A BAYESIAN APPROACH TO NONLINEAR DIFFUSION BASED ON A LAPLACIAN PRIOR FOR IDEAL IMAGE GRADIENT

Aleksandra Pižurica, Iris Vanhamel†, Hichem Sahli†, Wilfried Philips* and Antonis Katartzis†*

*Ghent University, Department Telecommunications and Information Processing
Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium
Email: Aleksandra.Pizurica@telin.ugent.be
†Vrije Universiteit Brussel, Department of Electronics and Informatics
Pleinlaan 2, 1050 Brussels, Belgium
Email: iuvanham@etro.vub.ac.be

ABSTRACT

We study the relationships between diffusivity functions in a nonlinear diffusion scheme and probabilities of edge presence under a marginal prior on ideal, noise-free image gradient. In particular we impose a Laplacian-shaped prior for the ideal gradient and we define the diffusivity function explicitly in terms of edge probabilities under this prior. The resulting diffusivity function has no free parameters to optimize. Our results demonstrate that the new diffusivity function, automatically, i.e., without any parameter adjustments, satisfies the well accepted criteria for the goodness of edge-stopping functions. Our results also offer a new and interesting interpretation of some widely used diffusivity functions, which are now compared to edge-stopping functions under a marginal prior for the ideal image gradient.

1. INTRODUCTION

Nonlinear diffusion filtering [1] has received and still attracts a considerable research attention in the field of image processing, the statistical signal and image analysis. Let $f(\mathbf{x}) = \{f^{(1)}(\mathbf{x}), f^{(2)}(\mathbf{x}), \ldots, f^{(M)}(\mathbf{x})\}$ denote a M-band image defined on a finite domain Ω, where $\mathbf{x} = (\mathrm{x}, \mathrm{y})$ is the position vector on $\mathbf{Z}^2$ defining the coordinates of image pixels. The following system of coupled partial differential equations (PDEs) produces a family of diffused images $u(\mathbf{x}, t)$, which are smoothed versions of $f(\mathbf{x})$:

$$\begin{cases} \partial_t u^{(r)}(\mathbf{x}, t) & = \mathrm{div}\left[c(\mathbf{x}, t)\nabla u^{(r)}(\mathbf{x}, t)\right] \quad \forall r = 1...M \\ u\,(\mathbf{x}, t = 0) & = f(\mathbf{x}) \\ \partial_{\mathbf{n}} u(\mathbf{x}, t) & = 0 \qquad \text{on } \delta\Omega \end{cases}$$

where ∂_t is the partial derivative over t, $c(\mathbf{x}, t)$ is the so-called *conductance coefficient*, which controls the diffusion process at each position $\mathbf{x}$; t being the continuous scale parameter. $\delta\Omega$ is the image boundary, with $\mathbf{n}$ denoting the normal direction to it.

In a *nonlinear* diffusion the conductance $c(\mathbf{x}, t)$ coefficient is chosen to vary spatially so as to encourage intra-region smoothing in preference to inter-region smoothing, and thus overcoming the distortion of region boundaries. Based on these requirements, in the widely adopted formulation by Perona and Malik [1] the conductance coefficient is a function of the gradient magnitude $|\nabla u(\mathbf{x}, t)|$:

$$c(\mathbf{x}, t) = g(|\nabla u(\mathbf{x}, t)|) \tag{1}$$

where $g(x)$ is the so-called *edge stopping* or *diffusivity* function. The diffusivity function is a non-negative function, which is usually decreasing with the increase of the gradient magnitude. It ensures, that important edges and region boundaries are less blurred than noise or low-contrast details. Some of the often used diffusivity functions include those proposed by Perona [1], denoted as *Lorentzian* (g_{Lor}) and *Le Clerc* (g_{Lec}), respectively, the *Tukey bi-weight* function (g_{Tuk}) [2] and the *Weickert* edge stopping function (g_{Wei}) [3]:

