Multiresolution Feature Extraction and Selection for Texture Segmentation

MICHAEL UNSER AND MURRAY EDEN, FELLOW, IEEE

Abstract—**This paper describes an approach to unsupervised segmentation of textured images. Local texture properties are extracted using local linear transforms that have been optimized for maximal texture discrimination. Local statistics (texture energy measures) are estimated at the output of an equivalent filter bank by means of a nonlinear transformation (absolute value) followed by an iterative Gaussian smoothing algorithm. This procedure generates a multiresolution sequence of feature planes with a half octave scale progression. A new feature reduction technique is then applied to the data and is determined by simultaneously diagonalizing scatter matrices evaluated at two different spatial resolutions. This approach provides a good approximation of Fisher's multiple linear discriminants and has the advantage of requiring no *a priori* knowledge. This new feature reduction methods appears to be an improvement on the commonly used Karhunen-Loève transform and allows efficient texture segmentation based on simple thresholding.**

Index Terms—**Feature reduction, filter bank, Karhunen-Loeve transform, local linear transform, (multiple or Fisher's) discriminants, segmentation, texture, threshold.**

I. Introduction

THE unsupervised segmentation of an image into regions that are homogeneous with respect to particular characteristics, such as gray tone or texture, is an important task in many image processing applications. This problem has received considerable attention during the last two decades [1]. Most segmentation techniques have been applied to the simple case for which the definition of "homogeneous" is based on a single gray-level characteristic. These methods are usually classified as thresholding [2], edge detection or gradient-based [3], region growing [4], and hierarchical (or pyramidal) schemes [5].

Some of these techniques have been extended for the use of multiple features and have been applied to the more difficult problem of segmentation based on texture. A natural generalization of applying thresholds to a single gray-level characteristic is segmentation based on clustering or classification in a multidimensional feature space [6]. Several edge oriented methods have been proposed [7], [8], these generally attempt to locate texture edges based on the computation of a multifeature gradient-like operator. Hierarchical approaches using pyramid node linking [9] or applying the split-and-merge algorithm to the cooccurrence matrix [10] have also been described. When compared to gray-level based approaches, these methods are faced with the following difficulties. First, an appropriate set of texture descriptors needs to be selected among the large number of features that may be proposed [11]. Second, textures are undefined at the single pixel level but are always associated with image regions. As a consequence, the image has to be subdivided into elementary regions whose size needs to be adjusted to maintain the statistical fluctuation of the texture features within an acceptable range while preserving spatial resolution insofar as possible. Finally, the dimensionality of the feature space is greater than one. That circumstance requires significantly greater complexity in decision making. Furthermore, it is more difficult, in the absence of *a priori* knowledge, to define satisfactory similarity or grouping criteria in a higher dimensional feature space, due to the likelihood of significant interdependence among the variables.

The approach described in this paper is an attempt to integrate simple and efficient solutions to the three difficulties that have been mentioned. At the most basic level, the choice of texture features computed from a bank of convolution masks is justified by the statistical properties of texture fields [14], as well as by recent findings on visual perception [13]. Such texture features have been used by several authors and have been found to compare favorably with texture measurements based on cooccurrence matrices [14]–[16]. The corresponding texture characterization is reasonably compact, easy to compute, and especially suited for parallel processing [17], [18]. The problem of spatial localization of texture properties is avoided in part by computing all local texture features within overlapping windows of several sizes. This is accomplished efficiently by using an iterative multiresolution approach similar to that described by Burt [19]. The last problem, which relates to the dimensionality of the feature space is solved by using effective feature reduction techniques. As an alternative to the classical Karhunen-Loève transform [20], we introduce an unsupervised feature reduction technique that diagonalizes scatter matrices at two different spatial resolutions. This new method provides a close approximation of Fisher's multiple linear discriminant functions [21], [22] that are optimal for region discrimination but can only be determined when the feature mean vectors and covariance matrices associated with the different image regions are known. In many

Manuscript received October 21, 1987; revised May 11, 1988. This work was supported in part by the N.I.H. Visiting Program.

The authors are with the Biomedical Engineering and Instrumentation Branch, National Institutes of Health, Bethesda, MD 20892.

IEEE Log Number 8928050.

cases, the initial set of texture features can be reduced to a single component that discriminates efficiently between the different texture regions. The obvious advantage of such a reduction is that it offers the possibility to use any of the segmentation techniques designed for the single gray tone characteristic as a subsequent step.

The different components of the resulting system, as well as the notation that is used throughout the presentation, are shown in Fig. 1. The highpass filter preprocessor largely eliminates the gray tone information but preserves the texture properties. For highly textured images the use of gray tone information is not essential and is not considered further in this study. Next, particular texture properties are determined by convolving the preprocessed image $\{x_{k,l}\}$ with a set of masks and computing local statistics from the successive application of a nonlinear operator (rectifier) and an iterative Gaussian smoothing procedure. A statistical justification of this approach as well as guidelines for the selection of suboptimal sets of convolution operators are given in Sections II-A–C. An efficient implementation of the smoothing algorithm that provides local texture feature planes with a half-octave scale progression is briefly described in Section II-D. The following module performs a linear feature reduction by using texture measurements at two successive levels of resolution. This technique, which is described in Section III, has a number of attractive properties such as its good approximation of the optimal supervised solution (multiple discriminants) and its ordering of the reduced components according to their discriminability. In the simple case of two texture regions, the local feature vector is reduced to a single component and an optimal threshold can be determined by using the method presented in Section III-D. Experimental results are presented and discussed in the final part of the paper.

II. Feature Extraction

A fruitful approach to texture characterization is based on the extraction of local texture properties by means of linear filtering operators [14]–[16], [23]–[26]. Recently, a unified treatment of most of these approaches has been proposed based on the notion of a local linear transform (LLT) [14]. In this formulation, the local texture properties are characterized by a set of statistics associated with the channel histograms at the output of a filter bank. An important aspect of this study was the definition of statistically optimal and suboptimal operators for texture analysis and classification. From the reported experimental evaluation and comparison with alternative approaches using correlation or cooccurrence-based measurements [27], it appears that advantages the LLT method may have for texture characterization make it particularly suitable for image segmentation by texture type. It performs texture classification well with a smaller number of texture measurements than are used by other approaches [14]. The method is robust, flexible and specifically designed for parallel and multichannel processing.

In this section, we briefly recall the major results related to this approach. We then describe a practical

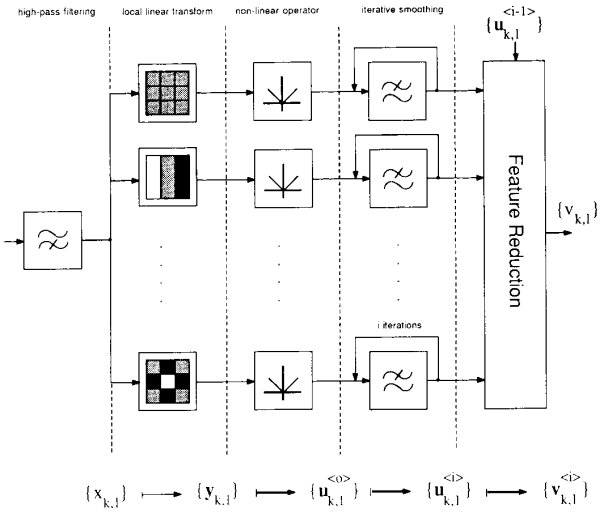

Fig. 1. Block diagram of an image processing system for the segmentation of textured images.

method for estimating local statistics over windows that are successively increased in size by a half-octave scale progression.

A. Local Linear Transform

The principle underlying this approach is to characterize the Nth-order probability density function (pdf) of the pixels in a restricted neighborhood by N first-order pdf's (or histograms) estimated along a set of suitably chosen axes. These projections are obtained by a local linear transformation.

