A multiresolution directional-oriented image transform based on Gaussian derivatives

Boris Escalante-Ramírez*, José L. Silván-Cárdenas*
Graduate Division, School of Engineering, National University of Mexico

ABSTRACT

In this work, a multi-channel model for image representation is derived based on the scale-space theory. This model is inspired in biological insights and it includes some important properties of human vision such as the Gaussian derivative model for early vision proposed by Young[19]. The image transform that we propose in this work uses similar analysis operators as the Hermite transform at multiple scales, but the synthesis scheme of our approach integrates the responses of all channels at different scales. The advantages of this scheme are: 1) both analysis and synthesis operators are Gaussian derivatives. This allows for simplicity during implementation. 2) The operator functions possess better space-frequency localization, and it is possible to separate adjacent scales one octave apart, according to Wilson's results on human vision channels[16]. 3) In the case of 2-D signals, it is easy to analyze local orientations at different scales. A discrete approximation is also derived from an asymptotic relation between the Gaussian derivatives and the discrete binomial filters. We show in this work how the proposed transform can be applied to the problem of image coding. Practical considerations are also of concern.

Keywords: Multiresolution, Gaussian derivatives, Hermite transform, binomial filters, visual system models, image coding.

1. INTRODUCTION

Human viewers often assess the quality of the results of image processing. In such cases, errors are significant only if perceived. Therefore, with a deep understanding of the human visual system (HVS) their properties can be exploited to achieve resulting images that *look* better.

The Gabor model is one of the most widely used in HVS modeling at the visual cortex level[1, 12]. A filter design-based approach has been also of interest in modeling the processing by the visual cortex. The Cortex transform[14] is one of these types of image transform whose analysis functions were designed to approximate the Gabor shape with settable frequency and orientation bandwidths. An alternative model proposed by Young in 1987 is based on the Gaussian derivatives (GDs)[19]. Young showed that these functions model the measured receptive field data more accurately than the Gabor functions do[18]. Like the receptive fields, both Gabor functions and GD are spatially local and they consist of alternating excitatory and inhibitory regions in a decaying envelope. However, the GD analysis is found to be more efficient because it takes advantage of the fact that GDs comprise an orthogonal basis if they belong to the same point of analysis. Gaussian derivatives can be interpreted as local generic operators in a scale-space representation described by the isotropic diffusion equation[7]. In a related work, the GDs have been interpreted as the product of Hermite polynomials and a Gaussian window[10], where windowed images are decomposed into a set of Hermite polynomials. Some mathematical models based on these operators at a single spatial scale have been described by several authors[2,8,10]. In the latter, the so called Hermite Transform, has been extended to the multiscale case[11], and has been successfully used in different applications such as noise reduction[2], coding[14], and motion estimation for the case of image sequences[3, 4].

The image transform introduced in this work differs from the Hermite transform in two major aspects. First, our image decomposition uses directional GDs at multiple scales instead of using 2-D separable GDs. Second, while the hierarchical implementation of the Hermite transform is done through a predictive process between successive layers [10], our approach simply adds together the layers generated by the transform. This reconstruction process arises because the layers are generated through a process of subtraction, the same way the Cortex transform does[14].

* boris@servidor.unam.mx; jlsc@verona.fi-p.unam.mx, phone +52-5-6223016; fax +52-5-6161073; http://verona.fi-p.unam.mx, DEPFI-UNAM, Secc. Eléctrica, Apartado Postal 70-256, México, D.F. 04510, México.

2. IMAGE STRUCTURE IN SCALE-SPACE

In order to obtain a full understanding of the concept of scale, we look at an image as a representation of a physical scene. Such a representation is influenced by physical quantities that can be either measured or estimated. The influence of the viewing distance, for instance, can be interpreted as a scaling of the intensity distribution of our representation. This means that when moving away from the observed object we increase the scale of our representation and, therefore, some detail on the perceived image gets lost, while getting closer reduces scale and then more detail can be perceived.

