Orthogonal Base Functions on a Discrete Two-Dimensional Region

Dr. ir. W. Philips

Department of Electronics and Information Systems (ELIS)
Medical Image and Signal Processing group (MEDISIP)
Institute Biomedical Technology (IBITECH)

ELIS Technical Report DG 91-20
January, 1992

This work was awarded the "Scientific Award Barco N.V. 1991"

Orthogonal Base Functions on a Discrete Two-Dimensional Region

Dr. ir. Wilfried Philips
Department of Electronics and Information Systems
St.-Pietersnieuwstraat 41
B9000 Gent
Belgium
tel: 32-9-264.33.85
fax: 32-9-264.35.94
E-Mail: philips@elis.rug.ac.be

January 10,1992
Revised: November 2, 1993

Abstract

Region Oriented Transform Coding (ROTC) is a new and promising image data compression method. This technique segments an image into different areas in which the image intensity is slowly varying. The image intensity in each segment is represented by a weighted sum of orthogonal base functions. Straightforward implementations of ROTC methods require significantly more computations than classical Block Transform Coding (BTC) methods. This report presents several fast algorithms that reduce the computational complexity of ROTC methods. The fast algorithms deal with the computation of both the orthogonal base functions and the coefficients of the weighted sum.

The report also presents some new theoretical results concerning the independence of the 'starting functions' from which the orthogonal functions are derived. One type of ROTC base, that allows a very fast ROTC implementation is discussed in more detail. It is shown that this base is 'weakly' separable and that it can be obtained from a set of one-dimensional orthogonal functions. Consequently, the corresponding ROTC coefficients can be obtained almost as fast as in classical BTC methods. The spatial properties of the base functions are described.

Contents

List of Figures

List of Tables

1 Introduction

The compression of still images is an important topic in image processing. Its aim is to reduce the number of bits needed to represent such an image, so that it can be stored or transmitted efficiently. A first application is the transmission of photographs over telephone lines or radio channels (e.g., mobile fax). The time required to transmit such an image is significant, since normally low-bandwidth channels are used, but it can be reduced considerably by using appropriate compression techniques. Other important applications of image compression are the storage of color images on floppy disks and electronic encyclopedia, e.g., on the new CDI- interactive video systems.

Block Transform Coding (BTC) methods [6, 1, 3] are among the most efficient compression techniques used today. They divide a single image into smaller rectangles. The image intensity in each of these rectangles is represented as a weighted sum of (usually orthogonal) base functions. The *coefficients* of the base functions in this sum are used to represent the image intensity[1]. BTC methods do not allow an exact image reconstruction since, for efficiency reasons, some of the coefficients are neglected and the others are quantized. However, the introduced errors can be kept as small as desired (at the expense of a lower compression). In practice, for an acceptable subjective image quality, the obtainable compression factor is limited to about 10[8]. This is because BTC methods are not adapted to the space varying characteristics of natural images. For example, they are not very efficient when it comes to reproducing intensity edges in the image. This problem has been recognized very early in the history of image coding. It lead to the idea that low frequency information, as found in large quasi-stationary image regions, and high frequency information such as image edges should be represented independently [9]. This basic insight was developed further into two distinct directions:

1. Sub-band Coding techniques [10, 11], in which filters separate different frequency bands. These bands are then coded independently. The paper of Graham [9] is an early example of this.

2. Region Oriented Transform Coding (ROTC) methods [7, 8, 13, 14, 15, 16, 17], which divide the image into regions or 'segments' of homogeneous grey-value, separated by intensity edges. The image value in

[1] The term 'coefficients' will be used in this meaning, unless explicitly noted.

the interior of the region is again represented as a combination of base functions.

In this report only ROTC methods are discussed. These techniques offer a very good visual image quality, even at high compressions where BTC methods fail (see section 3). The main topic is the representation of the image texture inside the regions and not the segmentation problem, which is amply discussed in literature [7, 18, 19].

For best results, the ROTC base should be orthogonal (see section 3). Since such an orthogonal base depends on the shape of the image segment to be coded, it has to be generated for every single region. Up till now, this was done by applying the Gram-Schmidt Orthogonalization (GSO) procedure [36, p. 218] to an ordered, non-orthogonal *starting base*, which normally consists of the polynomials[2] 1, x, y, x^2, xy, y^2, ... or the cosines 1, $\cos x$, $\cos y$, $\cos(2x)$, $\cos x \cos y$, $\cos(2y)$, ... The number of calculations involved, limits the practical usability of ROTC methods based on GSO. Section 5 presents an orthogonalization algorithm that is much faster than the GSO technique. Several versions of the algorithm are presented. These differ in the order in which the orthogonalizations are performed and produce different bases, which are useful in different circumstances. The algorithm works not only for polynomial and cosine base functions but also for more general bases, consisting of so-called warped polynomial functions [24]. Hence, it is applicable to the major types of starting bases used in practice.

GSO requires that the starting functions are linearly independent. Unless this can be guaranteed, numerical independence tests must be used in the orthogonalization procedure in order to reject dependent functions from the starting set. This problem has not (yet) been addressed in literature. Section 7 presents a starting base that is guaranteed to be independent. If this starting base is orthogonalized in a specific order, a complete orthogonal set of functions having very desirable properties is obtained. For instance, section 8 will show that these base functions are 'weakly' separable. This implies that not only the base functions, but also the coefficients in the weighted sum can be obtained by mostly one dimensional computations. This reduces the computational complexity of ROTC methods considerably. The spatial properties of the base functions, which are important for the theoretical study of ROTC, are summarized.

The set of independent starting functions just mentioned is not unique. Depending on the application, other starting sets, possibly ordered in differ-

[2]The reason for this choice of starting base is explained later.

ent ways, might be preferred. In section 6, non-numerical tests are presented for some commonly used orderings. These tests can determine whether or not a function in the starting set is linearly dependent on its predecessors in the set. The tests are not conclusive for every starting function, but they eliminate a lot of unnecessary calculations. If the image regions are restricted to a certain class, the non-numerical tests are applicable to all starting functions.

The ROTC methods described above treat the image segmentation and the representation of the image region as separate problems. Therefore, the segmentation is not necessarily optimal for the coding problem. Better results can be obtained by combining the segmentation and the coding step. Since this increases the total complexity of the ROTC methods, the need for fast algorithms is felt again. The fast orthogonalization algorithms presented in this report allow the recursive computation of both the orthogonal base functions and the coefficients for a given region from the base functions and coefficients of the subregions of this region. This is significantly faster than simply computing the base functions and the coefficients from scratch (which seems to be the only alternative). This subject will however not be treated in full. It will be discussed in another report.

2 Definitions and Notations

Let $\Omega \subset R^2$ be an arbitrary but finite set of points in the plane and let $\#\Omega$ denote the number of elements in Ω. In practice, Ω could be the set of all points on a rectangular sampling grid inside a closed contour C.

A function $f(x,y)$ on Ω associates a value $f(x,y)$ with every pair $(x,y) \in \Omega$. The values of f are irrelevant for any $(x,y) \notin \Omega$. Since $\#\Omega$ values have to be given to completely specify a function f, f can also be considered as a vector of dimension $\#\Omega$. Two functions are equal (on Ω) if the vectors of their values on Ω are identical. A set of functions $f_{m,n}(x,y)$ is *separable* on a set Ω if there exist functions $g_m(x)$ and $h_n(y)$ such that $f_{m,n}(x,y) = g_m(x)h_n(y)$, for all $(x,y) \in \Omega$. A set of functions $f_{n,m}(x,y)$ is *weakly separable* on a set Ω, if $f_{m,n}(x,y) = g_m(x,y)h_{m,n}(y)$ for all m,n and for all $(x,y) \in \Omega$. Of course, the roles of x and y can be interchanged.

Let $w(x,y)$ be a *weight function*. This is a function strictly positive on Ω (possibly $w(x,y) \equiv 1$). The *inner product* on Ω of two functions $\phi_p(x,y)$ and $\phi_q(x,y)$ is defined as $\ll \phi_p, \phi_q; w \gg_\Omega = \sum_\Omega w(x,y)\phi_p(x,y)\phi_q(x,y)$. If the weight function or the set are omitted in this notation, they default to

$w(x,y)$ and Ω. The functions ϕ_p and ϕ_q are orthogonal if $\ll \phi_p, \phi_q \gg = 0$. The *norm* of a function ϕ is defined as $\|\phi\| = \ll \phi, \phi \gg^{1/2}$.

The *set product* of $\Omega_x \subset R$ and $\Omega_y \subset R$ is defined as $\Omega_x \times \Omega_y = \{(x,y) : x \in \Omega_x, y \in \Omega_y\}$. A set Ω is *separable*, if it can be written in the form $\Omega = \Omega_x \times \Omega_y$. A *rectangular set* is a separable set of the form $\{x_i = x_0 + i\Delta_x : 0 \le i < N\} \times \{y_j = y_0 + j\Delta_y : 0 \le j \le M\}$.

If there is at least one point $(x,y) \in \Omega$ with $y = y_i$, the set $R_i = \{(x,y) \in \Omega : y = y_i\}$ is called a *row* of Ω. Similarly, $C_i = \{(x,y) \in \Omega : x = x_i\}$ is a *column* of Ω, if it is not empty.

Two discrete sets Ω and Ω' are said to have the same *structure* if there exist functions $f(x)$ and $g(y)$ such that for each point $(x,y) \in \Omega$, there exists a *unique* point $(x', y') \in \Omega'$ with $x = f(x')$ and $y = g(y')$. It is easily seen that 'has the same structure as' is an equivalence relation. Therefore, the structure of a region Ω can be meaningfully defined as the equivalence class to which Ω belongs. Loosely speaking, the structure of a set depends only on the way it is partitioned into rows and columns, but not on the precise location of these rows and columns. For instance, the structure of a set is not altered by a permutation of its rows or columns.

The function $p(x,y) = \sum_{i=0}^{M} \sum_{j=0}^{N} c_{i,j} x^i y^j$ is a bivariate polynomial of x-degree M and y-degree N. Its *total degree* is the degree of the univariate polynomial $p(x,x)$. A polynomial $q(x)$ is said to be of *precise degree* n, if it is of degree n and if its highest degree coefficient does not vanish. If $p(x,y)$ is a bi-variate polynomial then $p\Big(f(x), g(y)\Big)$ is called a warped polynomial. This terminology stems from the fact that such a function can be transformed into a polynomial by non-linearly scaling ('warping') both axes.

3 Block Transform Coding and ROTC

In Transform Coding (TC), an image is divided into a number of regions in which the image intensity is coded independently. The regions are rectangular blocks in Block Transform Coding (BTC), while more general (non-rectangular) regions are considered in ROTC. In each region Ω, the image intensity function $f(x,y)$ is represented as a weighted sum of base functions $P_{m,n}(x,y)$:

$$f(x,y) \approx \tilde{f}(x,y) = \sum_{m,n} A_{m,n} P_{m,n}(x,y). \tag{1}$$

The mapping that converts the image function $f(x,y)$ into the coefficients $A_{m,n}$ is called an image transform.

Usually, the quality of the approximation $\tilde{f}$ is measured by the Euclidean distance e^2 between f and $\tilde{f}$:

$$e^2 = \sum_{(x,y)\in\Omega} \left(f(x,y) - \tilde{f}(x,y)\right)^2. \tag{2}$$

or by the Mean Squared Error (MSE) $e^2_{\mathrm{mse}} = e^2/\#\Omega$ and the coefficients $A_{m,n}$ are chosen such that the MSE is minimized. Note that if the set of base functions $P_{m,n}$ in eq. (1) is complete, i.e., if it consists of $\#\Omega$ linearly independent functions, the minimal e^2 reduces to 0.

In general, the coefficients $A_{m,n}$ that minimize the MSE are obtained by solving a set of coupled equations. The number of equations equals the number of base functions in eq. (1) and can be very large. Also, the equations are usually ill-posed.

If the base functions $P_{m,n}(x,y)$ are orthonormal (i.e., orthogonal and normalized), the coupled nature of the equations disappears and the coefficients $A_{m,n}$ are found from

$$A_{m,n} = \ll f, P_{m,n} \gg = \sum_{(x,y)\in\Omega} f(x,y) P_{m,n}(x,y). \tag{3}$$

In addition, the squared reconstruction error is then given by

$$e^2 = \sum A^2_{m,n}, \tag{4}$$

where the sum runs over the neglected coefficients.

Therefore, the use of orthogonal base functions presents the following advantages:

1. The coefficients in the weighted sum are obtained independently, with fewer and numerically more stable computations.

2. Error controlled approximation algorithms, for which the number of terms in eq. (1) is variable, can be easily implemented. Since the squared reconstruction error for any given number of terms can be calculated from eq. (4), such algorithms can simply add additional coefficients until e^2 falls below a certain threshold. Each new coefficient can be calculated from eq. (3), while the coefficients that have already been found remain unchanged. For non-orthogonal base functions, the set of equations has to be solved all over again for every single coefficient that is added and the squared error is not so easily found.

In practice, the coefficients $A_{m,n}$ are quantized, i.e., represented with limited accuracy, and coded, e.g., by using variable length code words. This leads to an additional error in the reconstructed image, but at the same time the number of bits needed to store the coefficients is considerably reduced. Quantization and coding methods have been amply discussed in literature [6, 3, 1]. The techniques used are the same for both ROTC and BTC and will therefore not be treated in full here.

The Compression Factor CF is defined as $CF = n/(n_c + n_o)$, where n is the number of bits in the original image, n_c is the number of bits needed for the quantized and coded coefficients and n_o is the number of bits for any other information needed to reconstruct the image. An example of such overhead information is the description of the image regions in ROTC.

As mentioned above, BTC and ROTC differ in the way the image is divided into regions. The fact that non-rectangular regions are used in ROTC has a significant impact on the computational complexity and on the obtainable subjective image quality and coding efficiency. In order to compare ROTC and BTC, the main features of both methods are now described.

3.1 Block Transform Coding [6, 1, 3]

The image regions are equally sized rectangles ('blocks') $\Omega = \Omega_x \times \Omega_y$, where $\Omega_x = \{0, 1, \ldots, M\}$ and $\Omega_y = \{0, 1, \ldots, N\}$ [3]. Suppose that $\{p_m(x) : 0 \leq m \leq M\}$ and $\{q_n(y) : 0 \leq n \leq N\}$ are complete orthonormal bases on Ω_x and Ω_y respectively. It can be shown that $\{P_{m,n}(x, y) = p_m(x)q_n(y) : 0 \leq m \leq M, 0 \leq n \leq N\}$ is a complete orthonormal base for the region Ω. This means that the BTC base functions can always be chosen separable.

For non-separable base functions, M^2N^2 multiplications and additions are needed to compute all MN coefficients by eq. (3). This is because the computation of a single coefficient requires $\#\Omega = MN$ multiplications and additions. If the base functions are separable, the number of operations is considerably smaller, because then eq. (3) can be rewritten as

$$A_{m,n} = \sum_{x=0}^{M} p_m(x) \sum_{y=0}^{N} q_n(y) f(x, y), \tag{5}$$

or in matrix notation:

$$A = P^t F Q. \tag{6}$$

[3]Different origins are used for each block.