This formulation establishes a correspondence between the original image $\{x_{k,l}\}$ and a N channel multivariate sequence of local neighborhood vectors $\{\boldsymbol{x}_{k,l}\}$ defined for all spatial indices (k, l). The components of the local neighborhood vector $\boldsymbol{x}_{k,l}$ are the sequentially ordered gray level values belonging to an N point neighborhood centered around the spatial position indexed by (k, l). A local linear transform is defined by the matrix relationship:

$$\boldsymbol{y}_{k,l} = \boldsymbol{T}\boldsymbol{x}_{k,l} \tag{1}$$

where $\boldsymbol{T}$ is an $N \times N$ nonsingular transformation matrix.

This equation has two complementary interpretations. First, (1) may be thought of as a rotation in the local pixel space in which the neighborhood vector is defined, or equivalently, as a change of the original coordinate system. Second, since (1) is defined for every couple (k, l), it corresponds to the equation of a filter bank. In this interpretation, the local convolution masks—or elementary structure detectors—are obtained from the row-vectors of the transformation matrix $\boldsymbol{T}$ by using the spatial correspondence rule that defines the local neighborhood vector.

Following this transformation, a region of homogeneous texture is characterized by the set of channel histograms computed at the output of the filter bank [14]. Depending on our interpretation, these quantities may be viewed as a collection of one-dimensional projective

views of the Nth order probability density function of the local neighborhood vector, or as measurements of the relative strength of occurrence of some elementary spatial patterns in the texture field. A more compact (but less complete) representation is given by the variances of these histograms; these are often referred to as "texture energy measures" and were first used by Laws [15].

B. Transform Selection

It is not surprising that performance under this approach depends on the choice of the transformation matrix $\boldsymbol{T}$. The most trivial example is to consider the use the identity matrix or any of its permutations. This particular choice is the least favorable, because the statistics of the initial components of the local neighborhood vector are all identical and contain no neighborhood information.

The problem of the optimal selection of convolution masks has been discussed before; it requires that certain performance criteria for texture analysis and classification be defined [14]. The optimal solution for analyzing a given texture was shown to be the local Karhunen–Loève transform that diagonalizes the spatial covariance matrix. This transform has the remarkable property of producing the channel statistics that are the most different from one another; it also decorrelates the transformed coefficients, thereby justifying the approximation of the Nth order pdf by the product of N first-order pdf's. The ability of the corresponding set of masks to extract the different constituents of a given texture has been demonstrated by Ade [26]. Similarly, in the context of classification, the optimal transform that discriminates maximally between two textures has been determined as the one that simultaneously diagonalizes their spatial covariance matrices [14].

The use of these solutions, however, is restricted in practice because they are texture dependent. They are therefore not applicable to unsupervised texture segmentation. Fortunately, it has been demonstrated that almost equivalent performances could be obtained with suboptimal separable transforms such as the discrete sine (DST), cosine (DCT), Hadamard (DHT), and real even Fourier (DREFT) transforms. Experimental comparison of these transforms has shown that classification performance is not greatly affected by the exact shape of the convolution masks, suggesting that the method is quite robust. The DST was found to perform slightly better for odd dimensions of the local neighborhood (3×3, 5×5), while the DHT (or DREFT) gave slightly better results for even neighborhoods (2×2 and 4×4) [14], [17]. These suboptimal operators have some important advantages, including simplicity of implementation, the availability of fast algorithms and finally the guarantee that the performance is close to optimal for a large variety of textures.

C. Estimation of Local Statistics

Rather than estimating the channel histograms explicitly over elementary image regions and evaluating their corresponding moments, it is advantageous to compute local texture statistics from the succession of nonlinear operators and averaging filters. This procedure produces a sequence of local texture features $\{\boldsymbol{u}_{k,l}\}$ characterizing the texture properties in small regions centered on sample positions indexed by (k, l). An advantage of this formulation is that each component of the local feature vector defines an image, which, as such, can be displayed and processed using conventional methods.

1) Nonlinear Transformation: The purpose of the nonlinear transformation is to provide a transformed feature sequence $\{\boldsymbol{u}_{k,l}^{(0)}\}$ which leads to feature vectors with different mean values for distinct texture regions. This property is essential and is not usually satisfied by the rotated representations of the local neighborhood vector $\{\boldsymbol{y}_{k,l}\}$, especially for a high pass-filtered image within which all clusters are centered around the origin. This latter observation also provides a simple explanation as to why texture segmentation or classification procedures that consider the local individual pixel values as the components of a feature vector are relatively inefficient. It is therefore suitable to select a nonlinear transformation that converts differences in dispersion characteristics into differences in the region mean values. The consecutive application of a low-pass filter reduces the within-region variance and produces texture features that are local estimates of some related channel statistic.

In our system where the input image has been initially high-pass filtered, we have used the absolute value operator, which results in the evaluation of texture measurements that are local estimates of the channel mean deviations. The use of this particular set of features is motivated by the following reasons. First, the mean deviation (δ) provides a characterization of a distribution that is generally similar to that obtained with the standard deviation (σ) [28]. For a given parametric form of a probability density function, δ is usually proportional to σ; for example, for the Gaussian distribution this relation is: $\delta = \sqrt{\pi/2}\ \sigma$. The performance in texture classification is therefore very similar to that obtained with channel variances [17]. Second, the absolute value operator is extremely simple to implement and requires no scaling of the transformed images.

The performance in texture classification could be improved slightly by adding higher order statistics such as the kurtosis and skewness [14]. We feel, however, that this small improvement is generally not worth the added computational cost and the increase in the dimensionality of the feature space. It is usually more cost effective to increase both the number and the size of the convolution masks.

2) Iterative Smoothing: The use of a low-pass filter facilitates estimation of the local mean values of the transformed sequence, $\{\boldsymbol{u}_{k,l}^{(0)}\}$, assuming that the input signals are locally stationary. The impulse response of this filter may be viewed as an estimation window applied to the signal. The choice of a Gaussian window which is commonly used in image processing has several motivations. First, the corresponding low-pass filter provides an excellent compromise between the conflicting requirements of effective stop band rejection and good localization

properties: it is the real operator that has the minimal product of second moments in space and frequency [29]. In addition, when compared to an equivalent rectangular window, the Gaussian window tends to be less sensitive to nearby edges since it gives the greatest weight to the pixels near the sample position and progressively less weight to more distant points. Second, repeated convolution with Gaussian kernels is equivalent to a single convolution with a Gaussian kernel of greater size; this is a direct consequence of the central limit theorem [28]. This property provides a flexible way to increase the size of the observation window progressively by using iterative Gaussian filtering. Third, the isotropic Gaussian filter is the only real bidimensional operator that is both separable in the two principal directions and circularly symmetric. Accordingly, all spatial directions are treated equally even if an efficient implementation is based on successive one-dimensional filtering along the row and columns.

A fast iterative evaluation of Gaussian-like windows can be achieved by using the algorithm described by Burt based on the cascaded convolution of a kernel that is progressively expanded and filled with zeros to provide an octave scale progression [19]. We have modified the initial scheme slightly to accommodate for a half-octave scale progression which we found to be preferable for texture analysis. The corresponding sequence of operations is described in Table I. The successive levels of resolution are computed from a cascaded application of the basic "Gaussian-like" one-dimensional symmetrical operator

$$h_n(k) = a \cdot \delta(k) + b \cdot [\delta(k+n) + \delta(k-n)] + c \cdot [\delta(k+2n) + \delta(k-2n)] \quad (2)$$

along the rows and the columns. $\delta(k)$ is the one-dimensional unit impulse and a, b, and c are the weights of the impulse response, which are usually chosen as $a = 0.4$, $b = 0.25$, and $c = 0.05$ [19]. The condition $(a + 2b + 2c) = 1$, guarantees unity gain at zero frequency and thus insures that the output of the filter provides unbiased estimates of the channel means for locally stationary signals. The spatial variance σ_w^2 of the equivalent impulse response at each level of iteration can be determined by summing the second-order moments of the individual cascaded operators, since all impulse responses are normalized to unity ($\Sigma_{k=-\infty}^{+\infty} g_i(k) = 1$). The size of an equivalent (e.g., with the same second-order moment) window with constant weights is given by $M_{eq} = \sqrt{12}\ \sigma_w$. As shown in Table I, the progression of the size of the estimation window is exactly a factor $\sqrt{2}$ per iteration in the two principal directions.