The viewing angle, as the relative position of the vision system, is another physical factor that influences the representation of visual patterns within the scene. Such viewing angle makes image structures to be perceived with certain orientation, which can only be *seen* by operators that are sensitive to orientation changes. From this, it is clear the need to insert an orientation parameter in the scale-space representation to obtain an oriented scale-space representation.

In scale-space representations[17], a signal $L(x)$ is convolved with scaled versions of a normalized Gaussian to remove details from it, so that $L(x,s) = L(x)*G(x,s)$ is the signal representation at scale s, where $G(x,s)$ is the Gaussian with scale parameter s (one half of the Gaussian variance). The local structure of a signal is inferred from the derivatives of this scaled representation. Such derivatives can be obtained directly by convolving the signal with the scaled differential operators $G_n(x,s)$ for $n = 0,1,2,...$, where $G_n(x,s)$ denotes the n-th order derivative of $G(x,s)$.

In two dimensions, the Gaussian derivative operators are defined as $G_{m,n-m}(x,y,s) = G_m(x,s)G_{n-m}(y,s)$ for $m = 0,1,..,n$ and $n = 0,1,....$, where n is the total order of derivation. The set of n+1 two-dimensional derivatives of exact order n is enough to represent patterns with at most n simultaneous orientations[13]. For orientation analysis purposes it is more convenient to work with the rotated GD $G_{m,n-m}(x,y,s,\theta)=G_{m,n-m}(x\cos\theta + y\sin\theta,-x\sin\theta + y\cos\theta,s)$ as pointed out in[7] and [9].

One of the most appealing properties of these operators is that they can be calculated as a linear combination of the functions in the original coordinate system (such a formula is easily derived in the frequency domain). In particular, the directional GDs are expressed as

$$G_{n,0}(x,y,s,\theta)=\sum_{m=0}^{n}\alpha_{m,n-m}(\theta)G_{m,n-m}(x,y,s) \qquad (1)$$

for $n = 1,2,...$, with the angle function

$$\alpha_{m,n-m}(\theta)=C_n^m\cos^m\theta\,\sin^{n-m}\theta$$

for $m = 0,1\ldots n$. These filters have been successfully used in local orientation detection and discrimination[13] because they are directionally selective and satisfy the so-called steering property, which confirms that only n+1 different directional analyses are needed to completely describe a n-th order signal change in 2-D.

3. MULTI-CHANNEL MODEL

3.1 Case 1-D

In order to arrive to a result that shows the relationship among perceived luminance patterns at a given scale and the functions L_n at larger scales, we need to have an insight about the way the human vision system could carry out this process. Psychophysical findings show the existence of at least four channels at each point of the visual field at retinal level (more specifically in the ganglion cells). The spatial receptive fields of these channels all have approximately the shape of a Difference of Gaussians (DOG). This filter is defined from its impulse response function $DOG(x,s_1,s_2) = G(x,s_1)-G(x,s_2)$ for $s_2 > s_1$ and, in many applications, it can be approximated as the second-order GD.

Since the output of the DOG filter provides the information that must be added to the representation at scale s_2 in order to obtain the representation at the lower scale s_1, the signal can be expressed as the sum of DOG channels outputs. In the most general case we can write:

$$L(x)=\sum_{k=-\infty}^{\infty} L(x) * \mathrm{DOG}(x, s_{k-1}, s_k) \tag{2}$$

with $s_k > s_{k-1}$. Here, we will require the space parameter $\tau = (s_k - s_{k-1})/s_k$ to be constant for any integer index k, which is in agreement with psychophysical data[16]. These data support the existence in the human visual system of a number of distinct spatial frequency-tuned channels, approximately one octave apart ($\tau = 0.75$).