The dimensions of the matrices A, P, F and Q are $M \times N$, $M \times M$, $M \times N$ and $N \times N$ respectively. The matrices' components are:

$$\begin{aligned} F_{x,y} &= f(x,y) && (7) \\ P_{x,m} &= p_m(x) && (8) \\ Q_{y,n} &= q_n(y) && (9) \end{aligned}$$

Two matrix multiplications FQ and $P^t(FQ)$ are needed. The first requires MN^2 multiplications and additions and the second M^2N. Hence, the total complexity is $MN^2 + N^2M$, which is much smaller than the corresponding number for the non-separable case.

For specific image transforms, such as the widely used Discrete Cosine Transform, there exist fast algorithms [34] that further reduce the computational complexity of eq. (5).

From the preceding discussion, it is clear that block transform methods are computationally very efficient. Unfortunately, because the image blocks are coded independently, discontinuities can arise in the reconstructed image near the block boundaries. Such *block effects* become very objectionable at high compressions [31]. Furthermore, if the number of coefficients per image block is low, ringing effects can occur near edges in the image, because these cannot be accurately represented by only a few (low-pass) base functions. Consequently, at high compressions the reconstructed image looks very blurred and unnatural.

These problems can be partially alleviated by the use of variable block size methods, which use a smaller block size in regions of high detail [30, 33]. In large homogeneous areas a lower resolution is acceptable and therefore a larger block size is chosen there. Since the resolution is increased near the image edges, ringing effect are less likely to occur. Variable block size methods are not free from block effects, however and they do not increase the obtainable compression. An important research topic connected with this field is the recursive computation of the coefficients from an image block from the coefficients of its sub-blocks [30, 25].

3.2 Region Oriented Transform Coding

In Region Oriented Transform Coding (ROTC) [7, 8, 14, 15, 16, 17], the idea of a rectangular image division is abandoned altogether. Instead, the image is partitioned into non-rectangular regions in which the image intensity is homogeneous (i.e., slowly varying) and can be represented efficiently by a small number of base functions.

Given the method chosen to represent the image intensity inside the image regions, the segmentation should be such that an optimal compression factor is obtained at a given image quality. Conversely, the desired compression factor could be specified (e.g., if channel bandwidth is limited). Then, the segmentation process should optimize the image quality as determined by some (necessarily numerical) error criterion [32]. However, this approach requires that the segmentation and coding are considered together, as a combined optimization problem. Although such optimal segmentation methods have been investigated [7], they are generally considered to be too complex to be used in practice [17].

It should be noted that the recently developed method for recursive computation of BTC coefficients [25] can be extended for ROTC coefficients. The separable base of section 8 is especially suitable for such recursive computations. Due to time limitations, these techniques will not be discussed in this report.

Generally however, the segmentation and coding steps are performed independently of each other. In that case the segmentation method should produce maximal regions, not containing edges or other high-detail image features. In other words, region boundaries should coincide with image edges. Normally, some smoothness criterion guides the image partitioning. Most texture segmentation methods described in literature can be used. Such methods are based on e.g., region-growing [7], Voronoi polygons [18] or multiresolution techniques [19].

Again, the image intensity in the interior of the image regions is represented by the coefficients of a weighted sum of orthogonal base functions. Since these functions depend on the shape of the image region, a different base must be generated for every single region. Such an orthogonal base can be obtained by orthogonalizing a non-orthogonal *starting base*, e.g., by the Gram-Schmidt Orthogonalization (GSO) procedure. In practice, the fast algorithms of section 5 would be preferred. Because the slowly varying image intensity inside a region can be represented accurately using low degree polynomials or low-frequency cosines, the starting bases usually consist of these types of functions.

The properties of the orthogonal base depend on the type of starting functions and the orthogonalization order chosen. At present, not much is known about the influence of these choices on the properties of the orthogonal base. The knowledge of these properties is very important. For instance, the spatial distribution of the errors resulting from coefficient quantization or deletion can be estimated from the local amplitudes of the base functions. Also, the choice of starting functions can significantly influence the obtainable data compression. In the one dimensional case, this was already demonstrated on an ECG data compression application [24]. These one-dimensional results can easily be extended to the separable ROTC base of section 8. For instance, this allows the the errors introduced by coefficient quantization to be studied. It also brings the design of bases having specific properties, desirable for certain applications within reach. Such extensions will be treated in a later report. Section 9, describes the properties of the zero curves of one type of ROTC base in considerable detail. This demonstrates the ease with which one-dimensional results can be applied to the ROTC base.

3.3 Major Differences Between ROTC and BTC

There are a number of very important differences between region oriented transform coding and classical block transform coding:

1. ROTC requires a segmentation step, which is not present in BTC. Also, the *region information*, i.e., the description of the segmented image, must be coded along with the coefficient data, since it is needed to reconstruct the image. Normally this is done by coding the contours that separate the image regions, e.g., by chain codes or Fourier descriptors [35]. There is no need for this description to be error-free, as long as the same contours are used at the image coder and decoder.

2. The actual image intensity inside the different regions is represented much more efficiently in ROTC; since no high-frequency information has to be represented, a faithful reproduction can be obtained with less coefficients.

3. The subjective quality of ROTC-reconstructed images is much better than that of BTC-images, especially at high compressions [17]. The visually very important image contours are always very crisp in ROTC, since these are represented independently from the texture in the image regions. Objectionable block effects are almost never present. Discontinuous changes in image intensity between adjacent regions still occur, but normally they coincide with image edges, where they are less visible because of the masking effect.

4. ROTC methods require significantly more computations. Especially the generation of the orthogonal base is very time-consuming, since

it has to be repeated for every image region. The calculation of the ROTC coefficients is also slower, because the base functions are not separable in general (this problem is largely solved in section 8) and because there exist no fast DCT-like techniques.

The following small example illustrates some of these differences. An original image is shown in figure 1 (top image). In BTC, this image is divided into 8 image blocks. In each of these, the image intensity is represented by e.g., the base functions of the two-dimensional DCT, some of which are shown in figure 2. The reconstruction that results when 16 base functions are used in each block, is shown in figure 1 (middle image). In ROTC, the original image would be treated as two different regions, namely the ball itself and its background. The image at the bottom of figure 1 shows the ROTC reconstruction. The foreground region in that figure is represented by 8 ROTC coefficients and only a single coefficient is used for the background region. The base functions used for the 'foreground region' are shown in fig. 3.

Note that the 9 coefficient ROTC-approximation is visually much more pleasing than the BTC approximation which requires 128 coefficients. This is because in ROTC, the image is low-pass in both of the image regions. Therefore no ringing effects can occur. Also, the very disturbing block distortion in BTC, is totally absent in ROTC. An even better ROTC approximation would be obtained if the image was divided into three segments.

Of course, the boundary between the two regions must also be coded and this must be taken into account when comparing both methods. A quantitative comparison of ROTC and BTC methods is beyond the scope of this report, which concentrates mainly on computational and theoretical aspects of ROTC methods. For such information, the reader is referred to the appropriate literature.

4 Orthogonalizing the Starting Base

The orthogonal ROTC bases are obtained by orthogonalizing a set of non-orthogonal starting functions. Up till now, this has been done by the procedures of either Gram-Schmidt (GSO) [36, p. 218] or Householder [36, pp. 194–211].

GSO is now briefly described, not only for completeness, but also because the fast orthogonalization algorithms of section 5 generate the same base as GSO. Let $\mathcal{B} = \{\phi_0, \ldots, \phi_N\}$ be the set of starting functions. In a lot of

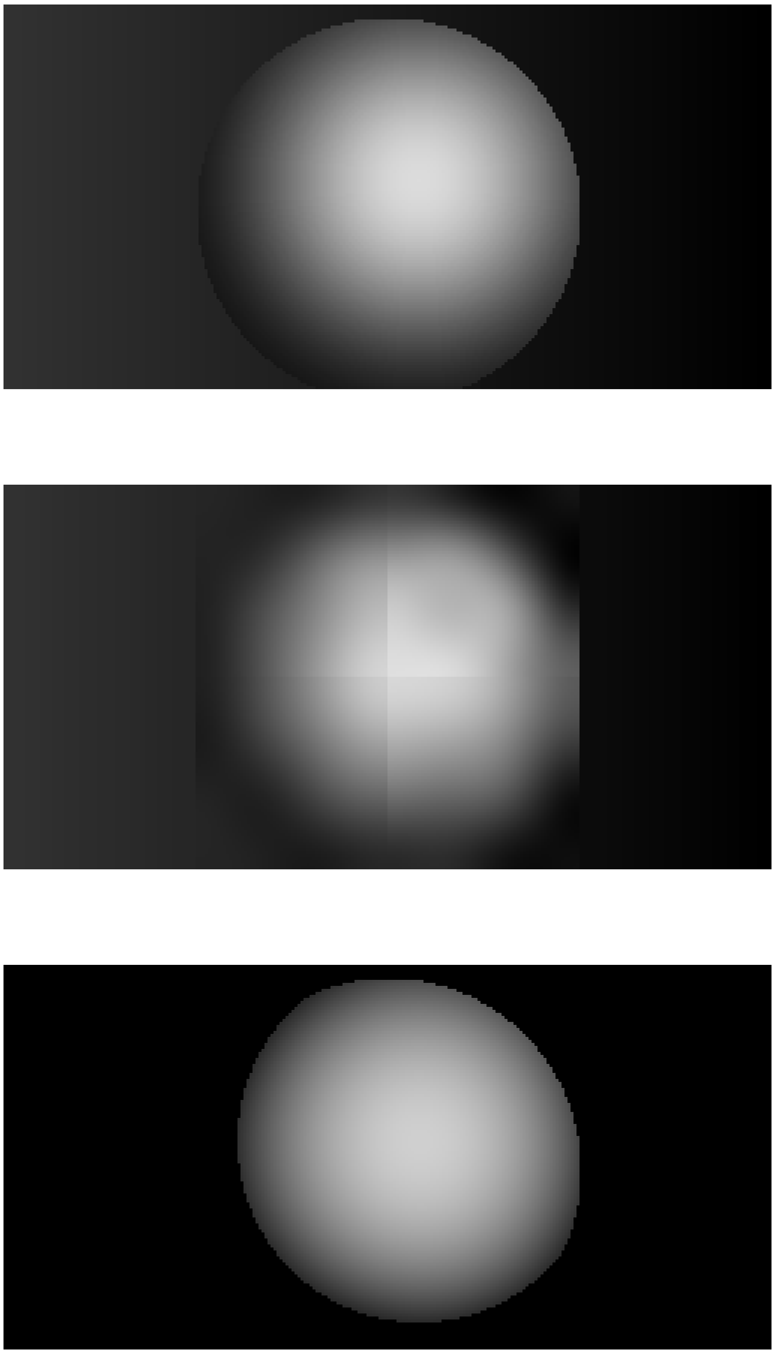

Figure 1: Example. a: The original image. b: DCT reconstruction, using 8 square blocks. Each block is represented by 16 coefficients. c: ROTC reconstruction. Eight coefficients are used for the foreground region and only a single one for the background.

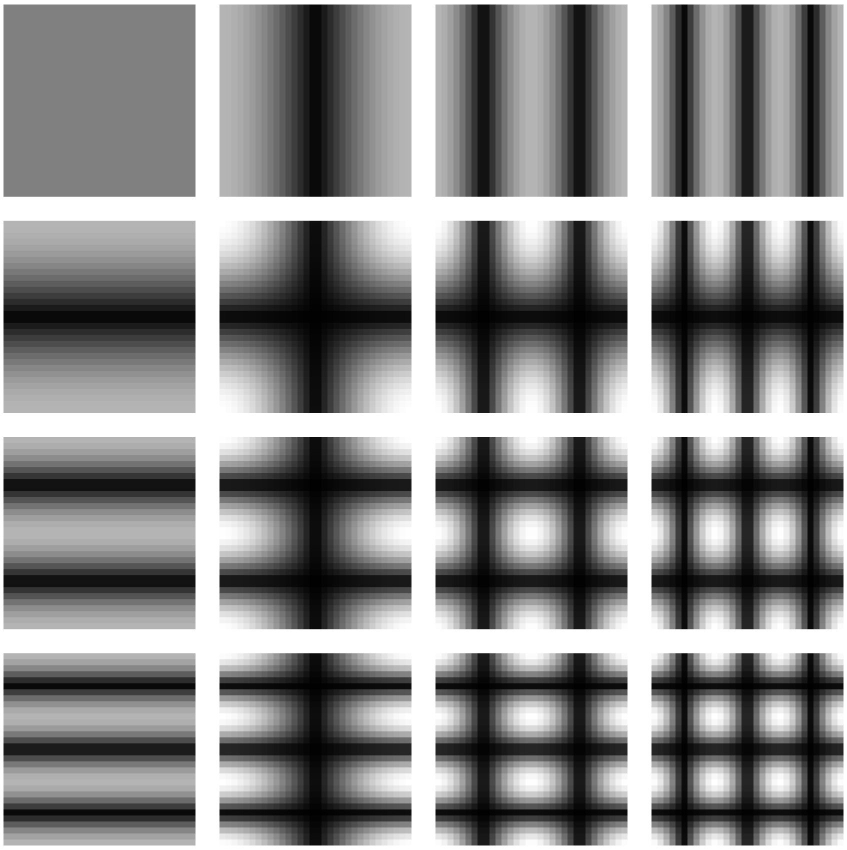

Figure 2: The DCT base functions used to compress the image of fig. 1.

