

"outwards." During grouping, we require that lines can be grouped together only if their gradient directions are toward each other, as shown in Fig. 1.

A. Partial Shape Completion

In real images, lines corresponding to all sides of a building are not always obtained. This is often due to poor contrast. Thus, the grouping may result in some partial boundary shapes of the buildings. We use deformable contours (snakes) [7] for partial shape completion. Deformable contours are active contours of deformable shape, where the final shape is determined by minimization of an energy function. In any approach using deformable contours, the initialization is important, which is two-folds, the initial shape and its spatial position. The spatial position of the initialization is very crucial, while the initial shape itself is not. In our approach, initialization is not a problem, since we have already obtained the partial shapes. The deformable contours are initialized at the locations of the partial shapes and then allowed to deform according to the following energy function:

Let (x_s, y_s) , $s = 1, 2 \dots N$ be the initial position of the snake. The energy function associated with the deformable contour is given by

$$\begin{aligned} E &= E_{snake} + E_{image} + E_{shape}; \\ E_{snake} &= \alpha_1 \sum_{s=1}^N (x_s'^2 + y_s'^2); \\ E_{image} &= -\beta_1 \sum_{s=1}^N |\nabla f(x_s, y_s)|; \quad E_{shape} \\ &= \gamma_1 \sum_{s=1}^N \rho_s^2 (1 - C_s) + \omega_1 C_s \end{aligned} \quad (7)$$

where $f(x_s, y_s)$ is the image intensity at (x_s, y_s) , ρ_s is the curvature at the site s , and C_s is a corner process defined between the snake sites (x_s, y_s) and (x_{s+1}, y_{s+1}) ,

$$\rho_s = \frac{x_s y_s'' - y_s x_s''}{(y_s'^2 + x_s'^2)^{\frac{3}{2}}}; \quad C_s = \{0, 1\}.$$

B. Results

Fig. 2 shows the LAX image that has many buildings of different sizes. Fig. 3 shows the lines with gradient directions. Gradient directions are represented by short projections at the centers of the lines. The result of the grouping is shown in Fig. 4. As seen in Fig. 4, many of the groupings display only partial shapes. The deformable contours approach is used to complete the partial shapes. Fig. 5 shows the final result after partial shape completion. Fig. 6 is a section of the aerial image made available in connection with the ARPA RADIUS project. Fig. 7 shows the lines with their gradients, Fig. 8 shows the result of grouping, and Fig. 9 shows the partial shape completion result.

V. SUMMARY

MRF models have been successfully used for many low-level image processing problems, but not many attempts have been made to use MRF's for high-level problems. Here, we have presented an MRF model for a high-level delineation problem. Using MRF models for high-level problems enables the specification of prior beliefs in a probabilistic framework. Our experiments show that the grouping based on MRF's provides good results on real aerial images.

REFERENCES

- [1] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Stat. Soc.*, vol. 36, pp. 192-326, 1974.
- [2] P. R. Cooper, "Parallel object recognition from structure," Tech. Rep. 301, Dept. of Comput. Science, Univ. of Rochester, 1989.
- [3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-6, pp. 721-741, 1984.
- [4] M. Herman and T. Kanade, "Incremental reconstruction of 3D scenes from multiple, complex images," *Artificial Intell.*, vol. 30, pp. 289-341, 1986.
- [5] A. Huertas and R. Nevatia, "Detecting buildings in aerial images," *Comput. Vision, Graphics Image Processing*, vol. 41, pp. 131-152, 1988.
- [6] R. B. Irvin and D. M. McKeown, "Method for exploiting the relationship between buildings and their shadows in aerial imagery," *IEEE Trans. Syst., Man Cybern.*, vol. 19, pp. 1564-1575, 1989.
- [7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vision*, vol. 4, pp. 321-331, 1988.
- [8] S. Krishnamachari and R. Chellappa, "Delineating buildings by grouping lines," Tech. Rep. CAR-TR 682, Univ. of Maryland, 1993.
- [9] T. Matsuyama and V. S. S. Hwang, *SIGMA: A Knowledge Based Aerial Image Understanding System*. New York: Plenum, 1990.
- [10] J. W. Modestino and J. Zhang, "A Markov random field model-based approach to image interpretation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 14, pp. 606-615, 1992.
- [11] V. Venkateswar and R. Chellappa, "A hierarchical approach to detection of buildings in aerial images," Tech. Rep. CAR-TR 567, Univ. of Maryland, 1991.
- [12] Y. T. Zhou, V. Venkateswar, and R. Chellappa, "Edge detection and linear feature extraction using a 2-D random field model," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, pp. 84-94, 1989.

