 = \delta a_0$.

given by

$$\mathbf{Y}'(\mathbf{x}, \boldsymbol{\delta}) = \begin{pmatrix} Y_0(\mathbf{x}) - \delta_0 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ Y_1(\mathbf{x}) - \delta_1 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ Y_2(\mathbf{x}) - \delta_2 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ Y_3(\mathbf{x}) - \delta_3 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ Y_4(\mathbf{x}) - \delta_4 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ Y_5(\mathbf{x}) - \delta_5 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ Y_6(\mathbf{x}) - \delta_6 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ Y_7(\mathbf{x}) + \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \end{pmatrix}, \quad (7.55)$$

where we employ ξ and v as the shorthand notations for

$$\begin{aligned} \xi(\mathbf{x}, \boldsymbol{\delta}) &= \delta_7 - Y_7(\mathbf{x}) + Y_6(\mathbf{x})\delta_6 + Y_5(\mathbf{x})\delta_5 + Y_4(\mathbf{x})\delta_4 + Y_3(\mathbf{x})\delta_3 + Y_2(\mathbf{x})\delta_2 \\ &\quad + Y_1(\mathbf{x})\delta_1 + Y_0(\mathbf{x})\delta_0, \end{aligned} \quad (7.56)$$

and

$$v(\boldsymbol{\delta}) = \delta_6^2 + \delta_5^2 + \delta_4^2 + \delta_3^2 + \delta_2^2 + \delta_1^2 + \delta_0^2 + 1. \quad (7.57)$$

In order to compute the Jacobian \mathbf{J} in equation (7.47), we have to obtain the partial derivatives of \mathbf{Y}' to $\boldsymbol{\delta}$. This is given for the case of δ_0 (the cases from δ_1 to δ_6 are similar):

$$\frac{\partial \mathbf{Y}'(\mathbf{x}, \boldsymbol{\delta})}{\partial \delta_0} = \begin{pmatrix} \frac{2\delta_0^2}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} - \frac{\delta_0 Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} - \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ \frac{2\delta_0 \delta_1}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} - \frac{\delta_1 Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} \\ \frac{2\delta_0 \delta_2}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} - \frac{\delta_2 Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} \\ \frac{2\delta_0 \delta_3}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} - \frac{\delta_3 Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} \\ \frac{2\delta_0 \delta_4}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} - \frac{\delta_4 Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} \\ \frac{2\delta_0 \delta_5}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} - \frac{\delta_5 Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} \\ \frac{2\delta_0 \delta_6}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} - \frac{\delta_6 Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} \\ -\frac{2\delta_0}{v(\boldsymbol{\delta})} \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} + \frac{Y_0(\mathbf{x})}{v(\boldsymbol{\delta})} \end{pmatrix}, \quad (7.58)$$

and δ_7 :

$$\frac{\partial \mathbf{Y}'(\mathbf{x}, \boldsymbol{\delta})}{\partial \delta_7} = \begin{pmatrix} \frac{-\delta_0}{v(\boldsymbol{\delta})} \\ \frac{-\delta_1}{v(\boldsymbol{\delta})} \\ \frac{-\delta_2}{v(\boldsymbol{\delta})} \\ \frac{-\delta_3}{v(\boldsymbol{\delta})} \\ \frac{-\delta_4}{v(\boldsymbol{\delta})} \\ \frac{-\delta_5}{v(\boldsymbol{\delta})} \\ \frac{-\delta_6}{v(\boldsymbol{\delta})} \\ \frac{1}{v(\boldsymbol{\delta})} \end{pmatrix}. \quad (7.59)$$

The difference vector $\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}, \boldsymbol{\delta})$ and its l_2 -norm are given by

$$\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}, \boldsymbol{\delta}) = \begin{pmatrix} \delta_0 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ \delta_1 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ \delta_2 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ \delta_3 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ \delta_4 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ \delta_5 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ \delta_6 \frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \\ -\frac{\xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})} \end{pmatrix} \quad \text{and} \quad \|\mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}, \boldsymbol{\delta})\|_2 = \frac{|\xi(\mathbf{x}, \boldsymbol{\delta})|}{\sqrt{v(\boldsymbol{\delta})}}. \quad (7.60)$$

The Jacobian matrix \mathbf{J} can be simplified by substituting expressions (7.58)-

(7.60) into equation (7.47):

$$\mathbf{J}(\boldsymbol{\delta}) = \begin{pmatrix} \sum_{\mathbf{x} \in \Omega} \text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta})) \frac{Y_0(\mathbf{x})v(\boldsymbol{\delta}) - \delta_0 \xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})^{3/2}} \\ \sum_{\mathbf{x} \in \Omega} \text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta})) \frac{Y_1(\mathbf{x})v(\boldsymbol{\delta}) - \delta_1 \xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})^{3/2}} \\ \sum_{\mathbf{x} \in \Omega} \text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta})) \frac{Y_2(\mathbf{x})v(\boldsymbol{\delta}) - \delta_2 \xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})^{3/2}} \\ \sum_{\mathbf{x} \in \Omega} \text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta})) \frac{Y_3(\mathbf{x})v(\boldsymbol{\delta}) - \delta_3 \xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})^{3/2}} \\ \sum_{\mathbf{x} \in \Omega} \text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta})) \frac{Y_4(\mathbf{x})v(\boldsymbol{\delta}) - \delta_4 \xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})^{3/2}} \\ \sum_{\mathbf{x} \in \Omega} \text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta})) \frac{Y_5(\mathbf{x})v(\boldsymbol{\delta}) - \delta_5 \xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})^{3/2}} \\ \sum_{\mathbf{x} \in \Omega} \text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta})) \frac{Y_6(\mathbf{x})v(\boldsymbol{\delta}) - \delta_6 \xi(\mathbf{x}, \boldsymbol{\delta})}{v(\boldsymbol{\delta})^{3/2}} \\ \sum_{\mathbf{x} \in \Omega} \frac{\text{sign}(\xi(\mathbf{x}, \boldsymbol{\delta}))}{\sqrt{v(\boldsymbol{\delta})}} \end{pmatrix}. \quad (7.61)$$

The incremental updates $\delta \mathbf{p}_G$ and $\delta \mathbf{p}_P$ are computed by substituting the expressions (7.53), (7.55) and (7.61) into the orthogonal distance regression solution given by equation (7.46). These updates iteratively improve the current registration parameters as given by equation (7.50). The performance of the proposed TLS method is evaluated in Section 7.5.3.

7.4 Fusion of irregularly spaced samples

After (proper) alignment, the LR images provide samples at non-uniform or irregular positions on the HR grid as illustrated in Figure 7.2. The conversion of these samples into samples placed on the regular HR grid is performed by interpolation or approximation techniques. This process is also called *fusion*, because it produces one HR image from multiple aligned LR images.

We give an overview of existing *spatial* fusion techniques for SR image reconstruction and discuss the powerful kernel regression techniques in more detail. We extend these kernel regression techniques in the total least square sense to cope with registration errors.

7.4.1 Survey of fusion techniques

From the interpolation point of view, there are two main strategies to process non-uniformly distributed samples: we can use the same interpolation kernel

everywhere and fit these kernels to the measurement data in a way that the reconstructed signal goes through the samples or we can define tailored basis functions (such as radial basis functions) that are better suited to the underlying non-uniform structure. Note that in higher dimensions the B-spline formalism is no longer applicable unless the grid is separable [Unser, 2000]. A more general approach is to use radial basis functions, which are closely related to splines as well, such as the membrane and thin-plate splines [Glasbey and Mardia, 1998].

In [Lertrattanapanich and Bose, 2002], the authors applied spatial tessellation and approximate each triangle patch in the Delaunay triangulation by a bivariate polynomial in order to reconstruct the HR image. Nguyen and Milanfar performed the reconstruction of non-uniformly sampled signals using wavelets in a multiresolution setting [Nguyen and Milanfar, 2000]. The main drawback of these interpolation techniques is the sensitivity to image noise and in addition, a conflict could arise if there are multiple noisy samples at the same position or very close to each other.

Iterative simulate-and-correct approaches to non-uniform interpolation are intuitively very simple. The most well-known iterative method is the Papoulis-Gerchberg algorithm [Gerchberg, 1974, Gerchberg, 1989, Papoulis, 1975] in which alternately, the known set of irregularly placed samples are projected onto the HR grid and an ideal low-pass filter is applied on the HR image to enforce bandlimitation. In the more general POCS algorithms, the ideal low-pass filter is substituted by other convex set operations (e.g. Gaussian blur). Iterative back-projection methods update the current estimated HR image by projecting the residual errors between the observed and the simulated LR images [Peleg et al., 1987, Keren et al., 1988]. The simulated LR images are simply obtained by resampling the current HR image.

A very fast and memory efficient way to aggregate multiple LR images into one HR image is the *shift-and-add* method (see also Section 5.3.2). This method assigns each pixel of the LR image to the nearest HR grid point after proper registration and upsampling as illustrated in Figure 4.8. If several samples are located on the same HR grid point, the HR pixel is estimated as the average or median value of these samples [Elad and Hel-Or, 2001, Farsiu et al., 2004]. Because the samples are snapped to the nearest grid points, the shift-and-add algorithm automatically generates additional positional errors on top of the registration errors. This effect adds another kind of correlated noise and artefacts to the reconstructed images such as undesirable and false zipper artefacts around edges. These errors can be reduced by choosing a large resolution enhancement factor, but as a result, the reconstructed HR image will have many more missing pixels due to an insufficient number of LR samples. In Section 5.3.2, we handle the problem of the missing pixels by assigning appropriate weights (i.e. according to the number of observation or samples) in the restoration step.

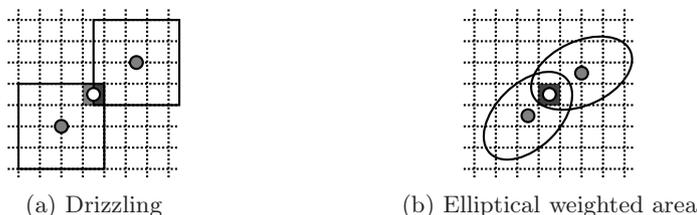


Figure 7.8: Illustration of the overlapping between the lattice area of the HR pixel (indicated by white circles) and the footprint of the samples (indicated by grey circles).

Another way to solve the problems of missing HR pixels is to enlarge the footprint of each sample of the LR images. The *variable-pixel linear reconstruction* algorithm, or informally known as *drizzling*, computes each HR pixel as the weighted average from all contributing surrounding samples [Fruchter and Hook, 2002]. A sample contributes to a HR pixel if the HR grid position is lying inside a square window around the sample, while the weight is determined by the degree of overlap between this square window and the area of the HR pixel lattice. An alternative to square windows are the use of adaptive ellipses, which results in *elliptical weighted area* (EWA) filtering techniques, where the ellipses are oriented according to the transformed LR grid [Jiang et al., 2003]. Both concepts interpret samples as tiny *waterdrops* (hence the term drizzling) raining on the HR grid (the coverage of these waterdrops are illustrated in Figure 7.8). Note that the shift-and-add method is a special case of drizzling techniques where the size of the square window is so small that each sample only covers one HR pixel.

In the drizzling and EWA fusion techniques, all HR pixels within the coverage of a sample receive the same weight no matter how far the HR pixel is lying from the sample position. Assigning weights in function of the spatial distance of between the HR pixel position and the sample position, results in the *Nadaraya-Watson* estimator [Nadaraya, 1964].

In [Pham et al., 2006], the authors used *structure adaptive normalized convolution*, which approximates the local signal by a set of basis functions such as the first-order polynomial basis. The greyvalues on the HR grid is then computed from the combination of these basis functions. In [Takeda et al., 2007], the authors proposed the use of *kernel regression* tools as a unified framework that combines the concepts of drizzling, EWA, Nadaraya-Watson estimator and normalized convolution methods. In the next sections, we will discuss these kernel regression techniques in more detail.

7.4.2 Kernel regression in the ordinary least square sense

We briefly describe the kernel regression method for solving the resampling problem in the ordinary least square sense as proposed Takeda et

al. [Takeda et al., 2007]. Suppose that we have to estimate the pixel value $f(\mathbf{x})$ at position \mathbf{x} on the HR grid. In the surrounding neighbourhood, we have a set of p noisy measurements g_i at irregularly sampled positions \mathbf{x}_i , the data measurement model is then given by:

$$g_i = f(\mathbf{x}_i) + n_i, \quad i = 1, \dots, p, \quad (7.62)$$

where $f(\cdot)$ is the unknown HR image, which also referred to as the *regression function* and n_i are independently and identically distributed zero-mean noise values. In a local neighbourhood, we can approximate the regression function by its local expansion of degree N . For example, we use the second order Taylor's series expansion ($N = 2$) of $f(\cdot)$, which is denoted by:

$$\begin{aligned} f(\mathbf{x}_i) &\approx f(\mathbf{x}) + \{\nabla f(\mathbf{x})\}^T (\mathbf{x}_i - \mathbf{x}) + \frac{1}{2} (\mathbf{x}_i - \mathbf{x})^T \{\mathcal{H}f(\mathbf{x})\} (\mathbf{x}_i - \mathbf{x}) \\ &\approx \beta_0 + \beta_1^T (\mathbf{x}_i - \mathbf{x}) + (\mathbf{x}_i - \mathbf{x})^T \beta_2 (\mathbf{x}_i - \mathbf{x}), \end{aligned} \quad (7.63)$$

where ∇ and \mathcal{H} are the gradient and Hessian operators, respectively. The coefficients of this polynomial are estimated by the following weighted least-squares optimization problem ($\hat{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2]$):

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^p [g_i - \beta_0 - \beta_1^T (\mathbf{x}_i - \mathbf{x}) - (\mathbf{x}_i - \mathbf{x})^T \beta_2 (\mathbf{x}_i - \mathbf{x})]^2 k_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x}), \quad (7.64)$$

which can easily be solved using Gauss-Newton algorithms and where $\hat{f}(\mathbf{x}) = \hat{\beta}_0$ is the estimated pixel value at the position \mathbf{x} on the HR grid, which we are looking for. The kernel function $k_{\mathbf{H}}(\cdot)$ (which has typically a Gaussian or exponential form) penalizes distances further away from the grid position and its strength is controlled by the smoothing matrix \mathbf{H} :

$$k_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x}) = |\mathbf{H}|^{-1} k(\mathbf{H}^{-1} [\mathbf{x}_i - \mathbf{x}]). \quad (7.65)$$

In case of $N = 0$, the solution corresponds to the Nadaraya-Watson estimator:

$$\hat{f}(\mathbf{x}) = \hat{\beta}_0 = \frac{\sum_{i=1}^p g_i k_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}{\sum_{i=1}^p k_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}. \quad (7.66)$$

This estimator only models locally flat signals, but does not model edges, ridges and blobs very well. On the other hand, the estimator given by equation (7.64) also takes these edges, ridges and blobs into account.

In most applications, the 2×2 smoothing matrix \mathbf{H} is equal to $h\mathbf{I}$ with h being the bandwidth parameter such that the kernel's footprint is isotropic. This is referred to as *classic kernel regression*. Adapting the kernel's footprint

locally according to the samples can prevent oversmoothing across edges in the same way as the geometry-driven tools discussed in Section 4.3.2. Anisotropic footprints of the kernel function $k_{\mathbf{H}}(\cdot)$ are illustrated in Figure 4.5. Therefore, the use of geometry-driven kernel functions is referred to as *steering kernel regression*. Such a modification requires at least two iterations in the estimation algorithm: in the first step, the construction of such oriented kernels employs the first-order derivatives of the unknown HR image, which can be estimated from the non-uniformly distributed samples via equation (7.64) as $\hat{\beta}_1$. In the subsequent steps, the desired pixel value $\hat{\beta}_0$ and the refined first-order derivatives $\hat{\beta}_1$ are estimated with the newly modified smoothing matrices. For a more detailed discussion of steering kernel regression, we refer the interested reader to the paper of Takeda et al. [Takeda et al., 2007].

We implemented the different kernel regression algorithms in 3D. In Figure 7.9, we give the results for a small synthetic experiment. From a 3D MRI volume, we simulate a fused slice by discarding 75% of the pixels. We reconstruct the missing pixels with classic kernel regression ($N = 0$ and $N = 2$) and steering kernel regression ($N = 2$ with 2 iterations) from the 25% remaining pixels in 3D. We can clearly see that the steering kernel regression outperforms the other methods both visually and in terms of PSNR.

7.4.3 Kernel regression in the total least square sense

The main drawback in the non-uniform resampling methods earlier mentioned is that these techniques do not take positional or registration errors into account. However, such errors are very common in practical SR applications, especially in the presence of severe image noise. To cope with these errors, we propose a novel kernel regression method in the TLS sense. The derivation of this algorithm is very similar to the derivation of the registration algorithm proposed in Section 7.3.2. Therefore, we omit some intermediate steps and give the formulas that are needed for the implementation. Our improved data measurement model of equation (7.62) is given by

$$g_i = f(\mathbf{x}_i + n_{\mathbf{x}_i}) + n_i, \quad i = 1, \dots, p, \quad (7.67)$$

where $n_{\mathbf{x}_i}$ is the relative positional error of $\mathbf{x}_i = (x_i, y_i)$ compared to the position $\mathbf{x} = (x, y)$ on the HR grid. $n_{\mathbf{x}_i}$ and n_i are assumed to be zero-mean distributed because the orthogonal distance regression minimizes the distance between the hyperplane and the measurement data. In case of the second order Taylor's series expansion (7.63), we have to estimate the coefficient vector β

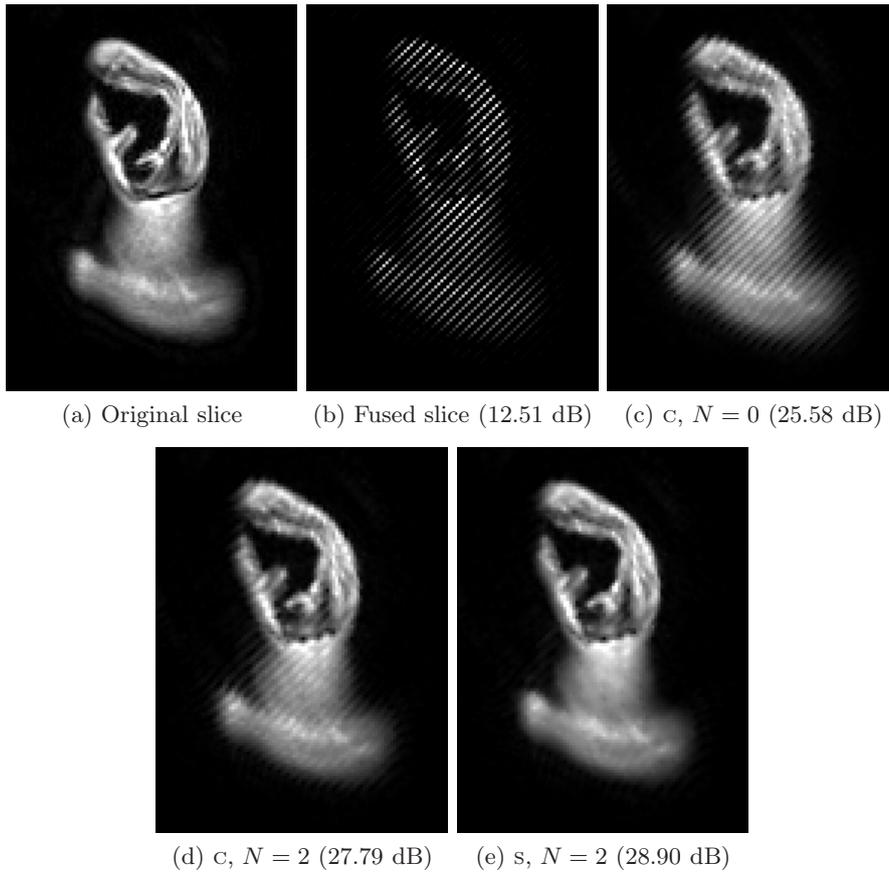


Figure 7.9: Illustration of several kernel regression results (c stands for *classic* and s stands for *steering*) on the proposed fusion experiment of a 3D MRI volume. The PSNR values are given in parentheses.

denoted by

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}) \\ \nabla_x f(\mathbf{x}) \\ \nabla_y f(\mathbf{x}) \\ \frac{1}{2}\mathcal{H}_{xx}f(\mathbf{x}) \\ \mathcal{H}_{xy}f(\mathbf{x}) \\ \frac{1}{2}\mathcal{H}_{yy}f(\mathbf{x}) \end{pmatrix}. \quad (7.68)$$

The estimation of these coefficients can be interpreted as a regression problem that fits the parameters to a hyperplane, given by the following implicit function:

$$h(\mathbf{Y}, \boldsymbol{\beta}) = \beta_0 - Y_0 + Y_1\beta_1 + Y_2\beta_2 + Y_3\beta_3 + Y_4\beta_4 + Y_5\beta_5 = 0. \quad (7.69)$$

Similarly to the orthogonal distance regression formulation in equation (7.46), the measurements of this hyperplane are given by the p non-uniformly distributed samples g_i :

$$\mathbf{Y}_i = \begin{pmatrix} Y_{0,i} \\ Y_{1,i} \\ Y_{2,i} \\ Y_{3,i} \\ Y_{4,i} \\ Y_{5,i} \end{pmatrix} = \begin{pmatrix} g_i \\ x_i - x \\ y_i - y \\ (x_i - x)^2 \\ (x_i - x)(y_i - y) \\ (y_i - y)^2 \end{pmatrix}. \quad (7.70)$$

The orthogonal projection \mathbf{Y}'_i on the hyperplane can be written in the following closed-form expression:

$$\mathbf{Y}'_i(\boldsymbol{\beta}) = \begin{pmatrix} Y_{0,i} - \beta_0 \frac{\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})} \\ Y_{1,i} - \beta_1 \frac{\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})} \\ Y_{2,i} - \beta_2 \frac{\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})} \\ Y_{3,i} - \beta_3 \frac{\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})} \\ Y_{4,i} - \beta_4 \frac{\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})} \\ Y_{5,i} - \beta_5 \frac{\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})} \end{pmatrix}, \quad (7.71)$$

where the terms ξ_i and v are a shorthand notation for

$$\xi_i(\boldsymbol{\beta}) = \beta_0 - Y_{0,i} + Y_{1,i}\beta_1 + Y_{2,i}\beta_2 + Y_{3,i}\beta_3 + Y_{4,i}\beta_4 + Y_{5,i}\beta_5, \quad (7.72)$$

and

$$v(\boldsymbol{\beta}) = \beta_5^2 + \beta_4^2 + \beta_3^2 + \beta_2^2 + \beta_1^2 + 1. \quad (7.73)$$

The l_2 -norm of the difference vector and the Jacobian matrix for each measurement are given by

$$d_{g,i} = \|\mathbf{Y}_i - \mathbf{Y}'_i(\boldsymbol{\beta})\|_2 = \frac{|\xi_i(\boldsymbol{\beta})|}{\sqrt{v(\boldsymbol{\beta})}} \quad (7.74)$$

and

$$\mathbf{J}_i(\boldsymbol{\beta}) = \begin{pmatrix} J_{0,i} \\ J_{1,i} \\ J_{2,i} \\ J_{3,i} \\ J_{4,i} \\ J_{5,i} \end{pmatrix} = \begin{pmatrix} \frac{\text{sign}(\xi_i(\boldsymbol{\beta}))}{\sqrt{v(\boldsymbol{\beta})}} \\ \text{sign}(\xi_i(\boldsymbol{\beta})) \frac{Y_{1,i}v(\boldsymbol{\beta}) - \beta_1\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})^{3/2}} \\ \text{sign}(\xi_i(\boldsymbol{\beta})) \frac{Y_{2,i}v(\boldsymbol{\beta}) - \beta_2\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})^{3/2}} \\ \text{sign}(\xi_i(\boldsymbol{\beta})) \frac{Y_{3,i}v(\boldsymbol{\beta}) - \beta_3\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})^{3/2}} \\ \text{sign}(\xi_i(\boldsymbol{\beta})) \frac{Y_{4,i}v(\boldsymbol{\beta}) - \beta_4\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})^{3/2}} \\ \text{sign}(\xi_i(\boldsymbol{\beta})) \frac{Y_{5,i}v(\boldsymbol{\beta}) - \beta_5\xi_i(\boldsymbol{\beta})}{v(\boldsymbol{\beta})^{3/2}} \end{pmatrix}. \quad (7.75)$$

The coefficient vector $\boldsymbol{\beta}$ is estimated iteratively according to the Gauss-Newton update rule:

$$\hat{\boldsymbol{\beta}}^{(j+1)} = \hat{\boldsymbol{\beta}}^{(j)} + \delta\hat{\boldsymbol{\beta}}^{(j)}, \quad (7.76)$$

where the incremental update $\delta\hat{\boldsymbol{\beta}}^{(j)}$ is given by the orthogonal distance regression solution, which includes the weights $w_i = k_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})$ of the kernel function:

$$\delta\hat{\boldsymbol{\beta}} = - \begin{pmatrix} \sum_{i=1}^p w_i J_{0,i}^2 & \dots & \sum_{i=1}^p w_i J_{0,i} J_{5,i} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^p w_i J_{5,i} J_{0,i} & \dots & \sum_{i=1}^p w_i J_{5,i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^p w_i J_{0,i} d_{g,i} \\ \vdots \\ \sum_{i=1}^p w_i J_{5,i} d_{g,i} \end{pmatrix}. \quad (7.77)$$

The performance of the proposed kernel regression algorithm is evaluated in Section 7.5.4.

7.5 Experimental results

7.5.1 Super-resolution in time

In some cases it is possible to perform super-resolution in the temporal domain if only *one* video stream is available. If the motion is periodic for instance, we can align the frames in time that contain the same motion pattern to achieve a higher temporal resolution. Motion blur is then removed by deconvolution in the temporal direction with a rectangular blur function as discussed in Section 7.1.4. This strategy is very similar to the non-local restoration methods discussed in Chapter 5, but it is now applied using periodic motion patterns in the temporal direction instead of repetitive structures in the spatial domain.

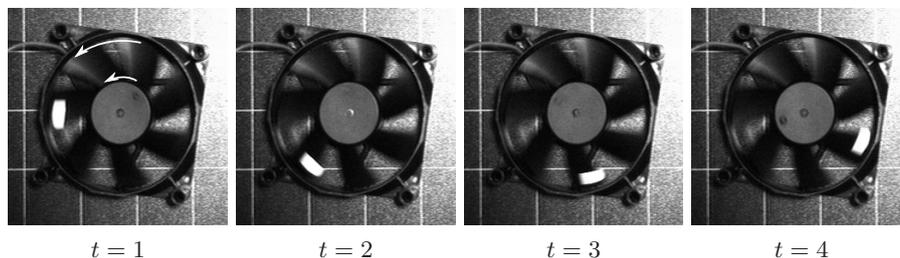


Figure 7.10: Illustration of a rotating fan.

As a small multi-frame motion deblurring experiment, we capture a rotating fan moving at two different speeds (one sequence at approximately 3 rotations per second during 3288 frames and the other at 8 rotations per second during 2047 frames) using the SONY DXC-9100P video camera. The first 4 frames of the first image sequence are given in Figure 7.10. From these images, we can clearly see the temporal aliasing effect: the individual blades do not move very much seemingly, however, the white sticker on one of the blades indicates the true large motion.

Motion deblurring of the fast rotating fan in the spatial domain requires a very complex algorithm, which in turn needs the segmentation of the individual blades and the knowledge of space-varying PSF's (because the velocity of the blades is varying in function of the distance to the axis as indicated by the arrows in Figure 7.10). Instead of applying motion deblurring in the spatial domain, we propose to perform SR reconstruction in the temporal direction to sharpen the rotating fan.

To perform super-resolution in time, we first increase the frame rate (or temporal resolution) tremendously ($50\times$) because there is not enough temporal correlation between successive frames due to the very large motion of the blades and because we need enough measurements/frames in the same exposure time interval (as illustrated in Figure 7.11). Instead of interpolation along the motion trajectories, we perform temporal SR by aligning other frames of the image sequence between the first frames. The temporal alignment is simplified in our case by tracking the centroid of the white sticker. The centroid follows a circular path, from which its spatial position can be linked to the relative position of each individual frame in time.

In the second step, we perform the image restoration (i.e. the actual motion deblurring) in a MAP framework using Tikhonov regularization (see Chapter 4). The PSF kernel in the degradation model is 3D: in the spatial direction, we use a small Gaussian blur with standard deviation $\sigma_b = 1$, while in the temporal direction, we employ a rectangular function that spans over 21 frames (i.e. the support of the kernel is $[-10, 10]$) for the slow motion (± 3 rotations per second)

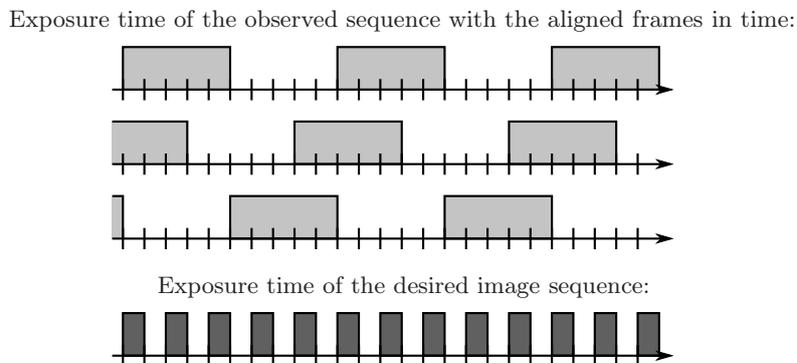


Figure 7.11: Exposure time of the observed images and the super-resolved images.

and over 69 frames ($[-34, 34]$) for the fast motion (± 8 rotations per second). The parameters have been selected based on trial and error to produce the visually most appealing results. We apply 100 steepest descent iterations with a small constant regularization parameter $\alpha = 0.5$. The result of this simple experiment is shown in Figures 7.12 and 7.13. We can see that edges of the individual blades are reconstructed very well in both cases, which demonstrates the effectiveness of the proposed approach with the temporal blur kernel. Note that the temporal blur kernel does not affect the restoration of the static scene.