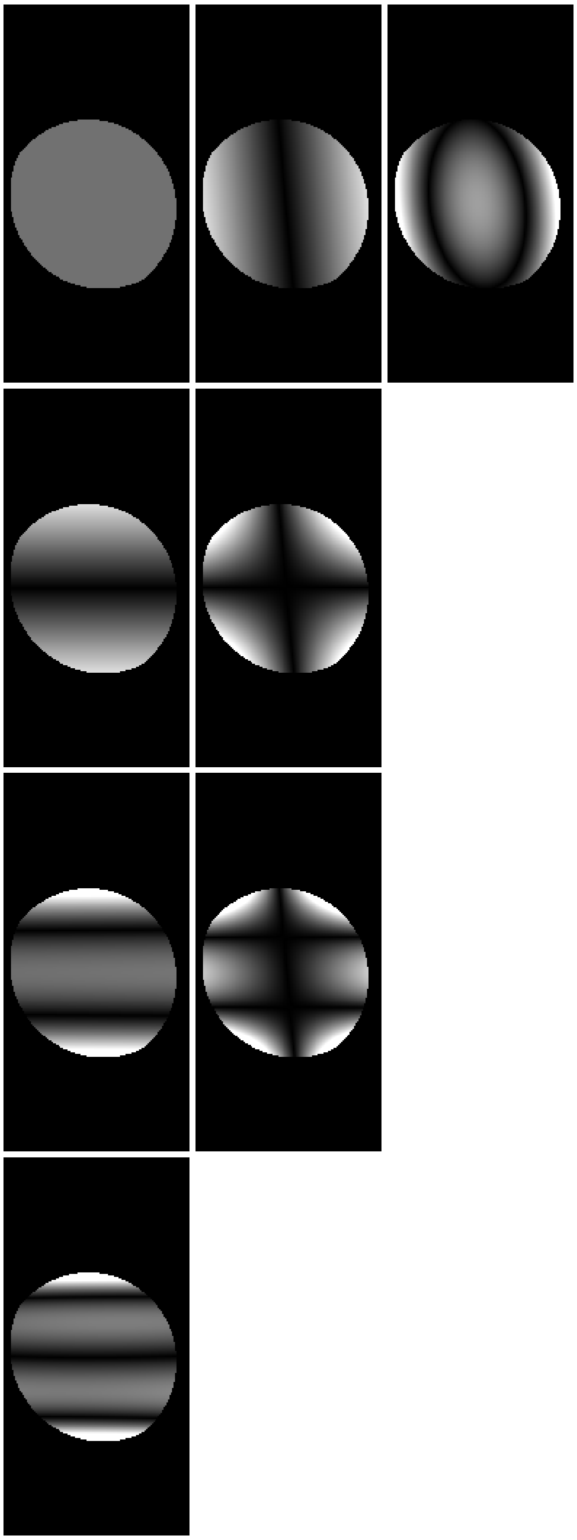

Figure 3: The ROTC base functions used to compress the image of fig. 1. The figure has been rotated over 90 degrees.

applications, only a small number of base functions is required. For instance, since the image intensity functions to be coded are normally slowly varying in ROTC, a small number of low-degree polynomials could be sufficient. Therefore, $\mathcal{B}$ is not assumed to be complete, i.e., it is not required that on Ω any function can be represented as a linear combination of the functions in $\mathcal{B}$.

First, assume that the starting functions in $\mathcal{B}$ are independent on Ω. GSO is a sequential procedure. It produces the functions $P_j = \sum_{i=0}^{j} c_{j,i}\phi_i$, $j = 0,\ldots,N$, where $c_{j,j} \neq 0$. Each P_j is orthogonal to the space S_{j-1} spanned by the predecessors ϕ_i, $i < j$ of ϕ_j. This space is also spanned by $P_0,\ldots P_{j-1}$.

The procedure starts by putting $P_0 = \phi_0/\|\phi_0\|$. The second orthogonal function P_2 is a combination of ϕ_0 and ϕ_1: $N_1 P_1 = \phi_1 - \gamma_{1,0}P_0$. The requirement that P_1 must be orthogonal to P_0 leads to $\gamma_{1,0} = \ll \phi_1, P_0 \gg$, while the normalization factor N_1 is chosen such that $\|P_1\| = 1$. Note that $\gamma_{1,0}P_0$ can be interpreted as the projection of ϕ_1 on the linear manifold spanned by P_0. In general, if n orthogonal functions $P_0,\ldots,P_{n-1}$ have been found, P_n is obtained by the following equation in which N_n is a normalization factor:

$$N_n P_n = \pi_n = \phi_n - \sum_{i=0}^{n-1} \ll P_i, \phi_n \gg P_i. \tag{10}$$

As for the case $n = 1$, $N_n P_n$ is the difference between ϕ_n and its projection on the space S_{n-1}. Eq. (10) shows that P_n is a combination of $\phi_0,\ldots,\phi_n$ only and that the coefficient of ϕ_n in this equation is always non-zero. Together, the generated functions form an orthogonal base $\mathcal{B}^* = \{P_0,\ldots,P_N\}$ for the space S_N.

Equation (10) is conceptually simple, but turns out to be not very stable numerically. A rearrangement of the calculation, known as Modified Gram-Schmidt Orthogonalization (MGSO), yields a much sounder procedure. In MGSO, eq. (10) is replaced by

$$\begin{aligned} \pi_{n,0} &= \phi_n - \ll P_0, \phi_n \gg P_0 \\ \pi_{n,i} &= \pi_{n,i-1} - \ll P_i, \pi_{n,i-1} \gg P_i \quad \text{for } 0 < i \leq n \\ N_n P_n &= \pi_n = \pi_{n,n} \end{aligned} \tag{11}$$

which has just the same computational complexity and generates the same P_n as GSO[4].

[4] It might seem that this description of MGSO differs from the one in [36]. It can be shown that both versions produce identical numerical results.

If the starting set $\mathcal{B}$ does not consist of independent functions, the procedure described above must be modified to account for such dependencies. If for a given m, ϕ_m is a linear combination of its predecessors ϕ_j, $j = 0, \ldots, m-1$, the right hand side of eq. (10) will vanish for $n = m$. The corresponding function $P_m = 0$ must then be excluded from the orthogonal base $\mathcal{B}^*$. Also, terms in P_m can be removed from eq. (10), when this is used to generate other P_n with $n > m$. However, for notational simplicity, we will put $P_m = 0$, such that eq. (10) does not have to be modified. Note that B^* remains a complete orthogonal base for the space S_N, which now has a lower dimensionality.

In practice, the right hand side of eq. (10) will not vanish completely, because of numerical inaccuracies in computing the base functions. Instead π_m is a random function of small amplitude. Therefore some test is needed to determine whether or not π_m should be considered to be the zero function. In practice, this test consists of computing the angle φ between ϕ_m and the space S_{m-1}. This angle is given by $\sin\varphi = \|\pi_m\|/\|\phi_m\|$. If φ is close to zero, it is decided that $\phi_m \in S_{m-1}$ and P_m is set to zero and excluded from $\mathcal{B}^*$.

Unfortunately, this test depends on the (rather arbitrary) choice of a threshold for the magnitude of φ. Because of numerical errors or a wrong choice of the threshold, the test can produce the wrong result. If a random function π_m is not identified as such, it will be normalized and included in (what is thought to be) the orthogonal base. Because of the normalization step, its amplitude will be significantly increased. Such a random function is undesirable in ROTC, because it is totally unsuitable for representing a homogeneous image intensity. In addition, because P_m appears in eq. (10) for $n > m$, the functions P_n, $n > m$ will also have random characteristics.

Even if the numerical independence test produces correct results, a lot of computation time is wasted for unnecessary evaluations of eq. (10). Therefore, in section 6 some non-numerical independence tests will be developed.

For further reference, the following, easily proved theorem is stated.

Theorem 1 *The orthogonal base function P_n can also be obtained by orthogonalizing $v_n = c\phi_n + \phi$, where $c > 0$ and where ϕ is any function in S_{n-1}. If $c < 0$, P_n is obtained up to a sign reversal.*

5 Fast Orthogonalization algorithms

The (modified) GSO method is computationally intensive because each of the functions in the starting base must be orthogonalized with respect to the space spanned by all of its predecessors. This section presents a new orthogonalization algorithm that produces the same base as GSO. It can be used for starting bases consisting of polynomials, warped polynomials or cosines. It is much faster because it does not orthogonalize the starting functions ϕ_n themselves, but other functions that are already orthogonal to most of the predecessors of ϕ_n. The new scheme is based on a recurrence [42, 38] which is an extension of the well-known three-term recursion relation [37, 41] for orthogonal polynomials. Therefore, it will be referred to as Polynomial Recursive Orthogonalization (PRO).

PRO can orthogonalize the starting functions in a few specific orders only. Fortunately, the orderings supported by PRO are the ones that occur in practice. To keep the derivations as general as possible, we denote the starting functions $\phi_{i,j}$.

The following types of starting functions are supported by PRO:

- Polynomials $\phi_{i,j} = x^i y^j$
- Warped Polynomials $\phi_{i,j} = \Big(f(x)\Big)^i \Big(g(y)\Big)^j$
- Cosines $\phi_{i,j} = \cos(iux)\cos(jvy)$

For further reference we define $X = \phi_{1,0}$ and $Y = \phi_{0,1}$. For polynomials, $X = x$ and $Y = y$. It is convenient to call the numbers i, j and $i+j$ the x-, y- and total degree, even if the base functions are not polynomials. For cosines, i and j can be interpreted as (normalized) spatial frequencies. It is always assumed that i and j are non-negative.

The base functions now have two indices as opposed to the one index of the previous section. This is only a notational difference. The theory of the preceding section still applies if the order in which the base functions are orthogonalized is specified. The functions $\phi_{m,n}$ are ordered by defining an ordering relationship on their index pairs (m,n). The notation $(i,j) \prec (k,l)$ means that $\phi_{i,j}$ is orthogonalized before $\phi_{k,l}$. The following ordering relations will be considered.

1. *Natural ordering:* $(i,j) \prec (k,l)$ if either $i+j < k+l$ or $i+j = k+l$ and $j < l$.
2. *Lexicographical ordering:* $(i,j) \prec (k,l)$ if either $i < k$ or $i = k$ and $j < l$.

The (lexicographical or natural) *predecessors* of $\phi_{k,l}$ are all those $\phi_{i,j}$ for which $(i,j) \prec (k,l)$. The notation $(i,j) \preceq (k,l)$ will also be used. Its meaning is self-evident.

Natural ordering sorts the functions $\phi_{i,j}$ by increasing total degree $i+j$, or by increasing y-degree j if they have the same degree. Lexicographical ordering sorts by increasing x-degree; only if two base functions have the same x-degree, the y-degrees are compared[5]. Of course, different orderings are almost trivially obtained, by interchanging x and y. Unless explicitly noted, 'natural' and 'rectangular' will refer to the orderings, as defined above. The orderings described above have the property that $(i,j) \prec (i',j)$ if $i < i'$ and $(i,j) \prec (i,j')$ if $j < j'$.

Let $\mathcal{B}$ be the ordered starting set consisting of functions $\phi_{i,j}$ and let $\mathcal{B}^*$ be the orthogonal base produced when GSO is applied to $\mathcal{B}$. Let $P_{i,j}$ be the function obtained by orthogonalizing $\phi_{i,j}$. As mentioned earlier, $P_{i,j}$ is a combination of only those $\phi_{k,l}$ in $\mathcal{B}$ for which $(k,l) \preceq (i,j)$. As in section 4, we put $P_{i,j} = 0$ if $\phi_{i,j}$ is lineraly dependent on its predecessors in $\mathcal{B}$. Such a $P_{i,j}$ is of course not included in $\mathcal{B}^*$. Note that if $\phi_{i,j} \notin \mathcal{B}$, $P_{i,j}$ is not defined. Therefore, whenever the notation $P_{i,j}$ is used, it is understood that there exists a corresponding $\phi_{i,j} \in \mathcal{B}$.

Let $S_{i,j}$ be the space spanned by $\mathcal{B}_{i,j} = \{\phi_{k,l} \in \mathcal{B} : (0,0) \preceq (k,l) \preceq (i,j)\}$. Of course, $S_{i,j}$ is also the space spanned by $\mathcal{B}^*_{i,j} = \{P_{k,l} \in \mathcal{B}^* : (0,0) \preceq (k,l) \preceq (i,j)\}$. Note that $S_{k,l} = S_{m,n}$, $\mathcal{B}_{k,l} = \mathcal{B}_{m,n}$ and $\mathcal{B}^*_{k,l} = \mathcal{B}^*_{m,n}$ if $(k,l) \prec (m,n)$ and if there are no $\phi_{i,j} \in \mathcal{B}$ for which $(k,l) \prec (i,j) \preceq (m,n)$. The PRO starting sets must have the following properties:

- PRO1: If $\phi_{m,n} \in \mathcal{B}$ then for all $i \leq m$ and $j \leq n$, $\phi_{i,j} \in \mathcal{B}$.
- PRO2: If $\phi_{m,n} \in \mathcal{B}_{m,n}$, at least one of the following must hold:
  - PRO2X: $m > 0$ and for all $\phi_{i,j} \in \mathcal{B}_{m-1,n}$, $X\phi_{i,j} \in S_{i+1,j}$.
  - PRO2Y: $n > 0$ and for all $\phi_{i,j} \in \mathcal{B}_{m,n-1}$, $Y\phi_{i,j} \in S_{i,j+1}$.

The following theorems, proved in appendix A, apply to a polynomial, warped polynomial or cosine PRO starting base $\mathcal{B}$:

Theorem 2 *Each of the orthogonal functions $P_{m,n}$ can be obtained by orthogonalizing $X\,P_{m-1,n}$ (if* PRO2X *applies) or $Y\,P_{m,n-1}$ (if* PRO2Y *applies), with respect to the predecessors of $P_{m,n}$ in $\mathcal{B}^*$.*

[5] This corresponds to the ordering used in dictionaries.

Theorem 3 *If $\phi_{m,n}$ satisfies* PRO2X, *$XP_{m-1,n}$ is orthogonal to all those $P_{i,j}$ in $\mathcal{B}^*$ for which $(i,j) \prec (m-2,n)$.*

Theorem 4 *If $\phi_{m,n}$ satisfies* PRO2Y, *$YP_{m,n-1}$ is orthogonal to all those $P_{i,j}$ in $\mathcal{B}^*$ for which $(i,j) \prec (m,n-2)$.*

The conditions of these theorems can be relaxed, but this is generally not worthwhile. Because of these theorems, the orthogonal functions $P_{m,n}$ can be generated by the following equation:

$$N_{m,n}P_{m,n} = \pi_{m,n} = XP_{m-1,n} - \sum^*_{(m-2,n)\preceq(i,j)\prec(m,n)} \alpha^{m,n}_{i,j} P_{i,j} \quad (12)$$

$$N'_{m,n}P_{m,n} = \pi'_{m,n} = YP_{m,n-1} - \sum^*_{(m,n-2)\preceq(i,j)\prec(m,n)} \beta^{m,n}_{i,j} P_{i,j}, \quad (13)$$

where $\alpha^{m,n}_{i,j} = \ll XP_{m-1,n}, P_{i,j} \gg$ and $\beta^{m,n}_{i,j} = \ll YP_{m,n-1}, P_{i,j} \gg$. The superscript * indicates that the sums run over all those index pairs (i,j) for which $\phi_{i,j} \in \mathcal{B}$ or, more specifically, for which $P_{i,j} \in \mathcal{B}^*$. The numbers $N_{m,n}$ and $N'_{m,n}$ are normalization factors, chosen such that $\|P_{m,n}\| = 1$. It is easily seen that $N_{m,n} = \alpha^{m+1,n}_{m-1,n}$ and $N'_{m,n} = \beta^{m,n+1}_{m,n-1}$.