In most real systems one is mainly interested in enhancing certain perceptually important patterns such as edges, lines, junctions, etc. and, at the same time, one wishes to attenuate or even eliminate the annoying component of the signal such as fine textures or noise. The fulfillment of this requirement is easier handled within a derivative analysis context because it allows computing robustly physical quantities related to the local structure of the visual patterns [2]. Thus, it is convenient to express the DOG filter in terms of the scaled derivative operators. This is achieved by expanding a Gaussian at scale s_{k-1} in Taylor series around scale s_k and applying the diffusion equation, which results in the following

$$\mathrm{DOG}(x, s_{k-1}, s_k)=\sum_{n=1}^{\infty} \frac{(-\tau\ s_k)^n}{n!} G_{2n}(x, s_k) \tag{3}$$

In practice, the number of terms of this series must be limited to a small number N (there are indications from neurophysiology studies that indicate that the visual system works with derivatives up to fourth order, however, we will work here with all the terms of the series. Particular cases can be further derived from this.

Notice that the expression (3) shows explicitly that derivatives of odd order do not contribute to the image description. However, it is desirable that the model, at least in the analysis stage, includes all the derivatives for a richer image structure description. We factorize the even-order GD filters in (3) by using the closure under convolution property of the GD and then we substitute (3) in (2). Furthermore, if we interpret convolutions as an inner product (the integral of the product of two operand functions from minus infinite to plus infinite respect to the variable x, which we denoted as $\langle L,H\rangle_x$), we finally can write the reconstruction process as

$$L(x)=\sum_{k=-\infty}^{\infty}\sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \tag{4}$$

$$D_n^{(k)}(x)=\left\langle \hat{L}_n^{(k)}(\xi), \hat{G}_n^{(k)}(x,\xi)\right\rangle_\xi \tag{5}$$

where the analysis stage

$$\hat{L}_n^{(k)}(\xi)=\left\langle L(x), \hat{G}_n^{(k)}(x,\xi)\right\rangle_x \tag{6}$$

is nothing but the projection of the input signal onto the analysis functions

$$\hat{G}_n^{(k)}(x,\xi)=\frac{1}{\sqrt{2s_k}} G_n^*\left(\frac{x}{\sqrt{2s_k}}-\xi\right) \tag{7}$$

for n = 1,2... and $s_k = (1-\tau)^{-k} s_0$ for k = 1,2,.. ,K. The notation $G_n^*(x)$ stands for the normalized filters $G_n(x,1/4)/(2^n n!)^{1/2}$.

Expression (6) reminds the Hermite Transform analysis introduced in[10] the difference being that here ξ represents a normalized coordinate instead of a discrete position. Sub-sampling can be only possible if an interpolation process with functions that differ from the analysis functions replaces expression (4). Our image transform also differs from Hermite transform in that the latter synthesizes the signal by means of a predictive process between successive layers[10], while here this is done through addition of all the layers. This reconstruction process can be done recursively by adding successively the details to the current scaled representation, that is $L^{(k-1)} = L^{(k)} + D^{(k)}$, where $D^{(k)}$ is the k-th DOG channel output and is referred to as the *detail function* at layer k.

3.2 Case 2-D

In this section we show how expressing the 2-D DOG channels in terms of GDs has a notable utility for local orientation analysis, (this is because the DOG is isotropic in 2-D).

The Taylor expansion of the two-dimensional DOG function can be written in terms of either 2-D separable GDs or in terms of a set of directional GDs. The first expansion leads to a separable image transform, which is nothing but the 1-D image transform of subsection 3.1 along each coordinate. In the second case we write the DOG filter as

$$\mathrm{DOG}(x,y,s_{k-1},s_k)=\sum_{n=1}^{\infty}\sum_{j=0}^{n}\frac{(-\tau\ s_k)^n}{\alpha_n n!}G_{2n,0}(x,y,s_k,\theta_j) \tag{8}$$