To verify if the reconstructed edges are really sharper, we apply Canny edge detection [Canny, 1986] with modifications as suggested by Fleck [Fleck, 1992], on the images as illustrated in Figure 7.14. The edge detection algorithm is followed by a mathematical morphological thinning step (or also called *mathematical morphological skeletonization*) to suppress the local non-maxima. In Figure 7.14, we can clearly see that we are able to detect more edges and also the shape of the individual blades becomes perceptible using the proposed method.

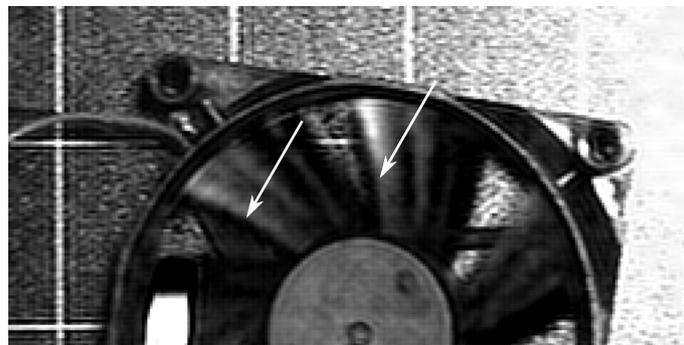
7.5.2 Geometric image registration

In this section, we perform a simple controlled experiment to measure the subpixel precision of various state-of-the-art geometric registration methods compared to several modifications of the gradient-based registration algorithm as discussed in Section 7.2.3.3 and Section 7.2.3.4. In this experiment, we restrict ourselves to the translational motion representation.

In order to compare the registration performance, we obtain the ground truth subpixel motion vectors starting from 5 HR images. From each HR image, 11 degraded 200×200 LR images (which correspond to 1 reference LR image and 10 pairwise registrations) are generated by successively shifting the HR image by integer pixels, cropping (to remove undefined border regions), performing



(a) Original image (± 3 rotations per second)

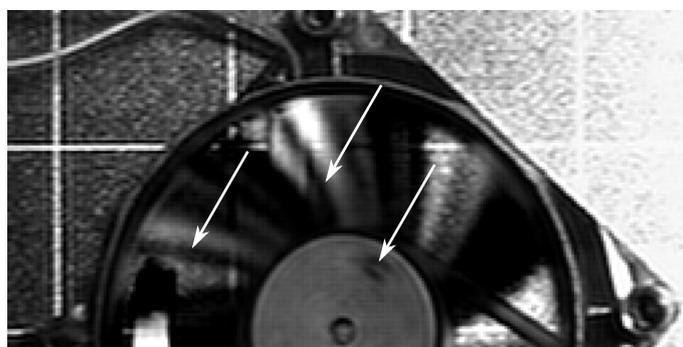


(b) Our motion deblurring result

Figure 7.12: Results of motion deblurring in time. The support of the kernel in the temporal direction is $[-10, 10]$.



(a) Original image (± 8 rotations per second)



(b) Our motion deblurring result

Figure 7.13: Results of motion deblurring in time. The support of the kernel in the temporal direction is $[-34, 34]$.

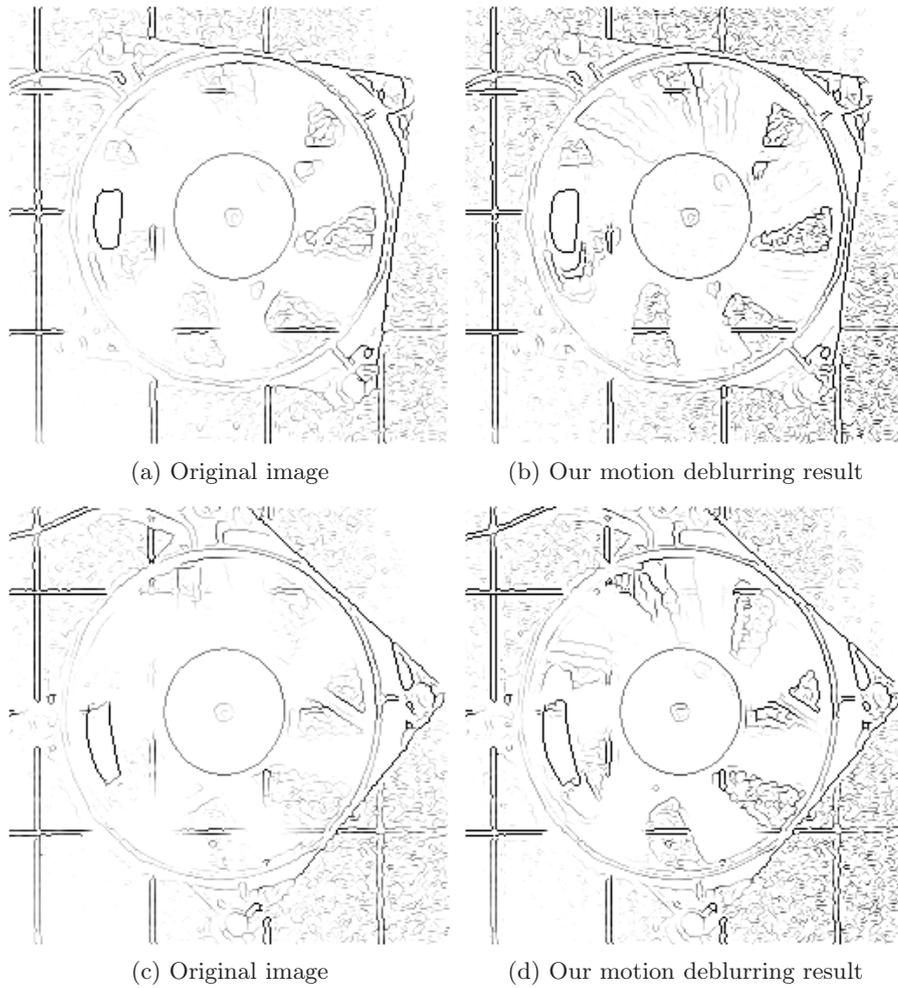


Figure 7.14: Results of Canny edge detection on the images produced by the experiments with (a)-(b) ± 3 rotations per second and (c)-(d) ± 8 rotations per second.



Figure 7.15: Reference LR image from the registration experiment ($\sigma_n = 10$).

decimation (where each LR pixel value is the average of a 4×4 block) and adding zero-mean white Gaussian noise (with standard deviation σ_n). Due to the decimation operation, the x - and y -offsets of the generated motion vectors are multiples of 0.25. An example of a simulated LR image is shown in Figure 7.15.

In this experiment, we compare the accuracy of various state-of-the-art registration methods such as the standard Lucas-Kanade gradient-based registration algorithm [Lucas and Kanade, 1981] (as given by equation (7.29)), a modified Levenberg-Marquardt gradient-based algorithm in a hierarchical polynomial spline pyramid [Thévenaz et al., 1998], a joint registration/reconstruction algorithm based on the principles of variable projections⁶ [Robinson et al., 2009], a frequency-based registration method performed on the aliasing-free part of the spectrum [Vandewalle, 2006] and the bias correcting gradient-based shift estimator [Pham et al., 2005].

We compare these registration methods to several modifications of the gradient-based motion estimation algorithm of [Lucas and Kanade, 1981] (all performed by 30 iterations within a 3-level Gaussian pyramid): a first simple modification is the use of the robust Huber loss function instead of the l_2 -norm in the cost

⁶The registration performance of this algorithm depends on the available image prior information for the reconstruction of the HR image. In this experiment, we assume that the HR image is locally smooth as in the Tikhonov regularization.



Figure 7.16: Blackman-Harris windowed sinc interpolation of Figure 7.15 as obtained by the G-SI method.

function as derived in equation (7.34). This modification is denoted as G-RH. A second modification is the use of Blackman-Harris windowed sinc resampling instead of the bilinear resampling scheme to generate the transformed image f in each iteration, which is denoted as G-SI. This modification is approximated by a LR-to-HR registration scheme, which is similar to the EM algorithm described in Section 7.2.3.4 and has a benefit that the computational complexity is much lower (the Blackman-Harris windowed sinc interpolation is only applied once on a separable grid and the bilinear resampling scheme can still be employed). The initialized HR image is illustrated in Figure 7.16.

A third modification is postprocessing the registration parameters with the bundle adjustment tool as given in equation (7.37) (based on the skew-antisymmetry and the Jacobi identity principles as illustrated in Figure 7.6). This modification is referred to as G-BA.

In the same spirit as the bundle adjustment tool, we optimize the group structures in the Lie algebra jointly with the registration parameters in the fourth modification as given in equation (7.38) and [Farsiu et al., 2005], which is denoted as G-JO.

As the last modification, we solve the joint registration and SR reconstruction problem by an EM algorithm that iteratively alternates between estimating the HR image (produced by the steering kernel algorithm from 11 LR images as described in Section 7.4.2) and the registration parameters as discussed



Figure 7.17: Steering kernel regression result as obtained by the proposed G-KR method.

in Section 7.2.3.4. The produced HR image is illustrated in Figure 7.17. In contrast to Figures 7.15 and 7.16, we are able to recognize some details such as the text above the gate. To diminish the computational complexity, the HR image is only reconstructed every 10 Gauss-Newton iterations of the gradient-based algorithm. This method is referred to as G-KR.

In Table 7.2, we show the average *root mean square error* (RMSE) registration accuracy between the 50 ground truth motion vectors and the estimated registration parameters for several noise variances. We observe that all of the algorithms provide a relative high degree of subpixel precision, which is sufficient to most practical SR applications with enlargement factors of 4 or 8. Robust loss functions (G-RH) do not contribute to the subpixel precision, however they are useful in the presence of scratches or object motions as illustrated in Figure 7.4. The choice of a better resampling operator (G-SI) has a positive influence on the accuracy and requires few extra computations (i.e. the interpolation of the reference image) and extra memory requirements to store the HR image. In this simple experiment, the incorporation of bundle adjustment (G-BA) or jointly optimizing the group structures in the Lie algebra (G-JO) improves the subpixel precision tremendously. The major drawback of both methods is the enormous computational load (i.e. $k(k-1)$ or 90 pairwise registrations are computed instead of 10), but these tools can be useful in the

Table 7.2: Average RMSE registration accuracy (in units of pixels) over 50 registrations in the presence of additive zero-mean white Gaussian noise (with standard deviation σ_n). The best RMSE result is written in bold letters for each noise level.

Registration method	$\sigma_n = 0$	$\sigma_n = 5$	$\sigma_n = 10$	$\sigma_n = 15$
[Thévenaz et al., 1998]	1.1e-2	1.2e-2	1.4e-2	2.1e-2
[Robinson et al., 2009]	1.7e-2	1.6e-2	2.0e-1	1.4e-1
[Vandewalle, 2006]	6.2e-3	1.2e-2	3.4e-2	5.6e-2
[Pham et al., 2005]	4.1e-3	4.9e-3	6.0e-3	8.7e-3
[Lucas and Kanade, 1981]	2.5e-2	2.6e-2	2.5e-2	2.5e-2
G-RH	2.4e-2	2.5e-2	2.5e-2	2.6e-2
G-SI	1.4e-2	1.5e-2	1.4e-2	1.5e-2
G-BA	1.1e-2	9.0e-3	9.9e-3	1.1e-2
G-JO	9.3e-3	9.1e-3	1.0e-2	1.1e-2
G-EM	2.9e-3	2.8e-3	4.2e-3	5.3e-3

presence of severe registration errors. The joint registration/reconstruction algorithm based on steering kernel regression (G-EM) delivers the most accurate registration parameters among all registration methods in our experiment. By reconstructing the SR image, the effect of frequency aliasing in the LR images is exploited efficiently for the motion estimation, which results in the highest subpixel precision.

7.5.3 Photometric image registration

In a first experiment, we perform a quantitative evaluation of the linear photometric registration in both the OLS sense (given by equation (7.41)) and the TLS sense (as described in Section 7.3.2). In Section 7.3.1, we have shown that the TLS solution⁷ produces consistent registration parameters, which is also illustrated in Figure 7.7. To compute the accuracy of the photometric parameters, we apply random linear greyscale modifications with gain parameters within $[0.8, 1.2]$ and bias parameters within $[-25, 25]$ to a 200×200 image and add zero-mean white Gaussian noise on both images. The RMSE accuracy is computed over 200 simulations for each noise level and is plotted in Figure 7.18. We can clearly see that the OLS solution of the photometric registration produces very bad results as the noise increases, while the TLS solution is very accurate and also robust to noise.

In a second controlled experiment, we compare the performance of the joint ge-

⁷Note that both svd-based solution [Markovsky and Van Huffel, 2007] and the orthogonal distance regression algorithm [Sullivan et al., 1994, Ahn et al., 2002] given in equation (7.46) produce exactly the same parameters.

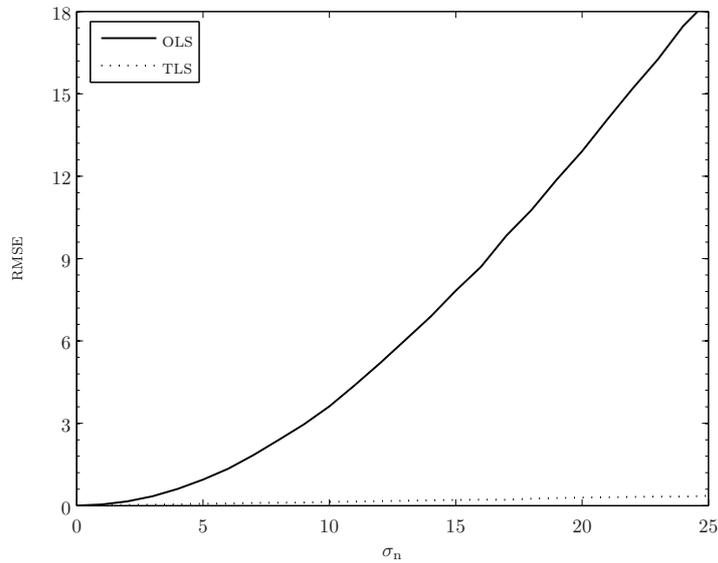


Figure 7.18: Average RMSE accuracy of the photometric registration parameters in function of additive zero-mean white Gaussian noise (with standard deviation σ_n).

ometric and photometric registration algorithms according to the affine/linear model (7.49). We simulate 20 degraded 500×500 LR images by successively applying a Tukey windowing function on the HR image (to prevent the influence of non-overlapping regions due to the spatial deformations), applying random photometric linear and geometric affine transformations, performing decimation (via averaging of 2×2 blocks) and adding zero-mean white Gaussian noise (with standard deviation σ_n). This process is repeated for each noise level. Examples of the degraded LR images are shown in Figure 7.19.

We compare the TLS solution as described in Section 7.3.2 to its OLS counterpart and the joint geometric linear and photometric affine registration algorithm of Bartoli [Bartoli, 2006], which operates in an inverse compositional gradient-based framework (see Section 7.2.3.3) using the ordinary least square metrics. The average RMSE accuracy is plotted in Figure 7.20 in function of the noise standard deviation. Again, the TLS solution produces the best results. In this experiment, the LR images are clipped at $[0, 255]$, which explains the relative performance difference between the TLS and OLS solution compared to the previous experiment (see Figure 7.18).

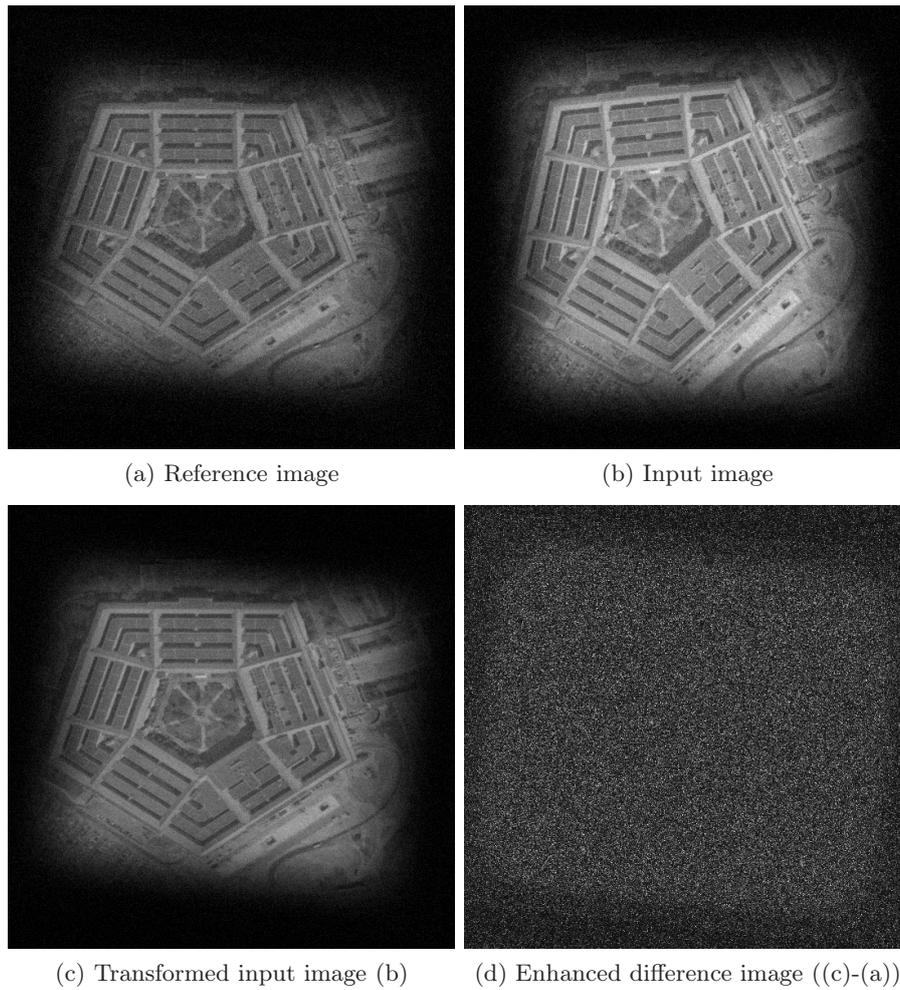


Figure 7.19: Illustration of joint photometric linear and geometric affine registration algorithm in the TLS sense ($\sigma_n = 10$).

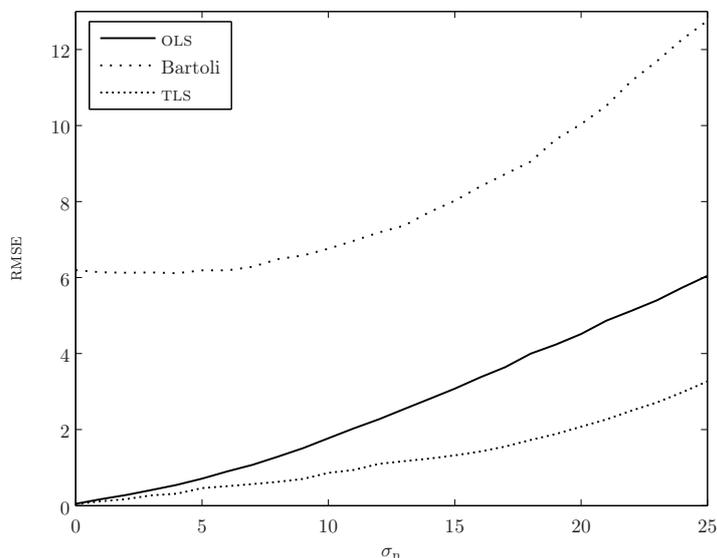


Figure 7.20: Average RMSE accuracy of all registration parameters in function of additive zero-mean white Gaussian noise standard deviation σ_n .

7.5.4 Image fusion

As a simple experiment, we have grabbed 30 LR images with the *Philips Inca Smartcam* in rather poor lighting conditions and we enlarge these images 4 times in each dimension. After applying gradient-based registration (G-RH), we compare various fusion algorithms (see Section 7.4.1 for a more detailed discussion) such as the Delaunay triangulation with a bicubic polynomial model [Lertrattanapanich and Bose, 2002], the POCs algorithm (with a Gaussian low-pass filter ($\sigma = 1$)), the shift-and-add method [Elad and Hel-Or, 2001, Farsiu et al., 2004], the structure adaptive normalized convolution [Pham et al., 2006] and the classic kernel regression algorithm [Takeda et al., 2007] (see Section 7.4.2 for a more detailed discussion). The image fusion results are shown in Figure 7.21.

We notice some heavy characteristic noise patterns in the Delaunay triangulation, POCs and shift-and-add fusion algorithms. The structure-adaptive normalized convolution results show less noise, but the noise amplitude changes near the edges of the characters (also referred to as *non-stationary noise*). The kernel regression method produces very smooth HR images, which are preferable to most deconvolution algorithms. These observations are explicitly illustrated in a more detailed view given in Figure 7.22.

We show the power spectral density from an image that is reconstructed by the

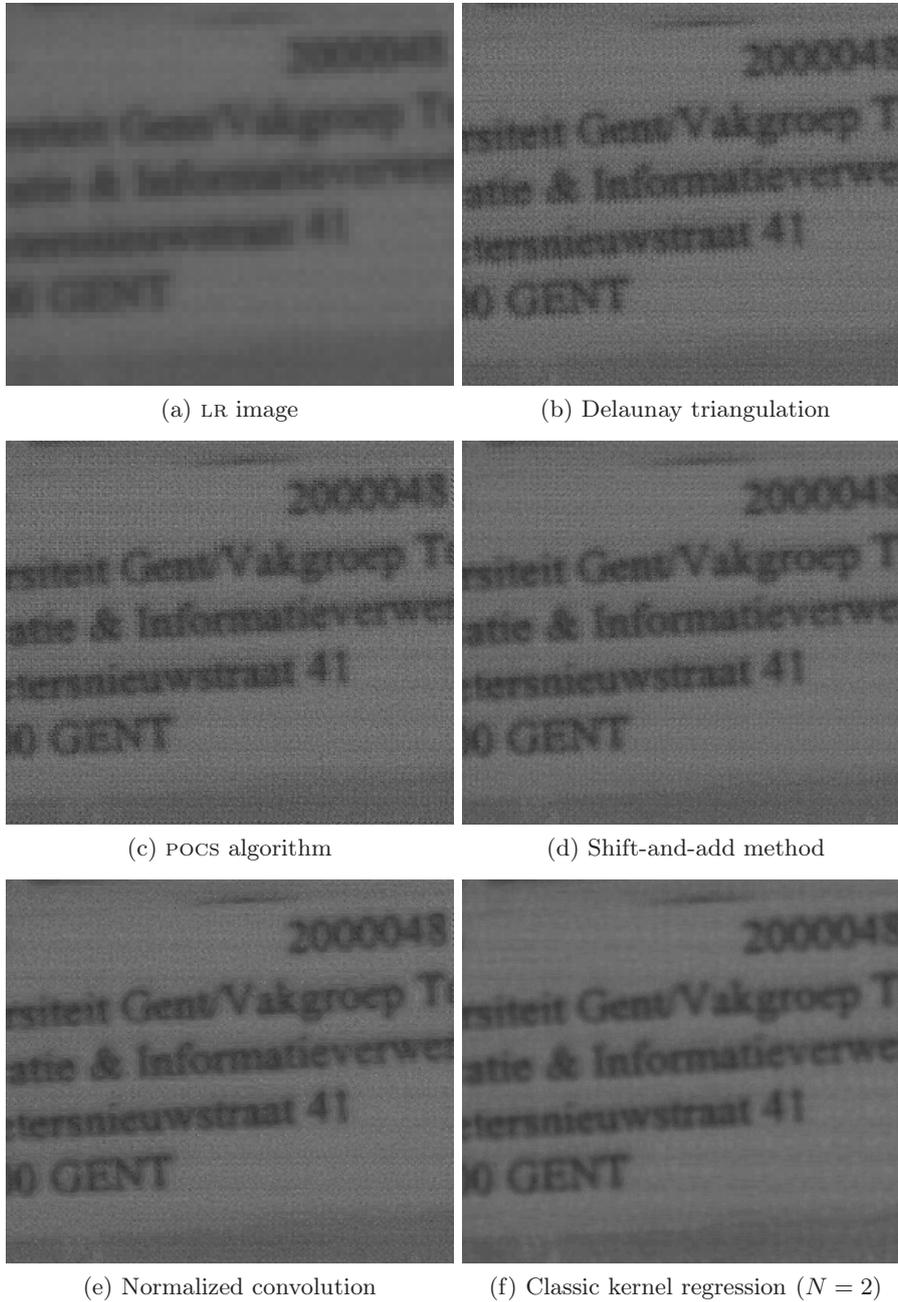
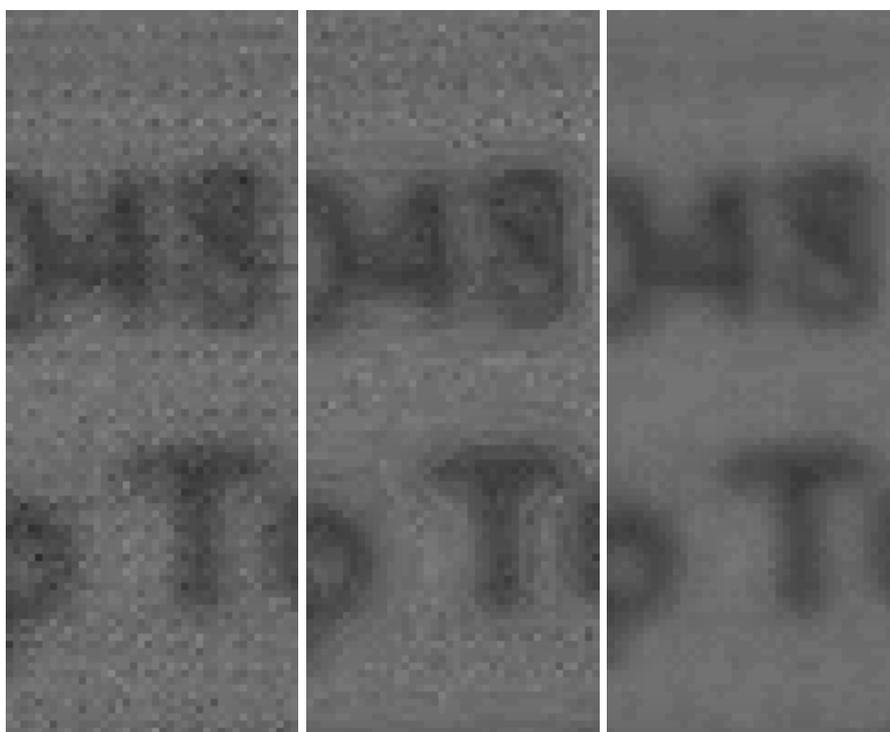


Figure 7.21: Illustration of several fusion algorithms over 30 images (all fused images are enhanced with the unsharp masking algorithm to emphasize the differences).



(a) Shift-and-add method (b) Normalized convolution (c) Kernel regression

Figure 7.22: Detailed view of the fusion results from Figure 7.21.

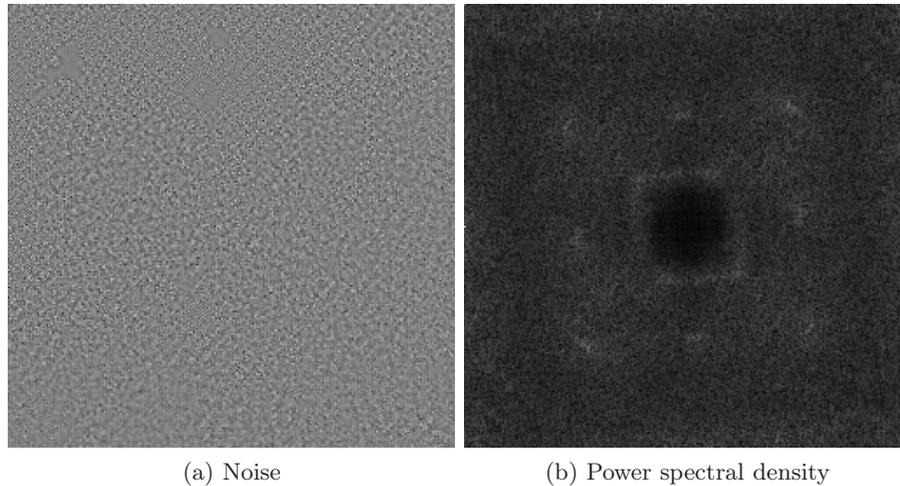


Figure 7.23: Illustration of the correlated noise pattern produced by the shift-and-add algorithm and its power spectral density.

shift-and-add fusion algorithm in Figure 7.23. This figure clearly demonstrates that the noise is *not* white, but correlated. The treatment of correlated noise in the SR framework is an interesting research topic for future work.

In Section 7.4.3, we proposed a new kernel regression algorithm in the TLS sense to handle positional or registration errors. We now determine which techniques (the OLS solution (referred to as KROLS) or the TLS solution (referred to as KRTLS)) are preferable under which conditions. We extract over 9 million 7×7 patches from the Kodak data set (see Figure 3.11). For each patch, we estimate the central pixel from its 48 neighbouring pixels (this is also referred to as the *leave-one-out* principle). These pixels are corrupted by additive zero-mean white Gaussian noise (with standard deviation σ_n) and suffer from random perturbations on the spatial coordinates (from a uniform distribution in the range of $[-\sigma_u, \sigma_u]$ for $\sigma_u \in [0, 2]$). From these pixel estimates, we compute the average RMSE accuracy for both classic and steering KROLS/KRTLS ($N = 2$ and 2 iterations for the steering version), which are shown in Figures 7.24, 7.25 and 7.26 for $\sigma_n = 0$, $\sigma_n = 10$ and $\sigma_n = 20$ respectively.

From this experiment, we can conclude that the steering versions produce slightly better RMSE results compared to the classic versions. However, there is a big difference in visual quality in some applications, especially when fewer samples are available (see for example Figure 7.9). For minor noise levels (e.g. $\sigma_n < 10$ or $\sigma_u < 0.8$), the OLS solution is preferable to the KRTLS method. In all other cases, the KRTLS method produces more accurate results. Note that the TLS solution also increases at a slower rate in function of σ_u .