The equation that would be used in GSO to generate $P_{m,n}$ is:

$$N''_{m,n}P_{m,n} = \pi''_{m,n} = \phi_{m,n} - \sum^*_{(0,0)\preceq(i,j)\prec(m,n)} \gamma^{m,n}_{i,j} P_{i,j}, \quad (14)$$

where $\gamma^{m,n}_{i,j} = \ll \phi_{m,n}, P_{i,j} \gg$. If we call the computation of a term like $\alpha^{m,n}_{i,j}P_{m,n}$ an orthogonalization operation (OO), then it is seen that far less OO's are needed in PRO. This is because $XP_{m-1,n}$ or $YP_{m,n-1}$ are already orthogonal to most of the $P_{i,j}$. Note that in PRO, just as in GSO, an independence test is needed if it cannot be guaranteed that the functions in the starting set are independent. The next section will present some non-numerical independence tests.

The numerical stability of equations (12) and (13) is increased by a modification similar to the one in MGSO. The modified equations are notationally complex and are therefore not given. The modifications do not change the computational complexity and should therefore be applied in practice.

Note that equations (12) and (13) are exactly the same for the cosines $\cos(iux)\cos(jvy)$ and the warped polynomials $\big(\cos(ux)\big)^i \big(\cos(vy)\big)^j$ because X and Y are the same in both cases. This implies that the orthogonal base generated by applying GSO to a cosine starting set is the same as that generated from the corresponding warped polynomial starting set.

We conclude this section with the description of two specific PRO bases (one for each ordering) and the procedures used to orthogonalize them. A third, very important PRO base is not presented until section 8, because it requires some additional theory.

5.1 Orthogonal Functions of Total Degree Not Exceeding N

The starting set $\mathcal{B} = \{\phi_{i,j} : i + j \leq N\}$ satisfies PRO1. PRO2X holds if $m > 0$ and PRO2Y if $n > 0$. The base is orthogonalized in natural order. Unless $m = 0$, $P_{m,n}$ is generated by eq. (12), which requires $2(m + n) - 1$ OO's in this case. The corresponding number of OO's in GSO is $(m + n)^2/2 + 3/2(m + n) + 1$, which is nothing but the number of predecessors of $\phi_{m,n}$ in $\mathcal{B}$ and which is considerably larger. $P_{m,n}$ can also be generated by eq. (13). Then, the number of required OO's is $2(m+n)+1$. This is slightly bigger than for eq. (12). Therefore eq. (13) is only used for $m = 0$. Figure 5 illustrates the orthogonalizations needed and fig. 4 shows some typical base functions.

Note that if there are dependencies in the base, the number of OO's is actually smaller, because only $P_{i,j} \in \mathcal{B}^*$ appear in the equations. Even in this case, PRO is faster than GSO. Table 1 list the number of multiplications and additions required to orthogonalize an independent starting set of total degree not exceeding N. The values in this table have been computed in [26]. It can be seen that the advantage of PRO over GSO increases with increasing N. Even if only 36 base functions are needed ($N = 8$), PRO is twice as fast as GSO.

5.2 Orthogonal Functions of Limited X and Y Degree

This time the starting set is $\mathcal{B} = \{\phi_{i,j} : 0 \leq i \leq M, 0 \leq j \leq N\}$ and the ordering is lexicographical. The base functions $P_{0,n}$, $n = 0, \ldots, N$ are generated by eq. (13), which requires only 2 OO's if $n > 1$. Eq. (12) is applied if $m > 1$ and requires $2N + 2$ OO's if $m > 1$ or $N + 1 + n$ OO's if $m = 1$. A major difference with the previous PRO base is that eq. (13) *cannot* be used if $m > 1$ [6]. This is unfortunate, because eq. (13) is very efficient. The number of OO's needed in GSO is $m(N + 1) + n$. which is

[6]This is not entirely correct. Eq. (13) can be used if $\mathcal{B}$ contains the starting base of section 7.

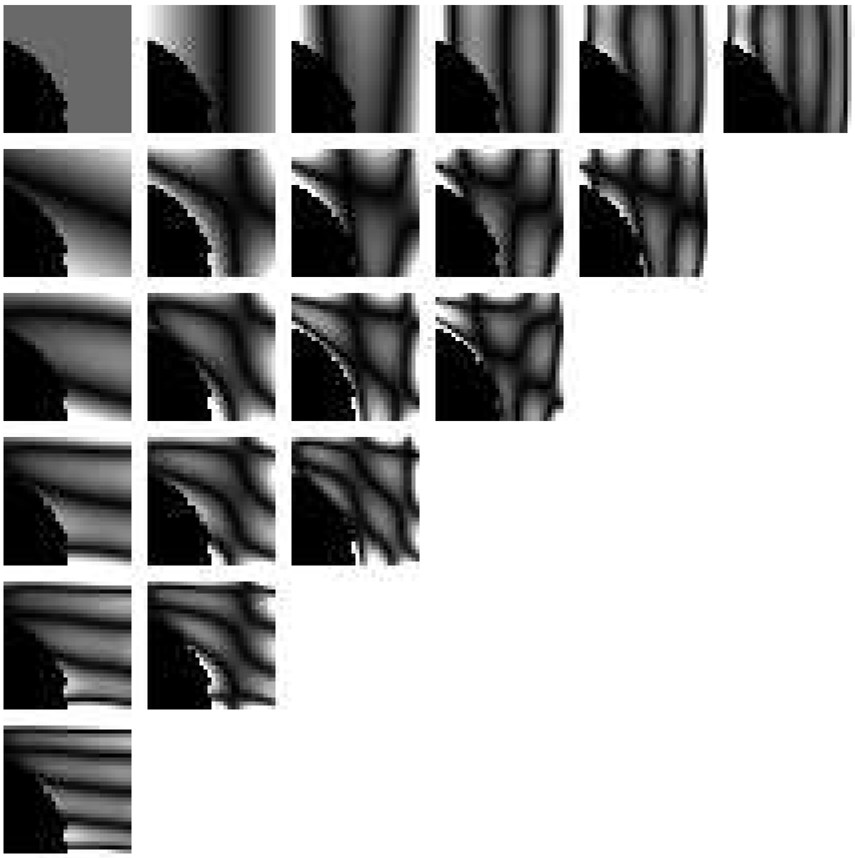

Figure 4: Natural ROTC base functions (of total degree not exceeding 5), for a typical image region.

	Computational Complexity			
	Multiplications		Additions	
N	GSO	PRO	GSO	PRO
2	42	43	36	32
4	240	181	225	152
8	2070	1001	2025	912
16	23 562	6 609	23 409	6 304
32	315 282	48 033	31 4721	46 912
64	4 603 170	366 401	4 601 025	362 112

Table 1: Total number of operations required per pixel, to calculate polynomial base functions of degree not exceeding N in natural order.

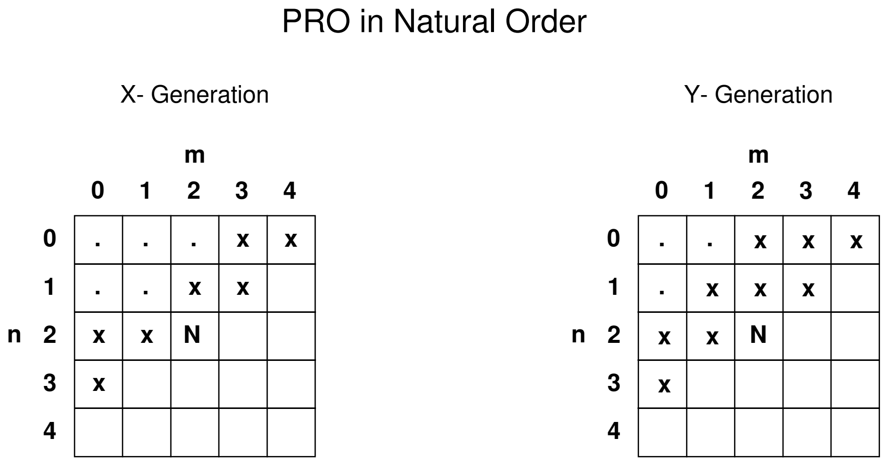

Figure 5: Natural base: Base functions (represented by x) needed in the orthogonalization of $xP_{m-1,n}$ (left) and $yP_{m,n-1}$ (right). The symbol N represents the base function to be generated. Both the base functions marked by an x and those marked by a dot would be needed in GSO.

PRO in Lexicographical Order

X- Generation

n \ m	0	1	2	3	4
0	.	.	x	x	
1	.	.	x	x	
2	.	x	x	N	
3	.	x	x		
4	.	x	x		

Y- Generation

n \ m	0	1	2	3	4
0	.				
1	x				
2	x				
3	N				
4					

Figure 6: Lexicographical base of maximal y-degree 4: Base functions (represented by x) needed in the orthogonalization of $xP_{m-1,n}$ (left) and $yP_{m,n-1}$ (right). The symbol N represents the base function to be generated. Both the base functions marked by an x and those marked by a dot would be needed in GSO.

again much larger than in PRO. It is easily seen that for the generation of the full base of the order of $M^2N^2/2$ OO's are needed in GSO and only $2MN^2$ in PRO. Figure 6 illustrates the orthogonalization of this base for $N = 4$.

6 Checking for Dependencies in the Starting Base

An important question that arises in both PRO and GSO orthogonalization is whether or not a given $\phi_{i,j}$ is linearly dependent on its predecessors in $\mathcal{B}$. This question can be solved numerically but, as explained in section 4, this has several disadvantages. This section shows that merely from the structure of the image region Ω, a lot of conclusions can be drawn about the (in)dependence of the functions $\phi_{i,j}$. Although for general image regions the tests are not applicable to all functions $\phi_{i,j}$, they significantly reduce the number of computations required by PRO and GSO, by eliminating at least some of the dependencies in the starting set.

Unless explicitly noted, the term 'predecessor' will refer to any predecessor of $\phi_{i,j}$, whether or not it belongs to the starting set. Since the phrases 'linearly dependent on its predecessors' and 'linearly independent of its predecessors' will often occur, they will be abbreviated as *LDP* and *LIP* respectively. The theorems in this section assume a natural or lexicographical ordering. First, the theorems are formulated for polynomials. The proofs are generally fairly easy but sometimes tedious. Therefore, only a few of them are given. A short example will demonstrate how the theorems can be applied in practice. Finally, it is shown that the theorems are also valid for warped polynomials and cosines, although with some mild restrictions.

6.1 Polynomials

For further reference, let $\mathcal{A}_{m,n} = \{\phi_{i,j} : 0 \leq i \leq m, o \leq j \leq n\} \setminus \{\phi_{m,n}\}$. The first two theorems enable a whole set of dependent functions to be found as soon as one has been found:

Theorem 5 *If $\phi_{m,n}$ is LDP then $\phi_{m+k,n+l}$ is also LDP if k and l are non-negative.*

Theorem 6 *If $\phi_{m,n}$ is a combination of the functions in $\mathcal{A}_{m,n}$ and $k, l \geq 0$, then $\phi_{m+k,n+l}$ is a combination of the functions in $\mathcal{A}_{m+k,n+l}$.*

The first theorem is of immediate use in the PRO orthogonalization of the natural base of section 5.1 because there all the predecessors of a functions $\phi_{i,j} \in \mathcal{B}$ are also in $\mathcal{B}$. This is not the case for the lexicographically ordered base of section 5.2[7]. For this base the second theorem can be useful.

The following theorems assure that some of the functions $\phi_{m,n}$ are LDP, while others are not. The conditions of these theorems depend only on the structure of the image region Ω.

Theorem 7 *$\phi_{i,j}$ is a combination of its (natural or lexicographical) predecessors and more precisely a combination of the functions in $\mathcal{A}_{i,j}$ if one of the following conditions is satisfied:*

- *LDP1: $i > M$ or $j > N$, where $N + 1$ and $M + 1$ are the number of rows and columns of Ω.*
- *LDP2: There exist sets Ω_1 and Ω_2, such that $\Omega = \Omega_1 \cup \Omega_2$ and $i > M_2$ and $j > N_1$, where $N_1 + 1$ is the number of rows of Ω_1 and $M_2 + 1$ the number of columns of Ω_2.*

Theorem 8 *If Ω contains a product set $\Omega_x \times \Omega_y \subset R^2$, then $\phi_{m,n}$ is LIP if $m < \#\Omega_x$ and $n < \#\Omega_y$.*

Theorem 9 *Let y_j, $j = 0, \ldots, N$ be the y-values of the rows R_j of the set Ω and x_i, $i = 0, \ldots, M$ the x-values of its columns C_i. Assume that these rows and columns have been indexed such that $\#R_i \leq \#R_j$ and $\#C_i \leq \#C_j$ if $i > j$ and define $\mathcal{B}_1 = \{\phi_{i,j} : 0 \leq j \leq N, 0 \leq i \leq \#R_j - 1\}$ and $\mathcal{B}_2 = \{\phi_{i,j} : 0 \leq i \leq M, 0 \leq j \leq \#C_i - 1\}$. If Ω is such that $(x_k, y_l) \in \Omega$ and $0 \leq i \leq k$, $0 \leq j \leq l$ imply that $(x_i, y_j) \in \Omega$ then $\mathcal{B}_1 = \mathcal{B}_2$ and $\phi_{i,j}$ is LDP if and only if $\phi_{i,j} \in \mathcal{B}_1 = \mathcal{B}_2$.*

Theorem 7 shows that never more than $(M+1)(N+1)$ starting functions must be checked for dependencies. The example in section 6.2 will demonstrate the application of the second part of this theorem. For approximately rectangular image regions Ω that contain a rectangle of size $m \times n$, theorem 8 determines mn independent functions. These functions can be orthogonalized by the algorithm presented in section 5.2. Although the mn functions do not form a complete base, they can be sufficient to represent the slowly varying intensity functions encountered in ROTC. Even image regions deviating substantially from a rectangular shape normally contain enclosed

[7] Unless $\mathcal{B}$ contains the independent starting base of section 7. See theorem 14.

Non-Numerical Independence Tests

The Set Ω

y \ x	0	1	2	3	4	5	6	7
0	x	x	x					
1		x	x				x	x
2	x		x			x		
3	x		x	x				
4								

LIP and LDP Functions

n \ m	0	1	2	3	4	5	6	7
0	x	x	x	x				0
1	x	x				0	0	0
2	x	x			0	0	0	0
3	x			0	0	0	0	0
4	0	0	0	0	0	0	0	0

Figure 7: Example. Left: the set Ω; its pixels are marked by an x. Right: Functions $\phi_{m,n}$ that are LIP (x) by theorem 8 or LDP (0) by theorem 7.

rectangles of a reasonable size. Therefore, theorem 8 is very important in practice.

The conditions of theorem 9 might seem a bit complicated. In fact, they state that Ω must have the form

$$\Omega = \{(x_i, y_j) : 0 \leq j \leq N, 0 \leq i \leq n_j\}, \tag{15}$$

where $n_j \leq n_i$ if $j > i$. In section 8, it is shown that the conclusion that $\phi_{i,j}$ is LIP iff $\phi_{i,j} \in \mathcal{B}_1$ holds for lexicographical ordering, *even if the conditions of theorem 9 are not satisfied.* Theorem 9 is proved in the appendix in the case where the conditions do apply. This proof is not superfluous because the generalization does not hold for natural ordering. In any case, the theorem is very important, for it presents a complete natural PRO starting base and shows that it is unique.

