Markovian regularization of Hermite transform based SAR-image classification

Penélope López-Quiroz^a, Boris Escalante-Ramírez^{*a,b}, José L. Silván-Cárdenas^b

^aUniversidad Nacional Autónoma de México,
DEPFI-UNAM, Apartado Postal 70-256, México, D.F., 04510

^bCentro de Investigación en Geografía y Geomática “Ing. Jorge L. Tamayo”, A.C.,
Cataratas No. 3, Jardines del Pedregal, México, D.F., 01900

ABSTRACT

A novel classification scheme for SAR images based on the perceptual classification of image patterns in the Discrete Hermite Transform domain has been developed. In order to obtain the DHT referred to a rotated coordinate system the set of coefficients of a given order are mapped through a unitary transformation based on the generalized binomial function. This representation allows a perceptual classification, including constant patterns (0-D), oriented structures (1-D), and non-oriented structures (2-D). Classification is based on light adaptation and contrast masking properties of the human vision. Finally, classification is improved by means of a probabilistic approach based on Markov Random Fields.

Keywords: SAR image classification, Hermite transform, binomial filters, Markov random fields, multiresolution analysis, steerable transforms, visual perception.

1. INTRODUCTION

Unsupervised image classification schemes often aim at finding differences among a wide variety of image characteristics such as texture, gray level, pattern structure, etc. Some of the more efficient approaches combine statistical methods, such as Markov Random Fields¹ with multiresolution analysis methods such as Wavelets². The former discriminates classes on the basis of a probabilistic model of the local grey level distribution, while the latter assumes that images are conformed by objects of different dimensions and contrasts which are uniquely represented in the scale space. Some of these multiresolution methods are inspired in important properties of human vision such as the Hermite Transform. This transform has been proposed as a robust model for image representation since it carries out similar operations to those present in the primary stages of the human visual system³. These operators are modeled as Gaussian derivatives and have been successfully used in several applications, including restoration⁴, coding⁵ and optic flow estimation⁶. They possess symmetry properties related to translation, rotation and scaling which are ideal for the analysis of the geometric properties of images^{7, 8}. The discrete approach of these operators based on binomial filters satisfies a number of recursive relationships that can be efficiently implemented using adders and delays only. Moreover, these discrete filters present interesting orthogonality properties. In this work we use the discrete version of the Hermite transform (DHT). For the case of SAR imagery, an algorithm for speckle reduction, also based on the Hermite Transform, is used before classification⁴. Our approach, concerned with visual perception states that psychophysical redundancy can be taken care of by eliminating the transform coefficients that do not contribute to the perceived image structure. Furthermore, we state that image structures can be 0D, 1D or 2D, depending on the information lying within the coefficients and certain perceptual thresholds. This paper also presents how this pattern classification can be regularized by means of Markov Random Fields. This approach allows a better estimation of structure parameters, especially in the case of 1D structures where orientation is a critical parameter.

* boris@servidor.unam.mx; phone +52-55-56223016; fax +52-55-561610173; <http://verona.fi-p.unam.mx/amovip.html>

2. THE DISCRETE HERMITE TRANSFORM

The Discrete Hermite Transform (DHT) maps the image data into a set of spectral coefficients from which it is possible to recover the original image:

Let $W_k = \sqrt{2^{-N} C_N^k}$ denote the *symmetric binomial window* and let

$$B_{n,k} = \sqrt{2^{-N}} \Delta^n \{C_{N-n}^{k-n}\},$$

for $k = 0, \dots, N$, denote the *binomial function* of order n , for $n = 0, \dots, N$, where $C_n^k = n!/(n-k)!$ are the binomial coefficients for $k = 0, \dots, n$ and

$$\Delta^n f_k = \sum_{j=0}^n (-1)^{n-j} C_n^j f_{k+j}$$

is the n -th order forward difference of the discrete sequence f_k . Then, the DHT is obtained by convolving the input signal with the binomial filters $D_{n,k} = W_n B_{n,N-k}$ and taking the outputs at points in a given lattice³.