Multichannel Techniques in Color Image Enhancement and Modeling

I. Pitas and P. Kiniklis

Abstract—We present novel multichannel methods in two target research areas. The first area is color image modeling. Multichannel AR models have been developed and applied to color texture segmentation and synthesis. The second area is color image equalization, which is performed on the three channels *RGB* simultaneously, using the joint pdf. Alternatively, equalization at the *HSI* domain is performed in order to avoid changes in digital image hue. A parallel algorithm is proposed for color image histogram calculation and equalization.

I. MULTICHANNEL AR MODELING

We use the three-channel 2-D AR model of the form

$$\hat{x}[i, j] = \sum_{\substack{m=0 \\ (m,n) \neq (0,0)}}^{p_1} \sum_{n=0}^{p_2} \mathbf{A}[m, n] \mathbf{x}[i-m, j-n] + \mathbf{w}[i, j] \quad (1)$$

in which the model coefficients $\mathbf{A}[m, n]$ are 3×3 arrays, and image pixels $\mathbf{x}[m, n]$ are vectors of length 3. The double summation is

Manuscript received February 18, 1993; revised September 30, 1994. The associate editor coordinating the review of this paper and approving it for publication was Prof. Rama Chellappa.

The authors are with the Department of Informatics, University of Thessaloniki, Thessaloniki, Greece.

Publisher Item Identifier S 1057-7149(96)00128-5.

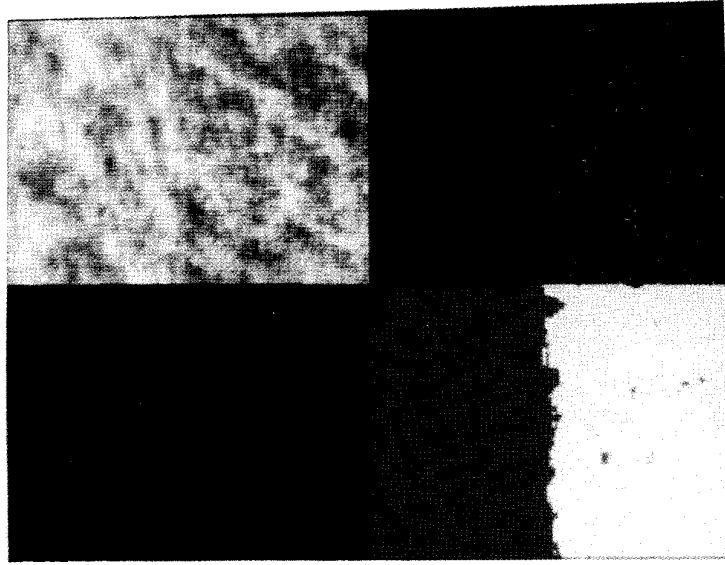


Fig. 1. (Upper left) Color image consisting of two different textured regions; (upper right) multichannel prediction error image; (lower left) prediction error image filtered by a 3×3 moving average filter; (lower right) result of thresholding image 2c.

over a quarter plane window. The three-channel 2-D Yule-Walker equations are

$$\sum_m \sum_n \mathbf{A}[m, n] \mathbf{R}_{xx}[i - m, j - n] = \begin{cases} \mathbf{P}_w, & \text{for } [m, n] = [0, 0] \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

The autocorrelation matrices are given by $\mathbf{R}_{xx}[k, l] = \mathcal{E}\{\mathbf{x}[i + k, j + l] \mathbf{x}^T[i, j]\}$. The solution of (2) gives the coefficients $\mathbf{A}[m, n]$ and the correlation matrix \mathbf{P}_w .