A consequence of theorem 9 is that any $\phi_{k,l} \notin \mathcal{B}_1$ is a combination of its predecessors *in* $\mathcal{B}_1$ *only.* This is because, according to theorem 9, such a function can be written as a combination of *all* its predecessors. Each of these predecessors is either contained in $\mathcal{B}_1$ or is a combination of all its own predecessors. The proof of the consequence then follows by induction.

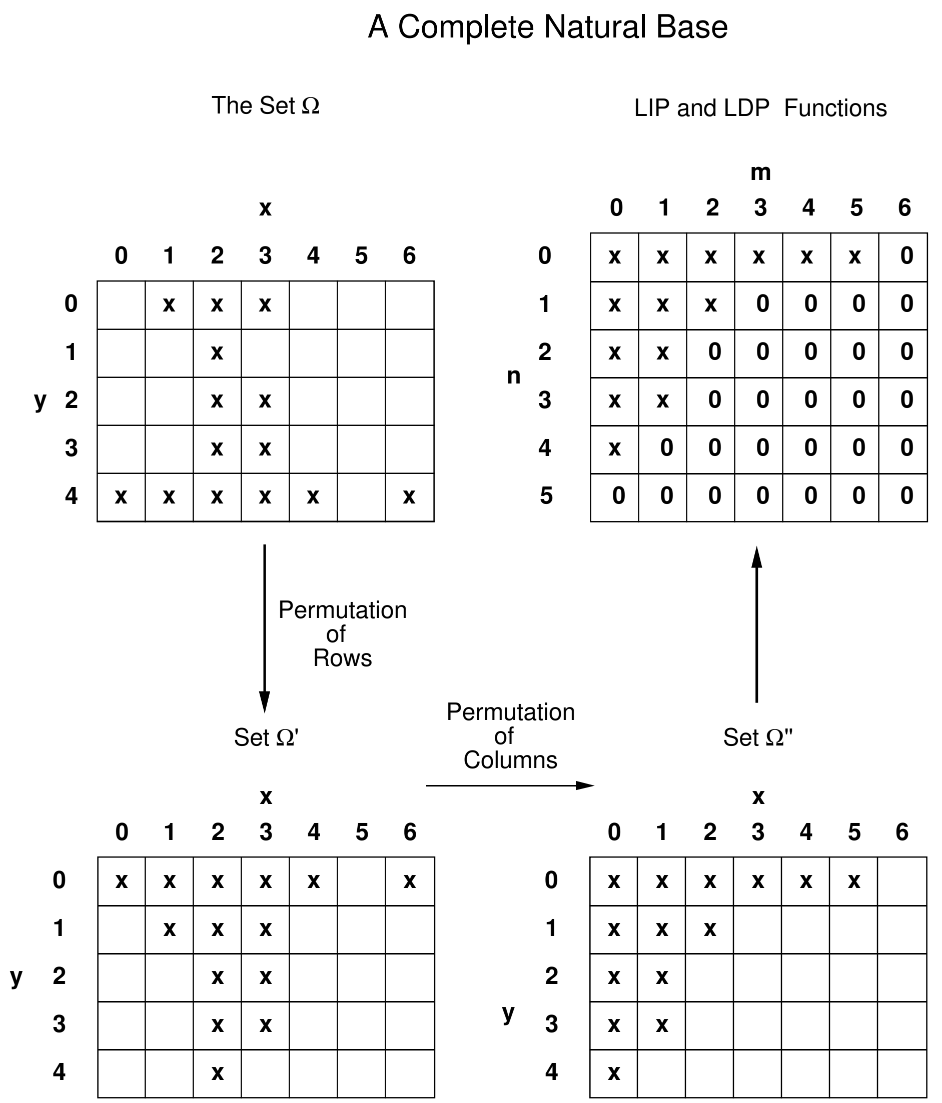

Figure 8: Example of the application of theorem 9. Top left: the set Ω; its pixels are marked by an x. Right: Functions $\phi_{m,n}$ that are LIP (x) or LDP (0). Bottom: by permuting the rows and columns of Ω, the LIP and LDP functions can be found.

6.2 Examples

The first example demonstrates how the theorems just presented can be used to determine whether a $\phi_{m,n}$ is LIP or LDP. The set Ω of figure 7 has four rows and seven columns. According to theorem 7, $\phi_{m,n}$ is LDP if $m > 3$ or $n > 6$. Therefore, we only consider $m \leq 3$ and $n \leq 6$. The functions marked by an x in figure 7 are guaranteed to be LIP by theorem 8. For instance, since $\Omega_x \times \Omega_y = \{0,2\} \times \{0,2,3\} \subset \Omega$, the functions $\phi_{m,n}$ are LIP if $0 \leq m \leq 1 = \#\Omega_x - 1$ and $0 \leq n \leq 2 = \#\Omega_y - 1$.

On the other hand, if we put $\Omega_1 = \{(6,1),(7,1)\}$ and $\Omega_2 = \Omega \setminus \Omega_1$ then Ω_1 has one row ($N_1 = 0$) and Ω_2 has 5 columns ($M_2 = 4$). Theorem 7 then shows that the functions $\phi_{m,n}$ are LDP if $m \geq 5$ and $n \geq 1$. Similar reasonings hold for the other functions marked by a zero in fig. 7.

At this point, nine independent functions have been found. Since there exist $13 = \#\Omega$ of them, only four more have to be found. These must be among the ten non-marked functions in fig. 7, but they can not be found by applying the theorems above.

The second example is shown in figure 8. This figure shows a set Ω for which the conditions of theorem 9 are fulfilled. The corresponding natural base functions (i.e., functions independent of their natural predecessors) are shown in the same figure. Any non-marked functions can certainly be written as a combination of their natural predecessors in the base $\mathcal{B}_1$ of marked functions. The fact that the conditions of theorem 9 are indeed satisfied, is shown by permuting the rows and columns of Ω. The permuted set Ω'' has the same structure as Ω and clearly satisfies the theorem.

6.3 Warped Polynomials and Cosines

The theorems of section 6.1 can be adapted for the warped polynomials $\big(f(x)\big)^i\big(g(y)\big)^j$. An essential but non-restrictive requirement is that the transform $\mathcal{T} : \Omega \to \Omega' : (x,y) \mapsto (x',y') = \big(f(x), g(y)\big)$ is bijective. If this requirement is fulfilled Ω and Ω' have the same structure.

If some mononomial $x^i y^j$ is (in)dependent of its predecessors on Ω then the same must hold for $x^i y^j$ on Ω', because the theorems of section 6.1 depend only on the structure of the image region. This implies that the corresponding warped mononomial $\big(f(x)\big)^i\big(g(y)\big)^j$ is (in)dependent of its predecessors on the set Ω, which is the set obtained by inversely warping Ω'. This is summarized in the following theorem:

Theorem 10 *The theorems of section 6.1 are also valid for the warped polynomials* $\Big(f(x)\Big)^i\Big(g(y)\Big)^j$ *if* $\forall(x_1, y_1), (x_2, y_2) \in \Omega : x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$ *and* $y_1 \neq y_2 \Rightarrow g(y_1) \neq g(y_2)$.

A similar theorem can be proved for cosines:

Theorem 11 *The theorems of section 6.1 are also valid for the cosines* $\cos(iux)\cos(jvy)$ *if* $\forall(x_1, y_1), (x_2, y_2) \in \Omega : x_1 \neq x_2 \Rightarrow \cos(ux_1) \neq \cos(ux_2)$ *and* $y_1 \neq y_2 \Rightarrow \cos(vy_1) \neq \cos(vy_2)$.

The conditions of this theorem can always be satisfied by choosing the fundamental spatial frequencies u and v large enough. The proof of this theorem follows immediately from the following 'equivalence' theorem:

Theorem 12 *Let* $\phi_{i,j} = \cos(iux)\cos(jvy)$ *and* $\psi_{i,j} = \Big(\cos(ux)\Big)^i\Big(\cos(vy)\Big)^j$. *The cosine* $\phi_{i,j}$ *is LIP (LDP) if and only if the warped polynomial* $\psi_{i,j}$ *is.*

The proof of theorem 12 will not be discussed here. However, it is based on the following properties of the cosine function [37]:

$$\Big(\cos(ux)\Big)^i = \sum_{k=0}^{i} C_{i,k}\cos(kux) \tag{16}$$

$$\cos(kux) = \sum_{i=0}^{k} D_{k,i}\Big(\cos(ux)\Big)^i, \tag{17}$$

where $C_{i,i}$ and $D_{i,i}$ are both non-zero. This theorem should not come as a surprise since it was already shown that cosine starting sets lead to the same orthogonal base as the corresponding warped polynomial starting sets.

6.4 Use of the Dependency Tests

The theorems presented in this section are most useful if the number of base functions n to be generated by PRO is of the order of $\#\Omega$. If $n \gg \#\Omega$, the n considered starting functions will almost certainly be independent. Of course, this conclusion is worthwhile on its own. As explained briefly in the following section, the independence tests are also very important in region oriented texture coding methods based on constrained iterative restoration techniques [45, 46, 47].

7 An Independent Starting Base

This section shows how a starting set consisting of independent functions can be selected for a given set or image region Ω. As explained before, the knowledge that the starting base is independent eliminates the need for numerical independence checks. The starting base of this section should also find applications in region oriented texture coding methods based on constrained iterative restoration techniques [45, 46]. These methods represent the image intensity in the different regions by non-orthogonal functions [47]. Since the functions used in [47] are not necessarily independent, generally more coefficients than needed are computed. This slows down the computations but it also leads to a redundant representation. These problems would not occur with the independent functions introduced here.

By orthogonalizing the starting base of this section in a specific order, base functions with some very interesting properties are obtained. This topic will be discussed in section 8. First, the starting base is described and its independence is proved for polynomials. Then, the results are extended for cosines and warped polynomials. Finally, the theory will be illustrated by a small example.

7.1 Theory

Let Ω consist of the $N+1$ different rows $R_i = \{(x,y) \in \Omega : y = y_i\}$, $i = 0, \ldots, N$. Without loss of generality, the rows can be indexed such that $\#R_i \geq \#R_j$, if $j \geq i$. Furthermore, define $\psi_j(y) = \prod_{k=0}^{j-1}(y - y_k)$ and $\psi_0(y) = 1$. Let $n_j = \#R_j - 1$.

The following theorem will now be proved:

Theorem 13 *The functions in $\mathcal{B} = \{x^i y^j : 0 \leq j \leq N, 0 \leq i \leq n_j\}$ are independent on Ω.*

Since the number of functions in $\mathcal{B}$ equals the number of points in Ω, theorem 13 implies that $\mathcal{B}$ is a complete base on Ω. This means that any function defined on Ω is a combination of the functions in $\mathcal{B}$. Note that the structure of Ω is sufficient to determine $\mathcal{B}$. Theorem 13 shows that the same base $\mathcal{B}$ can be used for all sets having the same shape.

The proof of theorem 13 is simplified by the following lemmas.

Lemma 1 *The functions $1, x, \ldots, x^{n_0}$ are linearly independent on Ω.*

Lemma 2 *The independence of the base $\mathcal{B}$ is equivalent with the independence of $\mathcal{B}' = \{\psi_j(y)x^i : 0 \le j \le N, 0 \le i \le n_j\}$.*

The proof of these lemmas can be found in the appendix.

Proof of theorem 13. Because of lemma 2, we only need to prove the independence of $\mathcal{B}'$. Denote $\mathcal{B}'_j = \{\psi_k(y)x^i : 0 \le k \le j, 0 \le i \le n_k\}$. Suppose that $\mathcal{B}'_{j-1}$ is independent on $\cup_{k=0}^{j-1} R_k$, for $j = J > 0$. We will prove that the same holds for $j = J + 1$. This, combined with lemma 1, will prove the theorem. Suppose that the contrary is true and that there exist numbers α_{ji}, not all zero, such that

$$\sum_{j=0}^{J} \sum_{i=0}^{n_j} \alpha_{ji} \psi_j(y) x^i = 0 \qquad \forall (x, y) \in \Omega. \tag{18}$$

Since $\psi_J(y)$ vanishes on $\cup_{k=0}^{J-1} R_k$, eq. (18) implies that $\sum_{j=0}^{J-1} \sum_{i=0}^{n_j} \alpha_{ji} \psi_j(y) x^i$ vanishes on $\cup_{k=0}^{J-1} R_k$. Since, by supposition, $\mathcal{B}'_{J-1}$ is independent on this set, this in turn implies that the coefficients α_{ji} vanish for $0 \le j \le J - 1$. Therefore, eq. (18) reduces to

$$\psi_J(y_J) \sum_{i=0}^{n_J} \alpha_{Ji} x^i = 0, \tag{19}$$

$\forall (x, y) \in \Omega$ and specifically $\forall (x, y) \in R_J$. Since $\psi_J(y_J) \ne 0$, eq. (19) implies that, unless all α_{Ji} are zero, there exists a polynomial of degree at most $n_J = \#R_J - 1$ that vanishes on R_J. This is of course impossible, as explained in the proof of lemma 1. Hence, we conclude that all α_{ji} must be zero, which is in contradiction with the assumption made above.

The base $\mathcal{B}$ of theorem 13 is the same as the base $\mathcal{B}_1$ of theorem 9 in section 6. However, the image regions considered here are totally arbitrary, unlike those of section 9. Therefore, the conditions of theorem 9 are not necessarily fulfilled. The following theorem shows that the main conclusion of theorem 9 still holds, even for these arbitrary regions.

Theorem 14 *$\phi_{i,j}$ is linearly independent of its lexicographical predecessors if and only if $\phi_{i,j} \in \mathcal{B}$, where $\mathcal{B}$ is the base of theorem 13.*

Note that the theorem does *not* guarantee that $\mathcal{B}_1 = \mathcal{B}_2$ (compare with theorem 9). As in section 7, it does however imply that any $\phi_{k,l} \notin \mathcal{B}$ can be written as a combination of its lexicographic predecessors in $\mathcal{B}$ only. The 'if' part of the theorem is nothing but theorem 13. The 'only if' part is proved in the appendix.