- $\star\ g_{Lor}(x) = \frac{1}{1+\frac{x^2}{k^2}}$ $\qquad \star\ g_{Lec}(x) = \exp\left[-\frac{x^2}{k^2}\right]$
- $\star\ g_{Tuk}(x) = \begin{cases} \left[1 - \frac{x^2}{5k^2}\right]^2 & \text{if } |x| < \sqrt{5}k \\ 0 & \text{else} \end{cases}$
- $\star\ g_{Wei}(x) = \begin{cases} 1 - \exp\left[\frac{-3.31488k^8}{x^8}\right] & \text{if } x \neq 0 \\ 1 & \text{else} \end{cases}$

The parameter k appearing in these functions is called the contrast parameter, because it controls the shape of the diffusivity function, balancing the degrees of inter-region smoothing and edge enhancement in the diffusion process. Black *et al* [2] showed that different edge-stopping functions in the anisotropic diffusion are closely related to the corresponding error norms and the so-called influence functions in the robust statistical estimation framework.

This research was supported by the Fund for the Scientific Research in Flanders (Fonds voor Wetenschappelijk Onderzoek Vlaanderen - FWO), project G.0376.03 Multiscale statistical models for image denoising and segmentation.

Since the first formulation of the nonlinear diffusion, the starting idea in defining edge-stopping functions is that they should have a smaller value for those gradient magnitudes that are more likely to present image edges. In this sense, the edge stopping functions *implicitly* express the probabilities of edge absence. To our knowledge no attempts were made so far to *explicitly* express the diffusivity function $g(x)$ as a probability that x presents no edge under a suitable marginal prior for ideal, noise-free image gradient.

In related works, several researchers have recently proposed stochastic formulations of *conductance coefficients* [4, 5], which is usually called "probabilistic" diffusion. Note that these approaches replace the classical diffusivity function of the image gradient by a stochastic conductance coefficient, which is not an explicit function of the image gradient.

In this paper, we return to the classical Perona-Malik formulation of the nonlinear diffusion, governed by an explicit function of the image gradient and we revisit this formulation within a Bayesian framework. By imposing a suitable prior distribution for ideal, noise-free image gradient, we derive the diffusivity function as a probability that the observed noisy gradient magnitude presents no edge of interest. In deriving the edge probabilities we make a parallel to the related recent developments in the wavelet processing [6]. Other researchers have recently studied connections between the nonlinear diffusion and the classical wavelet shrinkage functions, like hard- and soft-thresholding [7, 8]. In this paper, we point to new relationships between the two domains, employing the statistical image models that are initially developed for wavelet processing for deriving new diffusivity functions.

The paper is organized as follows. Section 2 presents the proposed Bayesian formulation of the edge-stopping function. We start from the specification of a suitable prior for noise-free gradient (Section 2.1) and then we formalize the diffusivity in terms of the corresponding edge probability (Section 2.2). The results are presented in Section 3, where we compare the shape of the resulting probabilistic diffusivity function to other, often used functions. We also present the results of the diffusion process for several representative images. Discussions and concluding remarks are given in Section 4.

2. PROPOSED APPROACH

The main idea of our approach is to express the diffusivity function as a probability that the observed gradient presents no edge of interest under a suitable marginal prior distribution for the noise-free gradient histogram.

2.1. Specifying a suitable prior for ideal gradient

Ideal, noise-free images typically contain large portions of relatively uniform regions that produce negligible gradient values. Therefore, for any reasonable gradient definition, the histogram of the noise-free gradient will be sharply peaked at zero. On the other hand, sharp edges and textured regions will generally produce some relatively large gradients, building in this way long tails of the gradient histogram (see the ideal gradient in Fig. 1(b) and its histogram in Fig. 2(a)).

In the wavelet literature, motivated by the above reasoning, a well-adopted prior for the marginal probability density function of the noise-free wavelet coefficients is the generalized Laplacian, also called generalized Gaussian distribution [9, 10]

$$p(x) = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\exp(-\lambda|x|^{\nu}), \tag{2}$$

where $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$ is the Gamma function, $\lambda > 0$ is the *scale* parameter and ν is the *shape* parameter. The two parameters of this prior are reliably estimated from the noisy histogram, provided that noise is additive white Gaussian. The variance σ^2 and the fourth moment m_4 of the generalized Laplacian signal corrupted by additive white Gaussian noise with standard deviation σ_n are [10]

$$\sigma^2 = \sigma_n^2 + \frac{\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})}, \quad m_4 = 3\sigma_n^4 + \frac{6\sigma_n^2\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})} + \frac{\Gamma(\frac{5}{\nu})}{\lambda^4\Gamma(\frac{1}{\nu})}. \tag{3}$$