The binomial functions satisfy a number of properties that can be expressed using the matrix notation $\mathbf{B} = [B_{i,j}]_{i,j=0,\dots,N}$, where each row represents a mask of analysis. Therefore, the outputs of the filtered signal by the binomial filters at a generic lattice point can be obtained by applying the matrix $\mathbf{D} = \mathbf{W}\mathbf{B}$ to a vector containing the samples of the signal, where $\mathbf{W} = \text{diag}(W_i)_{i=0,\dots,N}$ is a diagonal matrix of the window function.

The two most appealing properties of the binomial functions are the orthogonality $\mathbf{B}\mathbf{B} = \mathbf{I}$ and the symmetry $\mathbf{W}^2\mathbf{B} = \mathbf{B}^T\mathbf{W}^2$, which leads to the following identity for the binomial filters: $\mathbf{D}^T\mathbf{D} = \mathbf{W}^2$. Thus, if $\mathbf{F}$ is a matrix containing the data of a block of $(N+1) \times (N+1)$ pixels, then $\mathbf{G} = \mathbf{D}\mathbf{F}\mathbf{D}^T$ are the spectral coefficients of the DHT. The DHT is said to be a local expansion of the image. This is expressed by $\mathbf{W}^2\mathbf{F}\mathbf{W}^2 = \mathbf{D}\mathbf{G}\mathbf{D}^T$, which is the windowed block of data.

3. LOCAL ROTATION

The rotation in the DHT domain can be seen as a mapping of the spectral coefficients (for each point in the sampling lattice) through a linear transformation⁵. This mapping is done over the set of coefficients of the same total order by means of the *generalized binomial functions*

$$B_{n,k}^{(N,\theta)} = \tan^{k+n} \theta \Delta^n \{C_{N-n}^{k-n} \cos^{2k} \theta \sin^{N-2k} \theta\}$$

for $0 \leq \theta \leq \pi$. Notice that when the slant parameter θ is set to $\pi/4$ we get the symmetric case, i.e., $B_{n,k} = B_{n,k}^{\pi/4}$. If $\mathbf{g}_n$ for $n = 0, \dots, 2N$ denote the column vectors formed by the anti-diagonals of matrix $\mathbf{G}$, then the coefficients in a coordinate system that is rotated by an angle θ are given by $\mathbf{g}_n^{(\theta)} = \mathbf{R}^{(\theta)} \mathbf{g}_n$, with the rotation matrix $\mathbf{R}^{(\theta)} = \mathbf{W}\mathbf{B}^{(\theta)}\mathbf{W}^{-1}$. Since $\mathbf{R}$ is unitary, the inverse rotation is achieved by applying the transposed matrix $\mathbf{R}^T$. Hereafter, we refer to the rotated coefficients as the matrix $\mathbf{G}^{(\theta)}$ whose anti-diagonals are the vectors $\mathbf{g}_n^{(\theta)}$. In the remaining work we consider the angle $\theta = \arctan(G_{0,1}/G_{1,0})$, which is nothing but the gradient orientation.

4. PATTERN CLASSIFICATION

Since image content is not homogenous, it is convenient to separate the space of blocks into three classes according to the dimensionality of the underlying pattern. The dimensionality of a pattern is directly related to the number of coefficients needed to code the block. For blocks over a constant-luminance background all the AC-coefficients vanish. Therefore, only a DC-coefficient ($G_{0,0}$) is needed to code the block. For blocks containing a strongly oriented structure, only the coefficients $G_{i,0}^{(\theta)}$, for $i = 0, \dots, N$, are required to represent that pattern (for simplicity we omit the superscript hereafter). The third general class comprises all the non-oriented patterns like corners, junctions, dots, etc.

Figure 1. Construction of (a) light adaptation and (b) contrast masking thresholds.