An interesting case is to consider the input of this system to be white noise with variance σ_{no}^2. After i iterations of the smoothing procedure, the residual variance σ_{ni}^2 can be shown to be equal to

$$\sigma_{ni}^2 = \sigma_{no}^2/(2^i). \quad (3)$$

For a half-octave progression, the variance is thus decreased by a factor of two at each iteration. A similar proportionality can be observed empirically when iteratively smoothing the rectified outputs of the filter bank for most homogeneous texture fields, although the factor of proportionality is usually not precisely two. As mentioned earlier, this variance reduction property is useful for texture segmentation because it will usually improve class or region separability.

TABLE I
SEQUENCE OF CONVOLUTION OPERATIONS AND WINDOW CHARACTERISTICS AT SUCCESSIVE LEVELS OF AN ITERATIVE GAUSSIAN SMOOTHING PROCEDURE WITH A HALF-OCTAVE SCALE PROGRESSION.

iteration	Impulse response	Spatial variance σ_w^2	Equivalent size $M_{eq} = \sqrt{12}\,\sigma_w$
1	$g_1(k) = h_1(k)$	$\sigma_1^2 = 1.02$	$M_1 = 3.50$
2	$g_2(k) = g_1(k) * h_1(k)$	$2\,\sigma_1^2 = 2.04$	$\sqrt{2}\,M_1 = 4.95$
3	$g_3(k) = g_2(k) * h_1(k) * h_1(k)$	$4\,\sigma_1^2 = 4.08$	$2\,M_1 = 7$
4	$g_4(k) = g_3(k) * h_2(k)$	$8\,\sigma_1^2 = 8.16$	$2\sqrt{2}\,M_1 = 9.89$
5	$g_5(k) = g_4(k) * h_2(k) * h_2(k)$	$16\,\sigma_1^2 = 16.32$	$4\,M_1 = 14$
...	.....	....	....
2n	$g_{2n}(k) = g_{2n-1}(k) * h_{2^{n-1}}(k)$	$2^{2n-1}\,\sigma_1^2$	$2^{n-1}\sqrt{2}\,M_1$
2n+1	$g_{2n+1}(k) = g_{2n}(k) * h_{2^{n-1}}(k) * h_{2^{n-1}}(k)$	$2^{2n}\,\sigma_1^2$	$2^n\,M_1$

Since the sequences $\{u_{k,l}^{\langle i \rangle}\}$ are band-limited, the amount of computation and storage requirement may be reduced by resampling the data by a factor $\sqrt{2}$ at each iteration (or 2 every other iteration) and generating a Gaussian pyramid similar to that described in [19]. By interpolation, the data at any level of resolution can be recovered with minimal loss.

III. Feature Reduction and Decision

In this section, we investigate the use of linear feature reduction methods that compress the set of initial feature planes into many fewer components while retaining the information for optimal discrimination between texture regions. We first consider the case of supervised classification and review the properties of the multiple discriminant representation that provides an upper theoretical bound for a common criterion for class separability; namely, the ratio of the between-region to within-region variances [21], [22], [30]. In the case of unsupervised segmentation, a somewhat less satisfactory feature reduction is obtained with the Karhunen–Loève transform. As an alternative to this classical approach, we present a multiresolution feature reduction technique that provides a significantly better approximation of the optimal solution by combining the information of two different levels of spatial resolution. Finally, we describe briefly a thresholding technique that optimally selects a partition of a reduced one-component feature space by maximizing the ratio of the between and within-region variances.

A. Optimal Supervised Feature Reduction

Let R_0 denote the set of all spatial indexes contained within the boundaries of the image to be segmented. R_0 is assumed to be partitioned into r disjoint regions of different but homogeneous texture: $R_0 = R_1 \cup R_2 \cdots \cup R_r$.

The number of pixels in each of those regions are $n_1, \cdots, n_r$, respectively, and satisfy $n_1 + n_2 + \cdots n_r = n_0$ where n_0 is the total number of pixels. Given a sequence of local texture features $\{\boldsymbol{u}_{k,l}\}$, the decision procedure should be optimized to produce a partition $(R_1', R_2', \cdots, R_r')$ that is as close as possible to the ideal on $(R_1, R_2, \cdots, R_r)$. As pointed out previously, this task is simplified when the N dimensional feature vectors $\{\boldsymbol{u}_{k,l}\}$ are projected on a set of $M < N$ (ideally, $M = 1$) axes specified by the basis vectors $\boldsymbol{t}_1, \cdots \boldsymbol{t}_M$

$$\boldsymbol{v}_{k,l} = [\boldsymbol{t}_1 \cdots \boldsymbol{t}_M]^T \boldsymbol{u}_{k,l}. \tag{4}$$

In order to be efficient, this dimensionality reduction must globally preserve class (or texture region) separability. This condition implies that the components of $\{\boldsymbol{v}_{k,l}\}$ should individually be more discriminative than the original texture features.

Let $t_{k,l} = \boldsymbol{t}^T \boldsymbol{u}_{k,l}$ where $\boldsymbol{t}$ is a given vector of coefficients, denote an arbitrary linear combination of the original feature vector $\boldsymbol{u}_{k,l}$. Assuming that the texture regions are given, as might be the case in a controlled experiment, the between to within variance ratio (β) is computed as [30]

$$\beta = \frac{\boldsymbol{t}^T \boldsymbol{S}_b \boldsymbol{t}}{\boldsymbol{t}^T \boldsymbol{S}_w \boldsymbol{t}} \tag{5}$$

where $\boldsymbol{S}_b$ and $\boldsymbol{S}_w$ are the so called between-region and within-region scatter matrices of the initial sequence of feature vectors. These matrices are defined by

$$\boldsymbol{S}_w = \sum_{i=1}^{r} \sum_{(k,l)\in R_i} n_i (\boldsymbol{u}_{k,l} - \boldsymbol{m}_0)(\boldsymbol{u}_{k,l} - \boldsymbol{m}_0)^T \tag{6}$$

$$\boldsymbol{S}_b = \sum_{i=1}^{r} n_i (\boldsymbol{m}_i - \boldsymbol{m}_0)(\boldsymbol{m}_i - \boldsymbol{m}_0)^T \tag{7}$$

where $\boldsymbol{m}_0$ is the global mean vector and $\boldsymbol{m}_1, \cdots, \boldsymbol{m}_r$ are the mean feature vectors for each texture region

$$\boldsymbol{m}_i = \frac{1}{n_i} \sum_{(k,l)\in R_i} \boldsymbol{u}_{k,l} \quad (i = 0, 1, \cdots, r). \tag{8}$$

It is well known that the sum of the between and within scatter matrices is equal to the total scatter matrix $\boldsymbol{S}_t$ [30]

$$\boldsymbol{S}_t = \sum_{(k,l)\in R_0} (\boldsymbol{u}_{k,l} - \boldsymbol{m}_0)(\boldsymbol{u}_{k,l} - \boldsymbol{m}_0)^T = \boldsymbol{S}_w + \boldsymbol{S}_b. \tag{9}$$

In a practical segmentation problem, $\boldsymbol{S}_t$ is the only accessible measure because the determination of $\boldsymbol{S}_w$ and $\boldsymbol{S}_b$ requires that the texture regions be known.