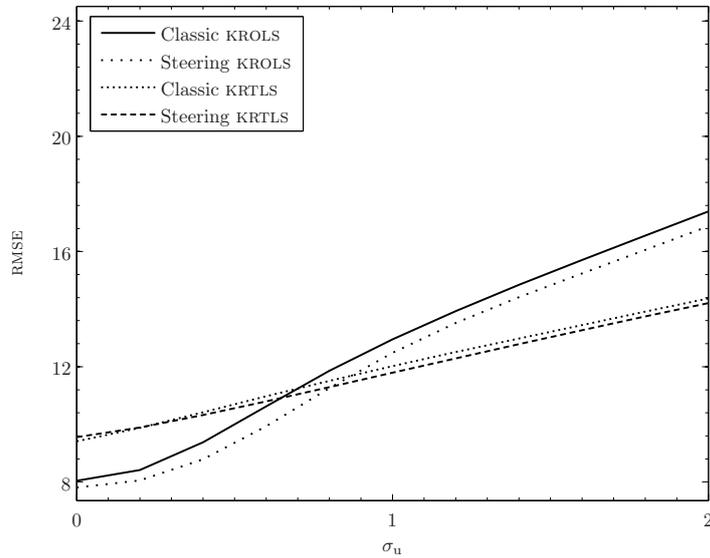


Figure 7.24: Average RMSE accuracy of the pixel value in function of the positional error (uniformly distributed in $[-\sigma_u, \sigma_u]$) with no added noise on the pixel values.

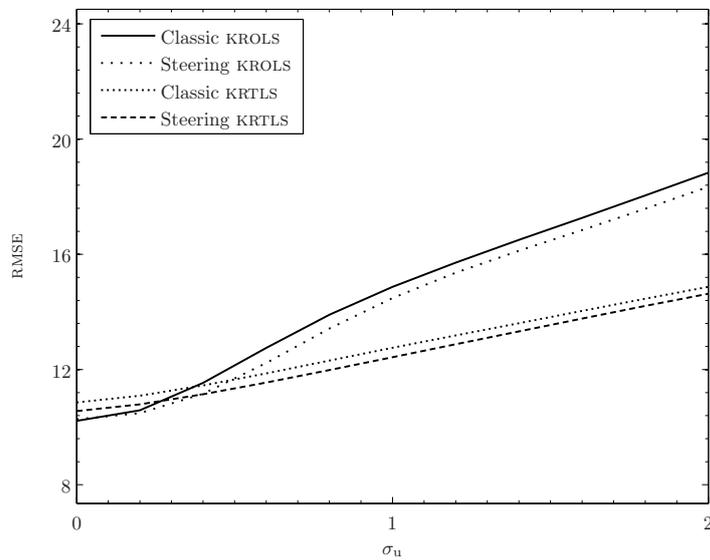


Figure 7.25: Average RMSE accuracy of the pixel value in function of the positional error (uniformly distributed in $[-\sigma_u, \sigma_u]$) in the presence of additive zero-mean white Gaussian noise ($\sigma_n = 10$).

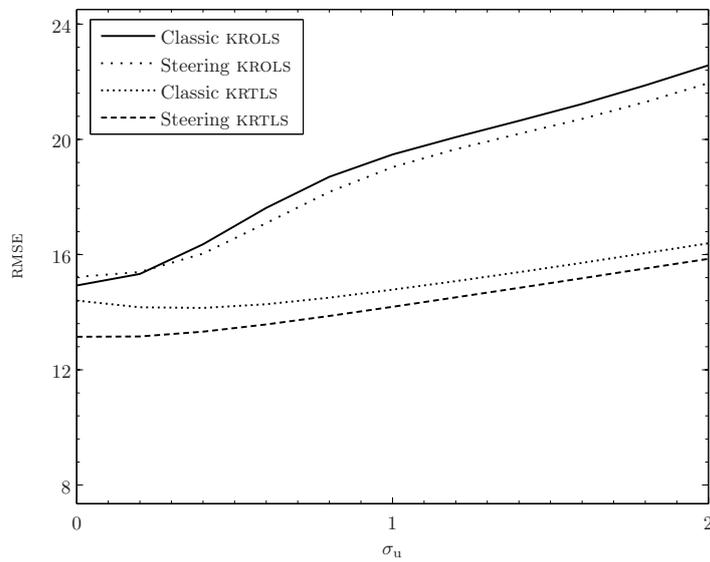


Figure 7.26: Average RMSE accuracy of the pixel value in function of the positional error (uniformly distributed in $[-\sigma_u, \sigma_u]$) in the presence of additive zero-mean white Gaussian noise ($\sigma_n = 20$).

In Figure 7.27, we give the visual difference between the OLS and TLS solution and we compare these results to the reference image (with additive zero-mean white Gaussian noise ($\sigma_n = 10$)). Random offsets ($\in [-2, 2]$) are added to the spatial coordinates to simulate registration errors. The TLS solution produces a much smoother image and reconstructs for example the vertical strips of the fence in a better way compared to the OLS result.

7.5.5 Practical super-resolution

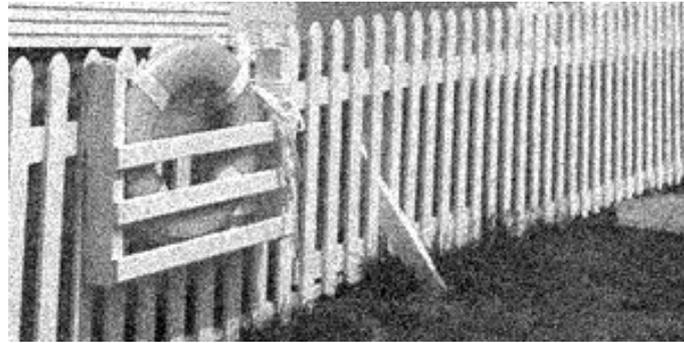
In this section, we show some SR image reconstruction results from real-world examples obtained by successive motion estimation, fusion and image restoration as shown in Figure 7.2. The proposed SR framework combines the gradient-based motion estimation algorithm with the robust Huber loss function (see Section 7.2.3.3), the classic kernel regression algorithm (see Section 7.4.2) and the joint deconvolution and denoising algorithm using the standard Tikhonov regularization (see Chapter 4). This SR algorithm is illustrated by the flow chart given in Figure 7.28.

In a first experiment, we process 30 LR text images from the *Philips Inca Smartcam* to produce a $4\times$ enlarged HR image (see also Figure 7.21). After shift-and-add fusion and classic kernel regression ($N = 2$), we apply joint deconvolution and denoising using the standard Tikhonov regularization scheme for zero-mean white Gaussian noise (see Chapter 4 for a more detailed discussion) within 100 iterations and with a Gaussian blur kernel ($\sigma_b = 3$).

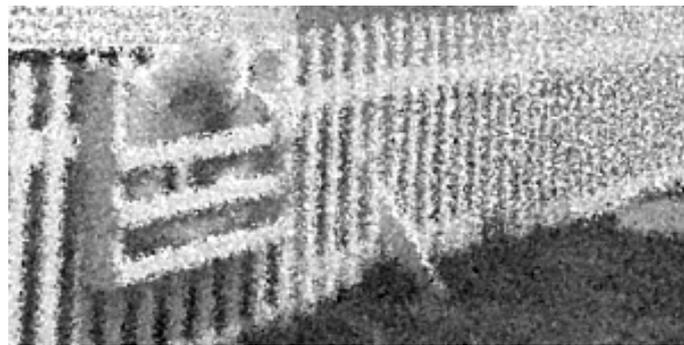
We compare the obtained SR results with the non-SR Blackman-Harris windowed sinc interpolation, iterative back-projection method (IBP) [Irani and Peleg, 1991], robust SR based on median shift-and-add scheme and restoration with a bilateral total variation image prior [Farsiu et al., 2004], fast SR based on median shift-and-add scheme and restoration with the robust Lorentzian loss function [Luong et al., 2006d].

As an objective evaluation, we apply optical character recognition on the reconstructed text images using two page-reading systems: Scansoft OmniPage 15.0 and ABBYY FineReader 8.0. The recognition is applied on the visible text, which contains 81 characters. Both visual results and character OCR accuracy results (see equation (6.15)) are given in Figure 7.29.

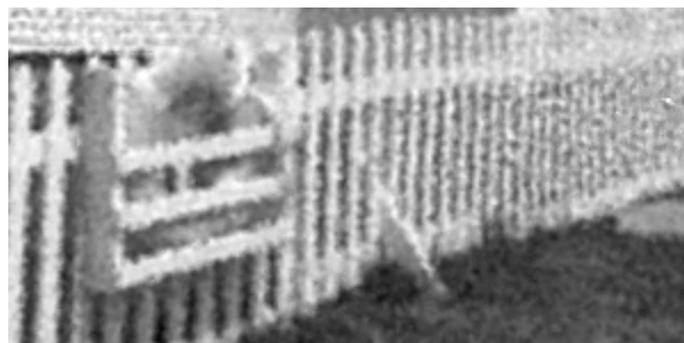
Blackman-Harris windowed sinc interpolation does not improve the image quality very much, while IBP SR method produces a blurred HR image despite the deconvolution. The SR results produced by [Farsiu et al., 2004] and [Luong et al., 2006d] are much sharper, however, both methods generate noticeable intensity staircase artefacts and painting effects. Applying the standard Tikhonov regularization for white noise does not remove the correlated noise entirely from the shift-and-add image. From all HR images shown in Figure 7.29, the combination of kernel regression and Tikhonov regularization delivers the image with the least number of artefacts and the



(a) Noisy reference image



(b) OLS solution



(c) TLS solution

Figure 7.27: Examples of the classic OLS and TLS solution in the presence of noise on the pixel values ($\sigma_n = 10$) and noise on the positions ($\sigma_u = 2$).

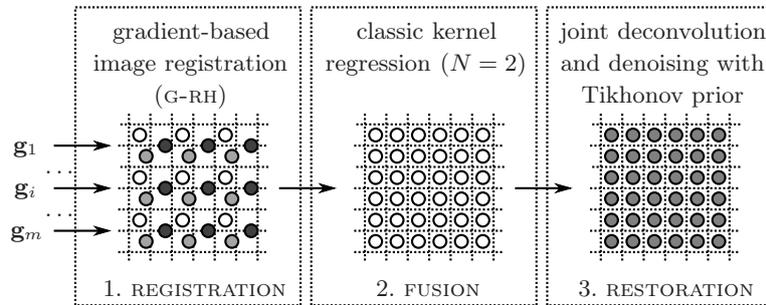


Figure 7.28: Schematic representation of the proposed SR algorithm.

highest readability in terms of objective character OCR accuracy measures.

In a second simple experiment, we apply the same fusion and restoration algorithms (with a Gaussian blur kernel of $\sigma_b = 2$) on the *car* sequence,⁸ where the car is moving from bottom to top during the whole image sequence (see Figure 7.30). Motion estimation is performed by simple block matching as a fast initialization (with 8×8 blocks given by equation (7.17)) at integer pixel accuracy followed by the gradient-based subpixel refinement per block.

The region of interest in Figure 7.30 is enlarged with a linear magnification factor of 8 and the results are shown in Figure 7.31. In most SR results, the make of car and the license plate become much more readable. The combination of kernel regression and Tikhonov regularization delivers again the best visual image quality. In this example, the effect of treating correlated noise as white noise in the joint deconvolution/denoising regularization (see for example shift-and-add and robust SR results) can clearly be noticed as annoying blob patterns across the whole HR image.

In the last experiment, we perform joint SR and deinterlacing (i.e. converting an interlaced image sequence into a progressive form). The interlaced image sequence provides us a considerable amount of frequency aliasing in the vertical direction. To retain the SR architecture as given in Figure 7.28, each interlaced frame is decomposed into two sub-images with only odd or even lines (which is referred to as *field*) as depicted in Figure 7.32. We perform SR image reconstruction on the decomposed image sequence, with the only difference being that the enlargement factor in vertical direction is now twice the enlargement factor in horizontal direction.

We apply the proposed deinterlacing/SR framework on the *football* sequence using shift-and-add fusion and kernel regression combined with Tikhonov regularization for joint deconvolution/denoising (with a Gaussian blur kernel of

⁸Downloadable from <http://www.ee.ucsc.edu/~milanfar/software/sr-datasets.html>.

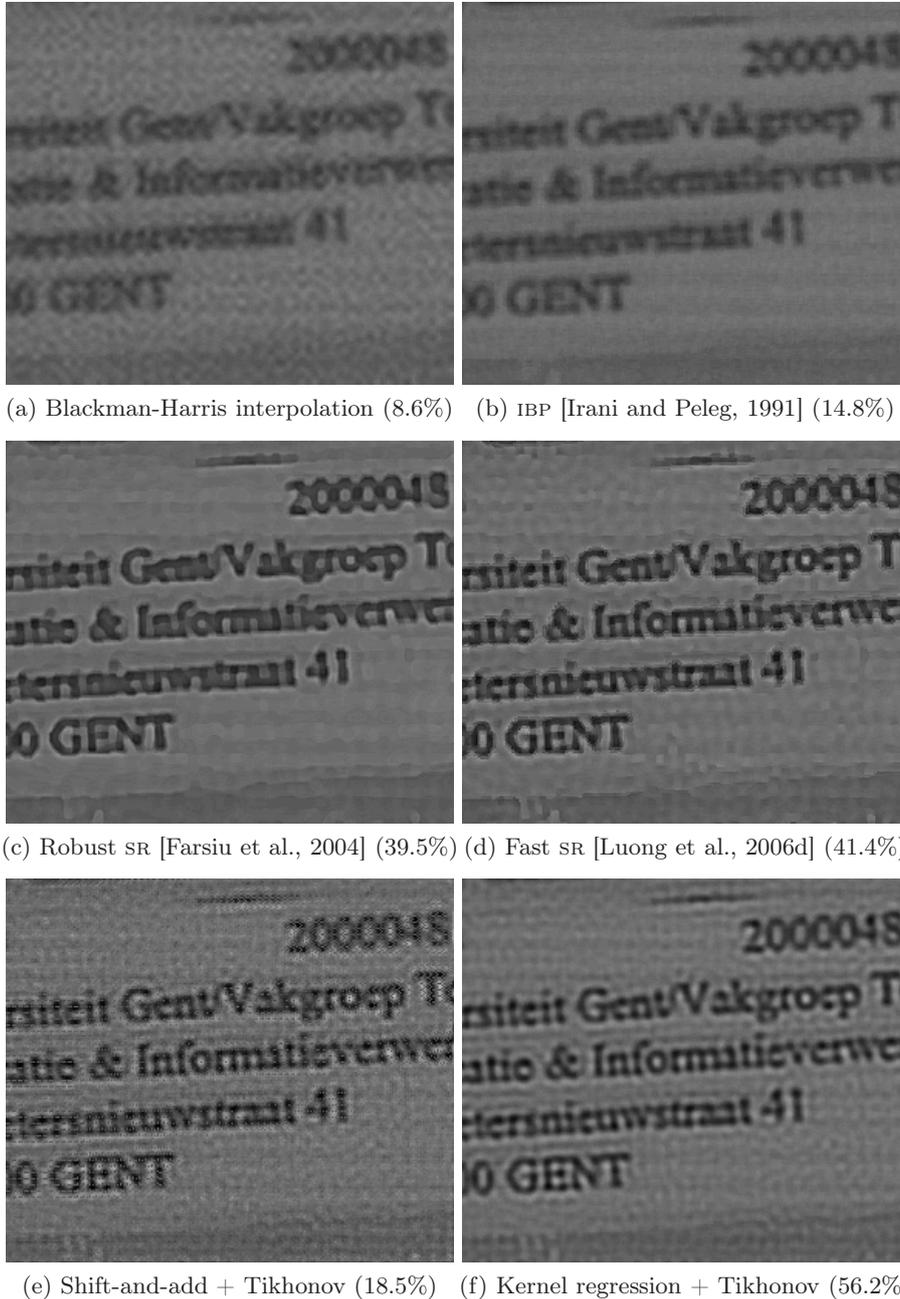


Figure 7.29: Super-resolution results of the *text* sequence with 30 images (all SR images are enhanced with the unsharp masking algorithm to emphasize the differences). The average character OCR accuracy is given in parentheses.



Figure 7.30: Frame 32 from the *car* sequence. The region of interest (ROI) is enlarged in Figure 7.31.

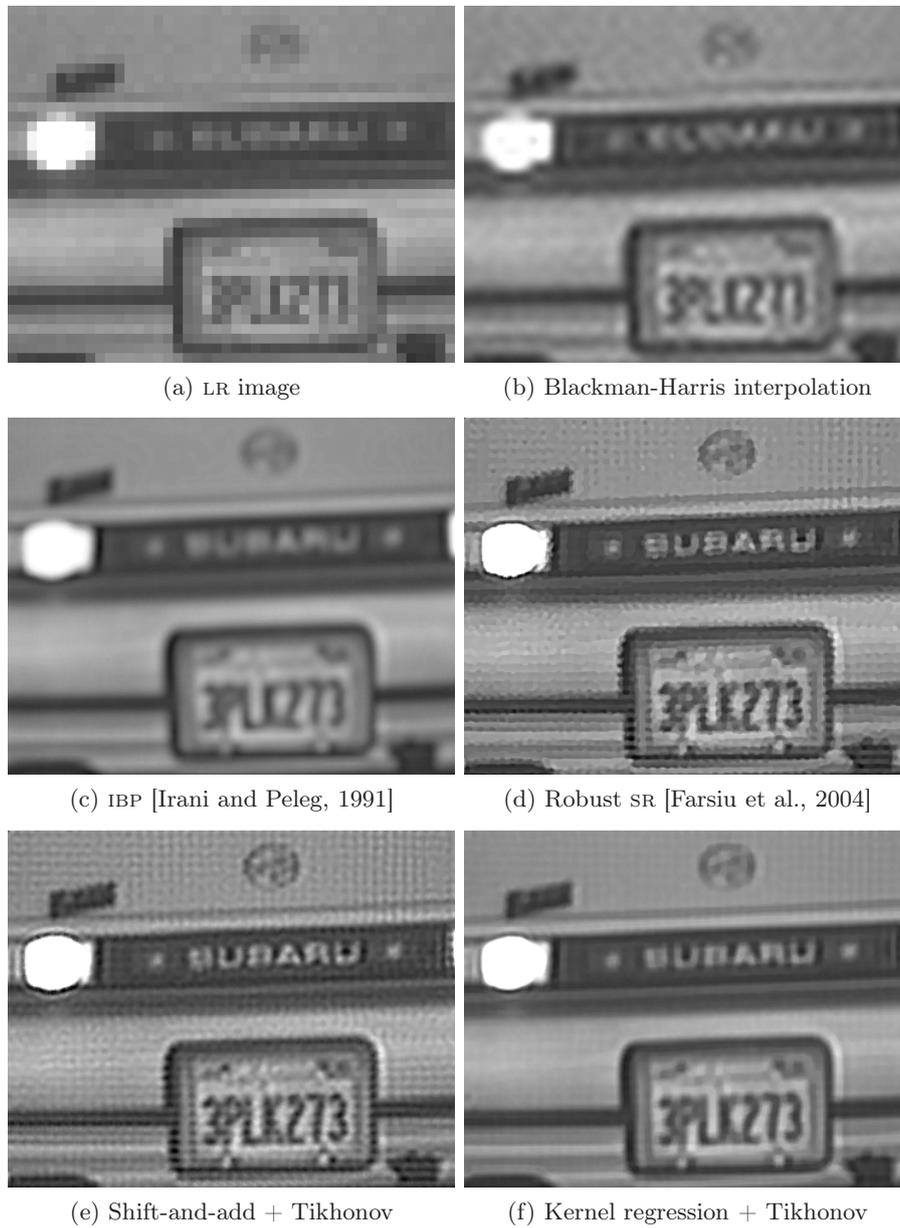


Figure 7.31: Super-resolution results of the *car* sequence with 64 images (all SR images are enhanced with the unsharp masking algorithm to emphasize the differences).

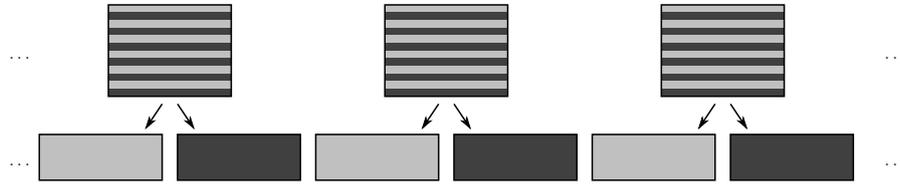


Figure 7.32: Principle of SR deinterlacing: each frame is decomposed into odd and even fields resulting in a new image sequence, which is feed into the SR algorithm given in Figure 7.2.



Figure 7.33: Frame 50 from the *football* sequence. The region of interest (ROI) is enlarged in Figure 7.34.

$\sigma_b = 2$ in 30 iterations). The original interlaced reference frame is shown in Figure 7.33 and the $4\times$ enlarged ROI results are shown in Figure 7.34. We compare the SR results with Blackman-Harris windowed sinc interpolation and a state-of-the-art wavelet deinterlacing technique, which performs joint spatio-temporal interpolation and denoising in the wavelet domain in a motion-compensated manner [Zlokolic et al., 2006], which is further enlarged 4 times with the Blackman-Harris interpolation.

The number on the helmet is clearly readable in the SR reconstructed images (number 82 corresponds to the number on the player's shirt), although the result produced by the shift-and-add method clearly suffers from correlated noise. The Blackman-Harris windowed sinc interpolation can not resolve the

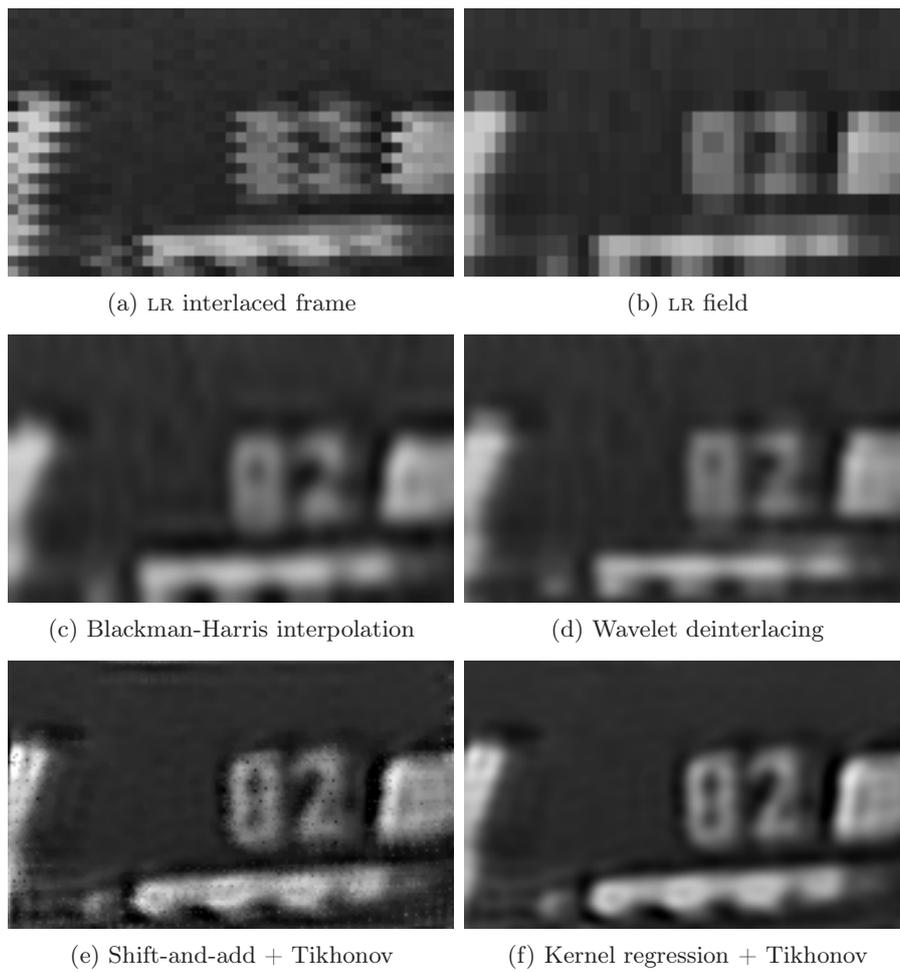


Figure 7.34: Super-resolution results of the *football* sequence with 50 images.

image details in such a way that the number becomes readable. Also applying wavelet deinterlacing does not improve the resolution drastically. This small example demonstrates that the SR framework is very suitable to deinterlacing problems and is even superior in reconstructing fine image details compared to standard deinterlacing/interpolation techniques.

7.6 Conclusion

Multi-frame super-resolution image restoration is quite a complex problem, which spans over several fields of image processing, such as motion estimation or image registration, image reconstruction from irregularly spaced samples, image deconvolution and denoising. In this chapter, we gave a broad overview of the existing spatial SR techniques, which can commonly be classified as frequency-domain, spatial-domain and motionless SR methods. Due to its relatively low computational load and low memory requirements, the standard three-step paradigm of the non-uniform interpolation SR approach is recommended in most practical applications. These three successive steps are image alignment (via subpixel motion estimation), image fusion (i.e. the conversion of the irregularly spaced samples into samples placed on the uniform HR grid) and HR image restoration (i.e. deconvolution and denoising).

Besides spatial resolution enhancement by SR techniques, it is also possible to perform SR in the temporal direction of the image sequence. Problems such as temporal aliasing and motion blur can be solved using SR methods. In Section 7.5.1, we evaluated a new multi-frame motion deblurring algorithm that exploits periodic motion patterns and applies deconvolution with a 3D PSF (where the component in temporal direction is related to the exposure time). The experimental results clearly show the effectiveness of the proposed approach. SR in the temporal direction has direct applications in frame rate conversion, e.g. from SD to HD, with the respective frame rates of 25 FPS and 50 FPS. An interesting future application for the proposed multi-frame motion deblurring algorithm is the restoration of video sequences coming from a vibrating camera because these sequences often contain a lot of periodic motion patterns.

In Section 7.2 and Section 7.3, we discussed various aspects of motion estimation and both geometric and photometric image registration. We proposed a new LR-to-HR gradient-based registration method that reconstructs the HR image via steering kernel regression. We evaluated the subpixel accuracy of several state-of-the-art shift estimators and concluded that the proposed algorithm produces the most accurate motion estimation.

For the photometric and joint geometric/photometric registration problem, we have introduced the use of the total least square framework in the proposed registration algorithms. In both cases, our methods produce more accurate and consistent registration parameters compared to the methods that use the ordinary least square approach, which are commonly employed in the literature.

In Section 7.4.3, we have proposed and derived a new kernel regression algorithm in the TLS sense to handle positional or registration errors. In case of moderate or heavy noise (on the pixel intensity values and/or on the spatial coordinates), the proposed method is more accurate in terms of RMSE compared to the standard kernel regression algorithms defined in the OLS sense.

Finally, we have demonstrated some SR results on real-world examples, where we noticed that correlated noise, from the image fusion and especially from the shift-and-add algorithm, provides a source of potential artefacts. We also showed that the SR framework is suitable to deinterlacing problems and is even superior in reconstructing the lost image information compared to deinterlacing/interpolation techniques.

Future work includes for instance the estimation of the noise covariance matrix to eliminate correlated noise properly in SR frameworks. A second challenging research topic is to develop real-time SR algorithms in which the focus is lying on visual pleasing images (i.e. with almost no artefacts) instead of exploring the limits of recovering spatial resolution. This will be important in many multimedia applications, where for example low-resolution video content in very bad quality (e.g. from the Internet or mobile phones) is displayed on high-resolution screens (e.g. HDTV). In addition, other sources of degradation such as compression and dynamic range should be taken into account, which makes the SR algorithm even more complex.

8

MRI resolution enhancement

Yes we can!
—Barack Obama

The possibility of applying super-resolution techniques to magnetic resonance imaging (MRI) and extending this way spatial resolution to a microscopic level is a challenging, but also a controversial problem with highly divided opinions among the leading researchers. In this work, we point out some limitations in the recent developments in super-resolution MRI reconstruction and we also argue that classical super-resolution cannot be applied in the Fourier encoded plane because of the complete absence of frequency aliasing during MRI acquisition. For this reason, *true* MRI resolution enhancement can only be obtained from information outside the k-space span.

In this chapter, we also introduce an elegant way to enhance the image resolution by multiple MRI images which are acquired over a rectangular k-space span. Thanks to image rotations, we are able to enlarge the k-space span and thus improve the image resolution. We propose a novel reconstruction algorithm that consists of a proper resampling scheme in the image domain followed by optimal fusion of multiple aligned k-space data. Simulations demonstrate the superiority of the proposed method, both quantitatively and qualitatively. We also demonstrate the effectiveness of the proposed scheme to real MRI data: improvements in image quality and spatial resolution can clearly be noticed in the reconstructed images. Fourier analysis also indicates the true spatial resolution improvement. This work led to a submitted journal publication [Luong et al., 2009] and a conference publication [Luong et al., 2008].

8.1 Introduction to resolution enhancement in MRI applications

For decades, various researchers have been developing numerous and diverse methods to shorten the acquisition time of magnetic resonance imaging (MRI) data while attempting to preserve image quality. A shorter acquisition time reduces possible motion artefacts caused by voluntary/involuntary subject movements. To mention a few techniques: fast low-angle shot imaging (FLASH) [Frahm et al., 1986, Peng et al., 2006], echo-planar imaging (EPI) [Landini et al., 2005, Mansfield, 1977], parallel imaging [Blaimer et al., 2004, Katscher et al., 2003, Larkman and Nunes, 2007], altering sampling trajectories and reconstructing undersampled k-spaces [Bernstein et al., 2004, Dologlou et al., 1996, Jackson et al., 1991, Pruessmann et al., 2001, Rokitta et al., 1999, Weiger et al., 2002], etc. Complementary to these techniques, the class of *resolution enhancement* techniques also plays an important role: it is obvious that low-resolution (LR) images require fewer measurements or sampling points than high-resolution (HR) images for a given *field of view* (FOV) which eventually results in much faster acquisitions. Sometimes motion artefacts are so severe that it prohibits imaging altogether. In these cases a resolution enhancement technique could be the enabling factor for previously infeasible imaging modalities.