To produce a partition of the block space we assume that each of these three classes can be well described by the following variables:

$$L = G_{0,0} \quad , \quad C = \left[\sum_{i=0}^N \sum_{j=0}^N G_{i,j}^2 - L^2 \right]^{1/2} \quad , \quad \Delta C = \left[C^2 - \sum_{i=1}^N G_{i,0}^2 \right]^{1/2}$$

which stand for mean luminance, contrast, and 1-D residual contrast respectively. Let the three classes be labeled as 0-D, 1-D and 2-D. The classification is done in two steps. First, the 0-D class is separated by means of the following comparison. If $C < k_0 C_{thr}(L)$, then 0-D, otherwise 1-D or 2-D, where $C_{thr}(L)$ is the *light adaptation threshold*⁹. This threshold can be inferred from the image of Fig. 1-(a). The image presents white Gaussian noise with mean varying linearly along the horizontal direction, while the variance varies linearly along the vertical direction. The curve plotted over the image represents a typical detection threshold. This threshold determines the limit beyond which local contrast becomes visually relevant. The light adaptation threshold was fitted by the following model

$$C_{thr} = C_{min} + \left| \frac{L^\alpha - L_{min}^\alpha}{L^\alpha + L_{min}^\alpha} \right|^{1/\alpha}$$

where the exponent α is in the interval $[0,1]$ (typically 0.6). C_{min} is the minimum contrast at luminance level L_{min} that is the level at which the eye has a maximum contrast sensibility. Fig. 1-(a) shows the theoretical curve of the previous equation and the measured curve along the image profile on the image.

Fig. 2 SAR ERS2 image over México City after speckle reduction

In the second step, we separate the 1-D blocks from the remaining blocks by means of the following comparison. If $\Delta C < k_1 \Delta C_{thr}(L, C)$ then 1-D, otherwise 2-D, where $\Delta C_{thr}(L, C)$ is the *contrast masking threshold*¹⁰. Ideally, for one-dimensional patterns all the energy is fully contained along the gradient orientation and the residual contrast becomes zero. In practice, however, there is a residual contrast different from zero even in the cases when oriented structures are well perceived. Contrast masking refers precisely to the reduction in the visibility of a component in the

image due to the presence of another. The contrast masking model used here is based on the works of Legge and Foley^{10,11}:

$$\Delta C_{thr} = \max(C_{thr}, C^{\beta} \cdot C_{thr}^{1-\beta})$$

where the exponent β is in the interval $[0,1]$. Note that if $\beta = 0$ no masking exists and the threshold is the constant C_{thr} . If $\beta = 1$ it results in what is commonly called the *Weber's law*. A typical value of β is 0.7. In Fig. 1-(b) we have plotted a stimulus image built with a sinusoid wave grating plus uncorrelated Gaussian noise. The amplitude of the sinusoid was varied linearly along the vertical axis, while the standard deviation of noise was varied linearly along the horizontal axis. In the figure it is also shown the theoretical curve and the measured values for the image profile shown to the left.

Fig. 3. Pattern classification at high resolution. Left: + (red) signs stand for 0D patterns, $\diamond$ (green) signs for 1D patterns, Δ (blue) signs for 2D patterns. Right: Local pattern orientations. Note that longest vectors correspond to 1D patterns

Fig. 4. Idem as Fig. 3 for medium resolution.

Results of the classification scheme were tested on a 1995 SAR-ERS2 image of Mexico City. Figure 2 shows this image after applying the speckle reduction technique based on the Hermit transform⁴. Test areas are indicated on the image. Figure 3 shows the results of pattern classification and pattern orientation estimation over test area 3. These results were obtained at high resolution, i.e., using a binomial analysis window of 3x3 pixels in the DHT. Figure 4 shows similar results at medium resolution with a 9x9 analysis window. Note that, in accordance with scale space, orientation estimation improves at lower resolution, while localization precision decreases.

5. MARKOVIAN REGULARIZATION

From the results of orientation estimation at high resolution shown in Fig. 3, it is clear that vector field estimation can be improved by regularizing the vector field with a probabilistic approach that takes into account neighborhood spatial relations. In order to estimate the true vector field from the output vector field we follow a similar approach to the problem of image restoration. Let us take a Markov Random Field (MRF) model for piecewise constant image restoration where the goal is to recover true pixel values f , from the observed image values d . In our case, let $d = \{d_1, \dots, d_m\}$ be an observation, i.e., a rectangular array of vector values, $f = \{f_1, \dots, f_m\}$ the set of sampled vectors, which is assumed to be a realization of a MRF, $L = \{l_1, \dots, l_M\}$ the label set composed of vector values and F the set of all possible labelings. According to the Hammersley-Clifford theorem and the theory for Bayes labeling of MRFs, an optimal estimate, known as the MAP estimate, is given by:

$$f^* = \arg \max_{f \in F} P(f | d) = \arg \max_{f \in F} \{p(d | f)P(f)\}$$

In the MAP-MRF labeling, $P(f | d)$ is the a-posteriori distribution of an MRF, and the a-posteriori probability is $P(f | d) \propto e^{-U(f | d)}$, where $U(f | d) = U(d | f) + U(f)$ is the a-posteriori energy. The MAP estimate can then be found by minimizing $f^* = \arg \min_f U(f | d)$, known as the a-posteriori energy.

For piecewise constant surfaces we can define the clique potentials using multi-level logistic (MLL), that is considering the clique potentials

$$V_c(f) = \begin{cases} 0 & \text{if all sites in } c \text{ have the same label} \\ -\xi_c & \text{otherwise} \end{cases}$$

where $\xi_c < 0$ is constant dependent on c .

The statement “all sites in c have the same label”, i.e. all $\{f_i | i \in c\}$ are the same, implies the entire smoothness of labels f on the clique c . Any violation of the entire smoothness incurs in a penalty of the positive number, i.e., $\xi_c < 0$. Since the more likely configurations are those with higher $P(f)$ or lower $U(f)$ values, the MLL model favors smooth f . A special case is such that V_c is nonzero only for the pair-site cliques and zero for all the other types. In this case,

$$V_c(f) = V_2(f_i, f_{i'}) = v_{20}[1 - \delta(f_i - f_{i'})]$$

where $\delta(\cdot)$ is the Kronecker delta function and v_{20} is the penalty for non equal labels on two-site cliques.

The a-priori energy is the sum of all clique potentials, this is,

$$U(f) = \sum_{i \in S} \sum_{i' \in N_i} v_{20}[1 - \delta(f_i - f_{i'})].$$

The likelihood energy $U(d|f)$ can be determined using a general observation model expressed as $d = \varphi(B(f))\Theta e$,

where B is a blurring effect, φ is a transformation, e is the sensor noise and Θ is an operator. These parameters are relevant for image restoration, however, in our case we assume no blurring, nor transformation nor noise, then $U(d|f)$ will be zero and $U(f|d)=U(f)$.

We need to minimize $U(f)$ and since L_d is discrete we deal with a combinatorial problem. The simplest algorithm is steepest local energy descent. ICM (Iterated Conditional Modes) is an algorithm that uses the “greedy” strategy in the iterative local maximization of probability. Given the data d and $f_{S-\{i\}}$ the set of labels at the sites in $S - \{i\}$, the algorithm sequentially updates each $f_i^{(k)}$ into $f_i^{(k+1)}$ by maximizing $p(d_i | f_i)P(f_i | N_i)$, the conditional (posterior) probability with respect to f_i . Maximizing the above is equivalent to minimizing the corresponding posterior potential using the following rule

$$f_i^{(k+1)} \leftarrow \arg \min V(f_i | d_i, f_{N_i}^{(k)}) \text{ where } V(f_i | d_i, f_{N_i}^{(k)}) = \sum_{i' \in N_i} V(f_i | f_{i'}^{(k)}) + V(d_i | f_i).$$

We consider $U(d|f) = 0$, so $V(d|f) = 0$, and the a-posteriori potential

$$V(f_i | d_i, f_{N_i}) = v_{20} \sum_{i' \in N_i} [1 - \delta(f_i | f_{i'})],$$

where $\sum_{i' \in N_i} [1 - \delta(f_i | f_{i'})]$ is the number of neighboring sites whose labels $f_{i'}$ differs from f_i . For discrete L , $V(f_i|d_i, f_{N_i})$ is evaluated for each $f_i \in L$ and the label causing the lowest $V(f_i|d_i, f_{N_i})$ value is chosen as true value for $f_i^{(k+1)}$. When applied to each label in turn, the above defines an updating cycle of ICM.