The fitness of the multichannel AR models to color image modeling is measured in terms of the prediction mean square error

$$\text{MSE} = \frac{1}{MN} \sum_i \sum_j \|\mathbf{x}[i, j] - \mathbf{x}_p[i, j]\|^2 \quad (3)$$

where $\mathbf{x}_p[i, j]$ is the predicted image

$$\mathbf{x}_p[i, j] = \sum_m \sum_n \mathbf{A}[m, n] \mathbf{x}[i - m, j - n]. \quad (4)$$

Several simulation experiments have been performed on the well-known color images LENNA, BABOON, and CAR. The multichannel prediction MSE has been found to be 5–30% lower than single-channel MSE. The conclusion from these experiments is that multichannel AR models describe better the color image than single-channel AR models. The price paid for better performance is the computational complexity for the calculation of the autocorrelation matrices and for the solution of the Yule-Walker equations.

We have employed multichannel AR models for segmentation of textured color regions, using techniques similar to the single-channel ones reported in the literature [2], [3]. Classical segmentation techniques fail in such cases. Let us suppose that we know the seeds of the textured region to be segmented from others. Its pixels can be used to estimate autocorrelation matrices. The corresponding Yule-Walker equations can be solved, and the resulting AR model can be applied to the entire image. The prediction error $\|\mathbf{x}[i, j] - \hat{\mathbf{x}}[i, j]\|$ at each image pixel can be used for segmentation by thresholding. The regions having similar (different) texture to the seed region give low (high) prediction error. In most cases, a lowpass filtering of the

prediction error image must be performed prior to its thresholding in order to eliminate spurious pixels. Multiple textured regions can be segmented using more than one seed regions. An application of color texture segmentation based on this method is shown in Fig. 2.

For color texture synthesis, a multichannel AR model is driven by a white Gaussian noise process $\mathbf{w}[i, j]$ having a known correlation matrix \mathbf{P}_w . When the matrices $\mathbf{A}[m, n]$ and \mathbf{P}_w have been derived from an area of random texture, by solving the corresponding Yule-Walker equations, the synthetic texture produced by these matrices will be similar to the original one.

II. MULTICHANNEL HISTOGRAM EQUALIZATION

BW histogram equalization techniques can be applied for color image enhancement in a component-wise manner [4]. However, such a method equalizes the marginal pdf's (histograms) of each primary color R, G, B , and not the joint pdf in the RGB cube, which is a 3-D distribution. Furthermore, it produces unacceptable shifts in hue. We propose a new method for multichannel equalization that equalizes the n -dimensional histogram of a n -channel image.

The histogram of an n -channel image is an empirical n -dimensional pdf. In the continuous case, a random vector $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ has the following pdf:

$$f(\mathbf{X}) = f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} \quad (5)$$

and cumulative distribution function

$$F(\mathbf{X}) = F(x_1, \dots, x_n) = P\{\mathbf{x}_1 \leq x_1, \dots, \mathbf{x}_n \leq x_n\}. \quad (6)$$

The conditional density of the RV's $\mathbf{x}_n, \dots, \mathbf{x}_{k+1}$, assuming $\mathbf{x}_k, \dots, \mathbf{x}_1$, is defined as

$$f(x_n, \dots, x_{k+1} / x_k, \dots, x_1) = \frac{f(x_1, \dots, x_k, \dots, x_n)}{f(x_1, \dots, x_k)}. \quad (7)$$

The corresponding distribution function is obtained by integration

$$\begin{aligned} F(x_n, \dots, x_{k+1} / x_k, \dots, x_1) \\ = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_{k+1}} f(z_n, \dots, z_{k+1} / x_k, \dots, x_1) dz_{k+1} \dots dz_n. \end{aligned} \quad (8)$$

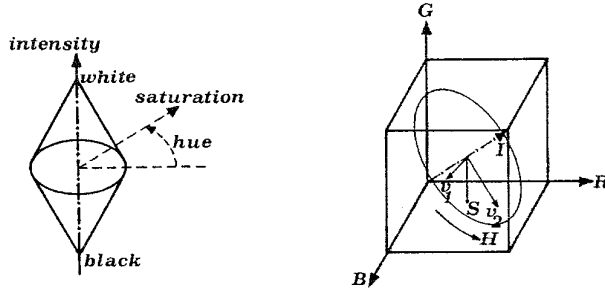


Fig. 2. HSI color coordinates.