The optimal transform that maximizes (5) is well known in the context of multiple discriminant analysis [22], [30] and is found by solving the generalized eigenvalue problem

$$\boldsymbol{S}_b \boldsymbol{t} = \beta \boldsymbol{S}_w \boldsymbol{t}. \tag{10}$$

The maximum of β is given by the largest eigenvalue and is achieved when the data are projected on the axis specified by the corresponding eigenvector. The complete set of eigenvectors with nonzero eigenvalue defines a non-orthogonal transform—the multiple discriminant transform (MDT)—that is optimal for a relatively large family of separability criteria defined for more than one component [31]. The dimensionality of this subspace is at most equal to $\min\{N, r - 1\}$, since the rank of $\boldsymbol{S}_b$ cannot exceed $r - 1$. An important property of the data representation obtained with the MDT is that it is invariant to any nonsingular linear transformation of the initial feature vector. It is worthwhile mentioning that this representation allows an exact and particularly efficient implementation of the minimum Mahalanobis distance classifier, which performs a minimum error (or Bayesian) classification for multivariate Gaussian distributed feature vectors with equal covariance matrices. This property is remarkable as the dimensionality of the MDT representation (for example, $M = 1$ for two texture regions) is generally much lower than the dimensionality of the original feature space.

Although the usefulness of the MDT in our problem is mainly theoretical, it will be used later on for comparing the performance of more practical methods that are discussed next.

B. Unsupervised Feature Reduction

Unlike the MDT representation, the methods discussed below require no *a priori* knowledge (e.g., the scatter matrices $\boldsymbol{S}_b$ and $\boldsymbol{S}_w$) and are therefore directly applicable to unsupervised segmentation problems.

1) The Karhunen-Loève Transform: The Karhunen-Loève transformation (KLT) provides a standard feature reduction method [20], [30]. It has been widely used in the context of multichannel image processing because of its global decorrelating properties. This transformation is defined from the eigenvectors of the total scatter matrix $\boldsymbol{S}_t$ given by (9). The KLT is optimal in the sense that it provides the minimum error projection of the data in a subspace of dimension $M < N$. However, there is no general guarantee that such a representation will be the best for class discrimination. In unsupervised classification problems, it is commonly used because representations such as MDT are generally not accessible and because it often achieves appreciable noise reduction.

2) Approximating the Multiple Discriminants Representation: An estimate of the MDT is obtained by taking advantage of the availability of feature planes at different spatial resolutions. Our method uses two sets of local feature planes $\{\boldsymbol{u}_{k,l}^{\langle i \rangle}\}$ and $\{\boldsymbol{u}_{k,l}^{\langle j \rangle}\}$ computed at iterations i and j ($i > j$) of the smoothing procedure. It is based on two hypotheses. The first one is that between-scatter matrices are almost the same at both resolutions: $\boldsymbol{S}_b^{\langle i \rangle} \cong \boldsymbol{S}_b^{\langle j \rangle}$. It appears to be a reasonable assumption for the only possible difference will derive from smoothing border regions, which usually account for a small portion of the total image area. The second hypothesis is that MDT is not changed significantly from one smoothing iteration to the other. This last condition is generally well satisfied, especially when i and j are not too different, for smoothing tends to reduce the within-region variance isotropi-

cally. We can then apply the following theorem which demonstrates that the MDT solution can be computed by simultaneously diagonalizing the total scatter matrices $S_t^{\langle i \rangle}$ and $S_t^{\langle j \rangle}$. It should be mentioned that this technique usually provides a good approximation of the MDT matrix but does not explicitly estimate the eigenvalues: $\{\beta^{\langle i \rangle}, i = 1, M\}$ or $\{\beta^{\langle j \rangle}, i = 1, M\}$.

Theorem: If t is a N dimensional vector such that

$$S_b^{\langle i \rangle} t = \beta^{\langle i \rangle} S_w^{\langle i \rangle} t = S_b^{\langle j \rangle} t = \beta^{\langle j \rangle} S_w^{\langle j \rangle} t \quad (11)$$

where $\beta^{\langle i \rangle}$ and $\beta^{\langle j \rangle}$ are two unknown nonzero scalars and where $S_b^{\langle i \rangle}$, $S_w^{\langle i \rangle}$, $S_b^{\langle j \rangle}$, and $S_w^{\langle j \rangle}$ are $N \times N$ positive definite matrices, then t also satisfies the following equation:

$$S_t^{\langle i \rangle} t = \lambda' S_t^{\langle j \rangle} t \quad (12)$$

where $S_t^{\langle i \rangle} = S_b^{\langle i \rangle} + S_w^{\langle i \rangle}$ and $S_t^{\langle j \rangle} = S_b^{\langle j \rangle} + S_w^{\langle j \rangle}$.

Proof: The two halves of (11) taken separately imply that

$$[S_b^{\langle i \rangle} + S_w^{\langle i \rangle}] t = (\beta^{\langle i \rangle} + 1) S_w^{\langle i \rangle} t \quad (13)$$

$$[S_b^{\langle j \rangle} + S_w^{\langle j \rangle}] t = (\beta^{\langle j \rangle} + 1) S_w^{\langle j \rangle} t. \quad (14)$$

Furthermore, it is straightforward to show that

$$S_w^{\langle i \rangle} t = (\beta^{\langle j \rangle} / \beta^{\langle i \rangle}) S_w^{\langle j \rangle} t$$

which may be directly substituted in (13). The next step is to isolate $S_w^{\langle j \rangle} t$ on the right-hand side of (13) which yields

$$S_w^{\langle j \rangle} t = \frac{(\beta^{\langle i \rangle} / \beta^{\langle j \rangle})}{(\beta^{\langle i \rangle} + 1)} [S_b^{\langle i \rangle} + S_w^{\langle i \rangle}] t$$

and to proceed in the same way for (14)

$$S_w^{\langle j \rangle} t = \frac{1}{(\beta^{\langle j \rangle} + 1)} [S_b^{\langle j \rangle} + S_w^{\langle j \rangle}] t.$$

We finally complete our proof by equating those two expressions and rewriting the resulting equation as

$$[S_b^{\langle i \rangle} + S_w^{\langle i \rangle}] t = \lambda' [S_b^{\langle j \rangle} + S_w^{\langle j \rangle}] t$$

where λ' is given by

$$\lambda' = \frac{1 + 1/\beta^{\langle i \rangle}}{1 + 1/\beta^{\langle j \rangle}}. \quad (15)$$

For practical purposes, note that the result of processing, $\{v_{k,l}\}$, is independent of the order in which the different operations (smoothing and linear transformation) are performed. This is a direct consequence of 1) the linearity of these operations, and 2) the parallelism by which the smoothing operators are applied to all channels simultaneously.

The solution of (12) defines the approximate multiple discriminant transform (AMDT). An efficient procedure to compute this solution is based on the property that the AMDT is independent of any linear nonsingular transformation of the feature vector. The sequence of operations is the following. First, at iteration j, we replace the original feature sequence by its standardized KLT where the rotated components are normalized by the square-root of their corresponding eigenvalue. At this stage, the scatter matrix of the transformed data is proportional to the identity matrix. Smoothing is then performed up to iteration level i. The AMDT is now given by the KLT of the corresponding sequence, since the succession of this and of the previous transformations diagonalizes both scatter matrices $S_t^{\langle i \rangle}$ and $S_t^{\langle j \rangle}$. The global transformation is thus given by:

$$T_{\text{AMDT}} = V_{ij} \cdot (\Lambda_j^{-1/2} U_j) = V_{ij} U_{sj} \quad (16)$$

where U_j is the matrix of eigenvectors of $S_t^{\langle j \rangle}$, Λ_j the diagonal matrix of corresponding eigenvalues, and V_{ij} the matrix of eigenvectors of the transformed scatter matrix $(U_{sj} S_t^{\langle i \rangle} U_{sj}^T)$. It can be verified that the row-vectors of T_{AMDT} satisfy (12).