where $\theta_j = \theta_o + j/(n+1)$ for $j = 0,.., n$. and the constants

$$\alpha_n=\sum_{j=1}^{n}\mathrm{sen}^{2n}\left(\frac{j\pi}{n+1}\right)$$

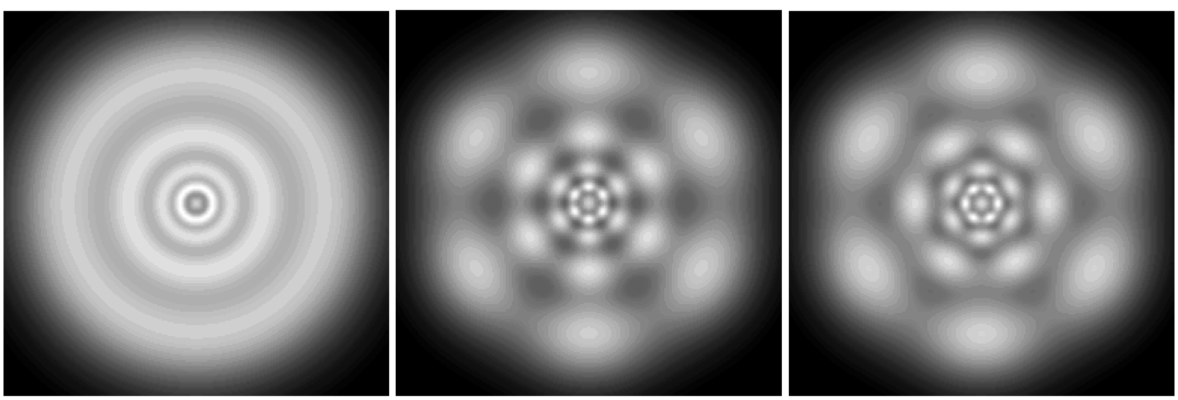
for $n = 0,1,....$, which leads to a multi-scale, multi-orientation image transform with similar expressions to those found in the previous subsection, but in 2-D with the following 2-D analysis functions

$$\hat{G}_n^{(j,k)}(x,y,\xi,\eta)=\frac{1}{2s_k}G_{n,0}^{*}\left(\frac{\mathbf{r}\cdot\boldsymbol{\alpha}_j}{\sqrt{2s_k}}-\xi,\frac{\mathbf{r}\cdot\boldsymbol{\beta}_j}{\sqrt{2s_k}}-\eta\right) \tag{9}$$

where $\mathbf{r} = (x,y)$, $\boldsymbol{\alpha}_j = (\cos\theta_j, \sin\theta_j)$, $\boldsymbol{\beta}_j = (-\sin\theta_j, \cos\theta_j)$ with $\theta_j = \theta_o + j/(n+1)$ for $j = 0,...,n$; $n = 1,2...$, and $s_k = (1-\tau)^{-k} s_0$ for k integer. The angle θ_o is a free parameter that can be set according to our convenience.

Apparently, this representation has a disadvantage in that the analysis functions of Eq. (9) are non-separable. In practice however, this is overcome by re-mapping the 2-D separable transform through Eq. (1) to carry out both analysis and synthesis stages of this multi-scale multi-orientation image transform.

Figure 1 shows the frequency distribution of four spatial frequency channels of the proposed transform for the following cases: a) the full transform, b) the simplified transform for image coding with three different orientations, i.e., 12 DOG channels plus a DC Gaussian channel, c) the simplified transform with three different orientations. Note in this case an orientation shift $\theta_o = \pi k/6$ for each k^{th} channel at every scale. This distribution presents a better coverage of the frequency space.



(a) (b) (c)

Figure 1. Frequency distribution of 4 DOG channels and DC Gaussian channel using (a) original definition, (b) first order approximation with θ_o=0 and (c) same approximation with $\theta_o = \pi k/6$ for the k^{th} channel.

3.3 Discrete formulation

Practical applications of our image transform require a formulation for discrete signals. We present here a discrete formulation regarding an asymptotic approximation of the GD operators by the binomial functions.

We approximate the continuous multi-scale representation by the set of filtered versions of the input signal with a binomial kernel $B[x,N]$ of length N+1 for $x = -N/2...N/2$, where N is a non-negative integer related to the scale through $N = 8s$ as it will be shown below. Then, the discrete counterpart of the GDs can be written as a forward difference of this binomial kernel, *i.e.*, $B_n[x,N]=\Delta_x^n B[x,N-n]$ for $n = 0,1,...,N$. Thus, we shall refer these operators to as the binomial differences (BDs). It can be shown that each BD tends asymptotically to the GD of the same order as N tends to infinite, that is, $B_n[x, N] \sim G_n(x, N/8), N \to \infty$, n = 0,1,2….