Many resolution enhancement techniques for single MRI images have been proposed in the literature. A comprehensive survey of existing linear image interpolation methods can be found in [Lehmann et al., 1999]. General non-linear image interpolation techniques try to avoid artefacts such as staircasing (i.e. jagged edges), blurring and/or ringing (see Chapter 3) or try to estimate the lost frequencies by training (see Section 5.1). Another class of techniques tries to recover the missing part of the k-space by extrapolating low frequency information to the high frequencies (e.g., using the Papoulis-Gerchberg algorithm or projections onto convex sets (POCS)) [Peng et al., 2006, Gerchberg, 1974, Gerchberg, 1989, Papoulis, 1975]. In the end, all these techniques incorporate prior knowledge (e.g., piecewise smoothness or the existence of sharp edges) to reconstruct the HR image in a spatially adaptive way. We note in this respect that these techniques treat high frequencies as missing information and estimate them from the available lower frequencies. Therefore, it is not certain that the features from the estimated frequencies correspond to *true* image features that would have been visible if the HR image was acquired at a higher resolution. For this reason, we say that these methods do not involve *true* resolution enhancement. This can be critical in some (clinical) applications, e.g., for the detection of small anomalies or tiny structures.

After several successful attempts of multi-frame super-resolution in camera-based applications [Irani and Peleg, 1991, Irani and Peleg, 1993,

Park et al., 2003], the ideas of incorporating multiple acquisitions to attain sub-pixel resolution in MRI appeared in the late 1990-ties (see the references in [Fiat, 2001]). The main idea is that multiple slightly different acquisitions (e.g., with the subject slightly translated) contain information that can help to reconstruct the high frequency components distorted by aliasing and blur. Therefore, techniques using multiple acquisitions should be superior to traditional reconstruction approaches that employ only a single image. This is true for camera-based super-resolution (SR) image reconstruction where the extra information is encoded as frequency aliasing. In [Roullot et al., 2000], the authors successfully combined several MRI acquisitions of anisotropically degraded resolutions (i.e. non-squared voxel sizes) to one HR volume. Unfortunately, the maximum achievable resolution is limited to the smallest voxel dimension over all acquisitions.

In [Peled and Yeshurun, 2001], a camera-based super-resolution method (the iterative back-projection method of Irani and Peleg [Irani and Peleg, 1993]) was applied to MRI data and demonstrated to yield an improved resolution of in-vivo diffusion weighted and diffusion tensor images. Shortly after this initial success, doubts were raised on the applicability of super-resolution to MRI in general due to the bandlimited nature of Fourier encoded (in the in-plane direction) data in MRI, which seemingly eliminates the possibilities of new information and super-resolution for classical Fourier encoding [Scheffler, 2002, Peled and Yeshurun, 2002]. Indeed, pixel shifts of bandlimited images result in a linear phase modulation of the Fourier coefficients, which eventually do not yield new information. For this reason, in several works, including [Greenspan et al., 2002, Kornprobst et al., 2002, Kornprobst et al., 2003, Peeters et al., 2004] the authors found that only inter-slice super-resolution is possible in MRI (because spatial frequencies in the inter-slice z -direction exhibit a less sharp cut-off which causes frequency aliasing to exist after sampling) while it is not possible to increase the resolution by means of pixel shifts within the slice. As a consequence, SR image reconstruction in the Fourier encoded plane did not appear to be an effective resolution enhancement technique.

Some of the most recent works [Carmi et al., 2006] contradicted these limitations and demonstrate in-plane resolution improvement for simulated data, however, the authors did not succeed to improve the resolution on real phantom data. Also in [Mayer and Vrscay, 2006, Mayer and Vrscay, 2007], the authors claim that new information in the frequency-encoded (FE) direction can be present in each acquisition if the subject is shifted in the FE direction prior to imaging. In this way, subject shifts cannot longer be represented as simple linear phase modulations.

To the best of our knowledge, all SR MRI papers only address translational shifts between the low-resolution images. In this chapter, we show that incorporating multiple image rotations can indeed improve the image resolution in the Fourier-encoded plane assuming that MRI data is acquired over a

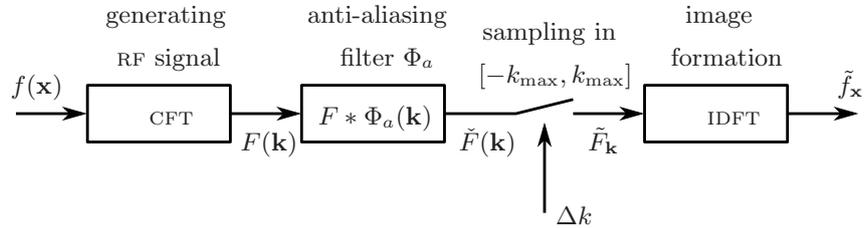


Figure 8.1: The schematic representation of the MRI image acquisition. The RF signal $F(\mathbf{k})$ is the continuous Fourier transform (CFT) of the hypothetical spatial signal $f(\mathbf{x})$. For more details concerning the physics of RF signals, we refer the reader to [Landini et al., 2005]. The ADC turns the RF signal into measurements $\tilde{F}_{\mathbf{k}}$ via proper sampling and the final MRI image $\tilde{f}_{\mathbf{x}}$ is obtained by the inverse discrete Fourier transformation.

rectangular k -space span. Our proof is both theoretical and practical.

In what follows in this chapter, we will discuss the recent developments in super-resolution MRI in Section 8.2. In Section 8.3, we introduce a new way of acquiring multiple images and describe a novel algorithm to reconstruct the HR image from these multiple LR images. Section 8.4 presents experimental results with both simulated and MRI data and quantitatively compares the proposed technique with other reconstruction methods. Section 8.5 discusses some potential applications and finally, Section 8.6 concludes this chapter.

8.2 On super-resolution magnetic resonance imaging

8.2.1 Basic concepts of k -space

In MRI acquisition, image information such as voxel intensity is not obtained directly in the spatial domain, but are acquired in the frequency space, which is often referred to as k -space. In the Fourier encoded imaging plane, the analog-to-digital converter (ADC) at the receiver side extracts values of the continuous radio-frequency (RF) signal $F(\mathbf{k})$ containing image frequency information at integer multiples of the frequency sampling period Δk . During the scan, the discrete k -space is filled with *raw* data line by line from the *readout* direction (or frequency-encoded direction). The position of each line in the vertical direction of k -space is determined by the phase-encoded magnetic gradient. The final MRI image $\tilde{f}_{\mathbf{x}}$ is related to the discrete k -space data $\tilde{F}_{\mathbf{k}}$ by the 2D inverse discrete Fourier transformation (IDFT). A schematic representation of the MRI acquisition is given in Figure 8.1.

By the Shannon sampling theorem [Shannon, 1949, Unser, 2000], the spatial resolution Δx of the final image is determined by the highest measured fre-

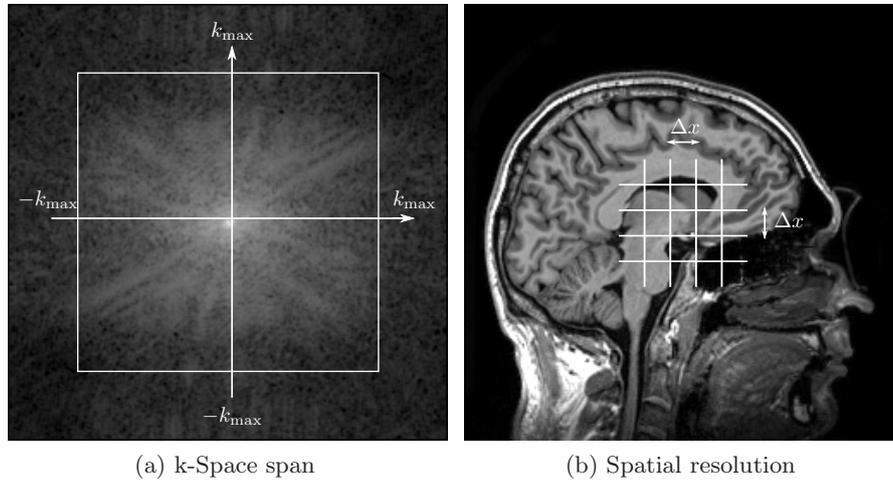


Figure 8.2: The span of k-space $[-k_{\max}, k_{\max}]$ (a) is inversely proportional to the spatial image resolution Δx (b).

quency k_{\max} ¹ and is given by the following expression:

$$\Delta x = \frac{1}{2k_{\max}}, \quad (8.1)$$

where we assume that the k-space span is symmetrical with respect to the origin, i.e., $k \in [-k_{\max}, k_{\max}]$. Following the duality principle of the Fourier transform, the spatial field of view (FOV) $[-x_{\max}, x_{\max}]$ is determined by the sampling rate $1/\Delta k$ in the frequency space. These relationships are illustrated for the 2D isotropic case in Figures 8.2 and 8.3.

In order to minimize the acquisition time, the sampling period Δk (in the phase encoding direction) is preferably chosen as large as possible for a fixed k-space span or fixed spatial resolution. This means that the FOV is to be chosen as small as possible. However, if the FOV does not contain the whole object, the discrete sampling in k-space results in aliasing (or *wrap-around* artefacts) as a consequence of violation of the Nyquist criteria. We will denote this form of aliasing as spatial aliasing (i.e. spatial information overlaps). This manifests in an image artefact where parts of the object outside the FOV wrap into the image. This spatial phenomenon is illustrated in Figure 8.4(b). Its dual counterpart (e.g. in camera-based applications with acquisitions in the spatial

¹From an analytic point of view, any support-limited object can be reconstructed perfectly from any interval of (continuous) k-space data using analytic continuation. This is possible because the Fourier transform of the object can be expressed as an analytic function over the whole k-space. In that sense, it should be possible to improve the resolution beyond the limitations of the k-space span. However, in practice, this would likely be very poorly conditioned.

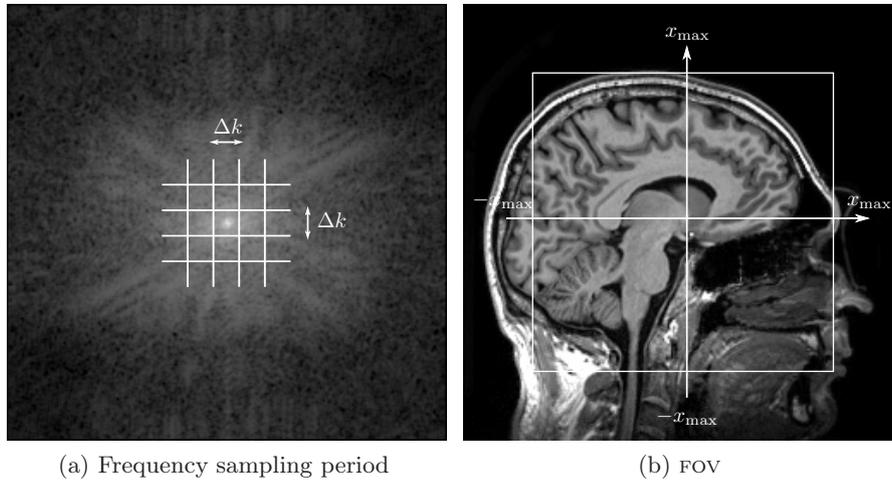


Figure 8.3: The sampling period Δk in k-space (a) is inversely proportional to the field of view $[-x_{\max}, x_{\max}]$ (b).

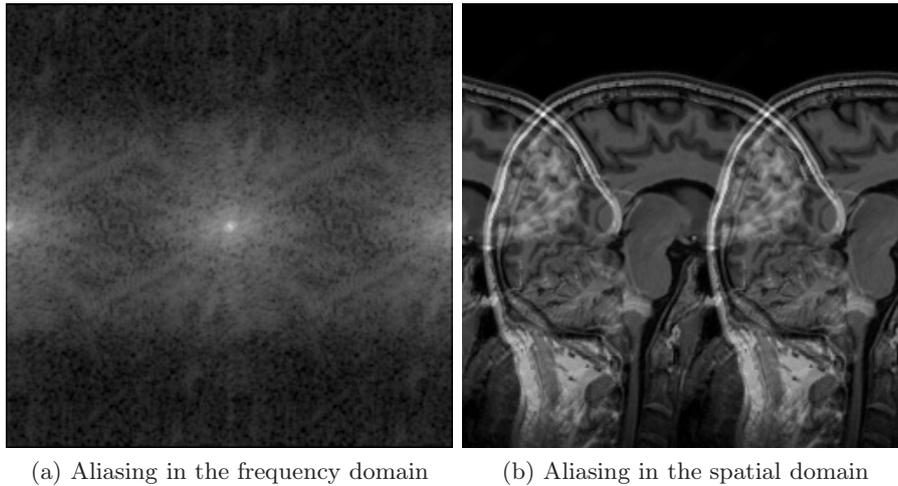


Figure 8.4: Aliasing in horizontal direction: (a) frequency aliasing (possible in camera-based applications) and (b) spatial aliasing (possible during MRI acquisition).

domain) is illustrated in Figure 8.4(a): the image suffers from frequency aliasing (i.e. replicated spectra overlap). To suppress aliasing in general, the sampling rate must be high enough to fulfil the Nyquist rate or a sufficient anti-aliasing low-pass filter Φ_a must be applied on the continuous k-space data $F(\mathbf{k})$ prior to sampling. The latter is equivalent to a multiplication with a windowing function in the spatial domain.

8.2.2 Super-resolution in MRI?

In several applications including satellite imaging, video applications and medical imaging, super-resolution (SR) image reconstruction has been proven useful where multiple images of the same scene or object can be obtained. Camera-based multi-frame super-resolution is a signal processing technique that combines a sequence of low-resolution (LR) noisy blurred images to produce a higher resolution image or sequence. In conventional imaging systems (as in camera-based applications), there is a natural loss of spatial resolution caused by optical distortions (camera lens blur, out of focus, diffraction limit, atmospheric blur, etc.), motion blur due to limited shutter speed, noise that occurs within the sensor or during transmission and insufficient sensor density [Park et al., 2003]. Thus, the acquired image usually suffers from blur, noise and (frequency) aliasing effects.

SR image reconstruction is possible if two conditions are fulfilled: the LR images must be subpixel shifted and the images must contain frequency aliasing (due to sampling) [Park et al., 2003]. Indeed, if there are no subpixel shifts (i.e. all pixel shifts are integer), each image contains exactly the same information in the absence of noise, except for a few extra rows and columns of pixel data near the image boundaries and therefore, no new information is available to reconstruct the missing high frequencies. If there is no frequency aliasing, then the observed object only contains bandlimited information and recovering high frequencies is not possible.

Unfortunately, in the process of recording an MRI image, there is no frequency aliasing possible because no frequency information is retrieved outside the span of k-space $[-k_{\max}, k_{\max}]$. However, in [Carmi et al., 2006], Carmi et al. achieved a resolution improvement for artificially generated images of different FOVs using a camera-based super-resolution scheme. They employed a general imaging model to solve the SR problem: each shifted LR image is degraded by a blur operator which is referred to as the point spread function (PSF), additive white noise and decimation (at different resolutions). These linear degradation operators are illustrated in Figure 4.2. In their paper, they assumed a box-type blur or Gaussian blur as PSF. For that reason, their model introduces frequency aliasing in the LR images, which is not possible during MRI acquisition because sampling occurs in k-space. That is why their results are successful on simulated data, but fail on real phantom data.

To demonstrate the influence of the PSFs on frequency aliasing, we compare the Fourier transforms of the box-type kernel and the Gaussian kernel with the one of the ideal low-pass filter in Figure 8.5. The ideal transfer function is constant and equal to 1 in the passband (denoted here by the interval $-\pi < \omega < \pi$ and the cutoff point or Nyquist frequency $\omega = \pi$) and zero in the stopband ($|\omega| > \pi$). The occurrence and the amplitudes of ripples and sidelobes in the stopband cause frequency aliasing effects: attenuated copies of the side lobes appear in the stopband. In Figure 8.5, we can clearly see that these high cutoff-

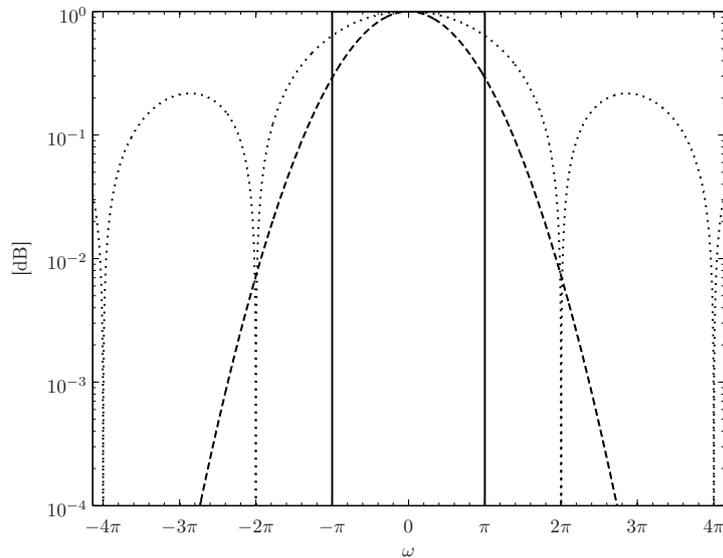


Figure 8.5: The logarithmic plot of the magnitude of the Fourier transforms of the box-type PSF (dotted), the Gaussian kernel PSF (dashed) and the ideal low-pass filter (solid) (in dB).

points ($> \pi$) provide us the source of frequency aliasing.

In their SR image reconstruction, Carmi et al. used images which are acquired at different spatial resolutions (or different FOVs given a fixed image size) [Carmi et al., 2006]. However, in case of the ideal transfer function, this will not yield new information for the estimation of the high frequencies beyond the maximal k_{\max} : all images acquired with a lower resolution compared to the image with the highest resolution cover a smaller span in k-space ($k_{\max}^* < k_{\max}$). It is clear that MRI images with lower resolutions can *not* contribute to the true resolution recovering process. On the other hand, employing multiple acquisitions can help to increase the overall image quality (i.e. peak-signal-to-noise ratio (PSNR)).

8.2.3 Does a k-space span limitation affect MRI resolution?

In [Stoch and Balcom, 2006], Stoch et al. investigated whether the span limitation $[-k_{\max}, k_{\max}]$ of the (discrete) k-space has an influence on the spatial MRI resolution and their conclusion is quite surprisingly:

“Does a k-space span limitation influence the experimental image or

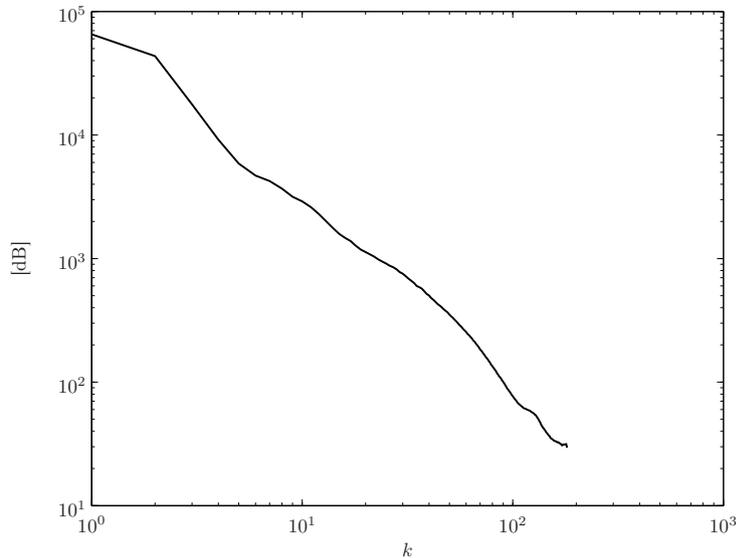


Figure 8.6: The radial profile of the 2D power spectrum of MPRAGE data (averaged over 176 different slices which are acquired at 1.0 mm resolution using a head coil within a 256×256 matrix).

not? The answer based on a continuous data approach is yes. The answer for discrete data is no.” [Stoch and Balcom, 2006]

Additionally, Stoch et al. ascribed the observed image distortion mainly to spatial aliasing (this is achieved by changing the FOV as described earlier in Section 8.2.1). To simplify the explanation we will omit such details in the rest of this chapter and we assume that the FOV contains the whole object and that the images do not suffer from spatial aliasing.

However, in their paper, Stoch et al. only based their conclusion on bandlimited functions in *discrete* k-space for which the frequency components beyond k_{\max} are zero (or almost zero). As long as the discrete Fourier coefficients beyond k_{\max} are zero, changing the k-space span does not affect the MRI resolution. But in MRI acquisition, the missing high frequencies in k-space are generally non-zero (otherwise perfect reconstruction is possible apart from noise). This claim is illustrated with a radial profile of real k-space data in Figure 8.6: from the decay ($\approx 1/k^{1.7}$) of the logarithmic magnitude, we can assume that non-zero coefficients must exist outside the k-space span. The same diminution can also be observed in natural images, in general, the power spectrum follows an $1/k^\alpha$ -decay ($\alpha \in [1.5, 3.0]$) [Field, 1987].

From a theoretical point of view, it is also impossible that images are bandlimited in the spatial domain (via FOV) as in the frequency domain simultaneously, see e.g. [Daubechies, 1992]. In other words, if the image is bandlimited in frequency space, its spatial support will be infinitely long and vice versa. If non-zero coefficients beyond a certain k_{\max} are discarded right before the IDFT, this will lead inherently to image blur. So, a k-space span limitation actually *does* affect MRI resolution in general and thus the answer on the question is definitely *yes*.

8.2.4 Subject shifts before acquisition

In [Carmi et al., 2006], Carmi et al. claimed that applying subpixel shift prior to filtering/sampling is different than a postprocessing mathematical manipulation. Therefore, high-resolution information can be introduced that makes SR image reconstruction feasible. This claim is later further investigated and validated by Mayer et al. [Mayer and Vrscaj, 2006, Mayer and Vrscaj, 2007]. The amount of HR information achieved at each shift is relatively small which may limit its practical use [Mayer and Vrscaj, 2007]. Because of the bandlimited nature of Fourier encoded data, we believe this new information does not contain any knowledge about the missing high frequencies.

Let us take a closer look at the mathematical acquisition model (illustrated in Figure 8.7) for FE signals used by Mayer et al. [Mayer and Vrscaj, 2006, Mayer and Vrscaj, 2007]: prior to imaging, the object is shifted on a subpixel scale. Since k-space data is measured over a finite frequency range, which can be considered as a *boxcar*-type bandlimited filter, the image is distorted by a convolution with a sinc kernel according to the Fourier convolution theorem. Before discretization by the ADC, the k-space signal is convoluted with an anti-alias filter, which attenuates intensities outside the FOV in the spatial domain. Finally, the *new information* is simply the difference between the two filtered spatial signals after proper registration.

We agree that applying subpixel shifts prior to imaging is different than applying a linear phase modulation as a postprocessing tool. However, it does not add any new information for the recovery of high frequencies [Mayer and Vrscaj, 2007], as no information outside the k-space span is acquired in the observation model. After bandlimited filtering, the spatial signals are exactly the same except for a shift. At this point, the k-space data of the shifted signal can still be derived from the k-space of the original signal, the only difference being a known phase shift. After reconstruction, the same image appear in both cases, but one of the images is shifted and because of the anti-aliasing in k-space, the images differ near the FOV boundaries (see Figure 8.7). The new information can be useful to avoid ringing near the image edges, but not for resolution enhancement. Note that the bandlimited filter in Figure 8.7 does not really exist in a standard implementation of an MRI scanner. It is only during the sampling process that the frequencies

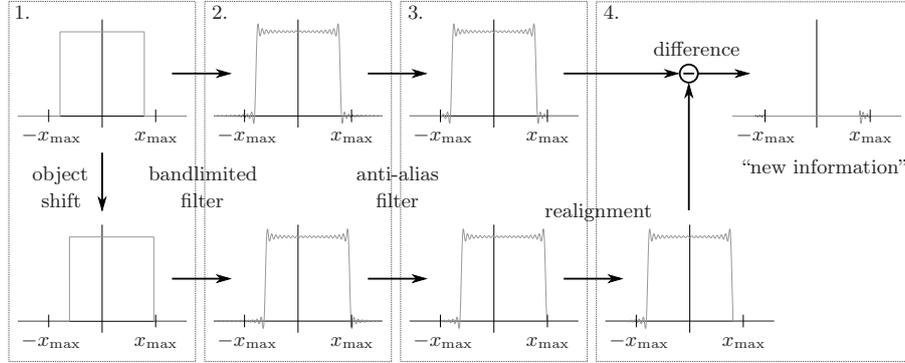


Figure 8.7: Mayer et al. [Mayer and Vrscaj, 2007]: schematic observation model of the MRI acquisition process in the spatial domain.

outside the k -space span are discarded (see Figure 8.1). Unfortunately, this will also not solve the resolution problem.

In the previous paragraphs we have pointed out some shortcomings in recent SR MRI developments and we have argued why SR image reconstruction cannot successfully make use of in-plane object translations. To reconstruct the true high image resolution, information beyond k_{\max} or prior knowledge is needed anyway. Despite all these negative indications, there is no need for pessimism. In the next section, we show that multiple acquisitions with different k -space and/or object rotations can contribute to the resolution enhancement of the MRI data.

8.3 Multi-frame MRI reconstruction in the Fourier encoded plane

8.3.1 Acquisition of low-resolution images

In order to describe how combining multiple MRI scans can improve the in-plane image resolution, we relate object motions to the acquired k -space data using Cartesian sampling schemes. We assume that the k -space signal prior to filtering/sampling is a multi-dimensional continuous signal $F(\mathbf{k})$. The scanner samples this signal on discrete positions within a multi-dimensional k -space span determined by k_{\max} .

We denote the n -dimensional k -space coordinates as $\mathbf{k} \in \mathbb{R}^n$ and the spatial coordinates as $\mathbf{x} \in \mathbb{R}^n$. The n -dimensional continuous Fourier transform $F(\mathbf{k})$ of the signal $f(\mathbf{x})$ is given by:

$$F(\mathbf{k}) = \int_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) e^{-2\pi j(\mathbf{k} \cdot \mathbf{x})} d\mathbf{x}, \tag{8.2}$$

where $\mathbf{k} \cdot \mathbf{x}$ is defined as the inner product between the two vectors. Analogously, the multi-dimensional inverse continuous Fourier transform is then denoted as:

$$f(\mathbf{x}) = \int_{\mathbf{k} \in \mathbb{R}^n} F(\mathbf{k}) e^{2\pi j(\mathbf{k} \cdot \mathbf{x})} d\mathbf{k}. \quad (8.3)$$

We suppose that the object undergoes an affine transformation. The new spatial coordinates \mathbf{y} are then related to the reference coordinates \mathbf{x} by the following linear transformation:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}, \quad (8.4)$$

where \mathbf{A} is an invertible $n \times n$ matrix and the image $f(\mathbf{x})$ is transformed into $g(\mathbf{x}) = f(\mathbf{y})$. We are interested in the influence of such a transformation on the k-space signal. Via the *change of variables* (with the introduction of the Jacobian), we can easily derive the Fourier transform $G(\mathbf{k})$ of the transformed image:

$$\begin{aligned} G(\mathbf{k}) &= \int_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{x}) e^{-2\pi j(\mathbf{k} \cdot \mathbf{x})} d\mathbf{x} \\ &= \int_{\mathbf{y} \in \mathbb{R}^n} f(\mathbf{y}) e^{-2\pi j(\mathbf{k} \cdot \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}))} |\mathbf{A}|^{-1} d\mathbf{y} \\ &= |\mathbf{A}|^{-1} e^{2\pi j(\mathbf{k} \cdot \mathbf{A}^{-1}\mathbf{b})} \int_{\mathbf{y} \in \mathbb{R}^n} f(\mathbf{y}) e^{-2\pi j(\mathbf{k} \cdot \mathbf{A}^{-1}\mathbf{y})} d\mathbf{y} \\ &= |\mathbf{A}|^{-1} e^{2\pi j(\mathbf{k} \cdot \mathbf{A}^{-1}\mathbf{b})} F((\mathbf{A}^{-1})^T \mathbf{k}), \end{aligned} \quad (8.5)$$

as a function of the Fourier data $F(\mathbf{k})$ of the same object at the reference position.

For simplicity, we will only discuss the case of isotropic voxels and orthogonal transformations, extension to the anisotropic and non-orthogonal cases is trivial. Because we are mainly interested in the movement of rigid objects, we briefly discuss the basic properties based on these three types of rigid body transformations (i.e. translation, rotation and scaling):

- translational shifts cause a linear phase modulation of the k-space data, i.e. multiplying Fourier coefficients by $e^{2\pi i(\mathbf{k} \cdot \mathbf{b})}$. The k-space span is invariant to translation, so, no new information for resolution enhancement is achieved by additional scans.
- a rotation by an angle θ causes a rotation in k-space by the same θ . The new k-space coordinates are related by $(\mathbf{A}^{-1})^T \mathbf{k}$ to the reference coordinates. In 2D we have:

$$\mathbf{A} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = (\mathbf{A}^{-1})^T. \quad (8.6)$$

Note that the determinant of the Jacobian is 1. However, the k-space spans of the transformed images now cover different parts of the k-space

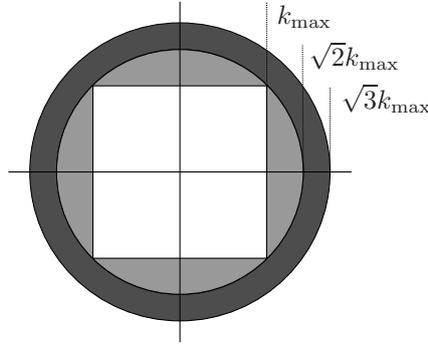


Figure 8.8: Visualization of the maximum k-space coverage by the reference low-resolution image and the high-resolution images obtained by 2D slice rotations and 3D volume rotations. The radii of the new k-space spans are given by $\sqrt{n}k_{\max}$.

in case the signal is sampled on a rectangular grid. The maximum coverages for slice and volume rotations are illustrated in Figure 8.8. The theoretical maximum resolution improvement in both horizontal and vertical direction is $\sqrt{n} - 1$ or about 41% (2D slice rotations) and 73% (3D volume rotations) with rotations in the full range of $[-45^\circ, 45^\circ]$.