Because the conditions of theorem 13 depend on the structure of Ω only, the results of this section are easily adapted for polynomials and cosines (see section 6.3). For warped polynomials, theorem 13 becomes:

Theorem 15 *If* $\forall(x_1, y_1), (x_2, y_2) \in \Omega : x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$ *and* $y_1 \neq y_2 \Rightarrow g(y_1) \neq g(y_2)$ *then the warped polynomials in*

$$\mathcal{B}_w = \left\{ \phi_{i,j} \left(f(x) \right)^i \left(g(y) \right)^j : 0 \leq j \leq N, 0 \leq i \leq n_j \right\}$$

are independent on Ω*. Under the same conditions,* $\phi_{i,j}$ *is linearly independent of its predecessors if and only if* $\phi_{i,j} \in \mathcal{B}_w$*.*

A Complete Lexicographical Base

The Set Ω

y \ x	0	1	2	3	4
0	x	x			
1		x	x		
2				x	x
3				x	x
4			x	x	
5		x	x		

LIP and LDP Functions (Usual Ordering)

n \ m	0	1	2
0	x	x	0
1	x	x	0
2	x	x	0
3	x	x	0
4	x	x	0
5	x	x	0
6	0	0	0

LIP and LDP Functions (Dual Ordering)

n \ m	0	1	2	3	4	5
0	x	x	x	x	x	0
1	x	x	x	x	0	0
2	x	x	x	0	0	0
3	0	0	0	0	0	0

Figure 9: Example: a complete lexicographical base. Left: a set Ω; its pixels are marked by an x. Middle: Functions $\phi_{m,n}$ (marked by x) that are included in the lexicographical base. Right: Functions in the dual base, obtained by interchanging the roles of x and y.

The following theorem presents an independent cosine base:

Theorem 16 *The functions in*

$$\mathcal{B}_c = \{\phi_{i,j}(x,y) = \cos(iux)\cos(jvy) : 0 \leq j \leq N, 0 \leq i \leq n_j\}$$

are independent on Ω *if* $\forall (x_1,y_1),(x_2,y_2) \in \Omega : x_1 \neq x_2 \Rightarrow \cos(ux_1) \neq \cos(ux_2)$ *and* $y_1 \neq y_2 \Rightarrow \cos(vy_1) \neq \cos(vy_2)$. *Under the same conditions,* $\phi_{i,j}$ *is linearly independent of its predecessors if and only if* $\phi_{i,j} \in \mathcal{B}_c$.

The final theorem of this section shows that the bases introduced in this section are PRO bases. It is proved in the appendix.

Theorem 17 *The bases* $\mathcal{B}$, $\mathcal{B}_w$ *and* $\mathcal{B}_c$ *of the theorems 13, 15 and 16 are PRO bases.* PRO2X *holds whenever* $m > 0$ *and* PRO2Y *holds whenever* $n > 0$.

7.2 Example

Figure 9 shows that the lexicographical base is very easily found. Of course, a second base $\{x^i, y^j : 0 \leq i \leq M, 0 \leq j < \#C_i\}$, where $M+1$ is the number of columns of Ω, is immediately obtained by interchanging the roles of x and y in theorem 13. For completeness, functions not in the base have been marked by a zero in fig. 9. All of these functions can be written as a combination of their lexicographical predecessors in the base.

8 A Separable Orthogonal ROTC Base [27]

Because of theorems 14, 15 and 16, the starting bases $\mathcal{B}$, $\mathcal{B}_w$ and $\mathcal{B}_c$ of the previous section can be orthogonalized in lexicographical order by PRO. This section discusses the properties of the resulting orthogonal base. Only polynomials and warped polynomials will be considered. This is because an orthogonal cosine PRO base can be obtained from the corresponding warped polynomial base (see section 5). It is assumed that the warping functions $f(x)$ and $g(y)$ satisfy the conditions of theorem 15.

Specifically, this section will show that the orthogonal base functions are of the form $P_{m,n}(x,y) = P_{m,0}(x,y)Q_{m,n}(y)$, i.e., they are weakly separable. Because of this, a dramatical speedup of the computation of both the base functions and the ROTC coefficients is possible. In addition, less memory is needed for storing a set of weakly separable functions than for general base functions. Another consequence of the separability is that the properties of the functions $P_{m,n}(x,y)$ can be derived from the properties of certain uni-variate (warped) orthogonal polynomials. This is very fortunate because the properties of uni-variate warped polynomials $P_n\big(f(x)\big)$, orthogonal to a weight function $w(x)$, have been determined in considerable detail [23, 24, 41]. Specifically, it is possible to estimate the local amplitudes and frequencies of the uni-variate orthogonal functions [37, 41]. By choosing an appropriate weight function $w(x)$ and warping function $f(x)$, the properties of the orthogonal functions can be changed significantly [24]. This has been exploited in ECG data compression methods based on Discrete Legendre Polynomials [22, 23] and warped orthogonal polynomials [24].

This section will also show that $P_{m,0}$ can be obtained without having to compute all of its predecessors in the orthogonal base $\mathcal{B}^*$. This is important because in a lot of practical applications only a limited number of base functions is required. Note that PRO methods in general do not provide this facility.

The zero crossings of the ROTC base functions can be easily localized by using the corresponding properties of the zeros of the uni-variate polynomials. Furthermore, it is possible to investigate the influence of the warping functions $f(x)$ and $g(y)$ and the (bi-variate) weight function $w(x, y)$ on the spatial properties (local amplitude and spatial frequencies) of the PRO functions. However, due to time limitations, this last subject will not be discussed in this report.

The separability of the PRO base functions and its consequences will also be treated in this section. Furthermore, it is shown that for rectangular image regions, the classical separable block transform base functions are obtained. The properties of the zero crossings are discussed in section 9. First, the relevant properties of the uni-variate orthogonal polynomials are described.

8.1 Uni-Variate Orthogonal Polynomials

Let $I = \{x_0, \ldots, x_N\}$ be a subset of R and let $w(x)$ be a weight function (i.e., a strictly positive function) on I. Furthermore, let $f(x)$ is a warping function such that for all $x_1, x_2 \in I$, $f(x_1) \neq f(x_2)$ if $x_1 \neq x_2$. Consider the polynomials $\phi_i(x) = x^i$ or the warped mononomials $\phi_i(x) = \left(f(x)\right)^i$. As in section 5, we put $X = \phi_0$. Let $p_n(x)$, $n = 0, 1, \ldots$ be the normalized uni-variate polynomials that result when $\phi_0, \phi_1, \ldots$ are orthogonalized (in that order), e.g., by GSO. Note that $\{\phi_i : 0 \leq i < \#I\}$ is a PRO starting set for the one-dimensional set I and that PRO2X is valid. In this case, the PRO equation (12) is nothing more than the well-known three-term recursion relation for orthogonal polynomials [37, 41]:

$$\eta_n p_n = (X - \xi_n) p_{n-1} - \zeta_n p_{n-2} \qquad \text{if } n \geq 1 \tag{20}$$

and $\eta_0 p_0 = 1$. Of course η_n, $n = 0, 1, \ldots$ are normalization factors. The numbers ξ_n and ζ_n are given by

$$\xi_n = \ll X p_{n-1}, p_{n-1}; w \gg_I \tag{21}$$

and

$$\zeta_n = \begin{cases} 0 & \text{if } n = 1 \\ \ll X p_{n-1}, p_{n-2}; w \gg_I & \text{if } n > 1. \end{cases} \tag{22}$$

The following properties are important for the remainder of this report:

- $p_n(x)$ is orthogonal on I to each polynomial in $f(x)$ of degree not exceeding n.
- The recursion (20) generates precisely $\#I$ (warped) polynomials. For $n > \#I$, the recursion produces the zero function.
- If $f(x) = x$ (i.e., for polynomials) or if $f(x)$ is a monotonous and continuous function, $p_n(x)$ has exactly n zeros in the smallest interval $[a, b] \subset R$ for which $I \subset [a, b]$.

In the following, it is assumed that the function $f(x)$ is continuous and monotonous. If $X = f(x) = x$, the functions $p_n(x)$ are nothing but the Discrete Legendre Polynomials associated with the set I. For $f(x) \neq x$, we will refer to the orthogonal functions as Warped Discrete Legendre Polynomials.

8.2 Computational Properties of the Orthogonal Functions

As already shown, PRO2Y holds for all $\phi_{m,n}$ for which $n > 0$. This means that eq. (13) is applicable for $n \neq 0$. Only two predecessors of PRO2Y appear in this equation if $n > 1$. Therefore eq. (13) can be rewritten as

$$N'_{m,n} P_{m,n} = Y P_{m,n-1} - \beta^{m,n}_{m,n-1} P_{m,n-1} - \beta^{m,n}_{m,n-2} P_{m,n-2} \tag{23}$$

This equation is also valid for $n = 1$ if we let $\beta^{m,1}_{m,-1} = 0$. Since $P_{m,1}, \ldots, P_{m,n}$ can all be generated by this equation, it is easily seen that $P_{m,n}$ must be of the form $P_{m,n}(x, y) = P_{m,0}(x, y) Q_{m,n}(y)$, where $Q_{m,n}(y)$ is a polynomial of precise degree n in y. A set of functions of the form $g_m(x, y) h_{m,n}(y)$ is called *class II separable* in [48]. We prefer to use the terminology *weakly separable* instead. More details about these and other kinds of separable functions can be found in [48].

The properties of the orthogonal functions $P_{m,n}$ now follow from the properties of $P_{m,0}(x, y)$ and $Q_{m,n}(y)$ respectively.

8.2.1 Properties of $P_{m,0}$

Let y_i, $0 \leq i \leq N$ be the y-coordinates of the rows R_i of Ω. The following important theorem holds for the base functions $P_{m,0}(x, y)$. It is proved in the appendix.

Theorem 18 *The base functions $P_{m,0}$ satisfy the orthogonality relation*

$$\sum_{x \in R_i} w(x, y_i) P_{m,0}(x, y_i) P_{m',0}(x, y_i) = K(y_i) \delta_{m,m'}.$$

This theorem shows that, on each of the rows R_i of Ω, the $M+1$ uni-variate (warped) polynomials $\psi_{i,m}(x) = P_{m,0}(x, y_i)$, $0 \le m \le M$ are orthogonal (but not normalized) with respect to the weight function $w(x, y_i)$. This is a very specific result that relates the bi-variate polynomials $P_{m,0}(x, y)$ to the uni-variate polynomials $\psi_{i,m}(x)$.

Let $p_{i,m}(x)$, $0 \le m < \#R_i$ be the Warped Discrete Legendre Polynomials associated with the row R_i of Ω. Since each of the $\psi_{i,j}(x)$ is a polynomial of precise degree j, theorem 18 implies that $\psi_{i,m}$ is orthogonal on R_i to *all* polynomials $p(x)$ of degree not exceeding m. Consequently, $\psi_{i,m}$ must be proportional to $p_{i,m}(x)$ if $m < \#R_i$. For $m \ge \#R_i$, $\psi_{i,m}(x)$ is orthogonal to *any* function $h(x)$ on R_i, because then it is orthogonal to the complete base $1, x, \ldots, x^{\#R_i - 1}$ on this set. This is only possible if $\psi_{i,m}(x)$ vanishes on R_i. The following theorem summarizes these results:

Theorem 19 *On each row R_i, $P_{m,0}(x, y)$ is proportional to the (Warped) Discrete Legendre Polynomial $p_{i,m}(x)$ for this row if $m < \#R_i$. For $m \ge \#R_i$, $P_{m,n}(x, y)$ vanishes on the row R_i. More precisely,*

$$\forall (x, y) \in R_i : P_{m,0}(x, y) = \begin{cases} s_m(y_i) p_{i,m}(x) & 0 \le m < \#R_i \\ 0 & \#R_i \le m \le M \end{cases},$$

where $s_m(y_i) \ne 0$.

The numbers $s_m(y_i)$ can be determined easily. This because the highest x-degree coefficient of $P_{m,0}(x, y)$ does not depend on y. Therefore, if $c_{i,m}$, $0 \le m < \#R_i$ are the highest degree coefficients of the (warped) polynomials $p_{i,m}(x)$, these coefficients must satisfy $c_{i,m} s_m(y_i) = c_m$, where c_m is a number that does not depend on i. The value of c_m is found by requiring that $P_{m,0}(x, y)$ is normalized on Ω.

From the previous discussion, it is clear that the polynomials $P_{m,0}(x, y)$ can be generated very quickly by the following procedure:

1. For each row R_i, generate the uni-variate (warped) polynomials $p_{i,m}(x)$, $0 \le m < \#R_i$ by the recursion (20).

2. Scale each of the $p_{i,m}(x)$ by appropriate factors $s_m(y_i)$ such that their highest x-degree coefficients match.

3. Define the function $g_m(x, y)$ by

$$g_m(x, y_i) = s_m(y_i) p_{i,m}(x) \qquad 0 \le i \le N.$$

The requested function $P_{m,n}$ is then given by

$$P_{m,0}(x,y) = g_m(x,y)/\|g_m(x,y)\|.$$

8.2.2 Properties of $Q_{m,n}$

Given $P_{m,0}$, the functions $P_{m,i}$, $i > 0$ could be obtained from eq. (23). In this case, two orthogonalization operations would be needed per base function. Each of these operations requires full bi-variate computations however. It will now be shown that the functions $Q_{m,n}(y)$ can be generated much quicker. This is because the orthogonality relation $\ll P_{m,n}, P_{m,n'} \gg = \delta_{n,n'}$ can be rewritten as

$$\sum_{(x,y)\in\Omega} w(x,y)P_{m,0}^2(x,y)Q_{m,n}(y)Q_{m,n'}(y) = \delta_{n,n'} \tag{24}$$

or

$$\sum_{y\in S(\Omega)} w_m(y)Q_{m,n}(y)Q_{m,n'}(y) = \delta_{n,n'}, \tag{25}$$

where we defined

$$S(\Omega) = \{y : \exists(x,y) \in \Omega\} = \{y_i : 0 \le i \le N\} \tag{26}$$

and

$$w_m(y) = \sum_{x:(x,y)\in\Omega} w(x,y)P_{m,0}^2(x,y) \qquad \forall y \in S(\Omega). \tag{27}$$

Note that $w_m(y)$ is always non-negative and that it is defined for all $(x,y) \in \Omega$. Since $P_{m,0}(x,y)$ vanishes on R_i if and only if $m \ge \#R_i$, the same must hold for $w_m(y)$. Of course, the value of $Q_{m,n}(y)$ is irrelevant if $w_m(y) = 0$ since then $P_{m,0}(x,y)$ is zero anyway. Let $M(m)+1$ be the number of base functions $P_{m,n}(x,y)$ that exist for a given m. It can be shown that for each m, there are exactly $M(m)+1$ rows R_i for which $w_m(y_i) > 0$.

Equation (25) shows that, for fixed m, the polynomials $Q_{m,n}(y)$, $n = 0, \ldots$ are orthogonal with respect to the weight function $w_m(y)$ on the set $S(\Omega)$. This means that they can be generated by the three-term recursion relation (20) for uni-variate orthogonal polynomials[8]. Considering that the values of these polynomials are relevant only for those y for which $w_m(y)$ does

[8] This is true despite the fact that $w_m(y)$ is not strictly positive.

not vanish, the number of (warped) polynomials produced by the recursion equals the number of values of y in $S(\Omega)$ for which $w_m(y)$ does not vanish. Therefore, precisely $M(m)+1$ functions $Q_{m,n}(y)$ are generated by the three-term recursion relation. This is of course just what is expected.