We use the previous limiting process in the approximation of the DOG filter by the difference of binomials DOB[x,N_1,N_2] = $B[x,N_1]$-$B[x,N_2]$ and then, we write the DOB channels in terms of the BD functions of even order with a space constant $\tau = (N_k-N_{k-1})/N_k > 0$. Based on that expansion the following expressions for the signal reconstruction can be found

$$L[x] = \sum_k \sum_{n=1}^{\tau N_k} \tau^n \Gamma^{(k)} D_n^{(k)}[x] \tag{12}$$

with

$$D_n^{(k)}[x] = \sum_y L_n^{(k)}[y] B_n^{(k)}[y-x] \tag{13}$$

where the factor

$$\Gamma^{(k)} = \prod_{i=1}^{n-1} \frac{1 - i/\tau N_k}{1 - i/N_k}$$

tends to one as k grows indefinitely (i.e, as the scale tends to infinite), and the signal analysis

$$L_n^{(k)}[x] = \sum_y L[y] B_n^{(k)}[x-y] \tag{14}$$

is carried out through the normalized binomial functions

$$B_n^{(k)}[x] = \sqrt{C_{N_k/2}^n}\, B_n[x, N_k/2]$$

Notice that this discrete approximation is not expressed in normalized coordinates because of the discrete domain. The rescaling of a discrete domain is not possible, but such behavior is emulated in practice by down-sampling the data. The reverse process involves an interpolation with functions that, in general, vary from the decimation filters. Such a process results in a pyramidal scheme. Another practical consideration is that we have to limit the scale sequence to a finite small set $k = k_1 \ldots k_2$, which requires to compute the low-pass residue (the information in the channels indexed with $k < k_1$) and the high-pass residues (the information in the channels indexed with $k > k_2$), similarly to what is done in the cortex transform[14].

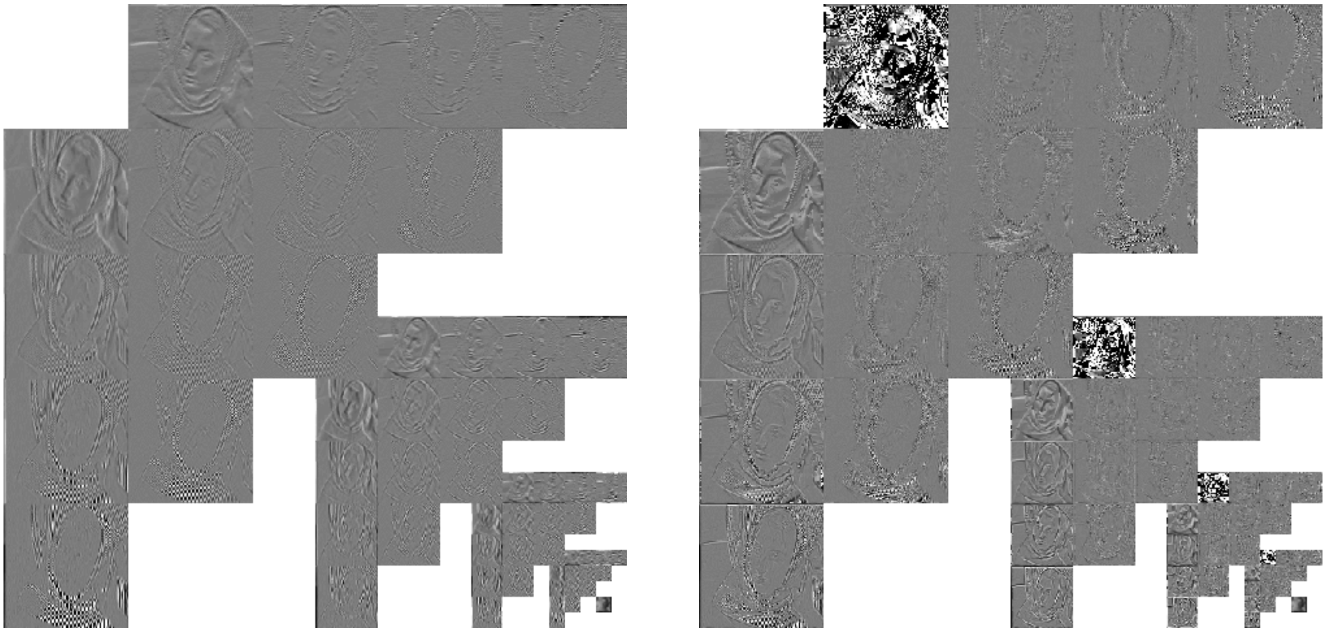
The extension to 2-D is immediate from the binomial kernel separability. However, a directional analysis is quite more complicated because of the discrete domain. Therefore we limit our analysis to the approximation of the directional GDs by Eq. (1) after replacing the GDs by the BDs.