The nonlinear pointwise transformation that equalizes the random vector $\mathbf{X} = [x_1, x_2, \dots, x_n]$ can be shown to be [5]

$$y_1 = F(x_1), y_2 = F(x_2/x_1), \dots, y_n = F(x_n/x_{n-1}, \dots, x_1). \quad (9)$$

If we want to transform $\mathbf{X} = [x_1, x_2, \dots, x_n]$ to $\mathbf{Z} = [z_1, z_2, \dots, z_n]$, where \mathbf{Z} must have a certain nonuniform joint pdf $f_Z(\mathbf{Z})$, we first derive the transformations $\mathbf{Y} = T(\mathbf{X})$ and $\mathbf{S} = G(\mathbf{Z})$ that equalize the random vectors \mathbf{X} and \mathbf{Z} . Then, we combine these two transforms into one

$$\mathbf{Z} = G^{-1}[T(\mathbf{X})]. \quad (10)$$

In the digital case, the multichannel histograms take the form

$$\begin{aligned} f(x_{1i}) &= \Pr\{x_1 = x_{1i}\}, \\ f(x_{1i}, x_{2j}) &= \Pr\{x_1 = x_{1i}, x_2 = x_{2j}\}, \dots \end{aligned} \quad (11)$$

where $0 \leq i, j, k, \dots \leq L$. L is the number of discrete levels per image channel, and x_{1i}, x_{2j}, \dots are possible values for the corresponding RV's. The multichannel histogram equalization takes the following form, by combining (9) and (11):

$$\begin{aligned} y_{1i} &= \sum_{m=0}^i f(x_{1m}), \quad y_{2j} = \sum_{m=0}^j \frac{f(x_{1i}, x_{2m})}{f(x_{1i})}, \\ y_{3k} &= \sum_{m=0}^k \frac{f(x_{1i}, x_{2j}, x_{3m})}{f(x_{1i}, x_{2j})}. \end{aligned} \quad (12)$$

Similar formulae can be derived for $n > 3$.

Although the RGB space provides a straightforward way to display color images, it is not always the best choice for processing and analysis. One of its disadvantages is its inability to handle the perceptual properties of colors, such as intensity, hue, or saturation. The HSI color space has primaries H (hue), S (saturation), and I (intensity) and corresponds directly to the perceptual attributes of colors. Fig. 2 shows the HSI coordinates and their relationship to the RGB cube. The intensity is measured along the diagonal of the RGB cube, i.e., the line segment from $(R, G, B) = (0, 0, 0)$ to $(1, 1, 1)$. Hue and saturation are polar coordinates in the plane perpendicular to the I diagonal. The transformation equations from the RGB to the HSI , and vice versa, can be found in [4].

Modification of one or more channels in an RGB image causes shifts in hue, but modification of the I or S components does not. This remark suggests the application of histogram equalization or modification only to I and/or S components. The cone shape of the HSI space suggests nonuniform densities for I and S if an overall uniform density is desired for the entire RGB cube. If this fact is not taken into account when image intensity is equalized, many pixels concentrate near the points $I = 0$ and $I = 1$. The limited color space provided near these points causes distortion of the colors.

The following pdf's fill the HSI color space uniformly, and they can be derived using geometrical concepts.

$$f_I(I) = \begin{cases} 12I^2 & \text{for } 0 \leq I \leq 0.5 \\ 12(1-I)^2 & \text{for } 0.5 \leq I \leq 1 \end{cases} \quad (13)$$

$$f_S(S) = 6S - 6S^2 \text{ for } 0 \leq S \leq 1 \quad (14)$$

$$f_{IS}(I, S) = \begin{cases} 6S & \text{for } S \leq 2I, I \in [0, 1/2] \\ 0 & \text{for } S > 2I, I \in [0, 1/2] \\ 6S & \text{for } S \leq 2(1-I), I \in [1/2, 1] \\ 0 & \text{for } S > 2(1-I), I \in [1/2, 1] \end{cases} \quad (15)$$

Transformations of the form (10) can be used for the modification of I and S either jointly or separately. It has been found in a variety of images that histogram modification toward (13) gives the best color image enhancement. Although modifications toward either (14) or (15) are mathematically correct, they usually lead to large saturation values that are not present in natural scenes.