C. Threshold Selection

We assume that the sequence of initial texture features $\{u_{k,l}\}$ has been reduced to a single component $\{v_{k,l}\}$ that is optimized for maximum discrimination between texture regions. In this case, the final decision can be reached by simple thresholding. In order to be consistent with the previous development, we describe a procedure for partitioning the feature space in a way that maximizes the ratio of the between-region and within-region variances.

Let us assume that $\{v_{k,l}\}$ has been quantified with $q_{\max}$ equidistant discrete levels which we denote by: $\{v_q, q = 1, \cdots, q_{\max}\}$. The histogram over the entire image is $\{N(v_q), q = 1, \cdots, q_{\max}\}$ where $N(v_q)$ is number of pixels with feature value v_q. In the simplest case of two texture regions, segmentation is performed on the basis of a single threshold T. Each value of T defines two complementary decision regions $Q_1(T) = \{q, v_q \leq T\}$ and $Q_2 = \{q, v_q > T\}$ corresponding to a specific partition of R. Following some simple algebraic manipulations, we find that the separability criterion defined by (5) may be expressed as a function of T

$$\beta(T) = \frac{2 n_1(T) n_2(T)}{n_1(T) + n_2(T)} \times \frac{[m_1(T) - m_2(T)]^2}{n_1(T) s_1^2(T) + n_2(T) s_2^2(T)} \quad (17)$$

where $n_i(T)$ $(i = 1, 2)$ are the number of pixels in the that regions defined by the threshold, $m_i(T)$ and $s_i^2(T)$, $(i = 1, 2)$ are the corresponding estimated means and variances. These quantities are computed from the gray-level histogram

$$n_i(T) = \sum_{q \in Q_i(T)} N(v_q), \qquad (i = 1, 2) \quad (18)$$

$$m_i(T) = \frac{1}{n_i(T)} \sum_{q \in Q_i(T)} N(v_q)\, v_q, \qquad (i = 1, 2) \quad (19)$$

$$s_i^2(T) = \frac{1}{n_i(T)} \sum_{q \in Q_i(T)} N(v_q) \cdot [v_q - m_i(T)]^2, \qquad (i = 1, 2). \quad (20)$$

It is thus possible to compute $\beta(T)$ for all admissible values of T and select the optimum threshold $T^*: \beta(T^*) = \max\{\beta(T)\}$. This computation is relatively inexpensive since the quantities defined by (18) to (20) may be updated recursively for successive values of T.

This approach may be generalized for more than two regions, although the exhaustive evaluation of the performance criterion for all possible threshold configurations is unrealistic when $r > 3$. When more than two thresholds are required, the optimization should probably be performed using iterative searching methods.

IV. Results and Discussion

A. Examples and Experimental Evaluation

For quantitative evaluation, we have used 128 × 128 test images with two predefined texture regions that were created by combining original texture images taken from [32]. The original 8 bits/pixel texture images were preprocessed to compensate for transducer nonuniformities and had their histograms equalized with 32 equiprobable gray-level values [14]. The use of such texture composites, as opposed to real world images, has at least the three following advantages. First, the image regions are unambiguously defined, which allows an objective performance evaluation. Second, due to preprocessing, the different texture regions have identical first-order statistics (gray-level histograms), which guarantees that segmentation is achieved using higher order texture properties exclusively. Third, the use of computer generated images allows us to vary important structural and geometrical parameters (e.g., use of different textures, size or the shape of the image-regions) in order to investigate their influence on the performance of the algorithm.

A typical illustration of the way the algorithm operates is given by the series of images in Fig. 2. All images that are shown in this and in the following figures had their gray scale linearly expanded to fill the range of the display. The test image created from the Brodatz D57 (grass) and D9 (paper) textures, and the masks defining the texture regions are displayed in Fig. 2(a) and (b), respectively. The input image was filtered using the local 2 × 2 discrete Hadamard transform and thereafter rectified. The corresponding channels are represented in Fig. 2(c1)–(c4). The first operator ($\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix}$) is a low-pass filter while the three others are vertical ($\begin{vmatrix} 1 & -1 \\ 1 & -1 \end{vmatrix}$), horizontal ($\begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix}$), and diagonal ($\begin{vmatrix} 1 & -1 \\ -1 & 1 \end{vmatrix}$) edge detectors, respectively. Two global indicators of individual channel performance have been included: η_i is the current relative energy contribution in channel i, and β_i is the corresponding ratio of the between- and within-region variance. This latter quantity was computed using the predefined regions displayed in Fig. 2(b), and provides an objective measure of the individual texture discrimination power of channel i (c.f. 3.1). The channel histograms are also represented.

For the initial sequence $\{u_{k,l}^{\langle 0 \rangle}\}$, the discrimination power is very small, although it can be seen that the texture regions differ slightly in their mean values. The effect of smoothing is to decrease the within-region feature

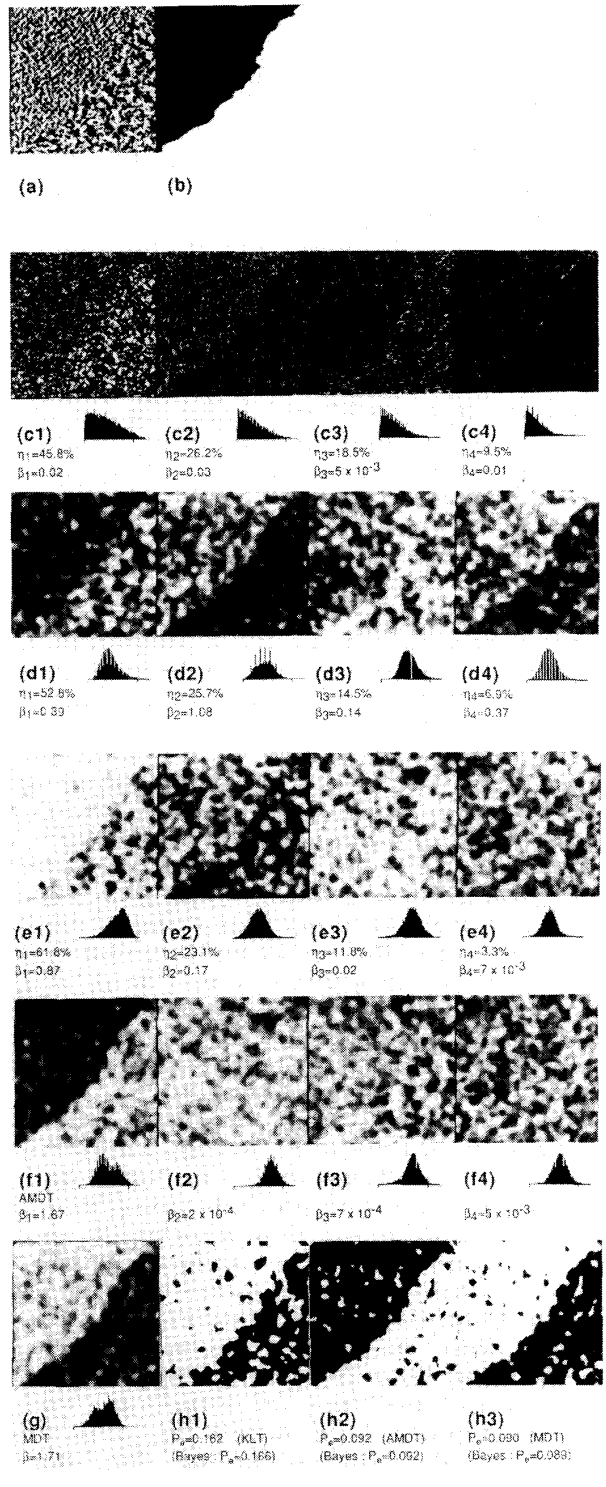

Fig. 2. Illustration of the segmentation algorithm. (a): 128 × 128 test image created using D57 and D9 Brodatz textures, (b) definition of image regions, (c1–4): filtered channels using masks obtained from the 2 × 2 DHT, (d1–4): feature planes at iteration 3 of the smoothing algorithm ($M_{eq} = 7$), (e1–4): KLT rotated feature planes, (f1–4) AMDT rotated feature planes, (g) MDT reduced feature sequence, (h1–3) binary result of segmentation obtained by thresholding the components displayed in (e1), (f1), and (g), respectively.