4. RESULTS

Experiments with natural images showed that if we use the BDs for both, decimation and interpolation, the resulting errors are not perceptually relevant when the sub sampling parameter is below the square root of the filter length. Then, if the sampling distances are power of two the more suitable filter lengths are $N_k = 4^k$ for k = 0,...,K (the top limit K is set according to the image resolution, *e.g.*, we use K = 4 for square images of 256 pixels). As in the cortex transform, the high order residue can be neglected without important image degradations. Figure 2 shows an example of image decomposition for the separable case without rotation (a) and with rotation (b). The reconstructed image and the error image for the last case are shown in Figure 3.

In other experiments we also used the approximation of the directional GDs. No important degradation of edges or lines was perceived. Moreover, when we take the first order approximation in the DOB expansion, we observe that the introduced error tends to enhance the contrast of the image. In fact, an ideal edge is reconstructed with some overshot resembling the Mach-bands effect.

In a more sophisticated scheme we set the angle θ_o to the gradient direction and we take only the directional analysis for j = 0. In this case we observe that oriented structures are quite well reconstructed, but non-oriented patterns such as dot textures present some artifacts.



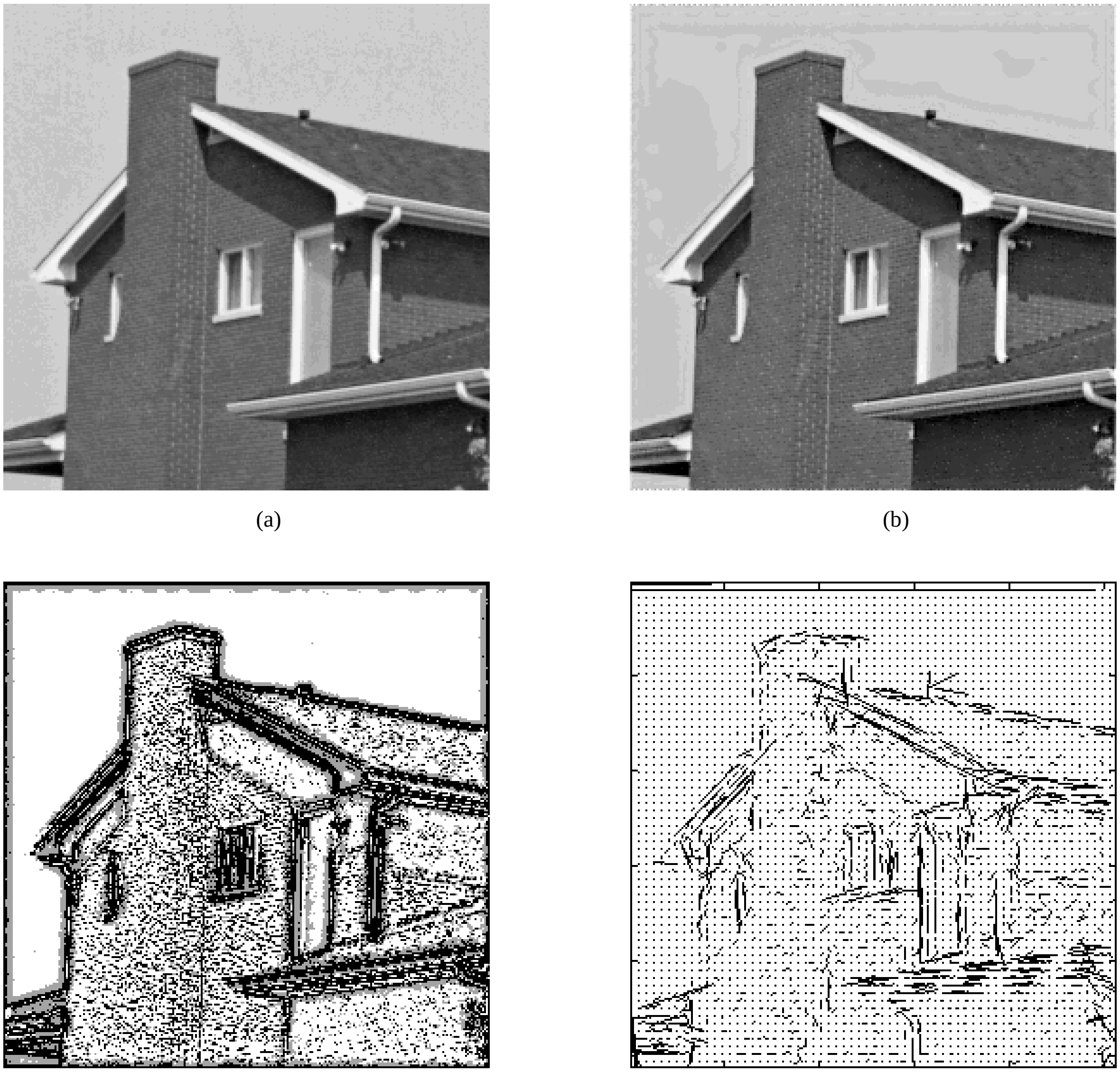
(a) (b)