The 3-D histogram matrix for a color image with 24 b/pixel has 16 777 216 elements and is very sparse. A usual way to represent a sparse matrix is to use a linear list. In our case, every element of this list holds four numbers: three indices (one for each R, G, B color component) and the number of occurrences of this triplet in the image. Only the triplets that are present in an image need to be stored in the list. List processing is easier when it is sorted based on the (R, G, B) triplets. In the following, an efficient algorithm for constructing a sorted RGB histogram list for a given image is described.

Let L represent the number of discrete levels for each component R, G, B (typically $L = 256$). We scan the entire image L times. The r th time corresponds to the level $R = r$ for the red component and has two steps. In the first step, we scan the image, select all the pixels having $R = r$, and construct a 2-D histogram matrix for the G and B components. In the second step, we scan this matrix, and we transfer all its nonzero elements to the corresponding elements of the histogram list having $R = r$. Since we scan the GB matrix in a certain order (e.g., row-wise), there is no need for any further sorting.

This algorithm can be easily parallelized on MIMD machines using the master-slave concept. The image is distributed from the master to every slave. The parts of the list corresponding to different ranges of R can be constructed independently in the slaves. These parts are consecutive and can be concatenated by the master processor. No time-consuming list merging is needed.

For multichannel histogram equalization, we have to transform only the (R, G, B) triplets that exist in the original image and, consequently, only the colors present in its histogram list. The transformation equations from the color (x_{1i}, x_{2j}, x_{3k}) to (y_{1i}, y_{2j}, y_{3k}) are given by (12). $f(\cdot)$ are image histograms, and the indices 1, 2, 3 correspond to R, G, B , respectively. We compute and store the histograms $f(x_{1i})$, $f(x_{1i}, x_{2j})$ of the original image using 1-D and 2-D arrays. We use a list to store the 3-D color histogram $f(x_{1i}, x_{2j}, x_{3k})$. Let M be the number of different values of red R appearing in the list. We traverse the histogram list M times, each corresponding to a different value $R = r$. In each pass, we apply (12) to compute two look-up tables for the transformation of the G, B components. The image can be equalized using these two look-up tables and the first equation of (12). Let us suppose that we have to transform the triplet (x_{1i}, x_{2j}, x_{3k}) . Since $f(x_{1m})$ and $f(x_{1i}, x_{2m})$ are known for every m , y_{1i} and y_{2j} can be easily calculated. Finally, we can compute the sum needed for evaluating y_{3k} by working only on the part of the sorted list that corresponds to $R = x_{1i}$. Thus, it is necessary to have access *only* to the part of the list that has $R = x_{1i}$ in order to transform (x_{1i}, x_{2j}, x_{3k}) . This fact facilitates the parallel implementation of this equalization algorithm on MIMD machines. The construction of the color histogram list can be followed

by equalization with no need to transfer the already computed parts of the list from the slaves to the master processor.

Both histogram list calculation and equalization algorithms have been implemented on a TELMAT T-NODE machine with 4, 8, or 16 transputers T800. In the case of 16 transputers, the speedup obtained over the serial computation ranges from 4.01 ($L = 64$) to 14.87 ($L = 256$). As expected, the speedup is best for 256 levels/component because the computational load is much larger than the communication load between processors. The same observations hold for parallel equalization using the color histogram list, where a speedup of 10.15 has been obtained for $L = 256$ and a farm of 16 transputers.

REFERENCES

- [1] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [2] C. W. Therrien, "Multichannel filtering methods for color image segmentation," in *Proc. IEEE Conf. Comput. Vision Patt. Recogn.*, 1985, pp. 637-639.
- [3] K. L. Rangasami and C. Ramalingam, "Estimation and choice of neighbors in spatial interaction models of images," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 60-72, Jan. 1983.
- [4] W. Niblack, *An Introduction to Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [5] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1984.

A Fast Learning Algorithm for Gabor Transformation

Ayman Ibrahim and Mahmood R. Azimi-Sadjadi

Abstract—An adaptive learning approach for the computation of the coefficients of the generalized nonorthogonal 2-D Gabor transform representation is introduced in this correspondence. The algorithm uses a recursive least squares (RLS) type algorithm. The aim is to achieve minimum mean squared error for the reconstructed image from the set of the Gabor coefficients. The proposed RLS learning offers better accuracy and faster convergence behavior when compared with the least mean squares (LMS)-based algorithms. Applications of this scheme in image data reduction are also demonstrated.