variance as illustrated in Fig. 2(d1)-(d4). They display the sequence of local texture features $\{u_{k,l}^{\langle 3\rangle}\}$, after 3 iterations of the Gaussian filter. The equivalent diameter of the spatial window is approximately 7. Channel No. 2 is the most discriminative with $\beta_2 = 1.08$ and has a histogram that is slightly bimodal. Fig. 2(e1)-(e4) show the rotated feature planes obtained using the KLT, ordered according to their relative energy contribution. In this particular case, the first eigencomponent turns out to be the most discriminative with $\beta_1 = 0.87$ but is not better than channel 2 taken on its own. The rotated feature planes obtained with the AMDT which simultaneously diagonalizes the scatter matrices at iteration 2 and 3 (e.g., $S_u^{\langle 2\rangle}$ and $S_u^{\langle 3\rangle}$) are represented in Fig. 2(f1)-(f4). The first component is the only interesting one with $\beta_1 = 1.67$ and provides an excellent approximation of the optimal MDT solution [Fig. 2(e)] computed using the procedure described in Section III-A, ($\beta^* = 1.71$). In this example, as well as in all other cases that we have considered using different Brodatz textures, the discrimination power of the first component of the AMDT solution was always found to be superior to any individual channel or to any component of the KLT. The texture regions usually appeared to be more nearly uniform and the modes in histogram were always more pronounced. The binary images obtained by thresholding the first components of the KLT, and AMDT and the theoretical MDT with a threshold calculated according to the technique described in Section III-A are displayed in Fig. 2(h1)-(h3), respectively. The corresponding probabilities of errors are shown. They are very close to the minimum values (given in parentheses) which could theoretically be obtained using an optimal Bayesian (minimum error) threshold. Again, the results obtained with the AMDT are very close to those obtained with the MDT and are superior to those of the KLT.

To quantitate the performances of the algorithm as a function of the size of the observation window, we have used the β_i's for the most discriminative channels associated with the initial representation, the KLT, and AMDT and the optimal MDT, which is used as a theoretical upper bound for performance. We have also computed the minimum error probabilities obtained by thresholding in an optimal way the first components of the different feature reduction techniques. These measures are represented as functions of the equivalent window size (diameter) of the Gaussian observation window whose increase with the number of iterations is given by Table I. Unless otherwise stated, the AMDT is computed by simultaneously diagonalizing the scatter matrices $S_u^{\langle i\rangle}$ and $S_u^{\langle i-1\rangle}$ obtained at two successive iterations. Fig. 3 displays these results for the test image in Fig. 2 in the case of a 4 channel (2 × 2 DHT) and 9 channel (3 × 3 DST) feature extraction. Interestingly enough, the reduced set of 2 × 2 operators performs nearly as well as the larger set of 3 × 3 operators. For small window sizes, the performances of the AMDT and the optimal MDT are almost undistinguishable and always significantly better than those of the KLT. For this particular test image, the optimal window size for minimum error thresholding can be seen to be approximately 14 (5 iterations). Beyond this limit, the performance tends to degrade progressively. Apparently, spatial averaging smears out interregional sharpness more than it diminishes the within-region variances. The interpretation is further supported by noticing that up to this optimal window size the increase in B/W ratio for all methods is approximately linear, as would be predicted by (3), until there is a change in tendency and that the curves saturate progressively. An important phenomenon, which has also been observed in several other cases not shown here and which is particularly clear in the third graph in Fig. 3, is that the approximation provided by the AMDT tends to be less satisfactory for spatial windows that are greater than the optimal size. This observation is not very surprising since the basic hypothesis on which the method is based, namely that $S_b^{\langle i\rangle} = S_b^{\langle i-1\rangle}$, can be expected to break down progressively as the averaging window is increased.

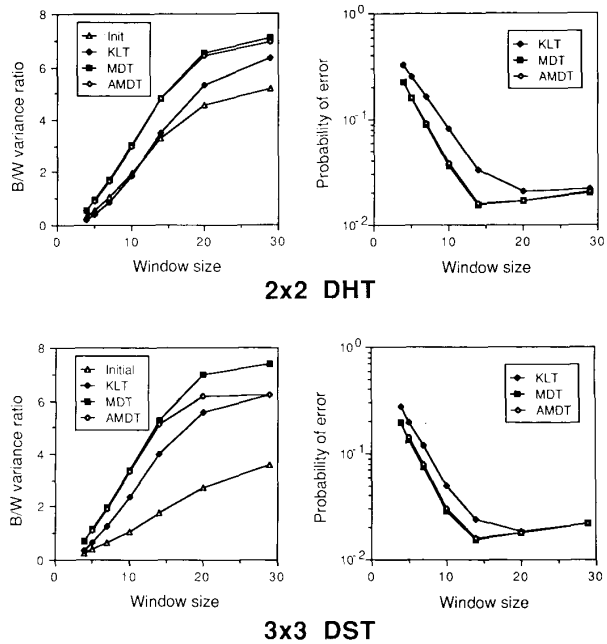

Fig. 3. Performance of the most discriminative component of various feature vector representations as a function of the size of the spatial estimation window for the test image in Fig. 2: (a)-(b) separability measures and segmentation errors when using 4 texture features computed from the 2 × 2 DHT; (c)-(d) separability measures and segmentation errors when using 9 texture features computed from the 3 × 3 DST.

Among the various texture composites that we have investigated, the previous example is representative of a situation in which the KLT performs quite well. However, as illustrated by the next example (Figs. 4 and 5), this is not always the rule. As in the previous case, the AMDT performs nearly as well as the optimal MDT. The performance of KLT, in contrast, is surprisingly bad (roughly 4-5 times worse). As it may be seen from Fig. 5, the optimal window size for minimum error thresholding with the AMDT (or MDT) is now $M_{eq} = 20$ corresponding to

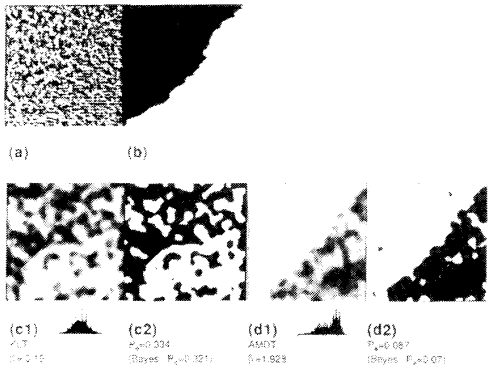

Fig. 4. Example of texture segmentation. (a) 128 × 128 test image created using D29 and D84 Brodatz textures, (b) definition of image regions, (c1–2) first component of the KLT and corresponding binary segmentation, (d1–2) first component of the AMDT and corresponding binary segmentation. Texture features: 4 mean deviations of the 2 × 2 DHT computed using 4 iterations of the smoothing procedure ($M_{eq} = 10$).