Figure 2. Four channels of the multi-scale decomposition and the low-pass residue. (a) Separable case, (b) Rotated case. The angle is displayed as an image in the place of $L_{0,1}$, which becomes zero after rotation.

a) b)

Figure3. (a) Reconstructed image from the rotated coefficients of Figure 2. (b) Reconstruction error.

A direct application of this representation model can be found in image coding. In this case the image is analyzed in order to determine the scale at which the local content can be better described as well as its optimal orientation. Results of this scheme are shown in figure 4. Original HOUSE image is shown in a). Reconstructed image is shown in b). c) shows a four gray level image representing the scale used to code every point of the image (black accounts for the smallest scale and white for the largest. d) shows the orientation used to code each point of the image.



(a) (b) (c) (d)

Figure 4.- Reconstruction results with HOUSE image (256x256 pixels) : (a) Original image, (b) Reconstructed image with the single-scale single-orientation model. Frequency distribution of Fig. 1c was used, (c) scale map in gray levels (black for smallest scale and white for largest scale) and (d) gradient orientation for the selected scale (each vector is weighted by the AC local energy)

5. CONCLUSIONS

The proposed transform represents an image representation model that analyses images by means of operators based on Gaussian derivatives and that possess important advantages over the similar representations, such as the Hermite Transform, namely better space-frequency distribution, simplicity and suitability for multiresolution local analysis. A simplified scheme proposed is based on a first order approximation of the DOG functions, four spatial channels and three different orientations. This scheme provides 12 channels oriented at different scales and orientations and presents frequency space coverage similar to that of the Cortex Transform[15]. This scheme can be directly applied to image coding. In this case the local content of the image is represented by only one spatial channel and one orientation. Results are encouraging and show that this transform can be used not only for coding but also for restoration, analysis, and parameter estimation. Of course, more experimentation is needed in order to come out with a competitive image coding scheme, however, we can conclude that the results shown here demonstrate the robustness of this image transform to information truncation. Such robustness is explained on the base of the redundancy among the various scales.