The three-term recursion relation is much quicker than the generation of $P_{m,n}(x,y) = P_{m,0}(x,y)Q_{m,n}(y)$ by eq. (23) since only uni-variate computations are involved. For instance, each inner product evaluation on Ω requires $\#\Omega$ multiplications and additions. On $S(\Omega)$ this number is reduced to $\#S(\Omega) = N+1$, which is generally much smaller. Of course, the computation of the weight functions $w_m(y)$ still requires bi-variate computations, but there are only $M+1$ of these functions to be calculated.

8.3 Computation of the ROTC Coefficients

Since the base functions are weakly separable, the ROTC coefficients can be computed much more quickly than in the general non-separable case. In general, the coefficients $A_{m,n}$ are computed from eq. (3) which, when written in full, becomes:

$$A_{m,n} = \sum_{(x,y)\in\Omega} P_{m,n}(x,y)f(x,y), \tag{28}$$

where $f(x,y)$ is the image intensity function on the region Ω. For the weakly separable base, this becomes:

$$\begin{aligned} A_{m,n} &= \sum_{y\in S(\Omega)} Q_{m,n}(y) \sum_{x:(x,y)\in\Omega} f(x,y)P_{m,0}(x,y) && (29)\\ &= \sum_{y\in S(\Omega)} Q_{m,n}(y)B_m(y) && (30)\end{aligned}$$

Only $M+1$ bi-variate computations are needed for the computation of $B_m(y)$, $0 \le m \le M$. Each of these has a complexity proportional to $\#\Omega$. All of the other computations are uni-variate and have a complexity proportional to $N+1$.

8.4 The Special Case of a Rectangular Image Region

Assume that the ROTC weight function $w(x,y) = w_1(x)w_2(y)$ is separable. For rectangular regions $\Omega = \Omega_x \times \Omega_y$, the polynomials $p_{i,m} = p_m$ do not depend on i, because

1. All rows are equal (up to an irrelevant difference in y-coordinate).

2. The weight functions $w(x, y_i) = w_1(x)w_2(y_i)$ for the different rows are proportional and the orthogonal polynomials are not changed if the weight function is scaled.

Since the numbers $s_m(y_i) = s_m$ are also independent of i, $P_{m,0}(x, y) = p_m(x)$ is a function of x only. More specifically, it is a uni-variate (Warped) Discrete Legendre Polynomial. It is also easily seen that all of the weight functions $w_m(y)$ must be equal. Consequently $Q_{m,n}(y) = Q_n(y)$ does not depend on m.

Therefore, we have shown that, given the conditions stated above, the base functions $P_{m,n}(x, y) = p_m(x)Q_n(y)$ are separable. This means that classical block transforms are comprised in the type of PRO base described in this section. In fact, it can be shown that this is also true for the PRO bases described in section 5.

8.5 Importance of the Separable Base

The orthogonal base described in this section has extremely attractive computational properties. Both the base functions and the ROTC coefficients can be determined using mostly one-dimensional (i.e., univariate) computations. Only a limited number of base functions can be computed if desired. In addition, it can be shown that this type of orthogonal base is very suitable for recursive methods that compute the base functions and coefficients for a given region from its sub-regions. For Block Transform Coding such a method was already presented in [25]. The method presented there can be easily extended for the PRO base of this section. The corresponding generalizations for other PRO bases (or GSO bases) are more complicated and not as fast since full bi-variate computations are involved.

9 Spatial Properties of the Separable Base Functions [28]

Almost all properties of the (warped) polynomials $P_{m,n}$ can be derived from those of $P_{m,0}(x, y)$ and $Q_{m,n}(y)$. In the previous section, it has been shown that $P_{m,0}(x, y)$ is a uni-variate (warped) orthogonal polynomial on each image row. Since $Q_{m,n}(x, y)$ is also a uni-variate (warped) orthogonal polynomial, the properties of the functions $P_{m,n}(x, y)$ can be found from the study of uni-variate (warped) orthogonal polynomials only. This is important, because much is known about the properties of such functions.

The influence of the warping functions $f(x)$ and $g(y)$ and of the weight function $w(x, y)$ on the separable orthogonal base can be easily studied. In the uni-variate case it was already shown that this study can lead to significant improvements in transform coding techniques [23, 24]. Such improvements are also possible in the bi-variate case. Unfortunately, due to time limitations this subject cannot be discussed here. Instead, this section will describe the simpler, but equally important, properties of the zero crossings of the orthogonal base functions.

Let $x_-(i) = \min\{x : (x, y) \in R_i\}$ and $x_+(i) = \max\{x : (x, y) \in R_i\}$. In other words $x_-(i)$ and $x_+(i)$ are the minimal and maximal x-coordinate of any point on the row R_i. The properties of the zero crossings of $P_{m,0}(x, y)$ for $m > 0$ are summarized in the following theorem:

Theorem 20 *There exist polynomials $\xi_{i,m}(y)$ of degree N such that*

$$\forall (x, y) \in \Omega : P_{m,0}(x, y) = c_m \prod_{i=0}^{m-1} \Big(x - \xi_{i,m}(y)\Big),$$

where c_m is a constant and where $x_-(i) \leq \xi_{i,m}(y_i) \leq x_+(i)$.

The proof of this theorem will not be given here. It is based on the fact that $P_{m,0}(x, y)$ is proportional to the uni-variate orthogonal polynomial $p_{i,m}(y)$ on the row R_i if $m < \#R_i$. Of course, it is well known that $p_{i,m}(y)$ has exactly m zeros in the interval $[x_-(i), x_+(i)]$. As explained in section 8, $P_{m,0}(x, y)$ vanishes on each row R_i for which $\#R_i \leq m$. This does not contradict the theorem. In fact, it shows that $\#R_i$ of the zero curves $x = \xi_{i,m}(y)$ pass through the points of R_i.

The zero properties of the orthogonal functions $Q_{m,n}(y)$ are even simpler. Let $y_-(m)$ be the smallest y for which there exist at least m points $(x, y) \in \Omega$. Similarly, let $y_+(m)$ be the largest y for which there exist at least m points $(x, y) \in \Omega$. The following theorem then holds:

Theorem 21 *If $Q_{m,n}(y)w_m(y)$ does not vanish identically on $S(\Omega)$, there exist numbers $y_{i,m,n}$ and non-zero constants $d_{m,n}$ such that*

$$Q_{m,n}(y) = d_{m,n} \prod_{i=0}^{n-1} (y - y_{i,m,n}),$$

where $y_-(m) \leq y_{i,m,n} \leq y_+(m)$.

The properties of the base functions $P_{m,n}(x,y) = P_{m,0}(x,y)Q_{m,n}(y)$ are clearly anisotropic. The zero-crossings of $Q_{m,n}(y)$ consist of lines parallel to the y-axes, while those of $P_{m,0}(y)$ are more general curves.

For a rectangular set Ω and a separable weight function $w(x,y)$, the polynomials $\xi_{i,m}(y)$ reduce to constants. In that case there is no anisotropy.

10 Conclusion and Further Research

This report has presented several techniques that diminish the computational burden of ROTC methods. The separable base of section 8 allows the ROTC coefficients to be computed almost as efficiently as in classical (separable) transform coding. It has been shown that the properties of univariate orthogonal base functions can be easily generalized for the separable base functions. This is an important breakthrough for the theoretical study of ROTC methods, which is now only in its infancy. For instance, the theory should make it possible to determine appropriate warping functions for which the ROTC base functions have a given, desired spatial distribution of local amplitude or frequency. The feasability and the success of such techniques have already been demonstrated in the uni-variate case. Such extensions will most certainly be the subject of further research. In combination with the previously introduced BTC recursive coefficient computation schemes, combined segmentation-coding schemes are also within reach. This facilitates another important improvement in ROTC methods.

Even for non-separable ROTC bases, the speedups achieved by the fast algorithms described in this report are considerable. In addition, the important problem of checking for dependencies in the starting base was partially solved in an elegant, non-numerical way. The further development of such tests and the study of the non-separable ROTC bases are also the subject of further research.

A Proofs of some theorems

A.1 Theorem 2

Suppose that PRO2X holds and consider the orthogonal function $P_{m,n}$. The existence of $P_{m,n}$ implies that $\phi_{m,n} \in \mathcal{B}$ and PRO1 implies that $\phi_{m-1,n}$ also belongs to $\mathcal{B}$. Therefore, the corresponding function $P_{m-1,n}$ exists. Because $P_{m-1,n}$ can be generated by GSO eq. (10) applies. Using the double index notations, that equation becomes:

$$N_{m-1,n} P_{m-1,n} = \phi_{m-1,n} - \sum_{(i,j)\prec(m-1,n)}^{*} c_{i,j}\phi_{i,j}, PRO \tag{31}$$

where the superscript * indicates that the sum runs over only those index pairs (i, j) for which $\phi_{i,j} \in \mathcal{B}$. From eq. (31) it follows that

$$N_{m-1,n} X P_{m-1,n} = X\phi_{m-1,n} - \sum\nolimits^{*}_{(i,j)\prec(m-1,n)} c_{i,j} X\phi_{i,j} \tag{32}$$

$$= X\phi_{m-1,n} - \sum\nolimits^{*}_{(i,j)\prec(m,n)} c'_{i,j}\phi_{i,j} \tag{33}$$

The last equality is a result of PRO2X.

For polynomials or warped polynomials $X\phi_{m-1,n} = \phi_{m,n}$ and therefore the previous equation reduces to

$$X P_{m-1,n} = (c\phi_{m,n} - \phi)/N_{m-1,n}, \tag{34}$$

where $c = 1$ and ϕ is a combination of only the predecessors $\phi_{i,j} \in \mathcal{B}$ of $\phi_{m,n}$. For cosines, $X\phi_{m-1,n} = 1/2\phi_{m,n} + 1/2\phi_{m-2,n}$ if $m > 1$, while $X\phi_{0,n} = \phi_{1,n}$. In this case eq. (34) holds with $c = 1/2$ or $c = 1$.

Because of eq. (34), theorem 1 applies, which means that orthogonalizing $XP_{m-1,n}$ or $\phi_{m,n}$ produces the same result. This proof can be repeated if only PRO2Y holds.

A.2 Theorems 3 and 4

Both proofs are similar. Only theorem 3 is proved. Let $(i, j) \prec (m - 2, n)$. If $\phi_{m,n}$ indeed satisfies PRO2X then $X\phi_{i,j} \in S_{i+1,j}$, because $(i, j) \prec (m - 1, n)$. The proof then follows from the fact that $P_{m-1,n}$ is orthogonal to the space $S_{i+1,j}$ again because $(i + 1, j) \prec (m - 1, n)$. In particular, $P_{m-1,n}$ is orthogonal to $X\phi_{i,j} \in S_{i+1,j}$. Mathematically, $\ll X\phi_{i,j}, P_{m-1,n} \gg = 0$, which is equivalent to $\ll \phi_{i,j}, XP_{m-1,n} \gg = 0$. This proves theorem 3.

A.3 Theorem 8

Suppose that the theorem does not hold. In that case

$$\phi_{m,n} = \sum_{i,j} \phi_{i,j} \qquad \forall (x, y) \in \Omega, \tag{35}$$

where the sum runs over the (rectangular or natural) predecessors of $\phi_{m,n}$. According to their indices (i,j), the predecessors $\phi_{i,j}$ can be classified into one of three groups:

- Group A: Predecessors with $i \leq m$ and $j \leq n$, but of course excluding $\phi_{m,n}$.
- Group B: Predecessors with $i < m$ and $j > n$.
- group C: Predecessors with $j < n$ and $i > n$. For rectangular ordering this group is empty.

It is easily seen that if $j > m$, there exist polynomials $p_j(y)$ of degree not exceeding n such that $\phi_{i,j}(x,y) = x^i p_j(y)$ for all $(x,y) \in \Omega_x \times \Omega_y$. Hence, on $\Omega_x \times \Omega_y$, functions of group B are combinations of the predecessors in group A. It can be verified, that the same holds for group C. Hence, the supposition (35) leads to

$$\phi_{m,n}(x,y) = \sum C_{i,j}\phi_{i,j}(x,y) \qquad \forall (x,y) \in \Omega_x \times \Omega_y, \tag{36}$$

where the sum only contains functions of group A. This contradicts the fact that the functions $\{\phi_{i,j} : 0 \leq m, 0 \leq n\}$ are independent on $\Omega_x \times \Omega_y$, if $m = \#\Omega_x$ and $n = \#\Omega_y$.

A.4 Theorem 9

Because of the conditions of the theorem, Ω can be described as

$$\begin{aligned} \Omega &= \{(x_i, y_j) : 0 \leq j \leq N, 0 \leq i < \#R_j\} & (37)\\ &= \{(x_i, y_j) : 0 \leq i \leq M, 0 \leq j < \#C_i\}. & (38) \end{aligned}$$

This implies that $\mathcal{B}_1 = \mathcal{B}_2$.

Now, let $\Omega_{x,k} = \{x_0, \ldots, x_k\}$ and $\Omega_{y,l} = \{y_0, \ldots, y_l\}$. Because of the structure of Ω, the following statements imply one another.

$$\phi_{k,l} \in \mathcal{B}_1 = \mathcal{B}_2 \tag{39}$$

$$(x_k, y_l) \in \Omega \tag{40}$$

$$\Omega_{x,k} \times \Omega_{y,l} \subset \Omega \quad . \tag{41}$$

Because of this and because $\#\Omega_{x,k} = k + 1$ and $\#\Omega_{y,l} = l + 1$, theorem 8 shows that $\phi_{k,l}$ is LIP if it belongs to $\mathcal{B}_1$.

Since $\#\mathcal{B}_1 = \#\Omega$, $\mathcal{B}_1$ is a complete base for the region Ω. Therefore, any $\phi_{k,l} \notin \mathcal{B}_1$ can be written as a combination of the functions in $\mathcal{B}_1$:

$$\phi_{k,l} = \sum_{j=0}^{N} \sum_{i=0}^{\#R_j-1} c_{i,j}\phi_{i,j} \tag{42}$$

We now show that this sum only contains predecessors $\phi_{i,j}$ of $\phi_{k,l}$. This will imply that $\phi_{k,l} \notin \mathcal{B}_1$ is LDP, which will complete the proof of the theorem.

Since $\phi_{k,l} \notin \mathcal{B}_1$, either $l > N$ or $k \geq R_l$. First suppose that $0 < l \leq N$ and define the polynomial

$$p(x) = x^{k-\#R_l+1} \prod_{i=0}^{\#R_l-1} (x - x_i) \tag{43}$$

Because of (37), $p(x)$ vanishes on the rows R_s, $s \geq l$. On the other hand, the polynomial

$$q(y) = \prod_{i=0}^{l-1} (y - y_l) \tag{44}$$

vanishes on the other rows R_s, $s < l$. Consequently, $p(x)q(y) = 0$ for all $(x, y) \in \Omega$. This conclusion remains valid for $l = 0$ if we put $q(y) = 1$ and for $l = N$ if we put $p(x) = x^k$.