Many researchers use a special case of the density (2) with $\nu = 1$, namely *Laplacian* or *double exponential* [11]

$$p(x) = Lap(x;\lambda) = (\lambda/2)\exp(-\lambda|x|) \tag{4}$$

because of its analytical simplicity, without incurring a significant performance degradation in image denoising and compression schemes. It can be shown that the parameter estimation from (3) now simplifies as

$$\lambda = [0.5(\sigma^2 - \sigma_n^2)]^{-1/2}. \tag{5}$$

with σ_n^2 and σ^2 the variances of noise and noisy signal, resp.

Note that each wavelet subband can be interpreted as a partial derivative of the smoothed image in a given direction [12] and thus the statistical models for wavelet coefficients can be largely applied to smoothed image gradients in the nonlinear diffusion. Here we present the results for both the generalized Laplacian (2) and the Laplacian (4) priors. The proposed methodology that follows is general and can be used with other priors as well. For example, other reasonable (highly kurtotic) candidates include Cauchy and α-stable priors.

In our particular implementation of the diffusion scheme, the gradient is calculated following [1], i.e., by taking the

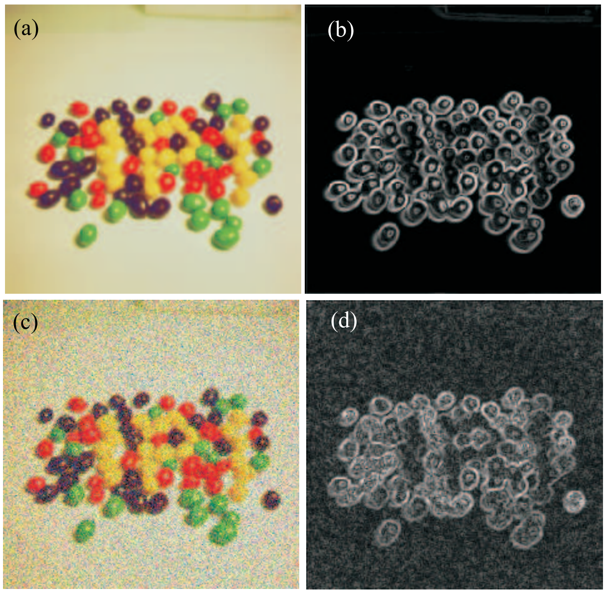

Fig. 1. (a) Noise free image, (b) its ideal gradient, (c) noisy contaminated image and, (d) its gradient.

absolute difference of neighboring pixel intensities on the 4-connected grid. For color images, we use the color distance between the neighboring pixels [13]. In Fig. 2, we illustrate the estimation of the Laplacian prior for the implemented gradient in our scheme. Fig. 2(a) and Fig. 2(b) show respectively the histograms of the gradient magnitudes for the ideal (noise-free) image in Fig. 1(a), and for the noisy image in Fig. 1(b). From the *noisy* histogram in Fig. 2(b), using (5), we estimate the parameter of the prior for *noise-free* data. The result is illustrated in Fig. 2(c), which shows the estimated Laplacian prior in comparison to the ideal, noise-free histogram. A generalized Laplacian prior, having two parameters, may achieve a slightly better fit, but the main concept remains the same.

2.2. The proposed probabilistic diffusivity function

A noise-free gradient magnitude that exceeds a specific threshold, T, corresponds to the edge-element *of interest*. We relate this threshold to the noise level: $T = \sigma_n$, where σ_n is the noise standard deviation in the observed gradient image. Let m denote the ideal, noise-free gradient magnitude and define the following two hypotheses: H_0: "*an edge-element of interest is absent*" and H_1: "*an edge-element of interest is present*" precisely as:

$$H_0 : m \leq \sigma_n \quad \text{and} \quad H_1 : m > \sigma_n. \tag{6}$$

We define the diffusivity function as

$$g(x) = A(1 - P(H_1|x)) \tag{7}$$

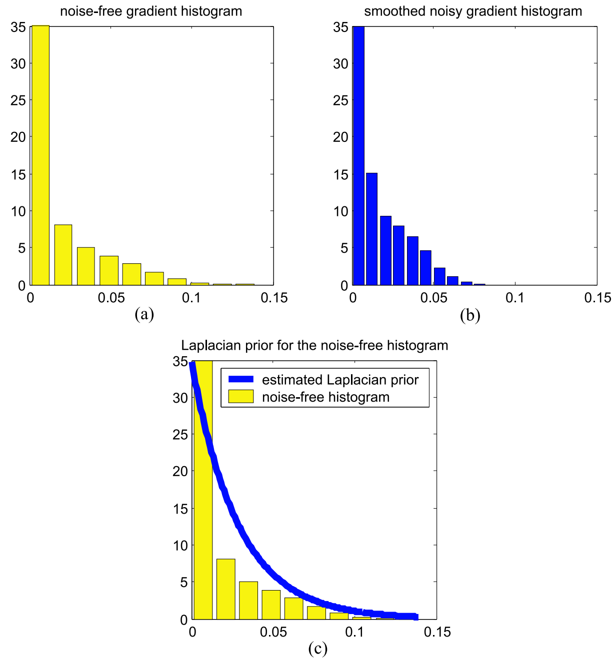

Fig. 2. (a) The histogram of the ideal, noise-free gradient magnitude from Fig. 1 (b), normalized to unit area. (b) The histogram of the noisy gradient magnitude from Fig. 1(d), normalized to unit area. (c) Laplacian prior for the noisy-free gradient magnitude, estimated from the noisy data (b).

where A is a normalizing constant. By choosing

$$A = 1/(1 - P(H_1|0)) \tag{8}$$

we ensure that $g(0) = 1$, because $P(H_1|0) = \min_x P(H_1|x)$ (see Fig. 3) and thus $1 - P(H_1|x)$ peaks at $x = 0$. The Bayes' rule yields

$$P(H_1|x) = \mu\eta/(1 + \mu\eta) \tag{9}$$

where $\mu = P(H_1)/P(H_0)$ and $\eta = p(x|H_1)/p(x|H_0)$ is the likelihood ratio. For the Laplacian prior (4) we have [6]: $\mu = \exp(-\lambda\sigma_n)/[1 - \exp(-\lambda\sigma_n)]$. Under the assumption of additive white Gaussian noise, the densities of the noisy gradients $p(x|H_0)$ and $p(x|H_1)$ follow from convolving the corresponding densities of the noise-free gradients with a normal density $N(0, \sigma_n)$. Let $Lap_c(x; \lambda)$ denote the central part of the Laplacian (for $x < \sigma_n$) and let $Lap_t(x; \lambda)$ denote the tails[1]. We have: $p(x|H_0) = Lap_c(x; \lambda) \star N(0, \sigma_n)$ and $p(x|H_1) = Lap_t(x; \lambda) \star N(0, \sigma_n)$. For more details on deriving η see [6], where the expressions are also analyzed for

[1] $Lap_c(x; \lambda) \propto Lap(x; \lambda)$ for $|x| < \sigma_n$ and $Lap_c(x; \lambda) = 0$ otherwise. $Lap_t(x; \lambda) = 0$ for $|x| < \sigma_n$ and $Lap_t(x; \lambda) \propto Lap(x; \lambda)$ otherwise.

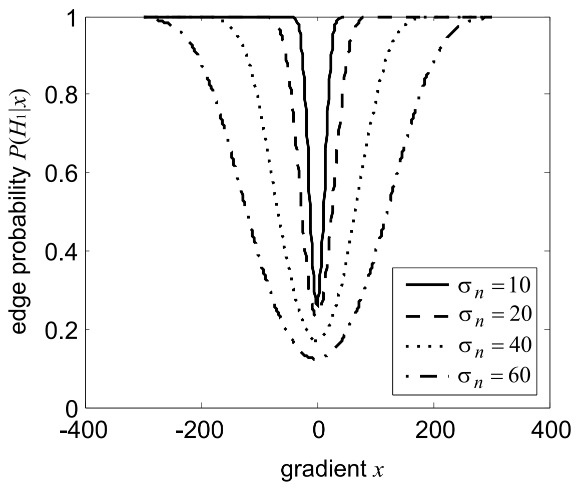

Fig. 3. Edge probability under the Laplacian prior with a fixed parameter λ (same as in Fig. 2, $\lambda = 0.0196$) and for different noise levels σ_n.

the generalized Laplacian prior. Full analytical formulation for the Laplacian prior is in [14]. For the diffusion experiments in this paper, we have calculated these convolutions numerically and stored η as a look-up-table.