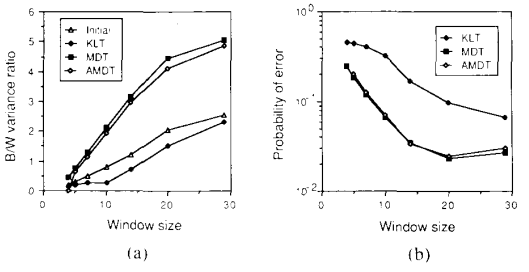

Fig. 5. Separability measures (a) and segmentation errors (b) obtained with the most discriminative component of various feature vector representations as a function of the size of the spatial estimation window for the test image in Fig. 4(a). Texture features: 4 mean deviations of the 2 × 2 DHT.

6 iterations of the smoothing algorithm. The comparatively poor performance of the KLT results in a much less satisfactory feature reduction and texture segmentation as shown in Fig. 4, which displays the reduced components of all three methods after 4 iterations ($M_{eq} = 10$) and their corresponding binary images obtained by applying the threshold maximizing the criterion defined by (17). For this particular image, we were also surprised to find that the most discriminative component of the KLT did not always correspond to the largest eigenvalue. For example, at iteration 2, we found the following values of the performance measures: $\beta_{KLT-1} = 0.012$, $\beta_{KLT-2} = 0.1177$, $\beta_{KLT-3} = 0.1859$, $\beta_{KLT-4} = 0.1914$, which have to be compared to ($\beta_{u1} = 0.005$, $\beta_{u2} = 0.0004$, $\beta_{u3} = 0.3012$, $\beta_{u4} = 0.2529$) for the initial representation and ($\beta_{AMDT-1} = 0.6483$, $\beta_{AMDT-2} = 0.013$, $\beta_{AMDT-3} = 0.017$, $\beta_{AMDT-4} = 0.011$) for the AMDT. These results clearly indicate a reversed tendency; the smallest eigenvalue of the scatter matrix corresponds to the most discriminative component! An explanation for this unexpected result lies in the fact that the most discriminative channels for this particular image (e.g., channel No. 3 and 4) only account for a relatively small proportion of the total energy and, as a consequence, do not contribute much to the first component of the KLT.

In order to investigate the dependence of the algorithm on geometrical parameters of the texture regions, we created test images with a "cross-like" central region of variable size using the same textures as in the first experiment. An example of such an image with a central region occupying 20 percent of the total area is shown in Fig. 6. The result of the unsupervised segmentation after 4 iterations of the smoothing procedure ($M_{eq} = 10$) using the KLT and the AMDT are displayed in Fig. 6(c) and (d), respectively. The effect of varying the size of the central region is illustrated by the graphs in Fig. 7. Once again, the performance of the MDT and AMDT are almost undistinguishable and always superior to that of the KLT. The general shape of the variance ratio curve in a given feature channel can be predicted if one assumes that the mean feature values in the different texture regions are approximately independent of the region size and that the variances are roughly the same over the entire image. In such a case, β can be shown to be proportional to $(A_1A_2)/(A_1 + A_2)$ where A_1 and A_2 are the areas of both texture regions. In the present example, this relation is relatively well satisfied (<2 percent relative error) for the most discriminative channel of the initial representation as well as for the MDT or AMDT. This also means that the relative superiority of the AMDT (or MDT) in performance over that of the initial representation is independent of the size of the central region. Obviously, this proportionality relationship does not hold for the KLT which tends to be less performant on both ends of the scale. The second graph in Fig. 6 shows the proportion of segmentation errors as a function of size of the central region and essentially compares the performance of the thresholding technique described in Section III-C with the optimal minimum error Bayesian strategy. The performance is close to optimal for a large central region but degrades progressively as the region becomes smaller. This observation is not surprising since the histograms are no longer bimodal for small texture regions and that it is then extremely difficult to select a satisfactory threshold. A simple way around this problem may be to compute the threshold adaptively over image regions of smaller size or to use a pyramid node linking approach as suggested in [9].

B. Discussion

The examples that we have considered indicate that the segmentation algorithm should perform quite well for a relatively large class of natural textures. The global performance of the procedure, however, has been found to depend on several factors. The most important one is structural and relates to the way the different image regions are differentiable based on their local texture properties. A second aspect is related to the geometry of the texture regions, mainly to the relative proportion of border points and also, to some extent, to the size of the regions.

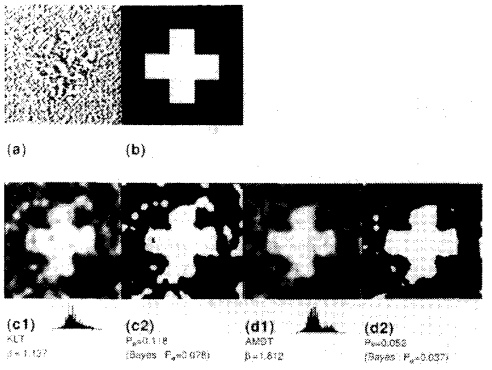

Fig. 6. Example of texture segmentation. (a) test image with a central cross (20 percent of total area) created using D57 and D9 Brodatz textures, (b) definition of image regions, (c1-2) first component of the KLT and corresponding binary segmentation, (d1-2) first component of the AMDT and corresponding binary segmentation. Texture features: 4 mean deviations of the 2 × 2 DHT computed using 4 iterations of the smoothing procedure ($M_{eq} = 10$).

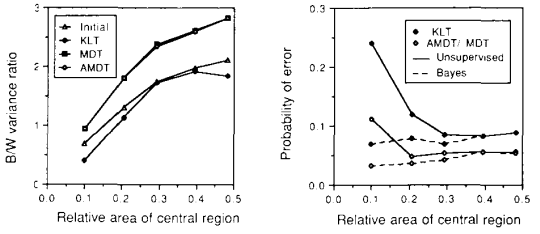

Fig. 7. Separability measures (a) and segmentation errors (b) obtained with the most discriminative component of various feature vector representations as a function of the size of central region for test images of the type displayed in Fig. 5(a). Texture features: 4 mean deviations of the 2 × 2 DHT computed using 4 iterations of the smoothing procedure ($M_{eq} = 10$). Threshold: unsupervised (maximum B/W ratio criterion) or Bayesian (minimum error).

From the evaluation of the performance of various feature reduction techniques under a variety of experimental conditions, it appears that the AMDT is generally superior to KLT in the sense of providing a much better approximation of the optimal MDT solution and therefore allowing a better segmentation. In addition, the relative performance of the KLT with reference to the MDT is significantly more sensitive to changes in structural and geometrical factors.

An essential parameter of the algorithm is the size of the estimation window. Increasing the size of the averaging window generally increases class separability but also deteriorates performance in border regions. It is therefore not surprising to find an optimal window size for a given image, resulting from a tradeoff between structural and geometrical picture properties, for which segmentation performance is optimal. Up to this limit, it is reasonable to assume that between-scatter matrix is not changed significantly by averaging; this explains the excellent performance of the AMDT for small window sizes. In the examples that have been considered, the KLT was usually found to perform poorly for small window sizes sometimes even less well than the original unrotated feature sequence. However, as the window size is increased the performance of the KLT tends to improve and generally surpasses those of the initial representation. This last result is consistent because averaging usually decreases the within-scatter matrix so that the total scatter arises mainly from the between-region differences.