In all cases, $p(x)$ is of precise degree k and $q(y)$ of precise degree l. Therefore, $p(x)q(y) = 0$ can be solved for $x^k y^l$, which means that

$$\phi_{k,l} = \sum_{x^i y^j \in \mathcal{A}_{k,l}} c_{i,j}\phi_{i,j}. \tag{45}$$

This concludes the proof.

A.5 Lemma 1

The first $\#R_0$ powers of x are independent on R_0. If this were not the case, there would exist a uni-variate polynomial of degree $n_0 = \#R_0 - 1$, that vanishes on the $\#R_0$ different points in R_0. This is impossible, since a polynomial of degree n can have at most n zeros. The proof is completed by remarking that if some functions are independent on a set S, they are also independent on any set containing S.

A.6 Lemma 2

To prove the equivalence of the independence of $\mathcal{B}$ and $\mathcal{B}'$, it suffices to show that any of the functions in $\mathcal{B}'$ can be expressed as a combination of the base functions in $\mathcal{B}$ and vice versa. The first part follows from the fact that $\psi_j(y)$ is a polynomial of precise degree j. Hence $\psi_j(y)x^i$ is a combination of the functions in the set $\{x^i y^k : 0 \leq k \leq j\}$, which are all part of $\mathcal{B}$, if $\psi_j(y)x^i \in \mathcal{B}'$. The second part can be proved similarly. Since $c_{k,k} \neq 0$, the set of equations

$$\psi_k(y) = \sum_{l=0}^{k} c_{kl} y^l \qquad 0 \leq k \leq j \tag{46}$$

can be solved for y^l, $0 \leq l \leq j$. The resulting solutions are of the form $y^l = \sum_{k=0}^{l} d_{lk}\psi_k(y)$, where $d_{ll} \neq 0$. Consequently, $y^j x^i$ is a combination of the functions in the set $\{\psi_k(y)x^i : 0 \leq k \leq j\}$ which is a subset of $\mathcal{B}$.

A.7 Theorem 14

Suppose that $\phi_{k,l} \notin \mathcal{B}$. It will have to be shown that $\phi_{k,l}$ is LDP (for lexicographic ordering). First suppose that $0 < l \leq N$ and let $\Omega' = \cup_{j=l}^{N} R_j$ and $\Omega'' = \Omega \setminus \Omega' = \cup_{j=0}^{l-1} R_j$. The rows R'_j of Ω' are given by $R'_j = R_{j+l}$.

Define $n'_j = \# R'_j - 1 = n_{j+l}$. According to theorem 13, $\mathcal{B}' = \{x^i y^j : 0 \leq j \leq N - l, 0 \leq i \leq n'_j\}$ is a complete base on Ω'. Hence, for any given k, there exist numbers $C_{i,j}$ such that

$$x^k = \sum_{j=0}^{N-l} \sum_{i=0}^{n'_j} C_{i,j} x^i y^j \qquad \forall (x,y) \in \Omega'. \tag{47}$$

On the other hand, the polynomial $\psi_l(y)$ vanishes on Ω''. This implies that

$$x^k \psi_l(y) = \sum_{j=0}^{N-l} \sum_{i=0}^{n'_j} C_{i,j} x^i y^j \psi_l(y) \qquad \forall (x,y) \in \Omega. \tag{48}$$

Since $\psi_l(y) = \sum_{s=0}^{l} d_{l,s} y^s$, where $d_{l,l} = 1$, eq. (48) can be rewritten as

$$x^k \psi_l(y) = \sum_{j=0}^{N-l} \sum_{i=0}^{n'_j} \sum_{s=0}^{l} C_{i,j} x^i y^{j+s} \psi_l(y) \qquad \forall (x,y) \in \Omega. \tag{49}$$

Each of the mononomials $x^i y^{j+s}$ appearing in this sum is a predecessor of $x^k y^l$. This is because $i \leq n'_j = n_{j+l} \leq n_l$, while $k > n'_0 = n_l$ since $x^k y^0 \notin \mathcal{B}'$. Therefore $i < k$, which proves that $x^k y^{s+l}$ is indeed a predecessor of $x^k y^l$. Consequently $x^k \psi_l(y) = \phi(x, y)$, where ϕ is a combination of the predecessors of $x^k y^l$ and $x^k y^l = \phi - \sum_{s=0}^{l-1} d_{l,s} x^k y^s$ is also a combination of those predecessors. This proves the theorem for $0 < l \leq N$. The proofs for the cases $l = 0$ and $l > N$ are similar.

A.8 Theorem 17

Let $m > 0$ and assume that $\phi_{m,n} \in \mathcal{B}$ and $\phi_{i,j} \in \mathcal{B}_{m-1,n}$. We now show that $X\phi_{i,j} \in S_{i+1,j}$. This will prove PRO2X. For (warped) polynomials $X\phi_{i,j} = \phi_{i+1,j}$, while for cosines $X\phi_{i,j} = c\phi_{i,j} + d\phi_{i-1,j}$ (see the proof of theorem 2). This means that $X\phi_{i,j} \in S_{i+1,j}$ if and only if $\phi_{i+1,j} \in S_{i+1,j}$, even for cosines. We now show that $\phi_{i+1,j} \in S_{i+1,j}$. This is trivially true if $\phi_{i+1,j} \in \mathcal{B}$. On the other hand, if $\phi_{i+1,j} \notin \mathcal{B}$ then theorems 14, 15 and 16 show that $\phi_{i+1,j}$ is a combination of its predecessors in $\mathcal{B}$, which again proves that $\phi_{i+1,j} \in S_{i+1,j}$.This concludes the proof.

A.9 Theorem 18

Any bi-variate polynomial of x-degree not exceeding $m - 1$ can be written as a combination of those functions $X^i Y^j \in \mathcal{B}$ for which $0 \leq i \leq m - 1$ (see theorem 14). Since $P_{m,n}$ is orthogonal on Ω to each of these functions, $P_{m,n}$ is in particular orthogonal to any bi-variate polynomial of x-degree not exceeding $M - 1$. Specifically, it is orthogonal to $r_j(y) P_{m',0}(x, y)$, where $m' < m$ and $r_j(y) = \prod_{k=0}^{N} (y - y_k)/(y - y_j)$. In mathematical form, this becomes:

$$\sum_{(x,y)\in\Omega} w(x, y) r_j(y) P_{m',0}(x, y) P_{m,0}(x, y) = 0 \tag{50}$$

or, since $r_j(y)$ vanishes on $\Omega \setminus R_j$:

$$r_j(y_j) \sum_{x:(x,y_j)\in R_j} w(x, y_j) P_{m',0}(x, y_j) P_{m,0}(x, y_j) = 0. \tag{51}$$

The conclusion of the theorem then follows, because $r_j(y_j) \neq 0$.

Acknowledgement

The author would like to thank the Belgian National Science Foundation, which provided the financial support for this work. He is also endebted to prof. M.C. Vanwormhoudt, for the many discussions on this work.

References

[1] R.J. Clarke, *Transform Coding of Images.* Academic Press, New York, 1985.

[2] A.N. Netravali and B.G. Haskell, *Digital Pictures, Representation and Compression.* Plenum Press, New York, 1988.

[3] A.N. Netravali and J.O. Limb, "Picture coding: a review," *Proc. IEEE*, vol. 68, no. 3, pp. 366–406, March 1980.

[4] A.K. Jain, "Image data compression: a review," *Proc. IEEE*, vol. 69, no. 3, pp. 349–389, March 1981.

[5] A.K. Jain, "Advances in mathematical models for image processing," *Proc. IEEE*, vol. 69, no. 5, pp. 502–528, May 1981.

[6] W.H. Chen and H. Smith, "Adaptive coding of monochrome and color images," *IEEE Trans. Commun.*, vol. 25, no. 11, pp. 1285–1292, November 1977.

[7] M. Kunt, M. Bénard, and R. Leonardi, "Recent results in high-compression image coding," *IEEE Trans. Circ. Syst.*, vol. 34, no. 11, pp. 1306–1336, Nov. 1987.

[8] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second-generation image-coding techniques," *Proc. IEEE*, vol. 73, no. 4, pp. 549–574, April 1985.

[9] D.N. Graham, "Image transmission by two-dimensional contour coding," *Proc. IEEE*, vol. 55, no. 3, pp. 336–346, March 1967.

[10] J.W. Woods and S.D. O'Neal, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1278–1288, October 1986.

[11] H.S. Malvar, "Lapped transform for efficient transform/subband coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 6, pp. 969–978, June 1990.

[12] K. Sauer, "Enhancement of low bit-rate coded images using edge detection and estimation," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 1, pp. 52–62, January 1991.

[13] S. Carlsson, C. Reillo and H. Zetterberg, "Sketch-based representation of grey values and motion information," in *From Pixels to Features* (J. Simon, ed.), North Holland: Elsevier Science Publishers, 1989.

[14] M. Gilge, "Coding of arbitrarily shaped image segments using moment theory," *Proc. EUSIPCO-88* (Grenoble), pp. 855–858, September 1988.

[15] M. Gilge, "Region-oriented transform coding (ROTC) of images," *Proc. ICASSP-90* (Albuquerque), April 1990, pp. 2245–2248.

[16] M. Gilge, "A generalized orthogonal transform for image coding," *Proc. Picture Coding Symposium* (Cambridge, Mass., USA), March 1990.

[17] M. Gilge, T. Engelhardt and R. Mehlan, "Coding of arbitrarily shaped image segments based on a generalized orthogonal transform," in *Signal Processing: Image Communication 1*, North Holland: Elsevier Science Publishers, 1989, pp. 153–180.

[18] M. Tüceryan and A.K. Jain, "Texture segmentation using Voronoi polygons," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 2, pp. 211–216, February 1990.

[19] L.M. Lifshitz and S.M. Pizer, "A multiresolution hierarchical approach to image segmentation based on intensity extrema," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 6, pp. 529–540, June 1990.

[20] J.F. Haddon and J.F. Boyce, "Image segmentation by unifying region and boundary information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 10, pp. 929–948, October 1990.

[21] J. Goutsias and J.M. Mendel, "Simultaneous optimal segmentation and model estimation of nonstationary noisy images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 9, pp. 990–998, September 1989.

[22] W. Philips and G. De Jonghe, "Polynomial approximation for data reduction on ECG's," in *Proc. MEDICON 89*, Patras, Greece, pp. 222-223.

[23] W. Philips and G. De Jonghe, "Data compression of ECGs by high-degree polynomial approximation," *IEEE Trans. Biomedical Engineering*, to be published April 1992.

[24] W. Philips, "Time-warped polynomials for signal representation and coding," submitted for EUSIPCO-92, Brussels, 1992.

[25] W. Philips, "Recursive computation of polynomial transform coefficients," *Electronic Letters*, vol. 27, no. 25, pp. 2337-2339, 1991.

[26] W. Philips, "A fast algorithm for the generation of orthogonal base functions on an arbitrarily shaped region," in *Proc. ASSP-92* (San Fransisco), 1992, to be published.

[27] W. Philips, "Choosing a linearly independent starting base in region oriented transformcoding (ROTC)," *Electronics Letters*, submitted.

[28] W. Philips, "On the properties of base functions that are orthogonal on a non-rectangular region," *PRORISC-92*, Houthalen, Belgium, 1992, submitted.

[29] W. Philips, "Adaptive orthogonal transform for image data compression," *EUSIPCO-92*, Brussels, 1992, submitted.

[30] P. Strobach, "Quadtree-structured recursive plane decomposition coding of images," *IEEE Trans. Signal Processing*, vol. 39, no. 6, pp. 1380–1397, June 1991.

[31] M. Miyahara and K. Kotani, "Block distortion in orthogonal transform coding–analysis, minimization and distortion measure," *IEEE Trans. Commun.*, vol. 33, no.1, pp. 90–96, January 1985.

[32] J.L. Mannos and D.J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory*, vol. 20, no.4, pp. 525–536, July 1974.

[33] Y.B. Yu and A.G. Constantinides, "Variable block size and position transform coding," in *Signal Processing IV: Theories and Applications*, North Holland: Elsevier Science Publishers, pp 1011–1013.

[34] Hou, "A fast recursive algorithm for computing the discrete cosine transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 10, pp. 1455–1461, October 1987.

[35] G.S. Zabele and J. Koplowitz, "Fourier encoding of closed planar boundaries," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 1, pp. 98–102, January 1985.

[36] G.H. Golub and C.F. Van Loan, *Matrix Computations.* John Hopkins University Press, Second Edition, 1991.

[37] G. Szegö, *Orthogonal Polynomials.* American Mathematical Society Colloquium Publications, vol. 23, Fourth Edition, 1975.

[38] A. Erdélyi, *Higher Transcendental Functions.* Bateman Manuscript Project, vol.2, 1981. Krieger Publishing Company.

[39] T. Carleman, "Über die approximation analytischer funktionen durch lineare aggregate von vorgegebenen potenzen," *Arkiv För Matematik, Astronomi och Fysik*, vol. 17, no. 9, pp. 1–29, 1923.

[40] M.J. Gottlieb, "Concerning some polynomials orthogonal on a finite or enumerable set of points," *American Journal Math.*, vol. 60, pp. 453–458, 1938.

[41] T.S. Chihara, "The three term recurrence relation and spectral properties of orthogonal polynomials," in *Orthogonal Polynomials: Theory and Practice* (P. Nevai, ed.), NATO-ASI Series, vol. 296, London: Kluwer, 1989.

P.H. Defert, "On first-degree multivariate polynomial approximation," *Journal Approx. Theory*, vol.35, pp. 381–387, 1982.

[42] D. Jackson, "Formal Properties of Orthogonal Polynomials in Two Variables," *Duke Mathematical Journal*, vol. 60, pp. 436–446, 1938.

[43] D. Jackson, "Problems of closest approximation on a two-dimensional region," *Am. Journal Math.*, vol. 60, pp. 436–446, 1938.

[44] D.N. Green and S.C. Bass, "Representing periodic waveforms with nonorthogonal basis functions," *IEEE Trans. Circ. Syst.*, vol. 31, no. 6, pp. 518–513, June 1984.

[45] R.W. Schafer, R.M. Mersereau and M.A. Richards, "Constrained iterative restoration algorithms," *Proc. IEEE*, vol. 69, no. 4, pp. 432–450, April 1981.

[46] V.T. Tom, T.F. Quatieri, M.H. Hayes and J.H. McClellan, "Convergence of iterative nonexpansive signal reconstruction algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, no. 5, pp. 1052–1058, October 1981.

[47] U. Franke, R. Mester and T. Aach, "Constrained iterative restoration techniques: a powerful tool in region oriented texture coding," in *Signal Processing IV: Theories and Applications*, North Holland: Elsevier Science Publishers, pp 1145–1148

[48] W.A. Porter and J.L. Aravena, "Fast 2-d signal processing," *Proc. IEEE*, vol. 78, no. 4, pp. 768–775, April 1990.

[49] J. Xu and Y-H Yang, "Generalized multidimensional orthogonal polynomials with application to shape analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 9, pp. 906–913, September 1989.