3. RESULTS

We compared the proposed diffusivity function to the ones given in Section 1, for which we optimized the diffusion performance by estimating the contrast parameter k according to [2]. For our method, the estimation of noise standard deviation σ_n is needed, and we use a median absolute deviation (MAD) estimator [12], where σ_n is estimated by the gradient MAD divided by 0.6745. All the reference functions belong to the class of the so-called *backward-forward* schemes [8], which are designed to smooth weak edges and enhance the strong ones. Our experiments showed that the new probabilistic diffusivity functions under both considered priors automatically (i.e., without any parameter adjustments) fit in the cluster of the reference edge-stopping functions with optimized values of the contrast parameter. Fig. 4 illustrates this for the case of the Laplacian prior and for the test image in Fig. 1(c). Similar behavior is observed on all the other tested images: the new function fits well in the cluster of the reference backward-forward diffusivities.

Fig. 4(a) demonstrates that the new probabilistic diffusivity function under the Laplacian prior resembles in shape the *Le Clerc* diffusivity. However the negative peaks of the proposed function occur usually at larger gradient magnitudes than for the *Le Clerc* one (see Fig. 4(b)), which indicates stronger edge enhancement capability. For the generalized Laplacian prior (not displayed) the shape is similar,

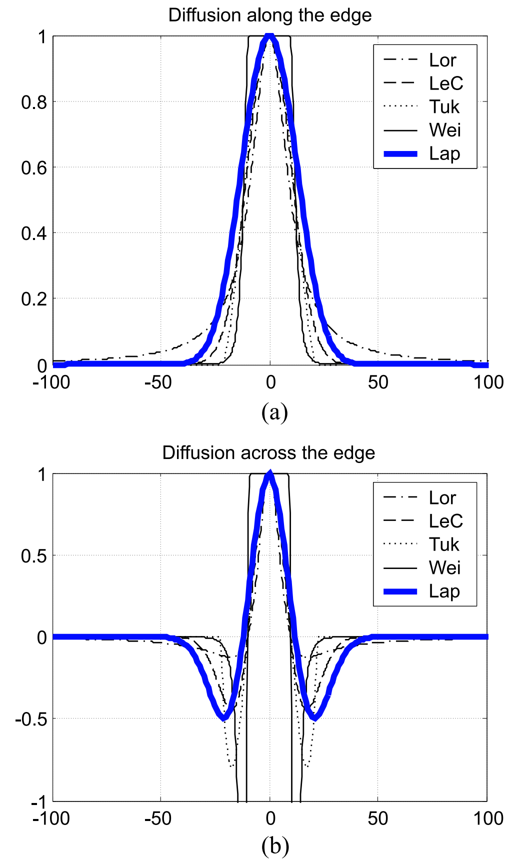

Fig. 4. (a) Diffusivity functions $g(x)$ (diffusion along side the edge) and (b) functions $[xg(x)]'$ (diffusion across the edge) for the image in Fig. 1.c. The proposed probabilistic diffusivity under the Laplacian prior (*Lap*) is displayed in comparison to the Lorentzian (*Lor*), LeClerc (*LeC*), Tukey bi-weight (*Tuk*) and Weickert (*Wei*) diffusivities.

but the negative peak in $[xg(x)]'$ is often even larger and occurs at a larger gradient. These properties are usually desired in the diffusion schemes, as they indicate a quick smoothing of the nearly uniform areas while maintaining and enhancing the strong edges.