Then, for the estimation of the true vector field problem, we propose to use the same methodology with the following considerations.

Let $d = \{d_1, \dots, d_m\}$ be an observation representing a rectangular array of vector angles values, $f = \{f_1, \dots, f_m\}$ the set of sampled vector angles, which is assumed to be a realization of a MRF and $L = \{l_1, \dots, l_M\}$ the label set, where labels are vector angle values. We then minimize $V(f_i | d_i, f_{N_i}) = \sum_{i' \in N_i} [1 - \delta(f_i | f_{i'})]$ where $\sum_{i' \in N_i} [1 - \delta(f_i | f_{i'})]$ is the number of neighboring sites whose labels $f_{i'}$ differs from f_i .

Figure 5 shows a crop of the original vector field and the regularized one as a result of applying the above method.

Fig.5 (a)Original vector field, (b) Regularized vector field

6. CONCLUSIONS

We have developed a method for SAR image classification based on the discrete Hermite transform. This transform is a computational efficient model that possesses interesting mathematical and psychophysical properties. In order for the algorithm to be efficient, speckle reduction should be applied before classification. 0D, 1D and 2D patterns are classified by comparing parameters extracted from DHT coefficients, such as mean luminance, contrast and 1D residual contrast with light adaptation and contrast masking thresholds. A Markov random field restoration approach is used to estimate the true vector field of oriented structures. Results of this classification scheme show that 0D patterns correspond to flat surfaces on the ground, such as water reservoirs and large flat surfaces. 1D patterns correspond to streets, large urban constructions, and boundaries between different regions. 2D patterns include inhabiting zones and irregular surfaces. Analysis of the evolution of classes along different spatial resolutions may help classifying zones accordingly: large objects at low resolutions while fine detail at higher resolutions.

ACKNOWLEDGMENTS

This work was partially supported by PAPIIT grant IN107101. SAR ERS2 images were provided by the European Space Agency under contract A03-318. We are also indebted to Centro GEO for providing support and means to carry out this work.

REFERENCES

1. R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields", IEEE Transactions on Acoustics, Speech and Signal Processing 33, 1985, pp. 959-963.
2. M. Unser, "Texture classification and segmentation using wavelet frames", IEEE Transactions on Image Processing 4, 1995, pp. 1549-1560.
3. J.-B. Martens, "The Hermite transform-theory", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 38, 1990, pp. 1595-1606.
4. P. Camarillo-Sandoval, A. Varela-López, B. Escalante-Ramírez "Adaptive Multiplicative-Noise Reduction in SAR Images with Polynomial Transforms", IGARSS '98, IEEE Geoscience and Remote Sensing Society, Proceedings, July 1998, pp. 1171-1173.
5. J. L. Silván-Cárdenas, B. Escalante-Ramírez, "Image coding with a directional-oriented discrete Hermite transform on a hexagonal sampling lattice", Applications of Digital Image Processing XXIV (A.G. Tescher, Ed.), Proceedings SPIE, vol. 4472, 2001, pp. 528-536.
6. B. Escalante-Ramírez, J. L. Silván-Cárdenas, "Motion Analysis and Classification with Directional Gaussian Derivatives in Image Sequences", Advanced Signal Processing Algorithms, Architectures, and Implementations X (F.T. Luk, Ed), Proceedings SPIE, vol. 4116, 2000, pp. 447-453.
7. J. J. Koenderink, A. J. van Doorn, "Generic neighborhood operators", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, 1992, pp. 597-605.
8. Z.-Q. Liu, R.M. Rangayyan, C.B. Frank, "Directional analysis of images in scale space". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, 1991, pp. 1185-1192.
9. H. B. Barlow, Handbook of Sensory Physiology, Dark and Light Adaptation: Psychophysics, (L. Hurvich and D. Jameson, Eds.), New York, 1972.
10. G. E. Legge and J. M. Foley, "Contrast masking in human vision," Journal of the Optical Society of America vol. 70 (12), 1980, pp. 1458-1471.
11. G. E. Legge, "A power law for contrast discrimination," Vision Research, vol. 21, 1981, pp. 457-467.