- in case of scaling (this corresponds with a change in FOV), the new k-space coordinates change inversely proportional to the scale factor s and the transformed Fourier signal is given by:

$$G(\mathbf{k}) = s^{-n}F(s^{-1}\mathbf{k}). \quad (8.7)$$

Additionally, the Fourier coefficients are scaled with s^{-n} . Note that the k-space span also proportionally changes with the scale factor in each dimension, but as discussed earlier in Section 8.2.1 and Section 8.2.2, the object must fit entirely into the FOV to avoid spatial aliasing and it is impossible that images with lower resolutions can contain extra high-resolution information.

Based on these observations, we can opt to acquire multiple slices or volumes of the same object with different rotations. Unfortunately, the proposed strategy can not be useful on all types of sequences: non-Cartesian circular sampling (e.g., spiral or radial trajectories) does not change the k-space spans by means of rotation.

8.3.2 The proposed reconstruction algorithm

There is an important difference between multi-frame SR image reconstruction schemes and the proposed multi-frame MRI reconstruction technique, namely the source of new information: in the proposed framework, high frequency

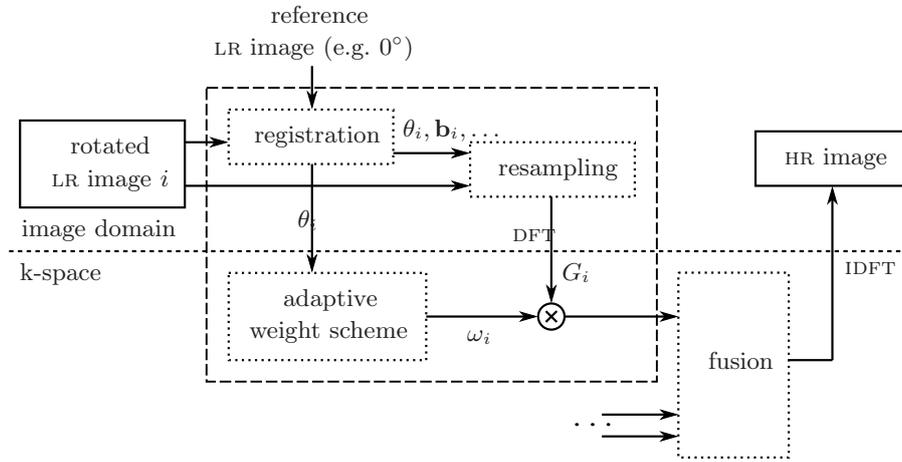


Figure 8.9: Outline of the proposed hybrid reconstruction algorithm.

information originates from different k-space coverages and not from frequency aliasing as in camera-based SR applications. Based on the k-space coverages, we propose a new reconstruction algorithm that is focused on the reconstruction of *true* k-space data (i.e. we do *not* create artificial frequencies).

The complete block scheme of our reconstruction algorithm is given in Figure 8.9. In case the transformation parameters (such as rotation θ_i , translational shifts \mathbf{b}_i , etc.) are unknown, they have to be estimated by a subpixel registration algorithm, as discussed in Section 7.2.3. The proposed algorithm consists mainly of two steps, namely proper resampling of k-space data and weighted fusion of the Fourier coefficients. We will discuss these components in more detail.

8.3.2.1 Resampling of k-space data

In order to combine the information from several rotated (and shifted) LR images, the images must be aligned to a reference image with subpixel accuracy. In Section 8.3.1, we have shown that a rotation of the object corresponds to the same rotation in k-space. A common approach to resample a regular Cartesian grid is to convolute the discrete data samples with the continuous impulse response of a reconstruction filter [Lehmann et al., 1999, Seppä, 2007]. Because of practical issues, finite-support interpolation kernels are employed in resampling tasks in contrast with the unlimited support of the ideal interpolation kernel (i.e. sinc function) (see Chapter 2). As a consequence, an inherent property of all traditional resampling schemes is that they smooth the data more or less. Because blurring attenuates the high frequencies and since most energy of the image is encoded in the lower frequencies, it is more effective to perform resampling in the spatial domain rather than employing regridding techniques in Fourier space [Jackson et al., 1991, O’Sullivan, 1985].

Additionally, resampling in the image domain handles object shifts automatically and can cope with higher order object transformations. After the discrete Fourier transformation, we have a set of aligned k-space data on the HR grid.

In the algorithmic implementation of (rotational) resampling from the LR grid to the HR grid, the estimated values are often computed in one step by the convolution with the reconstruction filter. However, frequency aliasing can be introduced because non-ideal interpolation kernels are employed. This can be avoided using a two-stage alias-free resampling scheme [Seppä, 2007]: in a first step, ideal sinc interpolation (i.e. zero-padding of Fourier coefficients) is applied to attenuate the influence of the replicated spectral patterns and in a second step, standard resampling is applied. In this way, we also avoid border problems (i.e. the loss of high frequency components) in k-space, which occur when image rotations are carried out on the same LR grid (see Section 2.2.4 for a more detailed discussion).

8.3.2.2 Weighted fusion of aligned k-space data

Suppose that we have a set of m scans with different rotation angles θ_i ($i = 1, \dots, m$). For simplicity, we will only discuss the 2D slice rotations. Extension to 3D rotations is fairly straightforward. So, for each position \mathbf{k} on the k-space grid, we have to estimate the Fourier coefficient $F(\mathbf{k})$ from m aligned coefficients $G_i(\mathbf{k})$. We formulate this as a standard weighted least-squares problem:

$$\hat{F}(\mathbf{k}) = \arg \min_{F(\mathbf{k})} \sum_{i=1}^m w_i(\mathbf{k}) (G_i(\mathbf{k}) - F(\mathbf{k}))^2. \quad (8.8)$$

The solution of this problem becomes simply a weighted average over the m samples:

$$\hat{F}(\mathbf{k}) = \frac{\sum_{i=1}^m w_i(\mathbf{k}) G_i(\mathbf{k})}{\sum_{i=1}^m w_i(\mathbf{k})}. \quad (8.9)$$

We choose the adaptive weights so to minimize the variance of the estimator in equation (8.9) and thus maximize the SNR in the presence of independently and identically distributed Gaussian noise. This yields [Van Trees, 1968]:

$$w_i(\mathbf{k}) = \begin{cases} 1 & \text{if } \mathbf{k} \text{ is in the support of } G_i, \\ 0 & \text{else.} \end{cases} \quad (8.10)$$

The final image is then obtained by the inverse discrete Fourier transform.

We have developed two versions of the proposed reconstruction algorithm: one with standard (one-stage) resampling (1SR) (i.e. jointly performing interpolation and rotation) and one with two-stage alias-free resampling (2SR).

The rotations in both reconstruction techniques are carried out by the 8-point Blackman-Harris windowed sinc kernel [Lehmann et al., 1999].

We note that there are some similarities with parallel imaging (PE) in MRI [Larkman and Nunes, 2007] and camera-based image mosaicing [Traka and Tziritas, 2003]: in PE, k-space data originating from different receiver channels have to be combined and each are weighted according to their spatial sensitivity. In image mosaicing, the FOV of the mosaic is extended by stitching several images with overlapping FOV's, a similar concept is now applied in k-space.

8.4 Results and evaluation

8.4.1 Simulated experiment

As a noise-free experiment, we start from a 512×512 high-resolution image phantom (see Figure 8.10). The k-space of the HR image has a circular footprint as indicated in Figure 8.8, so it is also the theoretical bound for our reconstruction framework. We generate 128×128 images by rotating the HR image successively with an offset $\Delta\theta = 1^\circ$ in the range $[0^\circ, 90^\circ]$. The rotation is performed by the 8-point Blackman-Harris windowed sinc kernel [Lehmann et al., 1999]. To avoid frequency aliasing which is necessary to model an MRI scanner, down-sampling is carried out as bandwidth truncation in the Fourier domain, i.e. only discrete samples in the k-space span are taken into account. As a result, we obtain 90 rotated noise-free LR images.

In our benchmark, we will include for instance zero-padding interpolation (or ideal sinc interpolation) of only the reference LR image and simple averaging over all aligned LR images. The averaging here is implemented as the combination of two-stage resampling (i.e. ideal sinc interpolation followed by resampling) and equiweighted fusion in the image domain. We will also demonstrate its effectiveness over the naive approach (denoted as naive averaging), i.e. rotation and fusion on the LR grid followed by ideal sinc interpolation as discussed in Section 8.3.2.1. We will also compare our reconstruction techniques with the classic kernel regression ($N = 2$) [Takeda et al., 2007] (see also Section 7.4.2). Kernel regression is a general resampling tool in the spatial domain that is often used in camera-based multi-frame super-resolution schemes.

Crop outs of the reconstruction results are shown in Figure 8.11. We can clearly see that our methods are visually superior to other reconstruction methods: they are very similar to the HR image, certainly when looking at the finest horizontal transitions at the left. The proposed fusion using two-stage resampling (2SR) also produces sharper results compared to the 1SR version. From both ideal sinc interpolation and naive averaging, we can not distinguish the

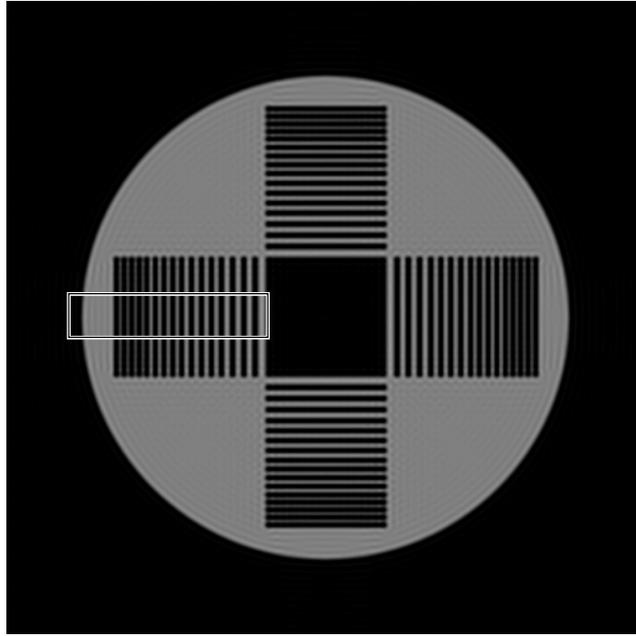


Figure 8.10: Original computer image phantom.

finest vertical lines, while other techniques do improve the true image resolution. This is once again explicitly demonstrated in Figure 8.12, where the images produced by averaging and kernel regression are enhanced using linear restoration filters for better visualization.

In Figure 8.13, we can clearly see the differences between several reconstruction methods: the plots of the proposed methods and especially of the 2SR version come very close to the one of the original HR image. Furthermore, averaging and kernel regression techniques produce profiles of very poor contrast. From the sinc interpolation we can not distinguish the individual peaks which underlines again the loss of spatial resolution.

If we take a closer look at the results produced by kernel regression in Figure 8.11, we can notice some zigzag artefacts: the vertical lines are not totally straight but contain some jaggies. This artefact arises from the fact that the kernel regression adapts itself to the pattern of the irregularly placed positions of the pixels or voxels. The zigzag artefact is then much more pronounced when the sampled data is not uniformly distributed, which is the case for plain rotations. The locations of the samples are plotted in Figure 8.14, which results in rosette-like structures.

We also compare the image quality of the reconstruction methods with the original HR image. The quantitative measures expressed in PSNR (dB) are given in Table 8.1. The PSNR values follow our subjective experiences and we can

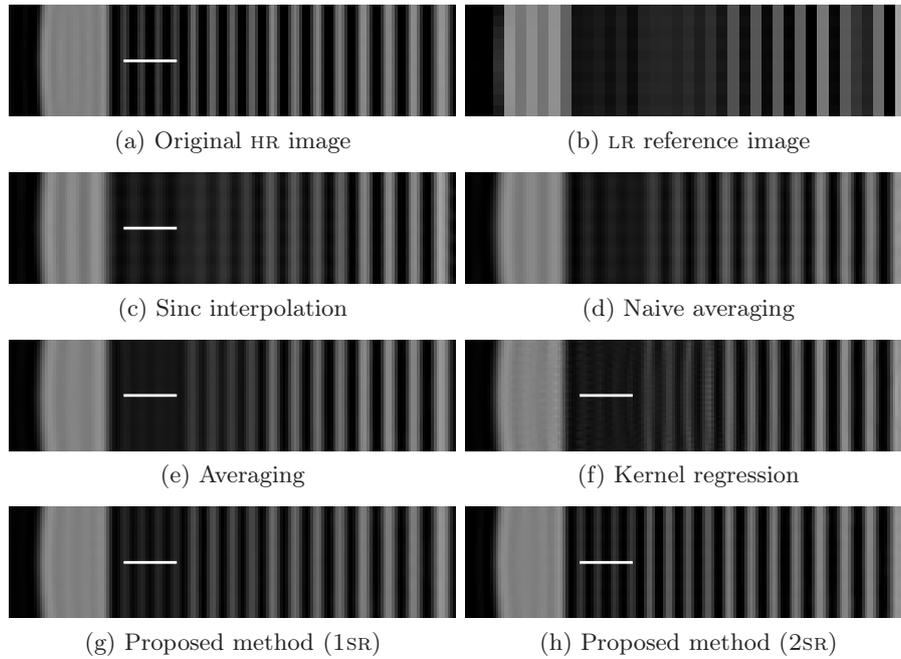


Figure 8.11: Detailed view of the reconstruction results on the region of interest (ROI) as indicated in Figure 8.10. The horizontal white lines represent the selected 1D profiles as illustrated in Figure 8.13.

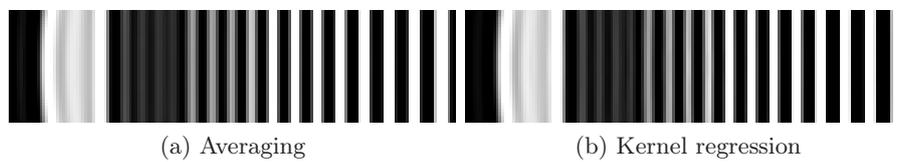


Figure 8.12: Enhanced visualization of the reconstructed results from Figure 8.11. The images are obtained after applying linear image enhancement techniques (a combination of vertical blurring and standard (isotropic) unsharp masking).

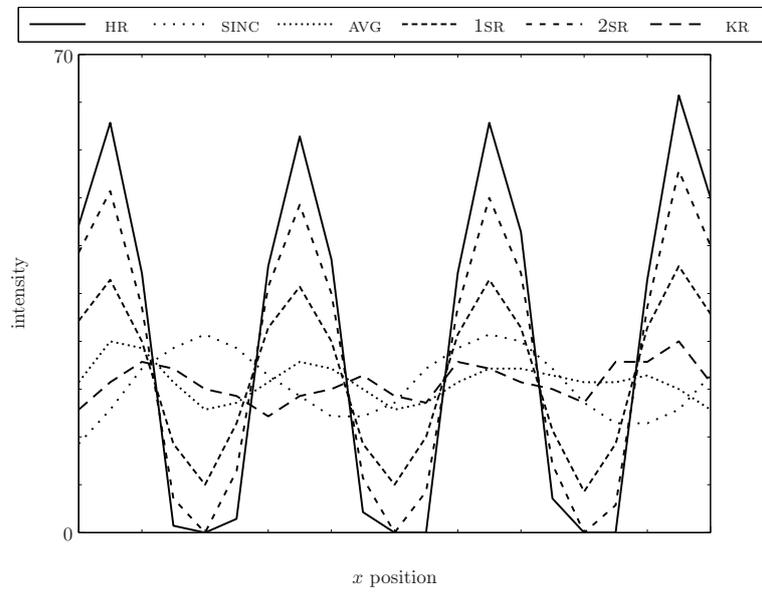


Figure 8.13: Comparison of several 1D profiles (as indicated in Figure 8.11): original HR image (HR), sinc interpolation (SINC), averaging (AVG), kernel regression (KR), proposed methods (1SR) and (2SR).

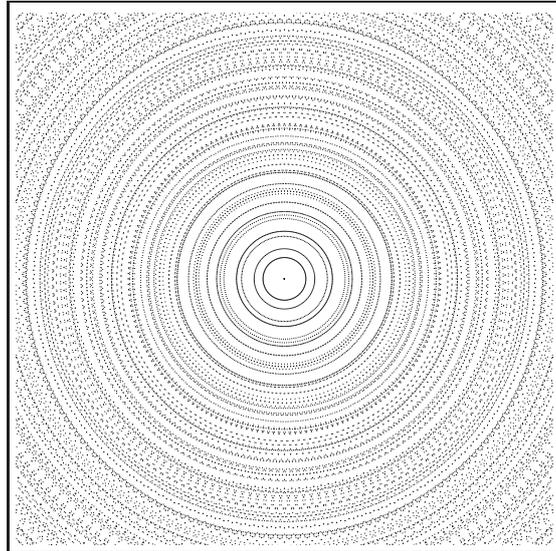


Figure 8.14: Sample positions plotted for several rotated regular grids.

Table 8.1: Image quality expressed in PSNR of several reconstruction methods. The best PSNR result is written in bold letters.

Method	PSNR (dB)
Sinc interpolation	25.90
Naive averaging	25.18
Averaging	29.98
Kernel regression	29.52
Proposed method (1SR)	31.79
Proposed method (2SR)	40.26

see that our method (2SR) outperforms all other techniques in image quality. We can also notice the loss in image quality of the naive averaging compared to the ideal sinc interpolation as the result of the loss of high frequencies.

8.4.2 MRI of a hardware resolution phantom

We perform a similar experiment with actual MRI data from a resolution phantom. All MRI data are acquired using a 3T SIEMENS Magnetom Tim Trio MRI scanner. Note that the elliptical filter² must be turned off during acquisition, otherwise substantial k-space data for resolution enhancement is discarded.

The resolution phantom (built from perspex) contains 11 rows of each 5 tubes with increasing sizes filled in a fluid with a concentration of 1.25g NiSO₄ · 6 H₂O per 1000g distilled H₂O (illustrated in Figure 8.15). The slices are obtained using a birdcage head coil and a FLASH sequence within a 384 × 384 mm FOV. Slice thickness was set at 5 mm, the repetition time (TR) and the echo time (TE) were set at 100 ms and 4.8 ms respectively. The HR scan has a 1.0 mm resolution (384 × 384 matrix), while the 181 LR rotated images are acquired at a resolution of 1.5 mm (resulting in 256 × 256 matrices) with an offset $\Delta\theta = 1^\circ$ in a range of $[-90^\circ, 90^\circ]$.

The reconstructed images are given in Figure 8.16. As a comparison with ground truth data, the HR image is enlarged with sinc interpolation for better visualization. In the HR MRI data, we can clearly distinguish 4 (and almost 5) smallest tubes on the top row. In the zero-padded LR MRI data on the other hand, we can not identify any individual tube on the top row, also, the distorted shapes reveal the loss in spatial resolution and image quality. In both kernel regression and naive averaging, none of the smallest tubes are distinguishable, but less artefacts are noticeable due to an increase of the PSNR.

²The elliptical filter is a postprocessing tool built into the (SIEMENS) MRI scanner that attenuates k-space data outside an elliptical or circular support.

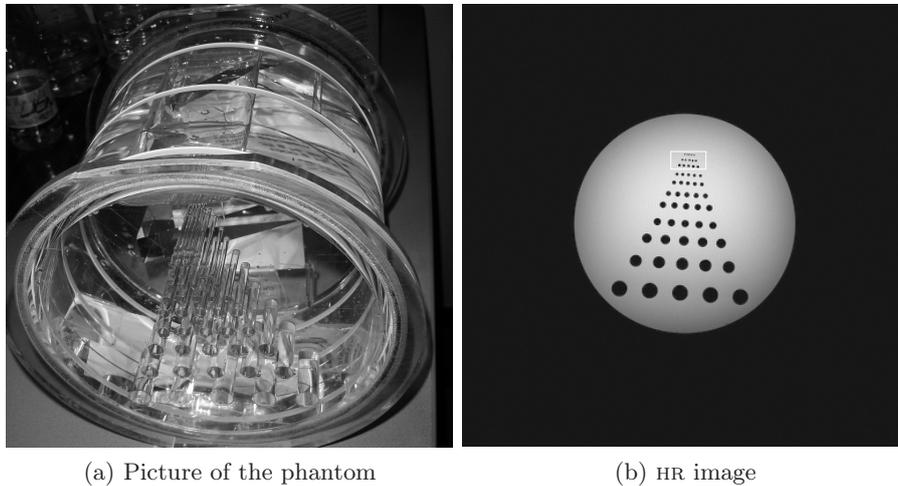


Figure 8.15: View of the resolution phantom and the 1×1 mm high-resolution acquisition.

With simple averaging, some structures in the top row can be noticed, but it is still very hard to recognize the individual tubes. Finally, in both versions of the proposed algorithm, we can actually distinguish 4 separated tubes which indicates resolution improvement on a real image. The 2SR version is visually much sharper and contains more ringing effects compared to the 1SR version, but still contains less ringing artefacts when compared to the zero-padded reference HR acquisition or sinc interpolation.

By means of the power spectra, we describe some properties of the different reconstruction methods in Fourier space. The plots of the squared magnitude of the Fourier transforms are shown in Figure 8.17 at a logarithmic scale due to the large dynamic range of power spectrum values. We can clearly see why naive averaging should be avoided: there is a notable suppression of frequencies compared to the reference LR image data, and no high frequencies are added unlike the correct averaging method. We also notice that the frequency magnitudes are fading out for higher k values in the averaging method. Kernel regression algorithms produce images with a lot of artificial high frequency components which do not necessarily correspond to the true high frequency data. The proposed methods reconstruct the k -space data very well in both cases, but in the 1SR version we can observe some frequency aliasing effects at the boundaries due to improper resampling as discussed before in Section 8.3.2.1. So, our methods outperform the other reconstruction techniques in both spatial and Fourier domains.

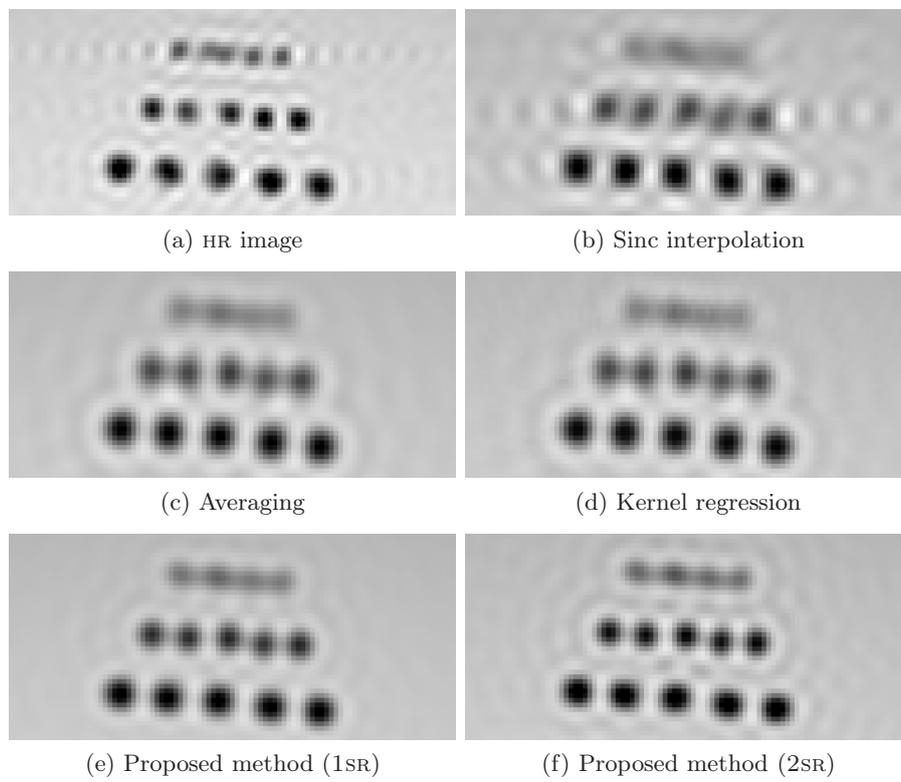


Figure 8.16: Detailed view of the reconstruction results on the region of interest (ROI) as indicated in Figure 8.15.

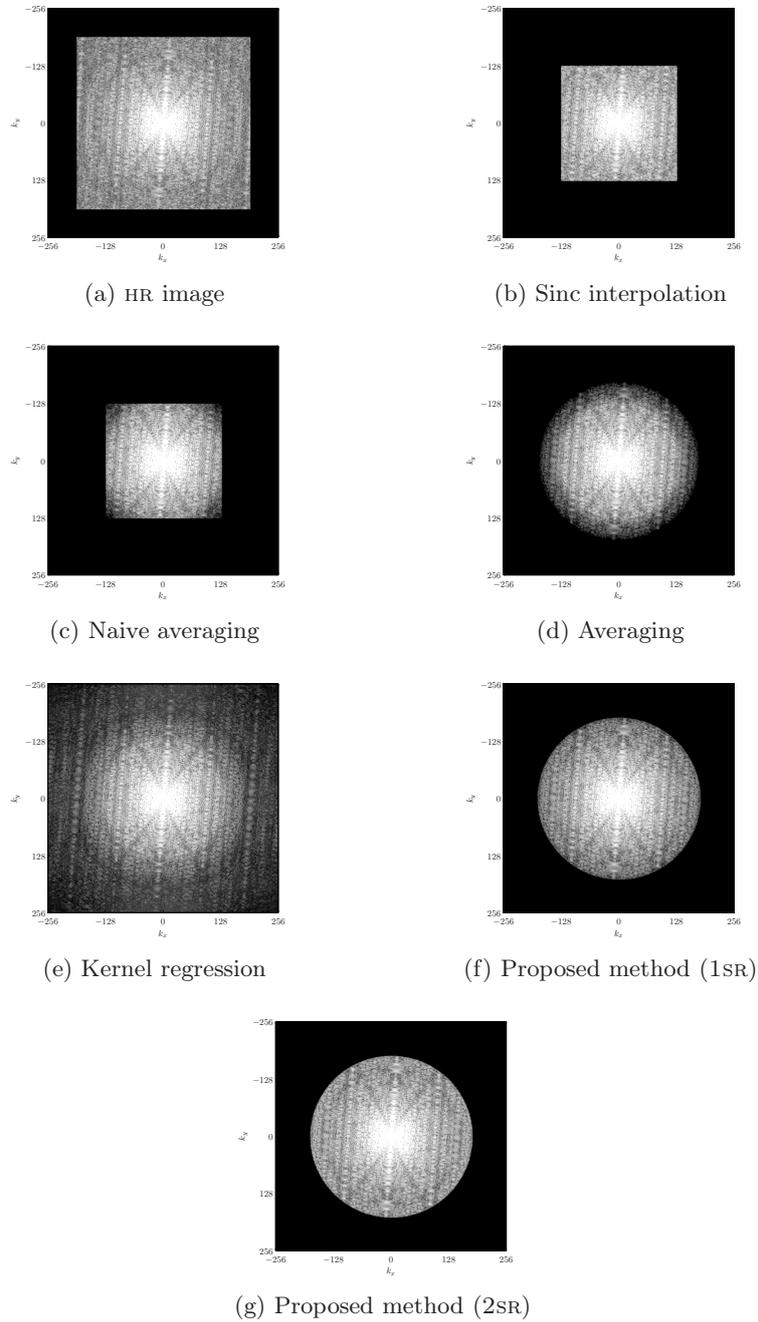


Figure 8.17: The power spectra produced by several methods (i.e. logarithmic plot of the squared magnitude of the Fourier transforms).

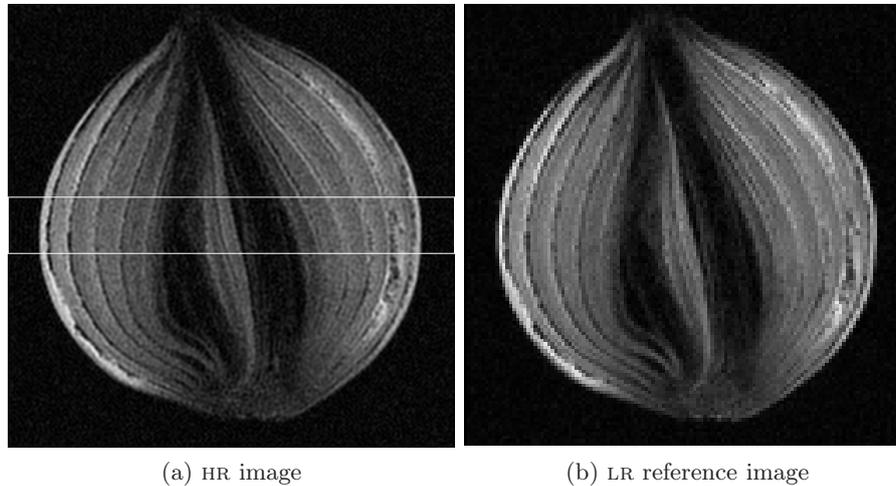


Figure 8.18: The HR and LR MRI scan of an onion with 0.44×0.44 mm and 0.66×0.66 mm resolution respectively.