Fig. 5 illustrates the diffusion results for four test images. The Lorentzian diffusion was terminated sooner which is mainly due to its weaker enhancement capabilities. The three other commonly used methods retain some artifacts. The latter might be resolved by diffusing longer, however this can result in the removal of important edges as well. The end-results under the generalized Laplacian are slightly better than using a simpler, Laplacian prior. Both of the proposed functions perform well with respect to the reference ones.

4. DISCUSSION AND CONCLUSIONS

We propose a Bayesian formulation of the diffusivity function by imposing a prior on noise-free gradient. All other probabilistic approaches give stochastic formulations for *conductivity* and *not* for the diffusivity function of the gradient.

Full analytical formulation of our diffusivity function under the analyzed Laplacian prior is available (the necessary expression for the likelihood ratio can be found in [14]) but is quite cumbersome. For the generalized Laplacian prior the situation is even more complex. It would be useful to examine other priors, like the Cauchy distribution.

Our current parameter estimation method assumes that the noise in the gradient image is additive white Gaussian, which does not necessarily hold for all possible gradient formulations even if the noise in the input image is white Gaussian. The validity of this assumption should be checked for different gradient formulations and if necessary, the parameter estimation needs to be updated accordingly. If we calculate the gradient using the orthogonal wavelet transform with the Haar wavelet (which is a moving difference operator) our assumptions are ensured for images with additive white Gaussian noise.

The performance of the proposed diffusivity functions under both analyzed priors compares favorably to the reference functions. The proposed diffusivity formulation automatically satisfies the well accepted criteria for the goodness of edge-stopping functions: on all the tested images, our new function lies in the middle of the cluster of a number of reference functions for which the contrast parameter was optimized to produce the best results. The shape of the proposed probabilistic diffusivity under the Laplacian prior resembles to the diffusivity based on the *Le Clerc* error-norm, while allowing a stronger enhancement of the edges. The diffusion results on both synthetic and natural scene color images are encouraging.

5. REFERENCES

[1] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, 1990.

[2] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 421–432, 1998.

[3] J. Weickert, *Anisotropic diffusion in image processing*, ECMI Series. Teubner-Verlag, Stuttgart, Germany, 1998.

[4] S.R. Arridge and A. Simmons, "Multi-spectral probabilistic diffusion using bayesian classification," in *Proc. Scale-Space '97*, 1997, pp. 224–235.

[5] Y. Bao and H. Krim, "Smart nonlinear diffusion: A probabilistic approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26(1), pp. 63–72, 2004.

[6] A. Pižurica and W Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Processing (In press)*.

[7] P. Mrázek, J. Weickert, and G. Steidl, "Correspondences between wavelet shrinkage and nonlinear diffusion," in *4th International Conference on Scale Space Methods in Computer Vision (Scale-Space 2003)*, L.D. Griffin and M. Lillholm, Eds., Isle of Skye-UK, 2003, vol. 2695 of *Lecture Notes in Computer Science*, pp. 1001–116, Springer.

[8] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Weber, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides," *SIAM Journal on Numerical Analysis*, vol. 42, no. 2, pp. 686–713, 2004.

[9] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. and Machine Intel.*, vol. 11, no. 7, pp. 674–693, 1989.

[10] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, Lausanne, Switzerland, 1996.

[11] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inform. Theory*, vol. 45, pp. 909–919, Apr. 1999.

[12] S. Mallat, *A wavelet tour of signal processing*, Academic Press, London, 1998.

[13] D. Androutsos, K.N. Plataniotis, and A.N. Venetsanopoulos, "Distance measures for color image retrieval," in *ICIP '98*, Chicago, Illinois - USA, 1998, pp. 770–774.

[14] A. Pižurica and W Philips, "On least square estimators under laplacian mixture priors," Technical Report, Ghent University, March 2005 available at http:\\telin.ugent.be\∼sanja.

Fig. 5. Original images (top) and the results using - top to bottom - Lorentzian ($T_{end} = 27.30$), Le Clerc ($T_{end} = 165$), Weickert ($T_{end} = 165$), Tukey-biweight ($T_{end} = 165$), novel with Laplacian prior ($T_{end} = 165$) and novel with generalized Laplacian prior ($T_{end} = 165$) diffusion functions. (a) and (b) synthetic images,(c) jelly beans and (d) natural scene image.