I. INTRODUCTION

The Gabor transform [1]-[11] is viewed as the optimum case of the short time Fourier transform (STFT) in which the window function is chosen to have a Gaussian shape. This choice of the window function in the 2-D Gabor elementary functions guarantees the lower bound of the joint uncertainty, i.e., the 2-D Heisenberg inequality, in the two conjoint spatial-frequency domains. The Gabor analysis is based on projecting a given signal/image onto a family of shifted and modulated Gaussian window functions, which are called the "Gabor elementary functions" or the "Gabor basis functions," and the corresponding projection coefficients are called the "Gabor transform

coefficients." The use of such a transform is motivated by the fact that Gabor elementary functions have optimal localization property [2], [3] in the joint time (or spatial) and frequency domains. This leads to optimal extraction of the textural information from the images, which is an important feature for pattern recognition, segmentation, and image analysis applications. Beside the optimal localization property, other benefits of the Gabor transform include compatibility with mammalian visual systems [2], [3] and energy packing capability, which leads to lower entropy in the transform domain [4]. The deficiency of the Gabor transform, however, is that the elementary functions are not orthogonal. As a result, there is no straightforward method available for extracting these transform coefficients. If they were orthogonal, the extraction of these coefficients could have been done easily by the simple inner product formula [4].

Many approaches have been proposed to find a method for extracting the Gabor transform coefficients [4]-[10]. Bastiaan [5] derived an analytic solution for the 1-D case based on the expansion onto another set of discrete functions that are biorthogonal to the Gaussian elementary functions. This method was extended to the 2-D case by Porat and Zeevi [7]. Daugman [4] proposed a three-layer neural network for extracting the Gabor coefficients. The learning of the neurons is accomplished using a least mean squares (LMS) type algorithm [11]. Teuner and Hosticka [8] presented an algorithm that computes the Gabor transform coefficients using the complex LMS algorithm. Recently, Wang *et al.* [9] proposed a method to calculate the Gabor transform coefficients based on the biorthogonal functions. They used the FFT algorithm for the computation of the Gabor transform coefficients. Yeo [10] proposed a method to calculate the coefficients by multiplying a constant complex matrix and the inverse of a sparse real matrix.

Generally, these methods are based on either finding an analytical solution [5]-[7], [9], [10] or solving a set of normal equations using the LMS algorithm [4], [8]. The analytical solution requires a significant number of computations, and further, the solution may never exist. On the other hand, the main shortcoming of the LMS-based approaches is that the choice of the step size results in a tradeoff between accuracy and speed of convergence [11]. The primary objective of this correspondence is to find a solution to these problems by introducing an adaptive learning for the Gabor transform computation. The proposed algorithm uses the recursive least squares (RLS) learning algorithm instead of the LMS, which converges to an optimal solution in only few iterations. After convergence is achieved, Gabor transform coefficients can be extracted at the weights of the adaptive system. This RLS-based learning algorithm offers better accuracy and faster convergence when compared with the LMS-based algorithm. In addition, it does not have the accuracy-speed trade-off problems of the LMS method and provides better numerical stability compared with the analytical solutions. Simulation results are presented that demonstrate the applications of this method for image dimensionality reduction areas.

II. TWO-DIMENSIONAL GABOR TRANSFORMATION USING RLS LEARNING RULE

The goal of the 2-D Gabor transform is to represent a digital image $f(x, y)$, where x and y represent spatial coordinates, either exactly or in some optimal sense (e.g., minimizing the mean squared error between the reconstructed image and the original image) by projecting it onto a set of 2-D Gabor elementary functions. For a finite extent image $f(x, y)$, $x = 0, 1, \dots, X - 1$; $y = 0, 1, \dots, Y - 1$ partitioned

Manuscript received May 1, 1994; revised March 7, 1995. The associate editor coordinating the review of this paper and approving it for publication was Prof. Rama Chellappa.

The authors are with Department of Electrical Engineering, Colorado State University, Fort Collins, CO 80523 USA.

Publisher Item Identifier S 1057-7149(96)00141-8.