8.4.3 MRI of an onion

An onion provides an interesting and challenging test case because of the existence of fine structures at various scales. The images are acquired using a knee coil, which has 8 receiver channels on board for parallel imaging. During the acquisitions, we turn the normalization filter on in order to compensate for the spatial sensitivities of the different receiver channels. A turbo spin-echo (TSE) T_2 -weighted sequence is employed within a 84×84 mm FOV. Slice thickness was set at 3 mm, TR was set at 800 ms and TE was set at 115 ms. The 91 LR rotated images are acquired at a resolution of 0.66 mm (resulting in 128×128 matrices) with 2 averages per scan and an in-between offset of 1 degree ($\Delta\theta = 1^\circ$) in a range of $[0^\circ, 90^\circ]$. The reference HR image is obtained with a resolution of 0.44 mm (192×192 matrix). Unfortunately, the adjusted sequence still produces a lot of noise as illustrated in Figure 8.18.

In Figure 8.19, we compare the results of several reconstruction methods on the ROI as indicated in Figure 8.18. When we take multiple low-resolution observations into account, we can clearly observe that the noise level is heavily reduced. In the first place, the improvement in image resolution is also somehow associated to the observed noise or signal-to-noise ratio: the ability to distinguish tiny structures or lines increases when noise decreases. Additionally, we can observe tinier structures in the results of our reconstruction methods compared to other resolution enhancement techniques, so extra resolution information is gained by extending the k-space span. Again, we can clearly notice some zigzag artefacts close to the rotation center in the kernel regression result. Fourier analysis of the reconstructed data reveals the same

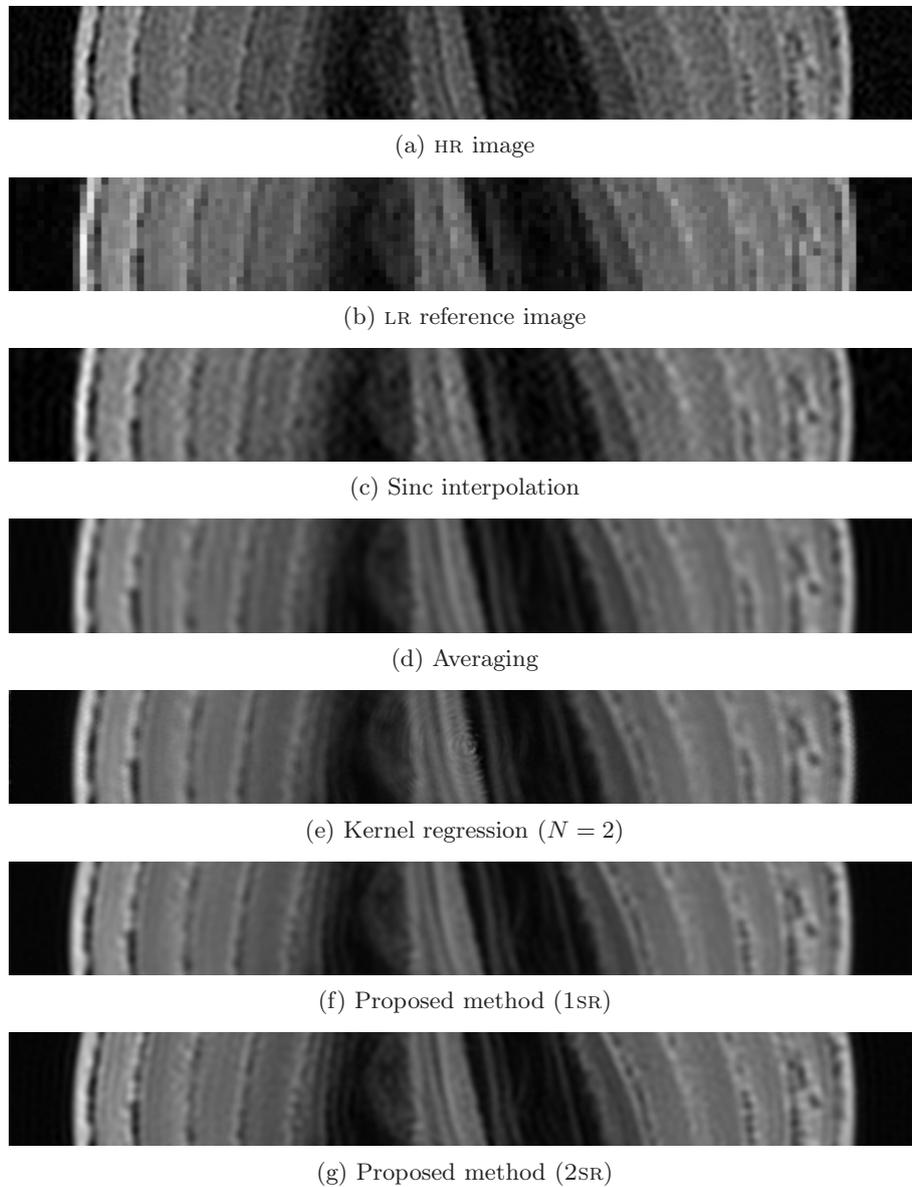


Figure 8.19: Detailed view of the reconstruction results on the region of interest (ROI) as indicated in Figure 8.18.

conclusion as for the reconstruction of the resolution phantom data illustrated in Figure 8.17.

8.5 Potential future applications

As we have shown in the previous sections, we can conclude that true image resolution enhancement can be achieved by combining multiple MRI data sets of rotated FOV's. From this fact, several potential applications can be pointed out.

A first application is the reduction of patient motion artefacts: we can combine several smaller LR scans instead of acquiring an HR scan with a longer acquisition time. Because low frequencies at the k-space center are measured each time once again, this results in a large total acquisition time. One way to speed up is to employ non-squared voxel sizes: the k-space is covered by rectangles instead of squares resulting in a better coverage with less overlapping measurements. In the limit, radial trajectories are a special case of our framework. The anisotropic voxels relate our work to *periodically rotated overlapping parallel lines with enhanced reconstruction* (PROPELLER) MRI,³ which is intentionally designed for reducing motion artefacts [Pipe, 1999]. Note that the proposed hybrid reconstruction algorithm performs resampling efficiently in the spatial domain instead of using regridding techniques in k-space, as for example in [Pipe, 1999].

The proposed hybrid reconstruction algorithm can be implemented very efficiently with only linear filtering operations. This offers a lot of advantage over the more exotic sampling strategies that avoid repeated sampling of the same k-values, such as *compressive sampling*, which undersamples the k-space at pseudo-random positions [Lustig et al., 2008]. Note that measuring the k-space at totally random positions is however not very beneficial to the efficiency of adjusting the magnetic gradients in the MRI scanner.

A second potential application of interest is for example diffusion tensor MRI and fiber tractography. For this technique, one T2-weighted image and diffusion weighted images of one slice are acquired in multiple directions, typically 30 or more. The diffusion weighted images only differ in contrast and contain similar information in the high frequencies. When rotating the FOV for the different diffusion weighted images, an HR image can be obtained when combining the images. The reconstruction algorithm should then be extended to handle images with different contrasts and, especially in case of EPI, varying distortions according to the rotation of the FOV. Other MRI techniques that require multiple acquisitions in any case, such as perfusion MRI and functional MRI, are also suitable for the proposed enhancement technique.

8.6 Conclusion

During the last decade, numerous papers deal with the spatial resolution enhancement of MRI and in particular, multi-frame super-resolution techniques have gained much interest in the MRI

³In SIEMENS MRI scanners, this is called BLADE MRI.

community. Despite the recent developments in super-resolution MRI [Carmi et al., 2006, Mayer and Vrscaj, 2006, Stoch and Balcom, 2006], we have pointed out some shortcomings of these approaches. Camera-based super-resolution can not be successful in Fourier encoded plane because frequency aliasing is absent in MRI acquisition. True image resolution enhancement is only possible if information outside the k-space span is obtained.

We have introduced an elegant way to enhance the image resolution by multiple MRI acquisitions: several rotated images are combined into one high-resolution image with an enlarged k-space span. We have proposed a novel reconstruction algorithm that consists of two parts: a proper resampling scheme in the image domain followed by optimal fusion of multiple aligned k-space data. We compare the proposed scheme with kernel regression, a general resampling tool [Takeda et al., 2007] and multi-frame averaging. Numerical simulations demonstrate the superiority of the proposed method, both quantitatively and qualitatively. The results also demonstrate the effectiveness of our scheme on real MRI data of a resolution phantom and an onion. In both cases, image quality has been improved and we can distinguish very small structures which is not possible using reconstruction techniques that only require a single image. Analyzing the Fourier data reveals that k-space data is indeed extended beyond k_{\max} and indicates that we have gained *true* spatial resolution.

One major drawback in our multi-frame MRI reconstruction framework is that the total acquisition time is still far too large to be applied in practical and clinical settings. In the proposed methodology, the same low-frequencies at the k-space center are measured each time once again. As mentioned in Section 8.5, a solution is to acquire non-squared voxel sizes as in (PROPELLER) MRI schemes, which is an alternative regridding technique. Future work consists of determining the optimal parameters (e.g. number of scans, rotation angles, the voxel dimensions, etc.) for a given acquisition time such that the maximum coverage of the k-space is obtained.

Another interesting research topic is to combine our resolution enhancement strategy in the Fourier encoded plane with classical SR MRI techniques in slice-direction. At last, it is also worthwhile to investigate our proposed scheme in other image representations such as steerable pyramids, which holds jointly spatial, frequency as well as directional image information [Simoncelli et al., 1992]. The reason is twofold: firstly, it avoids multiple transformations as it happens now and secondly, it can incorporate robustness against spatial outliers in an easy way (e.g. caused by local object motion).

9

Conclusion

Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.
—Winston Churchill

The main part of this dissertation addresses the improvement on the spatial resolution of images and video sequences. Experiments on both simulated and real data show the effectiveness of the different proposed techniques and its potential use in real applications. An overview of the proposed techniques is actually summarized in the decision chart shown in Figure 1.2.

In this chapter, we review our main contributions to image and video resolution enhancement techniques including the general class of image interpolation, restoration and super-resolution. This thesis is finally concluded with some directions for future research.

9.1 Review of our contributions

Linear interpolation methods are commonly proposed as a practical solution for the image resolution enhancement problem. Unfortunately, artefacts such as staircase, blur and ringing effects are inherit to these techniques. Therefore, we developed a new non-linear interpolation technique that eliminates these unwanted artefacts. The proposed interpolation algorithm sharpens edges by mapping the image level curves using constrained adaptive contrast enhancement techniques. To avoid the annoying amplification of jagged edges, the level curves are preprocessed by constrained isophote smoothing. Interpolation experiments show improvements of our method in both numerical PSNR results as well as in visual quality: the edges are much sharper, while staircase and ringing artefacts are heavily reduced. The proposed technique is a very good alternative for existing interpolation software or hardware.

In practice, a digital image suffers from several degradations such as blur, noise, compression artefacts, etc. Therefore, image restoration techniques are

proposed to take these degradation models into account. Among these numerous restoration methods, regularization techniques and especially the Bayesian methods are very powerful because they are able to include prior knowledge about the degradation and the hypothetical ideal image.

We introduced two new image colour priors, namely the adaptive bimodal colour prior and the multimodal colour prior. The adaptive bimodal colour prior assumes that the value of an edge pixel is a combination of the colours of two connected regions, each having a dominant colour distribution. The multimodal colour prior is proposed for images that normally just have a few dominant colours. Restoration results show the effectiveness and the visual superiority of the proposed Bayesian maximum a posteriori scheme to other interpolation/restoration schemes for images with a strong colour modality: noise and compression artefacts are removed very well and our method produces less annoying artefacts such as blur and jagged edges. The proposed restoration methods are found very suitable for the restoration of cartoons and cartoon movies, logos, maps, etc.

Common image restoration techniques only exploit the spatial redundancy in a local neighbourhood. We have demonstrated that the estimation of the restored pixel intensity can be based on information retrieved from the whole image, thereby exploiting the presence of similar patterns and features in the image, which we call repetitive structures. Therefore, the new approach is referred to as the non-local strategy, which is also related to the exemplar- and fractal-based approaches.

We presented a novel resolution enhancement scheme that exploits the repetitiveness. The new strategy brings more pixel information at our disposal, which leads to much better estimates of the unknown pixel values. Visual results show that the proposed non-local upscaling technique is superior and has less artefacts at very large magnifications to other resolution enhancement methods.

In document image processing applications, we can exploit the multiple occurrence of characters in the text. In order to take advantage of this repetitive behaviour in an efficient way, we divide the image into character segments. The character segmentation reduces the computation time drastically in two ways: the algorithm only has to focus on these regions of interest and the search space for possible matching candidates is enormously reduced. Matching between the character segments filters relevant information before the reconstruction. The proposed method combines the information from similar characters and reconstructs the high-resolution characters in the Bayesian framework with a text specific image prior. Experiments show that characters and symbols are reconstructed very well and additionally, OCR accuracy results show a significant improvement in comparison with other restoration methods. The proposed method is not restricted to font type or alphabet, therefore, it is also suitable to generic symbols such as musical notes, hieroglyphics or mathematical symbols. The same strategy can also be applied

in an exemplar-based search engine and in an efficient document compression scheme.

Multi-frame SR image restoration is quite a complex problem, which spans over several fields of image processing, such as motion estimation or image registration, image reconstruction from irregularly spaced samples, image deconvolution and denoising. In contrast to common restoration methods, it is possible to restore true high-frequency content via SR techniques by exploiting frequency aliasing. Due to its relatively low computational load and low memory requirements, the standard three-step paradigm of the non-uniform interpolation SR approach is recommended in most practical applications. These three successive steps are subpixel image alignment, image fusion and HR image restoration.

Problems such as temporal aliasing and motion blur can be solved by performing SR in the temporal direction. In some synthetic experiments, we evaluate the subpixel accuracy of several state-of-the-art shift estimators and conclude that the proposed LR-to-HR gradient-based registration method with steering kernel regression produces the most accurate subpixel information. For the photometric and joint geometric/photometric registration problem, we proposed the use of the total least square framework. The TLS solution produces in both cases more accurate and consistent registration parameters in comparison with the ordinary least square approach, which is commonly employed in the literature.

For the image fusion problem, we have proposed and derived the kernel regression algorithm in the TLS sense to handle positional or registration errors. The proposed method is more accurate and robust compared to the standard kernel regression algorithms in case of moderate or heavy noise and/or registration errors. We also have demonstrated the efficiency of the proposed SR scheme in real-world examples and deinterlacing problems.

In an extensive study, we have pointed out some limitations in the recent developments in super-resolution MRI reconstruction and we also have argued that classical super-resolution cannot be applied in the Fourier encoded plane because of the complete absence of frequency aliasing during MRI acquisition.

We introduced an elegant way to enhance the image resolution by multiple rotated MRI acquisitions. We have proposed a novel hybrid reconstruction algorithm that performs resampling in the image domain followed by fusion of multiple aligned k-space data. Numerical simulations demonstrate the superiority of the proposed method, both quantitatively and qualitatively. The results also demonstrate improvements on real MRI data of a resolution phantom and an onion. Analyzing the Fourier data reveals that we really have gained *true* spatial resolution. However, in practical and clinical settings, the MRI images should be acquired using non-squared voxel sizes as in PROPELLER MRI schemes.

9.2 Hints for future research

It is not our intention to repeat the many ideas for future work that already have been given at the end of each chapter. As an alternative, we choose to focus on three interesting and challenging topics.

A first interesting topic is how to perform joint super-resolution from images that are obtained in *different modalities* and/or under *different conditions*. We refer different modalities to the images that are captured by different devices, which is typical in for example remote sensing applications (e.g. multispectral images versus panchromatic images), medical applications (e.g. MRI versus CT or PET images) or depth cameras (e.g. HR regular images versus LR depth images). By different conditions, we understand for example varying illumination changes in uncontrolled experiments or different camera settings (e.g. with or without flash, automatic white balancing or different exposure times, which result in different dynamic ranges). Unfortunately, SR reconstruction can no longer be obtained by simply combining the pixel values of the different images. Instead of pixel values, we have to switch to higher level features such as edge information or to more complex (photometric) registration models that can deal with untrustworthy regions (e.g. due to clipping effects or saturation). Determining the set of useful features and their mutual relation in the different images is quite a challenging task.

A second topic of interest is how to define an accurate observation model for each application or acquisition device. The acquisition models used in this dissertation are quite generic in the sense that they are appropriate for an extremely broad class of images (i.e. dealing with blur, decimation and noise). In practice, these models should also take for instance compression and its architecture into account because most of the image and video material are stored in a compressed form. Artefacts produced by e.g. JPEG or JPEG2000 are in essence totally different and therefore, these compression schemes should also be tackled differently.

A related problem is how to find the right parameters automatically. Tools such as blur and noise identification become inevitable. The identification of the correlated noise produced by SR algorithms is for example an interesting research topic. A direct application can be found in a regular television set where various types of incoming video material should be processed to display on a single high-end screen. The quality of the incoming video can be high (e.g. HD streams or Blu-Ray disks), moderate (e.g. SD streams or DVD) or inferior (e.g. due to heavy compression in YouTube video's from the Internet). It is needless to say that the estimation of a correct and accurate observation model is a very complex and challenging task.

A third interesting topic is how to choose and incorporate prior knowledge for a specific application. Careful choice of the image prior models can result in impressive improvements in restoration quality and further developments in

this area should be encouraged. In this dissertation, we have already proposed some highly specific colour prior constraints for cartoons and text images, which lead to very good restoration results. Other applications where we can define specific prior knowledge (whether or not accomplished by training) include for example the restoration of faces, astronomical objects, etc.

Unfortunately, it is impossible to grasp the complexity and wealth of general image content in some simple analytic expressions. Therefore, the development of learning-based and non-local restoration techniques will be an important research topic. Another problem is the design and the choice of image priors, which does not only have to produce good restoration results (e.g. in terms of recognition), but also needs to result in visually pleasant images, i.e. without noticeable artefacts. A good example is the restoration of the texture of grass during soccer games on television, where the texture is lost at most places due to compression. It is not important that each stalk of grass can be recognized correctly, but it is important that the viewer has a natural impression of the soccer field. We believe that psycho-visual experiments can play an important role in the design of such natural image priors.

Bibliography

- [Aharon et al., 2006] Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322.
- [Ahn et al., 2002] Ahn, S., Rauh, W., Cho, H., and Warnecke, H.-J. (2002). Orthogonal distance fitting of implicit curve and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):620–638.
- [Allebach and Wong, 1996] Allebach, J. and Wong, P. (1996). Edge-directed interpolation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 707–710.
- [Allier et al., 2006] Allier, B., Bali, N., and Emptoz, H. (2006). Automatic accurate broken character restoration for patrimonial documents. *International Journal on Document Analysis and Recognition*, 8(4):246–261.
- [Altunbasak et al., 2002] Altunbasak, Y., Patti, A., and Mersereau, R. (2002). Super-resolution still and video reconstruction from MPEG-coded video. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(4):217–226.
- [Azzabou et al., 2007] Azzabou, N., Paragios, N., and Guichard, F. (2007). Image denoising based on adapted dictionary computation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 109–112.
- [Bab-Hadiashar and Suter, 1997] Bab-Hadiashar, A. and Suter, D. (1997). Optic flow calculation using robust statistics. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 988–993.
- [Baker et al., 2003a] Baker, S., Gross, R., and Matthews, I. (2003a). Lucas-Kanade 20 years on: a unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, The Robotics Institute, Carnegie Mellon University.
- [Baker et al., 2004a] Baker, S., Gross, R., and Matthews, I. (2004a). Lucas-Kanade 20 years on: a unifying framework: Part 4. Technical Report CMU-RI-TR-04-14, The Robotics Institute, Carnegie Mellon University.

- [Baker et al., 2003b] Baker, S., Gross, R., Matthews, I., and Ishikawa, T. (2003b). Lucas-Kanade 20 years on: a unifying framework: Part 2. Technical Report CMU-RI-TR-03-01, The Robotics Institute, Carnegie Mellon University.
- [Baker and Kanade, 1999] Baker, S. and Kanade, T. (1999). Super-resolution optical flow. Technical Report CMU-RI-TR-99-36, The Robotics Institute, Carnegie Mellon University.
- [Baker and Kanade, 2002] Baker, S. and Kanade, T. (2002). Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183.
- [Baker and Matthews, 2002] Baker, S. and Matthews, I. (2002). Lucas-Kanade 20 years on: a unifying framework: Part 1. Technical Report CMU-RI-TR-02-16, The Robotics Institute, Carnegie Mellon University.
- [Baker and Matthews, 2004] Baker, S. and Matthews, I. (2004). Lucas-Kanade 20 years on: a unifying framework. *International Journal of Computer Vision*, 56(3):221–255.
- [Baker et al., 2004b] Baker, S., Patil, R., Cheung, K., and Matthews, I. (2004b). Lucas-Kanade 20 years on: a unifying framework: Part 5. Technical Report CMU-RI-TR-04-64, The Robotics Institute, Carnegie Mellon University.
- [Banham and Katsaggelos, 1996] Banham, M. and Katsaggelos, A. (1996). Spatially adaptive wavelet-based multiscale image restoration. *IEEE Transactions on Image Processing*, 5(4):619–634.
- [Banham and Katsaggelos, 1997] Banham, M. and Katsaggelos, A. (1997). Digital image restoration. *IEEE Signal Processing Magazine*, 14(2):24–41.
- [Barron et al., 1994] Barron, L., Fleet, D., and Beauchemin, S. (1994). Performance of optical flow techniques. *International Journal on Computational Vision*, 12:43–77.
- [Bartoli, 2006] Bartoli, A. (2006). Groupwise geometric and photometric direct image registration. In *Proceedings of the 17th British Machine Vision Conference (BMVC)*, volume 1, pages 157–166.
- [Battiatto et al., 2002] Battiatto, S., Gallo, G., and Stanco, F. (2002). A locally adaptive zooming algorithm for digital images. *Image and Vision Computing*, 20:805–812.
- [Bennett et al., 2006] Bennett, E., Uyttendaele, M., Zitnick, C., Szeliski, R., and Kang, S. (2006). Video and image Bayesian demosaicing with a two color image prior. In *Proceedings of Seventh European Conference on Computer Vision (ECCV)*, volume 3951 of *Lecture Notes in Computer Science*, pages 508–521. Springer-Verlag.

- [Bern and Goldberg, 2000] Bern, M. and Goldberg, D. (2000). Scanner-model-based document image improvement. In *Proceedings of IEEE International Conference of Image Processing (ICIP)*, pages 582–585.
- [Bernstein et al., 2004] Bernstein, M., King, K., and Zhou, X. (2004). *Handbook of MRI Pulse Sequences*. Elsevier Academic Press.
- [Bertero and Boccacci, 1998] Bertero, M. and Boccacci, P. (1998). *Introduction to Inverse Problems in Imaging*. CRC Press.
- [Black and Anandan, 1996] Black, M. and Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104.
- [Black et al., 1998] Black, M., Sapiro, G., Marimont, D., and Heeger, D. (1998). Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–432.
- [Blaimer et al., 2004] Blaimer, M., Breuer, F., Mueller, M., Heidemann, R., Griswold, M., and Jakob, P. (2004). SMASH, SENSE, PILS, GRAPPA: How to choose the optimal method. *Topics in Magnetic Resonance Imaging*, 15(4):223–236.
- [Blake, 1989] Blake, A. (1989). Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):2–12.
- [Blu et al., 2001] Blu, T., Thévenaz, P., and Unser, M. (2001). MOMS: Maximal-order interpolation of minimal support. *IEEE Transactions on Image Processing*, 10(7):1069–1080.
- [Blu et al., 2004] Blu, T., Thévenaz, P., and Unser, M. (2004). Linear interpolation revitalized. *IEEE Transactions on Image Processing*, 13(5):710–719.
- [Borman, 2004] Borman, S. (2004). *Topics in Multiframe Superresolution Restoration*. PhD thesis, University of Notre Dame.
- [Bose and Boo, 1998] Bose, N. and Boo, K. (1998). Asymptotic eigenvalue distribution of block-Toeplitz matrices. *IEEE Transactions on Information Theory*, 44(2):858–861.
- [Bose et al., 1993] Bose, N., Kim, H., and Valenzuela, H. (1993). Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframe. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 269–272.
- [Brown, 1992] Brown, L. (1992). A survey of image registration techniques. *ACM Computing Surveys*, 24:326–376.

- [Brox, 2005] Brox, T. (2005). *From Pixels to Regions: Partial Differential Equations in Image Analysis*. PhD thesis, Saarland University.
- [Brox et al., 2008] Brox, T., Kleinschmidt, O., and Cremers, D. (2008). Efficient nonlocal means for denoising of textural patterns. *IEEE Transactions on Image Processing*, 17(7):1083–1092.
- [Bruhn et al., 2005a] Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T., and Schnörr, C. (2005a). Variational optical flow computation in real time. *IEEE Transactions on Image Processing*, 14(5):608–615.
- [Bruhn et al., 2005b] Bruhn, A., Weickert, J., and Schnörr, C. (2005b). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231.
- [Buades et al., 2005a] Buades, A., Coll, B., and Morel, J. (2005a). A non local algorithm for image denoising. In *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 60–65.
- [Buades et al., 2005b] Buades, A., Coll, B., and Morel, J. (2005b). A review of image denoising algorithms, with a new one. *SIAM Interdisciplinary Journal: Multiscale Modeling and Simulation*, 4(2):290–530.
- [Campisi and Egiazarian, 2007] Campisi, P. and Egiazarian, K., editors (2007). *Blind Image Deconvolution: Theory and Applications*. CRC Press.
- [Candès and Wakin, 2008] Candès, E. and Wakin, M. (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30.
- [Candocia, 2003] Candocia, F. (2003). Jointly registering images in domain and range by piecewise linear comparometric analysis. *IEEE Transactions on Image Processing*, 12(4):409–419.
- [Candocia and Mandarino, 2005] Candocia, F. and Mandarino, D. (2005). A semiparametric model for accurate camera response function modeling and exposure estimation from comparometric data. *IEEE Transactions on Image Processing*, 14(8):1138–1150.
- [Cannon et al., 1999] Cannon, M., Hochberg, J., and Kelly, P. (1999). Quality assessment and restoration of typewritten document images. *International Journal on Document Analysis and Recognition*, 2(2-3):80–89.
- [Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.
- [Capel and Zisserman, 2000] Capel, D. and Zisserman, A. (2000). Super-resolution enhancement of text image sequences. In *Proceedings of International Conference on Pattern Recognition (ICPR)*, pages 600–605.

- [Capel and Zisserman, 2003] Capel, D. and Zisserman, A. (2003). Computer vision applied to super resolution. *IEEE Signal Processing Magazine*, 20(3):75–86.
- [Carmi et al., 2006] Carmi, E., Liu, S., Alon, N., Fiat, A., and Fiat, D. (2006). Resolution enhancement in MRI. *Magnetic Resonance Imaging*, 24:133–154.
- [Casey and Lecolinet, 1996] Casey, R. and Lecolinet, E. (1996). A survey of methods and strategies in character segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):690–706.
- [Caspi and Irani, 2002] Caspi, Y. and Irani, M. (2002). Spatio-temporal alignment of sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11):1409–1424.
- [Chan et al., 2005] Chan, T., Esedoglu, S., Park, F., and Yip, A. (2005). *The Handbook of Mathematical Models of Computer Vision*, chapter Recent Developments in Total Variation Image Restoration. Springer-Verlag.
- [Chan et al., 2007] Chan, T., Ng, M., Yau, A., and Yip, A. (2007). Super-resolution image reconstruction using fast inpainting algorithms. *Applied and Computational Harmonic Analysis*, 23:3–24.
- [Chatterjee and Milanfar, 2008] Chatterjee, P. and Milanfar, P. (2008). A generalization of non-local means via kernel regression. In *Proceedings of SPIE Electronic Imaging*, volume 6814.
- [Chaudhuri and Joshi, 2005] Chaudhuri, S. and Joshi, M. (2005). *Motion-Free Super-Resolution*. Springer.
- [Cheeseman et al., 1994] Cheeseman, P., Kanefsky, B., Kraft, R., Stutz, J., and Hanson, R. (1994). Super-resolved surface reconstruction from multiple images. Technical Report FIA-94-12, NASA Ames Research Center, Moffett Field, CA.
- [Chiandussi and Ramponi, 1996] Chiandussi, S. and Ramponi, G. (1996). Nonlinear unsharp masking for the enhancement of document images. In *Proceedings of 8th European Signal Processing Conference (EUSIPCO)*, volume 1, pages 575–578.
- [Choudhury and Tumblin, 2003] Choudhury, P. and Tumblin, J. (2003). The trilateral filter for high contrast images and meshes. In *Proceedings of the 14th Eurographics Symposium on Rendering*, pages 186–196.
- [Chung et al., 2002] Chung, A., Wells III, W., Norbash, A., and Grimson, W. (2002). Multi-modal image registration by minimising Kullback-Leibler distance. In Dohi, T. and Kikinis, R., editors, *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2489 of *Lecture Notes in Computer Science*, pages 525–532. Springer-Verlag.

- [Comaniciu and Meer, 2002] Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619.
- [Coupé et al., 2008] Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C. (2008). An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Transactions on Medical Imaging*, 27(4):425–441.
- [Crow, 1977] Crow, F. (1977). The aliasing problem in computer-generated shaded images. *Communications of the ACM*, 20(11):799–805.
- [Şendur and Selesnick, 2002] Şendur, L. and Selesnick, I. (2002). Bivariate shrinkage with local variance estimation. *IEEE Signal Processing Letters*, 9:438–441.
- [Dabov et al., 2007] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095.
- [Dabov et al., 2008a] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2008a). Image restoration by sparse 3D transform-domain collaborative filtering. In *Proceedings of SPIE Electronic Imaging*, volume 6812.
- [Dabov et al., 2008b] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2008b). A nonlocal and shape-adaptive transform-domain collaborative filtering. In *Proceedings of International Workshop on Local and Non-Local Approximation in Image Processing*, pages 179–186.
- [Dalley et al., 2004] Dalley, G., Freeman, W., and Marks, J. (2004). Single-frame text super-resolution: a bayesian approach. In *Proceedings of IEEE International Conference of Image Processing (ICIP)*, pages 3295–3298.
- [Danielyan et al., 2008a] Danielyan, A., Foi, A., Katkovnik, V., and Egiazarian, K. (2008a). Image upsampling via spatially adaptive block-matching filtering. In *Proceedings of 16th European Signal Processing Conference (EU-SIPCO)*.
- [Danielyan et al., 2008b] Danielyan, A., Foi, A., Katkovnik, V., and Egiazarian, K. (2008b). Image and video super-resolution via spatially adaptive block-matching filtering. In *Proceedings of International Workshop on Local and Non-Local Approximation in Image Processing*, pages 53–60.
- [Das Gupta et al., 2007] Das Gupta, M., Rajaram, S., Huang, T., and Petrovic, N. (2007). Ordinal regression based subpixel shift estimation for video super-resolution. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9. Article ID 85963.