ACKNOLEGMENTS

This work was supported by grants CONACyT 27533-A, DGAPA-PAPIIT IN107101 and INCO-DC-961646-AMOVIP.

REFERENCES

1. Bastiaans, M, "Gabor's signal expansion and degrees of freedom of a signal", *Opt. Acta*, **vol. 29**, pp 1223-1229, 1982.
2. Camarillo-Sandoval, P., Varela-López, A., Escalante-Ramírez, B., "Adaptive Multiplicative-Noise Reduction in SAR Images with Polynomial Transforms", *Proceedings of the IGARSS´98, IEEE Geoscience and Remote Sensing Society*, pp. 1171-1173. 1998.
3. Escalante Ramírez, B, Silván-Cárdenas, J.L., "Optical-Flow Estimation by means of Local Projection Analysis with the Radon-Hermite Transform", *Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, F.Preteux, A. Mohammad-Djatari, E. Dougherty, Eds., **vol. 3816**, pp. 121-129, Proceedings SPIE, 1999.
4. Escalante-Ramírez, B., Silván-Cárdenas, J.L., "Motion Analysis and Classification with Directional Gaussian Deivatives in Image Sequences", *Advanced Signal Processing Algorithms, Architectures, and Implementations X*, F.T. Luk, Ed., **vol. 4116**, pp. 447-453, Proceedings SPIE, 2000.
5. Haddad, R.A., Akansu, A.N, "A new orthogonal transform for signal coding" [On line], *IEEE Transactions on Acoustics, Speech and Signal Processing*, **vol. 36-9**, pp. 1404 –1411, 1998. <http://iel.ihs.com:80/cgi-bin/>.
6. Hendee, W. R., Wells, P. N. T., *The perception of visual information*. Springer-Verlag, New York Inc., 1993.
7. Koenderink, J. J., van Doorn, A. J., "Generic neighborhood operators". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **vol. 14-6** , pp. 597 –605, 1992
8. Liu, Z.-Q., Rangayyan, R.M., Frank, C.B. "Directional analysis of images in scale space". [On line] *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **vol. 13 11**, pp. 1185 –1192, 1991 <http://iel.ihs.com:80/cgi-bin/>.
9. Martens, J.-B., "Local orientation analysis in images by means of the Hermite transform", *IEEE Transactions* on *Image Processing*, **vol. 6-8**, pp. 1103 –1116, 1997.
10. Martens, J.-B., "The Hermite transform-theory", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **vol. 38 9**, pp. 1595 –1606, 1990.
11. Martens, J.-B. "The Hermite transform-applications", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **vol. 38-9**, pp. 1607 –1618, 1990.
12. Porat, M., Zeevi, "The generalized Gabor scheme of image representation in biological and machine vision", *IEEE Trans. Pattern Analysis Mach, Intel.*, **vol 10**, pp. 452-467, 1988.
13. Shizawa, M.; Iso, T. "Direct representation and detecting of multi-scale, multi-orientation fields using local differentiation filters", *Proceedings CVPR '93, Computer Vision and Pattern Recognition*, IEEE Computer Society, pp. 508 –514, 1993.
14. Silván-Cárdenas, J.L., Ángeles-Meza, M.P., Escalante-Ramírez, B. "Radon-Hermite Analysis Applied to Image Coding", *Proceedings of the IEEE International Symposium on Industrial Electronics, ISIE'99*, **vol. 3**, pp. 1204-1207, 1999.
15. Watson, A. "The cortex transform: Rapid computation of simulated neural images", *Comput. Vision, Graph., Image Processing*, **vol 39**, pp. 311-327, 1987.
16. Wilson, H. R., Bergen, J.R., "A four mechanism model for spatial vision". *Vision Research*, **vol 19**. pp 19-32.
17. Witkin, A., "Scale-space filtering: A new approach to multiscale description", *Image Understanding 1984*, **ch. 3**, pp. 79-95, 1984.
18. Young, R. "Oh say, can you see?. The physiology of vision", *Proc. SPIE*, **vol.1453**, pp. 92-723, 1991.
19. Young, R. "The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles", General Motors Res. Labs., Rep. 4920.