- [Datsenko and Elad, 2007] Datsenko, D. and Elad, M. (2007). Example-based single image super-resolution: A global MAP approach with outlier rejection. *Journal of Multidimensional Systems and Signal Processing*, 18(2-3):103–121.
- [Daubechies, 1992] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (SIAM).
- [Dauwe et al., 2008] Dauwe, A., Goossens, B., Luong, H., and Philips, W. (2008). A fast non-local image denoising algorithm. In *Proceedings of SPIE Electronic Imaging*, volume 6812.
- [Davis, 1998] Davis, J. (1998). Mosaics of scenes with moving objects. In *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 354–360.
- [De Leersnyder et al., 2006] De Leersnyder, E., De Smet, P., Luong, H., De Kinder, J., and Philips, W. (2006). 3D-image processing: Analysis of ball-point lines on paper. In *Proceedings of the 4th Conference Of The European Document Experts Working Group*, page 18, The Hague, The Netherlands.
- [De Smet, 2002] De Smet, P. (2002). *Segmentation and Analysis of Digital Image Sequences*. PhD thesis, Universiteit Gent (UGent).
- [De Smet et al., 2005] De Smet, P., De Bock, J., and Luong, H. (2005). On the reconstruction of strip-shredded documents. In *Proceedings of 1st annual IEEE BENELUX/DSP Valley Signal Processing Symposium (SPS-DARTS)*, pages 185–188, Antwerp, Belgium.
- [De Smet et al., 2004] De Smet, P., Rooms, F., Luong, H., and Philips, W. (2004). Do not zero-put: an efficient homespun MPEG audio layer II decoding and optimization strategy. In *Proceedings of ACM Multimedia*, pages 376–379, New York, USA.
- [De Witte, 2007] De Witte, V. (2007). *Colour Morphology with Application to Image Magnification*. PhD thesis, Universiteit Gent (UGent).
- [De Witte et al., 2006] De Witte, V., Schulte, S., Kerre, E., Ledda, A., and Philips, W. (2006). Morphological image interpolation to magnify images with sharp edges. In *Proceedings of International Conference On Image Analysis And Recognition (ICIAR)*, volume 4141 of *Lecture Notes in Computer Science*, pages 381–393. Springer-Verlag.
- [Debevec and Malik, 1997] Debevec, P. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th annual Conference on Computer Graphics and Interactive Techniques*, pages 369–378.

- [Dempster et al., 1977] Dempster, A., Lairde, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1–38.
- [Diebel and Thrun, 2005] Diebel, J. and Thrun, S. (2005). An application of Markov random fields to range sensing. In *Proceedings of Conference on Neural Information Processing Systems (NIPS)*.
- [Dodgson, 1997] Dodgson, N. (1997). Quadratic interpolation for image resampling. *IEEE Transactions on Image Processing*, 6(9):1322–1326.
- [Dologlou et al., 1996] Dologlou, I., Van Ormondt, D., and Carayannis, G. (1996). MRI scan time reduction through non-uniform sampling and SVD-based estimation. *Signal Processing*, 55:207–219.
- [Donaldson and Myers, 2005] Donaldson, K. and Myers, G. (2005). Bayesian super-resolution of text in video with a text-specific bimodal prior. *International Journal on Document Analysis and Recognition*, 7:159–167.
- [Ebrahimi and Vrscay, 2007] Ebrahimi, M. and Vrscay, E. (2007). Solving the inverse problem of image zooming using “self-examples”. In *Proceedings of International Conference On Image Analysis And Recognition (ICIAR)*, volume 4633 of *Lecture Notes in Computer Science*, pages 117–130. Springer-Verlag.
- [Ebrahimi and Vrscay, 2008a] Ebrahimi, M. and Vrscay, E. (2008a). Multi-frame super-resolution with no explicit motion estimation. In *Proceedings of International Conference on Image Processing, Computer Vision and Pattern Recognition (IPCV)*, pages 455–459.
- [Ebrahimi and Vrscay, 2008b] Ebrahimi, M. and Vrscay, E. (2008b). Self-similarity in imaging, 20 years after “fractals everywhere”. In *Proceedings of the International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*, pages 165–172.
- [Elad, 2002] Elad, M. (2002). On the origin of the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing*, 11(10):1141–1151.
- [Elad and Aharon, 2006] Elad, M. and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745.
- [Elad and Feuer, 1997] Elad, M. and Feuer, A. (1997). Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12):1646–1658.
- [Elad and Feuer, 1999] Elad, M. and Feuer, A. (1999). Super-resolution reconstruction of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):817–834.

- [Elad and Hel-Or, 2001] Elad, M. and Hel-Or, Y. (2001). A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. *IEEE Transactions on Image Processing*, 10(8):1187–1193.
- [Eren et al., 1997] Eren, P., Sezan, M., and Tekalp, A. (1997). Robust, object-based high-resolution image reconstruction from low-resolution video. *IEEE Transactions on Image Processing*, 6(10):1446–1451.
- [Farsiu et al., 2005] Farsiu, S., Elad, M., and Milanfar, P. (2005). Constrained, globally optimal, multi-frame motion estimation. In *Proceedings of IEEE Workshop on Statistical Signal Processing*, pages 1396–1401.
- [Farsiu et al., 2006a] Farsiu, S., Elad, M., and Milanfar, P. (2006a). Multiframe demosaicing and super-resolution of color images. *IEEE Transactions on Image Processing*, 15(1):141–159.
- [Farsiu et al., 2006b] Farsiu, S., Elad, M., and Milanfar, P. (2006b). Video-to-video dynamic super-resolution for grayscale and color sequences. *EURASIP Journal on Advances in Signal Processing*, 2006:1–15. Article ID 61859.
- [Farsiu et al., 2004] Farsiu, S., Robinson, M., Elad, M., and Milanfar, P. (2004). Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344.
- [Feghali and Vincent, 2004] Feghali, R. and Vincent, A. (2004). Computation of discontinuous optical flow fields based on spatiotemporal bilateral filtering. In *Proceedings of SPIE Electronic Imaging*, volume 5308, pages 1325–1332.
- [Ferreira, 1994] Ferreira, P. (1994). Interpolation and the discrete Papoulis-Gerchberg algorithm. *IEEE Transactions on Signal Processing*, 42(10):2596–2606.
- [Fiat, 2001] Fiat, D. (2001). Method of enhancing MRI signal. US Patent Application no. 6,294,914 B1, filed September 3, 1997, issued September 25, 2001.
- [Field, 1987] Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2394.
- [Fleck, 1992] Fleck, M. (1992). Some defects in finite-difference edge finders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):337–345.
- [Foi et al., 2007] Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images. *IEEE Transactions on Image Processing*, 16(5):1395–1411.

- [Forney, 1973] Forney, G. (1973). The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278.
- [Frahm et al., 1986] Frahm, J., Haase, A., and Matthaei, D. (1986). Rapid three-dimensional MR imaging using the FLASH technique. *Journal of Computer Assisted Tomography*, 10(2):363–368.
- [Fransens et al., 2004] Fransens, R., Strecha, C., and Van Gool, L. (2004). A probabilistic approach to optical flow based super-resolution. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 12, pages 191–191.
- [Freeman et al., 2002] Freeman, W., Jones, T., and Pasztor, E. (2002). Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):55–65.
- [Fridrich et al., 2003] Fridrich, J., Soukal, D., and Lukáš, J. (2003). Detection of copy-move forgery in digital images. In *Proceedings Digital Forensic Research Workshop*.
- [Fruchter and Hook, 2002] Fruchter, A. and Hook, R. (2002). Drizzle: A method for the linear reconstruction of undersampled images. *Publications of the Astronomical Society of the Pacific*, 114(792):144–152.
- [Galatsanos and Katsaggelos, 1992] Galatsanos, N. and Katsaggelos, A. (1992). Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*, 1(3):322–336.
- [Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.
- [Gerchberg, 1974] Gerchberg, R. (1974). Super-resolution through error energy reduction. *Optica Acta*, 21(9):709–720.
- [Gerchberg, 1989] Gerchberg, R. (1989). Superresolution through error function extrapolation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(10):1603–1606.
- [Gevrekci and Gunturk, 2007] Gevrekci, M. and Gunturk, B. (2007). Super-resolution under photometric diversity of images. *EURASIP Journal on Advances in Signal Processing*, 2007:1–12. Article ID 36076.
- [Gharavi-Alkhansari et al., 1997] Gharavi-Alkhansari, M., Denardo, R., Tenda, Y., and Huang, T. (1997). Resolution enhancement of images using fractal coding. *Proceedings of SPIE, The International Society for Optical Engineering*, 3024(2):1089–1100.

- [Gilboa et al., 2002] Gilboa, G., Sochen, N., and Zeevi, Y. (2002). Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing*, 11(7):689–703.
- [Glasbey and Mardia, 1998] Glasbey, C. and Mardia, K. (1998). A review of image warping methods. *Journal of Applied Statistics*, 25:155–171.
- [Golub et al., 1979] Golub, G., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223.
- [Golub and Van Loan, 1980] Golub, G. and Van Loan, C. (1980). An analysis of the total least squares problem. *SIAM Journal of Numerical Analysis*, 17:883–893.
- [Goossens et al., 2008a] Goossens, B., Luong, H., Pižurica, A., and Philips, W. (2008a). An improved non-local denoising algorithm. In *Proceedings of International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*.
- [Goossens et al., 2008b] Goossens, B., Luong, H., Pižurica, A., and Philips, W. (2008b). Space-variant characterization of image noise in computed tomography. In *Proceedings of Liège Imaging Days*.
- [Goossens et al., 2007] Goossens, B., Pižurica, A., and Philips, W. (2007). Noise removal from images by projecting onto bases of principle components. In *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 4678 of *Lecture Notes in Computer Science*, pages 190–199. Springer-Verlag.
- [Goossens et al., 2008c] Goossens, B., Ruzic, T., Luong, H., Pižurica, A., and Philips, W. (2008c). Adaptive non-local means filtering of images corrupted by colored noise. In *Proceedings of 9Th FirW Ph.D Symposium*, Ghent, Belgium.
- [Govindu, 2004] Govindu, V. (2004). Lie-algebraic averaging for globally consistent motion estimation. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 684–691.
- [Greenspan et al., 2002] Greenspan, H., Oz, G., Kiryati, N., and Peled, S. (2002). MRI inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, 20:437–446.
- [Grossberg and Nayar, 2003] Grossberg, M. and Nayar, S. (2003). Determining the camera response from images: What is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1455–1467.
- [Guerrero-Colón et al., 2008] Guerrero-Colón, J., Mancera, L., and Portilla, J. (2008). Image restoration using space-variant gaussian scale mixtures in overcomplete pyramids. *IEEE Transactions on Image Processing*, 17(1):27–41.

- [Hadamard, 1923] Hadamard, J. (1923). *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. H. Milford, Oxford University Press.
- [Haindl and Filip, 2002] Haindl, M. and Filip, J. (2002). Fast restoration of colour movie scratches. In *Proceedings of the 16th International Conference on Pattern Recognition*, pages 269–272.
- [Hansen and O'Leary, 1993] Hansen, P. and O'Leary, D. (1993). The use of the L-curve in the regularization of ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503.
- [Hardie et al., 1997] Hardie, R., Barnard, K., and Armstrong, E. (1997). Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transaction on Image Processing*, 6(12):1621–1633.
- [Hirakawa and Parks, 2006] Hirakawa, K. and Parks, T. (2006). Joint demosaicing and denoising. *IEEE Transactions on Image Processing*, 15(8):2146–2157.
- [Hobby and Ho, 1997] Hobby, J. and Ho, T. (1997). Enhancing degraded document images via bitmap clustering and averaging. In *Proceedings of the 4th International Conference on Document Analysis and Recognition (ICDAR)*, pages 394–400.
- [Honda et al., 1999] Honda, H., Haseyama, M., and Kitajima, H. (1999). Fractal interpolation for natural images. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 657–661.
- [Horn and Schunck, 1980] Horn, B. and Schunck, B. (1980). Determining optical flow. Technical Report A.I. Memo No. 572, Massachusetts Institute of Technology.
- [Hunt, 1977] Hunt, B. (1977). Bayesian methods in nonlinear digital image restoration. *IEEE Transactions on Computers*, 26(3):219–229.
- [Immerkær, 2001] Immerkær, J. (2001). Use of blur-space for deblurring and edge-preserving noise smoothing. *IEEE Transaction on Image Processing*, 10(6):837–840.
- [Ince and Konrad, 2008] Ince, S. and Konrad, J. (2008). Occlusion-aware optical flow estimation. *IEEE Transactions on Image Processing*, 17(8):1443–1451.
- [Irani and Peleg, 1991] Irani, M. and Peleg, S. (1991). Improving resolution by image registration. *Graphical Models and Image Processing*, 53(3):231–239.
- [Irani and Peleg, 1993] Irani, M. and Peleg, S. (1993). Motion analysis for image enhancement: Resolution, occlusion and transparency. *Journal of Visual Communication and Image Representation*, 4(4):324–335.

- [Jackson et al., 1991] Jackson, J., Meyer, C., Nishimura, D., and Macovski, A. (1991). Selection of a convolution function for Fourier inversion using gridding. *IEEE Transactions on Medical Imaging*, 10(3):473–478.
- [Jacquin, 1992] Jacquin, A. (1992). Image coding based on a fractal theory of iterated contractive image transformations. *IEEE Transactions on Image Processing*, 1(1):18–30.
- [Jeannin, 1996] Jeannin, S. (1996). On the combination of a polynomial motion estimation with a hierarchical segmentation based video coding scheme. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 489–492.
- [Jensen and Anastassiou, 1995] Jensen, K. and Anastassiou, D. (1995). Sub-pixel edge localization and the interpolation of still images. *IEEE Transactions on Image Processing*, 4(3):285–295.
- [Jeon et al., 2006] Jeon, B.-W., Park, R.-H., and Yang, S. (2006). Resolution enhancement by prediction of the high-frequency image based on the Laplacian pyramid. *EURASIP Journal on Applied Signal Processing*, 2006:1–11.
- [Jiang and Moloney, 2002] Jiang, H. and Moloney, C. (2002). A new direction adaptive scheme for image interpolation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 369–372.
- [Jiang et al., 2003] Jiang, Z., Wong, T.-T., and Bao, H. (2003). Practical super-resolution from dynamic video sequences. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 549.
- [Jiji et al., 2007] Jiji, C., Chaudhuri, S., and Chatterjee, P. (2007). Single frame image super-resolution: Should we process locally or globally? *Journal of Multidimensional Systems and Signal Processing*, 18:123–152.
- [Kalifa et al., 2003] Kalifa, J., Mallat, S., and Rouge, B. (2003). Deconvolution by thresholding in mirror wavelet bases. *IEEE Transactions on Image Processing*, 12(4):446–457.
- [Kanungo and Haralick, 1993] Kanungo, T. and Haralick, R.M. and Phillips, I. (1993). Global and local document degradation models. In *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, pages 730–734.
- [Katkovnik et al., 2005] Katkovnik, V., Egiazarian, K., and Astola, J. (2005). A spatially adaptive nonparametric regression image deblurring. *IEEE Transactions on Image Processing*, 14(10):1469–1478.
- [Katsaggelos, 1990] Katsaggelos, A. (1990). A multiple input image restoration approach. *Journal of Visual Communication and Image Representation*, 1:93–103.

- [Katsaggelos, 1991] Katsaggelos, A., editor (1991). *Digital Image Restoration*, volume 23 of *Springer Series in Information Sciences*. Springer-Verlag.
- [Katscher et al., 2003] Katscher, U., Börnert, P., Leussler, C., and Van Den Brink, J. (2003). Transmit SENSE. *Magnetic Resonance in Medicine*, 49:144–150.
- [Keller, 1976] Keller, J. (1976). Inverse problems. *The American Mathematical Monthly*, 83(2):107–118.
- [Keren et al., 1988] Keren, D., Peleg, S., and Brada, R. (1988). Image sequence enhancement using sub-pixel displacements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 742–746.
- [Kervrann and Boulanger, 2006] Kervrann, C. and Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. *IEEE Transactions on Image Processing*, 15(10):2866–2878.
- [Kervrann et al., 2007] Kervrann, C., Boulanger, J., and Coupé, P. (2007). Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In *Proceedings of International Conference on Scale Space and Variational Methods (SSVM)*, pages 520–532.
- [Kia et al., 1998] Kia, O., Doermann, D., Rosenfeld, A., and Chellappa, R. (1998). Symbolic compression and processing of document images. *Journal of Computer Vision and Image Understanding*, 70(3):335–349.
- [Kim and Sikora, 2005] Kim, J. and Sikora, T. (2005). Hybrid recursive energy-based method for robust optical flow on large motion fields. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 129–132.
- [Kim et al., 1990] Kim, S., Bose, N., and Valenzuela, H. (1990). Recursive reconstruction of high resolution image from noisy undersampled multiframe. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(6):1013–1027.
- [Kimmel et al., 2000] Kimmel, R., Malladi, R., and Sochen, N. (2000). Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129.
- [Kinebuchi et al., 2001] Kinebuchi, K., Muresan, D., and Parks, T. (2001). Image interpolation using wavelet-based hidden markov trees. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1957–1960.
- [Kirsch, 1996] Kirsch, A. (1996). *An Introduction To The Mathematical Theory of Inverse Problems*. Springer.

- [Klette and Zamperoni, 1996] Klette, R. and Zamperoni, P. (1996). *Handbook Of Image Processing Operators*. John Wiley & Sons.
- [Kokaram, 1993] Kokaram, A. (1993). *Motion Picture Restoration*. PhD thesis, University of Cambridge.
- [Kornprobst et al., 2003] Kornprobst, P., Peeters, R., Nikolova, M., Deriche, R., Ng, M., and Van Hecke, P. (2003). A superresolution framework for fMRI sequences and its impact on resulting activation maps. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, pages 117–125. Springer-Verlag.
- [Kornprobst et al., 2002] Kornprobst, P., Peeters, R., Vieville, T., Malandain, G., Mierisova, S., Sunaert, S., Faugeras, O., and Van Hecke, P. (2002). Superresolution in MRI and its influence in statistical analysis. Technical report, Institut National De Recherche En Informatique Et En Automatique (INRIA).
- [Kotel'nikov, 2000] Kotel'nikov, V. (2000). *Modern Sampling Theory: Mathematics and Applications*, chapter Reprint (1933): On the Transmission Capacity of the "Ether" and Wire in Electrocommunications, pages 27–45. Birkhäuser.
- [Krawczyk-Stańdo and Rudnicki, 2007] Krawczyk-Stańdo, D. and Rudnicki, M. (2007). Regularization parameter selection in discrete ill-posed problems - the use of the U-curve. *International Journal of Applied Mathematics and Computer Science*, 17(2):157–164.
- [Kundur and Hatzinakos, 1996] Kundur, D. and Hatzinakos, D. (1996). Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64.
- [Kwon et al., 2003] Kwon, O., Sohn, K., and Lee, C. (2003). Deinterlacing using directional interpolation and motion compensation. *IEEE Transactions on Consumer Electronics*, 49(1):198–203.
- [Kybic and Unser, 2003] Kybic, J. and Unser, M. (2003). Fast parametric elastic image registration. *IEEE Transactions on Image Processing*, 12(11):1427–1442.
- [Landini et al., 2005] Landini, L., Positano, V., and Santarelli, M., editors (2005). *Advanced Image Processing in Magnetic Resonance Imaging*. CRC Press/Taylor & Francis Group.
- [Lange et al., 1989] Lange, K., Little, R., and Taylor, J. (1989). Robust statistical modeling using the t-distribution. *Journal of the American Statistical Association*, 84(408):881–896.
- [Larkman and Nunes, 2007] Larkman, D. and Nunes, R. (2007). Parallel magnetic resonance imaging. *Physics in Medicine and Biology*, 52:15–55.

- [Ledda, 2006] Ledda, A. (2006). *Mathematical Morphology in Image Processing*. PhD thesis, Universiteit Gent (UGent).
- [Ledda et al., 2005] Ledda, A., Luong, H., De Witte, V., Philips, W., and Kerre, E. (2005). Image interpolation using mathematical morphology. In *Proceedings of 6th FirW PhD Symposium*.
- [Ledda et al., 2006a] Ledda, A., Luong, H., De Witte, V., Philips, W., and Kerre, E. (2006a). Image interpolation using mathematical morphology. In *Proceedings of 2nd annual IEEE International Conference on Document Image Analysis for Libraries (DIAL)*, pages 258–367.
- [Ledda et al., 2006b] Ledda, A., Luong, H., Philips, W., De Witte, V., and Kerre, E. (2006b). Greyscale image interpolation using mathematical morphology. In *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 4179 of *Lecture Notes in Computer Science*, pages 78–90. Springer-Verlag.
- [Ledda et al., 2008] Ledda, A., Luong, H., Philips, W., De Witte, V., and Kerre, E. (2008). Binary image interpolation based on mathematical morphology. In *Proceedings of the 16th European Signal Processing Conference (EUSIPCO)*.
- [Lee and Kang, 2003] Lee, E. and Kang, M. (2003). Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration. *IEEE Transactions on Image Processing*, 12(7):826–837.
- [Lee, 1983] Lee, J.-S. (1983). Digital image smoothing and the sigma filter. *Computer Vision, Graphics and Image Processing*, 24:255–269.
- [Lee et al., 1996] Lee, S., Lee, D., and Park, H. (1996). A new methodology for gray-scale character segmentation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):1045–1050.
- [Lee et al., 1995] Lee, X., Zhang, Y., and Leon-Garcia, A. (1995). Information loss recovery for block-based image coding techniques - a fuzzy logic approach. *IEEE Transactions on Image Processing*, 4:259–273.
- [Lehmann et al., 1999] Lehmann, T., Gönner, C., and Spitzer, K. (1999). Survey: Interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–1075.
- [LeQuang et al., 1995] LeQuang, D., Zaccarin, A., and Caron, S. (1995). Object-oriented coding using successive motion field segmentation and estimation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 207–210.
- [Lertrattanapanich and Bose, 2002] Lertrattanapanich, S. and Bose, N. (2002). HR image from multiframe by Delaunay triangulation: a synopsis. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 869–872.

- [Li and Doermann, 1999] Li, H. and Doermann, D. (1999). Text enhancement in digital video using multiple frame integration. In *Proceedings of ACM Multimedia*, pages 19–22.
- [Li et al., 2008] Li, W., Yuan, Y., and Yu, N. (2008). Detecting copy-paste forgery of JPEG image via block artifact grid extraction. In *Proceedings of the International Workshop on Local and Non-Local Approximation in Image Processing*, pages 121–126.
- [Li and Orchard, 2001] Li, X. and Orchard, M. (2001). New edge-directed interpolation. *IEEE Transactions on Image Processing*, 10(10):1521–1527.
- [Liang et al., 2005] Liang, J., Doermann, D., and Li, H. (2005). Camera-based analysis of text and documents: a survey. *International Journal on Document Analysis and Recognition*, 7:84–104.
- [Lin and Shum, 2004] Lin, Z. and Shum, H.-Y. (2004). Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):83–97.
- [Liu et al., 1998] Liu, C., Rubin, D., and Wu, Y. (1998). Parameters expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, 85(4):755–770.
- [Lowe, 1999] Lowe, D. (1999). Object recognition from local scale-invariant features. In *Proceedings of the 7th International Conference on Computer Vision (ICCV)*, pages 1150–1157.
- [Lucas and Kanade, 1981] Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679.
- [Lüke, 1999] Lüke, H. (1999). The origins of the sampling theorem. *IEEE Communications Magazine*, 37(4):106–108.
- [Luong et al., 2004a] Luong, H., De Smet, P., De Geyter, M., and Philips, W. (2004a). Hierarchical browsing of digital video sequences. In *Proceedings of the 4th IEEE Benelux Signal Processing Symposium (SPS)*, pages 97–100.
- [Luong et al., 2004b] Luong, H., De Smet, P., and Philips, W. (2004b). Super resolution image reconstruction in digital video sequences. In *Proceedings of 5th FTW PhD Symposium*. Faculty of Engineering, Ghent University.
- [Luong et al., 2005] Luong, H., De Smet, P., and Philips, W. (2005). Image interpolation using constrained adaptive contrast enhancement techniques. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 998–1001.

- [Luong et al., 2009] Luong, H., Fieremans, E., Deblaere, K., Aelterman, J., Goossens, B., Pižurica, A., Michiels, J., and Philips, W. (2009). MRI resolution enhancement in the Fourier encoded plane. *IEEE Transactions on Medical Imaging (submitted)*.
- [Luong et al., 2008] Luong, H., Fieremans, E., Deblaere, K., Aelterman, J., Goossens, B., Pižurica, A., and Philips, W. (2008). An in-plane MRI resolution enhancement technique. In *Proceedings of Radiological Society of North America*, page 1063, Chicago, Illinois, USA.
- [Luong et al., 2006a] Luong, H., Foubert, A., and Philips, W. (2006a). Robust video mosaicing for benthic habitat mapping. In *Proceedings of the First International Conference on Computer Vision Theory and Applications (VISAPP)*, volume I, pages 306–310.
- [Luong et al., 2004c] Luong, H., Gautama, S., and Philips, W. (2004c). Automatic registration of synthetic aperture radar (SAR) images. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3864–3867.
- [Luong et al., 2007] Luong, H., Goossens, B., and Philips, W. (2007). Image upscaling using global multimodal priors. In Blanc-Talon, J., Philips, W., Popescu, D., and Scheunders, P., editors, *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 4678 of *Lecture Notes in Computer Science*, pages 473–484, Delft, The Netherlands. Springer-Verlag.
- [Luong et al., 2006b] Luong, H., Ledda, A., and Philips, W. (2006b). An image interpolation scheme for repetitive structures. In *Proceedings of International Conference On Image Analysis And Recognition (ICIAR)*, volume 4141 of *Lecture Notes in Computer Science*, pages 104–115. Springer-Verlag.
- [Luong et al., 2006c] Luong, H., Ledda, A., and Philips, W. (2006c). Non-local image interpolation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 693–696.
- [Luong et al., 2006d] Luong, H., Lippens, S., and Philips, W. (2006d). Practical and robust super resolution using anisotropic diffusion for under-determined cases. In *Proceedings 2nd annual IEEE BENELUX/DSP Valley Signal Processing Symposium (SPS-DARTS)*, pages 139–142, Antwerp, Belgium.
- [Luong and Philips, 2005] Luong, H. and Philips, W. (2005). Sharp image interpolation by mapping level curves. In *Proceedings of Visual Communications and Image Processing (VCIP)*, pages 2012–2022.
- [Luong and Philips, 2006] Luong, H. and Philips, W. (2006). Digital reconstruction of degraded low resolution images. In *Proceedings of 7th FirW Ph.D Symposium*, Ghent, Belgium.

- [Luong and Philips, 2007a] Luong, H. and Philips, W. (2007a). Non-local text image reconstruction. In *Proceedings of International Conference On Document Analysis And Recognition (ICDAR)*, volume 1, pages 546–550, Curitiba, Brazil.
- [Luong and Philips, 2007b] Luong, H. and Philips, W. (2007b). Reconstruction of low-resolution images using adaptive bimodal priors. In Campilho, A. and Kamel, M., editors, *Proceedings of International Conference On Image Analysis And Recognition (ICIAR)*, volume LNCS 4633 of *Lecture Notes in Computer Science*, pages 69–80, Montréal, Canada. Springer-Verlag.
- [Luong and Philips, 2008] Luong, H. and Philips, W. (2008). Robust reconstruction of low-resolution document images by exploiting repetitive character behaviour. *International Journal on Document Analysis and Recognition (IJDAR)*, 11(1):39–51.
- [Lustig et al., 2008] Lustig, M., Donoho, D., Santos, J., and Pauly, J. (2008). Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82.
- [Mahmoudi and Sapiro, 2005] Mahmoudi, M. and Sapiro, G. (2005). Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE Signal Processing Letters*, 12(12):839–842.
- [Mallat, 1989] Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.
- [Mancas-Thillou and Mirmehdi, 2007] Mancas-Thillou, C. and Mirmehdi, M. (2007). *Digital Document Processing: Major Directions and Recent Advances (Advances in Pattern Recognition)*, chapter An Introduction to Super-Resolution Text, pages 305–327. Springer-Verlag.
- [Mandelbrot, 1982] Mandelbrot, B. (1982). *The Fractal Geometry of Nature*. W. H. Freeman.
- [Mann, 2000] Mann, S. (2000). Comparametric equations with practical applications in quantigraphic image processing. *IEEE Transactions on Image Processing*, 9(8):1389–1406.
- [Mansfield, 1977] Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):55–58.
- [Markovsky and Van Huffel, 2007] Markovsky, I. and Van Huffel, S. (2007). Overview of total least-squares methods. *Signal Processing*, 87:2283–2302.
- [Marziliano and Vetterli, 2000] Marziliano, P. and Vetterli, M. (2000). Reconstruction of irregularly sampled discrete-time bandlimited signals with unknown sampling locations. *IEEE Transactions on Signal Processing*, 48(12):3462–3471.

- [Matas et al., 2000] Matas, J., Koubaroulis, D., and Kittler, J. (2000). Colour image retrieval and object recognition using the multimodal neighbourhood signature. In *Proceedings of the 6th European Conference in Computer Vision*, volume 1, pages 48–64.
- [Matthews and Namazi, 1995] Matthews, K. and Namazi, N. (1995). Simultaneous motion parameter estimation and image segmentation using the EM algorithm. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 542–545.
- [Mayer and Vrscay, 2006] Mayer, G. and Vrscay, E. (2006). Mathematical analysis of "phase ramping" for super-resolution magnetic resonance imaging. In *Proceedings of International Conference on Image Analysis and Recognition (ICIAR)*, volume 4141 of *Lecture Notes in Computer Science*, pages 82–93. Springer-Verlag.
- [Mayer and Vrscay, 2007] Mayer, G. and Vrscay, E. (2007). Measuring information gain for frequency-encoded super-resolution MRI. *Magnetic Resonance Imaging*, 25:1058–1069.
- [McNamee et al., 1971] McNamee, J., Stenger, F., and Whitney, E. (1971). Whittaker's cardinal function in retrospect. *Mathematics of Computation*, 25(113):141–154.
- [Meijering, 2002] Meijering, E. (2002). A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proceedings of the IEEE*, 90(3):319–342.
- [Meijering et al., 1999] Meijering, E., Niessen, W.J., P. J., and Viergever, M. (1999). Quantitative comparison of sinc-approximating kernels for medical image interpolation. In Taylor, C. and Colchester, A., editors, *Proceedings of Medical Image Computing and Computer-assisted Intervention (MICCAI)*, volume 1679 of *Lecture Notes in Computer Science*, pages 210–217. Springer-Verlag.
- [Meijering et al., 2001] Meijering, E., Niessen, W., and Viergever, M. (2001). Quantitative evaluation of convolution-based methods for medical image interpolation. *Medical Image Analysis*, 5(2):111–126.
- [Mémin and Pérez, 2002] Mémin, E. and Pérez, P. (2002). Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46(2):129–155.
- [Mikolajczyk and Schmid, 2004] Mikolajczyk, K. and Schmid, C. (2004). Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86.
- [Milanfar et al., 2006] Milanfar, P., Farsiu, S., Elad, M., and Robertson, M. (2006). Robust reconstruction of high resolution grayscale images from a

- sequence of low-resolution frames (robust gray super-resolution). US Patent Application no. 2006/0291751 A1, filed December 12, 2005, issued December 28, 2006.
- [Morbée et al., 2007] Morbée, M., Tessens, L., Luong, H., Prades-Nebot, J., Pižurica, A., and Philips, W. (2007). A distributed coding-based content-aware multi-view video system. In *Proceedings of 1st ACM/IEEE International Conference On Distributed Smart Cameras*, pages 355–362, Vienna, Austria.
- [Morse and Schwartzwald, 1998] Morse, B. and Schwartzwald, D. (1998). Isophote-based interpolation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 227–231.
- [Morse and Schwartzwald, 2001] Morse, B. and Schwartzwald, D. (2001). Image magnification using level-set reconstruction. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 333–341.
- [Muresan, 2005] Muresan, D. (2005). Fast edge directed polynomial interpolation. In *Proceedings of International Conference on Image Processing (ICIP)*, pages 990–993.
- [Muresan and Parks, 2001] Muresan, D. and Parks, T. (2001). Optimal recovery approach to image interpolation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 848–851.
- [Muresan and Parks, 2004] Muresan, D. and Parks, T. (2004). Adaptively quadratic (AQua) image interpolation. *IEEE Transactions on Image Processing*, 13(5):690–698.
- [Nadaraya, 1964] Nadaraya, E. (1964). On estimating regression. *Theory of Probability and its Applications*, 9(1):141–142.
- [Nakagaki and Katsaggelos, 2003] Nakagaki, R. and Katsaggelos, A. (2003). A VQ-based blind image restoration algorithm. *IEEE Transactions on Image Processing*, 12(9):1044–1053.
- [Navarro, 2001] Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88.
- [Neelamani et al., 2004] Neelamani, R., Choi, H., and Baraniuk, R. (2004). ForWard: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing*, 52(2):418–433.
- [Nguyen and Milanfar, 2000] Nguyen, N. and Milanfar, P. (2000). A wavelet-based interpolation-restoration method for superresolution (wavelet super-resolution). *Circuits Systems Signal Process*, 19(4):321–338.
- [Nosratinia, 2001] Nosratinia, A. (2001). New kernels for fast mesh-based motion estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(1):40–51.

- [Nyquist, 2002] Nyquist, H. (2002). Reprint (1928): Certain topics in telegraph transmission theory. *Proceedings of the IEEE*, 90(2):280–305.
- [Orchard et al., 2008] Orchard, J., Ebrahimi, M., and Wong, A. (2008). Efficient nonlocal-means denoising using the SVD. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 1732–1735.
- [O’Rourke and Stevenson, 1995] O’Rourke, T. and Stevenson, R. (1995). Improved image decompression for reduced transform coding artifacts. *IEEE Transactions on Circuits, Systems and Video Technology*, 5:490–499.
- [Osher and Rudin, 1990] Osher, S. and Rudin, L. (1990). Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27:919–940.
- [Oskoui-Fard and Stark, 1988] Oskoui-Fard, P. and Stark, H. (1988). Tomographic image reconstruction using the theory of convex projections. *IEEE Transactions on Medical Imaging*, 7(1):45–58.
- [O’Sullivan, 1985] O’Sullivan, J. (1985). A fast sinc function gridding algorithm for Fourier inversion in computer tomography. *IEEE Transactions on Medical Imaging*, 4(4):200–207.
- [Özkan et al., 1994] Özkan, M., Tekalp, A., and Sezan, M. (1994). POCS-based restoration of space-varying blurred images. *IEEE Transactions on Image Processing*, 3(4):450–454.
- [Papenberg et al., 2006] Papenberg, N., Bruhn, A., Brox, T., Didas, S., and Weickert, J. (2006). Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158.
- [Papoulis, 1975] Papoulis, A. (1975). A new algorithm in spectral analysis and band-limited extrapolation. *IEEE Transactions on Circuits and Systems*, 22(9):735–742.
- [Papoulis, 1977] Papoulis, A. (1977). Generalized sampling expansion. *IEEE Transactions on Circuits and Systems*, 24(11):652–654.
- [Park and Lee, 2008] Park, J.-S. and Lee, S.-W. (2008). An example-based face hallucination method for single-frame, low-resolution facial images. *IEEE Transactions on Image Processing*, 17(10):1806–1816.
- [Park et al., 2003] Park, S., Park, M., and Kang, M. (2003). Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36.
- [Patti et al., 1997] Patti, A., Sezan, M., and Tekalp, A. (1997). Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Transactions on Image Processing*, 6(8):1064–1076.

- [Peeters et al., 2004] Peeters, R., Kornprobst, P., Nikolova, M., Sunaert, S., Vieville, T., Malandain, G., Deriche, R., Faugeras, O., Ng, M., and Van Hecke, P. (2004). The use of super-resolution techniques to reduce slice thickness in functional MRI. *International Journal of Imaging Systems and Technology*, 14(3):131–138.
- [Peled and Yeshurun, 2001] Peled, S. and Yeshurun, Y. (2001). Superresolution in MRI: Application to human white matter fiber tract visualization by diffusion tensor imaging. *Magnetic Resonance in Medicine*, 45:29–35.
- [Peled and Yeshurun, 2002] Peled, S. and Yeshurun, Y. (2002). Superresolution in MRI - perhaps sometimes. *Magnetic Resonance in Medicine*, 48(2):409.
- [Peleg et al., 1987] Peleg, S., Keren, D., and Schweitzer, L. (1987). Improving image resolution using sub-pixel motion. *Pattern Recognition Letters*, 5(3):223–226.
- [Peng et al., 2006] Peng, H., Sabati, M., Lauzon, L., and Frayne, R. (2006). MR image reconstruction of sparsely sampled 3D k-space data by projection-onto-convex sets. *Magnetic Resonance Imaging*, 24:761–773.
- [Perona and Malik, 1990] Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639.
- [Pham, 2006] Pham, T. (2006). *Spatiotonal Adaptivity in Super-Resolution of Under-sampled Image Sequences*. PhD thesis, Technische Universiteit Delft.
- [Pham et al., 2005] Pham, T., Bezuijen, M., Van Vliet, L., Schutte, K., and Luengo Hendriks, C. (2005). Performance of optimal registration estimators. In *Proceedings of SPIEE Visual Information Processing XIV*, volume 5817, pages 133–144.
- [Pham et al., 2006] Pham, T., Van Vliet, L., and Schutte, K. (2006). Robust fusion of irregularly sampled data using adaptive normalized convolution. *EURASIP Journal on Applied Signal Processing*, 2006:1–12. Article ID 83268.
- [Pipe, 1999] Pipe, J. (1999). Motion correction with PROPELLER MRI: Application to head motion and free-breathing cardiac imaging. *Magnetic Resonance in Medicine*, 42:963–969.
- [Pires and Aguiar, 2005] Pires, B. and Aguiar, P. (2005). Featureless global alignment of multiple images. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 57–60.
- [Pižurica, 2002] Pižurica, A. (2002). *Image Denoising Using Wavelets and Spatial Context Modeling*. PhD thesis, Universiteit Gent (UGent).

- [Pižurica and Philips, 2006] Pižurica, A. and Philips, W. (2006). Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising. *IEEE Transactions on Image Processing*, 15(3):654–665.
- [Pižurica et al., 2005] Pižurica, A., Vanhamel, I., Sahli, H., Philips, W., and Katartzis, A. (2005). A Bayesian approach to nonlinear diffusion based on a Laplacian prior for ideal image gradient. In *Proceedings of 13th IEEE Workshop On Statistical Signal Processing*, pages 477–482.
- [Polidori and Dugelay, 1995] Polidori, E. and Dugelay, J.-L. (1995). Zooming using iterated function systems. In *Proceedings of NATO ASI Conference on Fractal Image Encoding and Analysis*.
- [Portilla and Simoncelli, 2003] Portilla, J. and Simoncelli, E. (2003). Image restoration using Gaussian scale mixtures in the wavelet domain. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 965–968.
- [Portilla et al., 2003] Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351.
- [Press et al., 1988] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1988). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- [Price and Hayes III, 1998] Price, J. and Hayes III, M. (1998). Resampling and reconstruction with fractal interpolation functions. *IEEE Signal Processing Letters*, 5(9):228–230.
- [Protter et al., 2009] Protter, M., Elad, M., Takeda, H., and Milanfar, P. (2009). Generalizing the non-local-means to super-resolution reconstruction. *IEEE Transactions on Image Processing*, 18(1):36–51.
- [Pruessmann et al., 2001] Pruessmann, K., Weiger, M., Börnert, P., and Boesiger, P. (2001). Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic Resonance in Medicine*, 46:638–651.
- [Rajan and Chaudhuri, 2002] Rajan, D. and Chaudhuri, S. (2002). Generation of super-resolution images from blurred observations using an MRF model. *Journal of Mathematical Imaging Vision*, 16:5–15.
- [Ratakonda and Ahuja, 1998] Ratakonda, K. and Ahuja, N. (1998). POCS based adaptive image magnification. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 203–207.
- [Rav-Acha and Peleg, 2005] Rav-Acha, A. and Peleg, S. (2005). Two motion-blurred images are better than one. *Pattern Recognition Letters*, 26:311–317.

- [Reddy and Chatterji, 1996] Reddy, B. and Chatterji, B. (1996). An FFT-based technique for translation, rotation and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8):1266–1271.
- [Reibman and Schaper, 2006] Reibman, A. and Schaper, T. (2006). Subjective performance evaluation of super-resolution image enhancement. In *2nd International Workshop on Video Processing and Quality Metrics (VPQM)*.
- [Rhee et al., 2000] Rhee, I., Martin, G., Muthukrishnan, S., and Packwood, R. (2000). Quadtree-structured variable-size block-matching motion estimation with minimal error. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(1):42–50.
- [Rhee and Kang, 1999] Rhee, S. and Kang, M. (1999). Discrete cosine transform based regularized high-resolution image reconstruction algorithm. *Optical Engineering*, 38(8):1348–1356.
- [Rice, 1996] Rice, S. (1996). *Measuring the Accuracy of Page-Reading Systems*. PhD thesis, University of Nevada.
- [Robertson and Stevenson, 2001] Robertson, M. and Stevenson, R. (2001). Temporal resolution enhancement in compressed video sequences. *EURASIP Journal on Applied Signal Processing*, 2001(4):230–238.
- [Robinson et al., 2009] Robinson, D., Farsiu, S., and Milanfar, P. (2009). Optimal registration of aliased images using variable projection with applications to super-resolution. *The Computer Journal*, 52(1):31–42.
- [Robinson and Milanfar, 2004] Robinson, D. and Milanfar, P. (2004). Fundamental performance limits in image registration. *IEEE Transactions on Image Processing*, 13(9):1185–1199.
- [Robinson and Milanfar, 2005] Robinson, D. and Milanfar, P. (2005). Bias minimizing filter design for gradient-based image registration. *Signal Processing: Image Communication*, 20:554–568.
- [Robinson and Milanfar, 2006] Robinson, D. and Milanfar, P. (2006). Statistical performance analysis of super-resolution. *IEEE Transactions on Image Processing*, 15(6):1413–1428.
- [Rokitta et al., 1999] Rokitta, M., Zimmerman, U., and Haase, A. (1999). Fast NMR flow measurements in plants using FLASH imaging. *Journal of Magnetic Resonance*, 137:29–32.
- [Romberg, 2008] Romberg, J. (2008). Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):14–20.
- [Rooms, 2005] Rooms, F. (2005). *Nonlinear Methods in Image Restoration Applied to Confocal Microscopy*. PhD thesis, Universiteit Gent (UGent).

- [Roullot et al., 2000] Roullot, E., Herment, A., Bloch, I., Nikolova, M., and Mousseaux, E. (2000). Regularized reconstruction of 3D high-resolution magnetic resonance images from acquisitions of anisotropically degraded resolutions. In *Proceedings of International Conference on Pattern Recognition*, volume 3, pages 350–353.
- [Rudin and Osher, 1994] Rudin, L. and Osher, S. (1994). Total variation based image restoration with free local constraints. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 31–35.
- [Saint-Marc et al., 1991] Saint-Marc, P., Chen, J.-S., and Medioni, G. (1991). Adaptive smoothing: a general tool for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):514–529.
- [Sawhney et al., 1998] Sawhney, H., Hsu, S., and Kumar, R. (1998). Robust video mosaicing through topology interference and local to global alignment. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 2, pages 103–119.
- [Schechtman et al., 2005] Schechtman, E., Caspi, Y., and Irani, M. (2005). Space-time super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):531–545.
- [Scheffler, 2002] Scheffler, K. (2002). Superresolution in MRI? *Magnetic Resonance in Medicine*, 48:408.
- [Schmid et al., 2000] Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172.
- [Schroeter et al., 1998] Schroeter, P., Vesin, J.-M., Langenberger, T., and Meuli, R. (1998). Robust parameter estimation of intensity distributions for brain magnetic resonance images. *IEEE Transactions on Medical Imaging*, 17(2):172–186.
- [Schultz and Stevenson, 1996] Schultz, R. and Stevenson, R. (1996). Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, 5(6):996–1011.
- [Schultz and Stevenson, 1997] Schultz, R. and Stevenson, R. (1997). Bayesian estimation of subpixel-resolution motion fields and high-resolution video stills. In *Proceedings of IEEE International Conference on Image Processing*, volume 3, pages 62–65.
- [Schutten et al., 1998] Schutten, R., Pelagotti, A., and De Haan, G. (1998). Layered motion estimation. *Philips Journal of Research*, 51(2):253–267.
- [Seppä, 2007] Seppä, M. (2007). High-quality two-stage resampling for 3-D volumes in medical imaging. *Medical Image Analysis*, 11:346–360.

- [Serra, 1982] Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.
- [Sethian, 1999] Sethian, J. (1999). *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press.
- [Shahram and Milanfar, 2006] Shahram, M. and Milanfar, P. (2006). Statistical and information-theoretic analysis of resolution in imaging. *IEEE Transactions on Information Theory*, 52(8):3411–3437.
- [Shannon, 1949] Shannon, C. (1949). Communication in the presence of noise. *Proceedings of Institute of Radio Engineers*, 37(1):10–21.
- [Shekarforoush et al., 1996] Shekarforoush, H., Berthod, M., Zerubia, J., and Werman, M. (1996). Sub-pixel bayesian estimation of albedo and height. *International Journal of Computer Vision*, 19(3):289–330.
- [Shekarforoush et al., 2002] Shekarforoush, H., Zerubia, J., and Berthod, M. (2002). Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3):188–200.
- [Shen et al., 2007] Shen, H., Zhang, L., Huang, B., and Li, P. (2007). A MAP approach for joint motion estimation, segmentation, and super resolution. *IEEE Transactions on Image Processing*, 16(2):479–490.
- [Shi and Karl, 2008] Shi, Y. and Karl, W. (2008). A real-time algorithm for the approximation of level-set-based curve evolution. *IEEE Transactions on Image Processing*, 17(5):645–656.
- [Shum and Szeliski, 2000] Shum, H.-Y. and Szeliski, R. (2000). Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2):101–130.
- [Simoncelli et al., 1992] Simoncelli, E., Freeman, W., Adelson, E., and Heeger, D. (1992). Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607.
- [Smith and Brady, 1997] Smith, S. and Brady, J. (1997). SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78.
- [Smolić et al., 1999] Smolić, A., Sikora, T., and Ohm, J. (1999). Long-term global motion estimation and its application for sprite coding, content description, and segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(8):1227–1242.
- [Stark, 1988] Stark, H. (1988). Theory of convex projection and its application to image restoration. In *Proceedings of IEEE International Symposium on Circuits and Systems*, volume 1, pages 963–964.

- [Stark and Oskoui-Fard, 1989] Stark, H. and Oskoui-Fard, P. (1989). High resolution image recovery from image-plane arrays, using convex projections. *Journal of Optical Society of America A*, 6(11):1715–1726.
- [Stark, 2000] Stark, J. (2000). Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing*, 9(5):889–896.
- [Stewart, 1999] Stewart, C. (1999). Robust parameter estimation in computer vision. *SIAM Reviews*, 41(3):513–537.
- [Stiller, 1997] Stiller, C. (1997). Object-based estimation of dense motion fields. *IEEE Transactions on Image Processing*, 6(2):234–250.
- [Stiller and Konrad, 1999] Stiller, C. and Konrad, J. (1999). Estimating motion in image sequences: A tutorial on modeling and computation of 2D motion. *IEEE Signal Processing Magazine*, 16(4):70–91.
- [Stoch and Balcom, 2006] Stoch, G. and Balcom, B. (2006). Does a k-space span limitation affect MRI resolution? *Concepts in Magnetic Resonance Part B (Magnetic Resonance Engineering)*, 29B(2):55–61.
- [Su and Willis, 2004] Su, D. and Willis, P. (2004). Image interpolation by pixel level data-dependent triangulation. *Computer Graphics*, 23(2):189–201.
- [Sullivan et al., 1994] Sullivan, S., Sandford, L., and Ponce, J. (1994). Using geometric distance fits for 3-D object modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(12):1183–1196.
- [Takeda et al., 2007] Takeda, H., Farsiu, S., and Milanfar, P. (2007). Kernel regression for image processing and reconstruction. *IEEE Transactions on Image Processing*, 16(2):349–366.
- [Taylor and Dance, 1998] Taylor, M. and Dance, C. (1998). Enhancement of document images from cameras. In *Proceedings of SPIE Document Recognition*, volume 3305, pages 230–241.
- [Teelen and Veelaert, 2005] Teelen, K. and Veelaert, P. (2005). Computing the uncertainty of transformations in digital images. In *Proceedings of SPIE Vision Geometry XIII*, pages 1–12.
- [Tekalp et al., 1992] Tekalp, A., Ozkan, M., and Sezan, M. (1992). High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 3, pages 169–172.
- [Tessens et al., 2007] Tessens, L., Ledda, A., Pižurica, A., and Philips, W. (2007). Extending the depth of field in microscopy through curvelet-based

- frequency-adaptive image fusion. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 861–864.
- [Thévenaz et al., 1998] Thévenaz, P., Ruttimann, U., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing*, 7(1):27–41.
- [Thouin and Chang, 2000] Thouin, P. and Chang, C. (2000). A method for restoration of low-resolution document images. *International Journal on Document Analysis and Recognition*, 2:200–210.
- [Tikhonov et al., 1990] Tikhonov, A., Goncharsky, A., Stepanov, V., and Yagola, A. (1990). *Numerical Methods For The Solution of Ill-Posed Problems*. Kluwer Academic Publishers.
- [Tom and Katsaggelos, 1995] Tom, B. and Katsaggelos, A. (1995). Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 539–542.
- [Tom and Katsaggelos, 1996] Tom, B. and Katsaggelos, A. (1996). An iterative algorithm for improving the resolution of video sequences. In *Proceedings of SPIE Conference on Visual Communication and Image Processing (VCIP)*, pages 1430–1438.
- [Tomasi and Manduchi, 1998] Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Proceedings of International Conference on Computer Vision*, pages 839–846.
- [Tonazzini et al., 2004] Tonazzini, A., Vezzosi, S., and Bedini, L. (2004). Analysis and recognition of highly degraded printed characters. *International Journal on Document Analysis and Recognition*, 6:236–247.
- [Traka and Tziritas, 2003] Traka, M. and Tziritas, G. (2003). Panoramic view reconstruction. *Signal Processing: Image Communication*, 18:465–481.
- [Triggs et al., 1999] Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A. (1999). Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag.
- [Trussell and Hartwig, 2002] Trussell, H. and Hartwig, R. (2002). Mathematics for demosaicking. *IEEE Transactions on Image Processing*, 11(4):485–492.
- [Tsai and Huang, 1984] Tsai, R. and Huang, T. (1984). *Advances in Computer Vision and Image Processing*, volume 1, chapter Multiframe Image Restoration and Registration, pages 317–339. JAI Press Inc.

- [Tschumperlé, 2006] Tschumperlé, D. (2006). Fast anisotropic smoothing of multi-valued images using curvature-preserving PDE's. *International Journal of Computer Vision*, 68(1):65–82.
- [Tschumperlé and Deriche, 2005] Tschumperlé, D. and Deriche, R. (2005). Vector-valued image regularization with PDE's: A common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):506–517.
- [Tuytelaars and Mikolajczyk, 2008] Tuytelaars, T. and Mikolajczyk, K. (2008). Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280.
- [Tuytelaars and Van Gool, 2004] Tuytelaars, T. and Van Gool, L. (2004). Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 1(59):61–85.
- [Ukkonen, 1995] Ukkonen, E. (1995). On-line construction of suffix trees. *Algorithmica*, 14(3):249–260.
- [Unser, 1999] Unser, M. (1999). Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6):22–38.
- [Unser, 2000] Unser, M. (2000). Sampling - 50 years after Shannon. *Proceedings of the IEEE*, 88(4):569–587.
- [Unser et al., 1993a] Unser, M., Aldroubi, A., and Eden, M. (1993a). B-spline signal processing: Part I - theory. *IEEE Transactions on Signal Processing*, 41(2):821–832.
- [Unser et al., 1993b] Unser, M., Aldroubi, A., and Eden, M. (1993b). B-spline signal processing: Part II - efficient design and applications. *IEEE Transactions on Signal Processing*, 41(2):834–848.
- [Van De Ville, 2001] Van De Ville, D. (2001). *Lineaire, Niet-Lineaire en Vaaglogische Beeldinterpolatietechnieken*. PhD thesis, Universiteit Gent (UGent).
- [Van Eekeren et al., 2007] Van Eekeren, A., Schutte, K., Oudegeest, O., and Van Vliet, L. (2007). Performance evaluation of super-resolution reconstruction methods on real-world data. *EURASIP Journal on Advances in Signal Processing*, 2007:1–11. Article ID 43953.
- [Van Trees, 1968] Van Trees, H. (1968). *Detection, Estimation and Modulation Theory. Part I: Detection, Estimation, and Linear Modulation Theory*. John Wiley and Sons, Inc.
- [Vandewalle, 2006] Vandewalle, P. (2006). *Super-Resolution From Unregistered Aliased Images*. PhD thesis, École Polytechnique Fédérale de Laussane (EPFL).

- [Vansteenkiste, 2007] Vansteenkiste, E. (2007). *Quantitative Analysis of Ultrasound Images of the Preterm Brain*. PhD thesis, Universiteit Gent (UGent).
- [Vincent, 2005] Vincent, E. and Laganière, R. (2005). Detecting and matching feature points. *Journal of Visual Communication and Image Representation*, 16(1):38–54.
- [Volovelsky et al., 2005] Volovelsky, K., Golan, M., and Goldberg, N. (2005). Super-resolution image processing. International Patent Application no. WO 2005/072060 A2, filed February 1, 2004, issued August 11, 2005.
- [Vrhel, 2005] Vrhel, M. (2005). Color image resolution conversion. *IEEE Transactions on Image Processing*, 14(3):328–333.
- [Šroubek and Flusser, 2005] Šroubek, F. and Flusser, J. (2005). Multichannel blind deconvolution of spatially misaligned images. *IEEE Transactions on Image Processing*, 14(7):874–883.
- [Wang and Adelson, 1993] Wang, J. and Adelson, E. (1993). Layered representation for motion analysis. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, pages 361–366.
- [Wang et al., 2006] Wang, J., Guo, Y., Ying, Y., Liu, Y., and Peng, Q. (2006). Fast non-local algorithm for image denoising. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 1429–1432.
- [Wang and Ward, 2007] Wang, Q. and Ward, R. (2007). A new orientation-adaptive interpolation method. *IEEE Transactions on Image Processing*, 16(4):889–900.
- [Weickert, 1998] Weickert, J. (1998). *Anisotropic Diffusion in Image Processing*. ECMI Series. Teubner-Verlag.
- [Weickert and Schnörr, 2001] Weickert, J. and Schnörr, C. (2001). Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14:245–255.
- [Weiger et al., 2002] Weiger, M., Pruessmann, K., Österbauer, R., Börnert, P., Boesiger, P., and Jezzard, P. (2002). Sensitivity-encoded single-shot spiral imaging for reduced susceptibility artifacts in BOLD fMRI. *Magnetic Resonance in Medicine*, 48:860–866.
- [Woods et al., 2006] Woods, N., Galatsanos, N., and Katsaggelos, A. (2006). Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images. *IEEE Transactions on Image Processing*, 15(1):201–213.
- [Yang et al., 1995] Yang, Y., Galatsanos, N., and Katsaggelos, A. (1995). Projection-based spatially adaptive reconstruction of block-transform compressed images. *IEEE Transactions on Image Processing*, 4(7):896–908.

- [Yang and Yan, 2000] Yang, Y. and Yan, H. (2000). An adaptive logical method for binarization of degraded document images. *Pattern recognition*, 33:787–807.
- [Yang et al., 2000] Yang, Y., Yan, H., and Yu, D. (2000). Content-lossless document image compression based on structural analysis and pattern matching. *Pattern recognition*, 33:1277–1293.
- [Yaoping and Chengke, 1998] Yaoping, Y. and Chengke, W. (1998). A novel video coding scheme using delaunay triangulation. *Journal of Visual Communication and Image Representation*, 9(1):80–86.
- [Yaroslavsky, 1985] Yaroslavsky, L. (1985). *Digital Picture Processing*. Springer-Verlag.
- [Ye et al., 2005] Ye, G., Pickering, M., Frater, M., and Arnold, J. (2005). A robust approach to super-resolution sprite generation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 897–900.
- [Yoo and Jeong, 2002] Yoo, H. and Jeong, J. (2002). Direction-oriented interpolation and its application to de-interlacing. *IEEE Transactions on Consumer Electronics*, 48(4):954–962.
- [Youla and Webb, 1982] Youla, D. and Webb, H. (1982). Image restoration by the method of convex projections: Part 1 - theory. *IEEE Transactions on Medical Imaging*, 1(2):81–94.
- [Young, 2006] Young, S. (2006). Method of super-resolving images. US Patent Application no. US 2006/0159369 A1, filed January 19, 2005, issued July 20, 2006.
- [Yu and Bajaj, 2004] Yu, Z. and Bajaj, C. (2004). A fast and adaptive method for image contrast enhancement. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 1001–1004.
- [Zhao and De Haan, 2003] Zhao, M. and De Haan, G. (2003). Content-adaptive video up-scaling. In *Proceedings of 9th annual Conference of the Advanced School for Computing and Imaging (ASCI)*, pages 151–156.
- [Zhao et al., 2002] Zhao, M., Leitão, J., and De Haan, G. (2002). Towards an overview of spatial up-conversion techniques. In *Proceedings of IEEE International Symposium on Consumer Electronics (ISCE)*, pages 13–16.
- [Zheng and Kanungo, 2001] Zheng, Q. and Kanungo, T. (2001). Morphological degradation models and their use in document image restoration. In *Proceedings of IEEE International Conference of Image Processing (ICIP)*, volume 1, pages 193–196.

- [Zibetti and Mayer, 2006] Zibetti, M. and Mayer, J. (2006). Outlier robust and edge-preserving simultaneous super-resolution. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 1741–1744.
- [Zibetti and Mayer, 2007] Zibetti, M. and Mayer, J. (2007). A robust and computationally efficient simultaneous super-resolution scheme for image sequences. *IEEE Transactions on Circuits Systems and Video Technology*, 17(10):1288–1300.
- [Zimmer et al., 2008] Zimmer, S., Didas, S., and Weickert, J. (2008). A rotationally invariant block matching strategy improving image denoising with non-local means. In *Proceedings of International Workshop on Local and Non-Local Approximation in Image Processing*, pages 135–142.
- [Zitová and Flusser, 2003] Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000.
- [Zlokolica, 2006] Zlokolica, V. (2006). *Advanced Nonlinear Methods for Video Denoising*. PhD thesis, Universiteit Gent (UGent).
- [Zlokolica et al., 2006] Zlokolica, V., Pižurica, A., and Philips, W. (2006). Wavelet-based joint video de-interlacing and denoising. In *SPIE Optics East*, volume 6383, Boston, USA.
- [Zomet and Peleg, 2000] Zomet, A. and Peleg, S. (2000). Efficient super-resolution and applications to mosaics. In *Proceedings of International Conference on Pattern Recognition (ICPR)*, volume 1, pages 579–583.
- [Zomet et al., 2001] Zomet, A., Rav-Acha, A., and Peleg, S. (2001). Robust super resolution. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 645–650.