Unlike the KLT, and AMDT is independent of any nonsingular linear transformation of the local texture feature vectors. This invariance property has some particularly useful practical implications. Instead of dealing with the original feature vector sequence, we can replace it at each iteration with a standardized and rotated version obtained by using the KLT and subsequently rescaling each component with a normalized total variance. As stated earlier (e.g., Section III-B), this method computes all successive AMDT's at no more cost than the KLT and decreases storage requirements. This approach is also quite favorable from the point of view of roundoff errors since the data in each channel is rescaled for full dynamic range before averaging. Among other advantages is the possibility of discarding unpromising components at early stages of processing.

From this computational scheme, it appears that the only difference between our feature reduction method and a comparable system that would compute KLT's at successively decreasing spatial resolutions, is in the standardization option preceding a smoothing iteration. Some insight in the way the AMDT operates is gained by observing that Gaussian smoothing reduces the within-region variance by a factor that is approximately the same for each channel, usually leaving the between-region variance unchanged. If we start with the same energy in each channel, after Gaussian filtering the components with the greatest contributions are those which have the greater between-region variance. It follows that the AMDT tends to put more weight on channels which, according to our hypotheses, are assumed to be the most discriminative. The prewhitening transformation is intended to maximize this effect. A direct consequence, however, is that the method breaks down as soon as the decrease of the between-region variance exceeds that of the within-region variance. As observed in our experimental examples, this might happen when the averaging window is increased to the point at which it has a size close to the smallest dimension of a texture region in the image. An easy remedy is to cease standardizing the rotated decorrelated components after a specified number of iterations while continuing with the standard KLT instead.

An advantage of the AMDT over the KLT is that the eigensolutions of (12) are almost always ordered according to their discriminative power. This is not necessarily the case for the KLT, as observed in the second example. A demonstration of this property of the AMDT is possible if, in addition to the two basic assumptions of Section III-B, we assume that the ratios between all paired separability measures at iteration steps i and j are constant:

$\beta_k^{\langle i \rangle}/\beta_k^{\langle j \rangle} = \alpha \geq 1$, $k = 1, \cdots, M$. For a half octave scale progression, the proportionality factor α is approximately 2. By using (15), the eigenvalues of (12), $\{\lambda_k', k = 1, M\}$, are thus related to their corresponding optimal between-within variance ratios $\{\beta_k^{\langle i \rangle}, k = 1, M\}$ by

$$\lambda_k' = \frac{\beta_k + 1}{\beta_k + \alpha}. \tag{21}$$

Since $\partial\lambda'/\partial\beta > 0$ for any β, λ' is a strictly increasing function of β', which implies that the λ_k''s will be ordered according to their decreasing $\beta_k^{\langle i \rangle}$'s.

To illustrate the performance of our procedure, we have only shown examples with two image regions. When there are more than two distinct textures in an image, the AMDT described in Section III-B is still applicable and we have had no difficulty in using it with 3–4 regions of approximately the same size. We have also found that the efficiency of this approach relates to the size in pixels of these regions and we are still actively pursuing the matter. For images with more than two regions, the number of components contributing to class separability is usually greater than 1 and can be determined by retaining only those with eigenvalues greater than a threshold $1/\alpha$, which is the expected energy reduction obtained with noise only and is obtained by substituting $\beta = 0$ in (21). Segmentation based on simple thresholding of a single reduced component, as described in Section III-C, is then usually not the most efficient approach and more refined clustering algorithms operating in a higher dimensional space should be used instead.

V. Conclusion

This paper has described an approach for the segmentation of textured images. The first step of processing involves the extraction of pixel neighborhood properties using a local linear transformation, which is equivalent to processing the image with a bank of FIR filters. Suboptimal convolution operators are obtained from the basis vectors of the local discrete sine or Hadamard transforms. Local texture "energy" measures corresponding to various levels of resolution are thereafter estimated using an iterative Gaussian smoothing procedure with a half octave scale progression. Linear feature reduction is achieved by diagonalizing simultaneously the scatter matrices at two successive level of spatial resolution. This procedure enables a satisfactory segmentation of a variety of textured images by using a single reduced component. This approach should be useful in many practical applications, on account of its good performance, its flexibility and its highly parallel structure.

The multiresolution feature reduction technique that has been presented has been shown to be more powerful than the conventional Karhunen-Loève transform. It usually provides a closer approximation to the optimal linear multiple discriminant functions, especially for small estimation windows. Furthermore, it produces a reduced feature set which, unlike the KLT, is ordered with decreasing discrimination power.

Insofar as segmentation is concerned, we have considered only the simplest approach based on the thresholding of a reduced component. The threshold is selected by maximizing an objective criterion of separability. The current approach could be potentially improved by using more sophisticated schemes such a coarse-to-fine strategy to refine the location of the texture edges or relaxation labeling techniques. An open problem that is still under investigation is the unsupervised determination of the optimal window size for minimum error segmentation.

Acknowledgment

We thank Dr. U. Ruttimann for his encouragements and helpful discussions. We also thank Dr. B. Trus for his help and DRCT for the use of its Image Processing Facility.

References

[1] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comput. Vision, Graph., Image Processing*, vol. 29, pp. 100–132, 1985.
[2] J. S. Weska, "A survey of threshold selection techniques," *Comput. Graph., Image Processing*, vol. 7, pp. 259–265, 1978.
[3] W. A. Perkins, "Area segmentation of images using edge points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, pp. 8–15, 1980.
[4] S. Zucker, "Region growing: Childhood and adolescence," *Comput. Graph. Image Processing*, vol. 5, pp. 382–399, 1976.
[5] P. C. Chen and T. Pavlidis, "Image segmentation as an estimation problem," *Comput. Graph. Image Processing*, vol. 12, pp. 153–172, 1980.
[6] G. B. Coleman and H. C. Andrews, "Image segmentation by clustering," *Proc. IEEE*, vol. 67, pp. 773–785, May 1979.
[7] S. Grinaker, "Edge based segmentation and texture separation," in *Proc. 5th Int. Conf. Pattern Recog.*, Miami Beach, FL, Dec. 1–4, 1980, pp. 776–780.
[8] D. Wermser, "Unsupervised segmentation by use of a texture gradient," in *Proc. 7th Int. Conf. Pattern Recog.*, Montreal, P.Q., Canada, 1984.
[9] M. Pietikäinen and A. Rosenfeld, "Image segmentation by texture using pyramid node linking," *IEEE Trans. Syst., Man., Cybernet.*, vol. SMC-11, pp. 822–825, 1981.
[10] P. C. Chen and T. Pavlidis, "Segmentation by texture using a co-occurrence matrix," *Comput. Graph. Image Processing*, vol. 10, pp. 172–182, 1979.
[11] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, pp. 786–804, 1979.
[12] D. H. Hubel and T. N. Wiesel, "Brain mechanism of vision," *Scientific Amer.*, vol. 241, no. 3, pp. 130–141, Sept. 1979.
[13] B. Julesz and R. Bergen, "Textons, the fundamental elements in preattentive vision and perception of textures," *Bell Syst. Tech. J.*, vol. 62, no. 6, pp. 1619–1645, July–Aug. 1983.
[14] M. Unser, "Local linear transforms for texture measurements," *Sig. Processing*, vol. 11, no. 1, pp. 61–79, July 1986.
[15] K. I. Laws, "Textured image segmentation," Ph.D. dissertation Rep. 940, Image Processing Inst., Univ. Southern California, Jan. 1980.
[16] M. Pietikäinen, A. Rosenfeld, and L. S. Davis, "Experiments with texture classification using averages of local pattern matches," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 421–426, 1983.
[17] M. Unser, "Description statistique de textures: Application à l'inspection automatique," Ph.D. dissertation 534, Swiss Fed. Inst. Technol., Lausanne, Switzerland, 1984.
[18] W. E. Blantz, J. L. C. Sanz, and D. Petkovic, "Control-free low-level image segmentation: Theory, architecture, and experimentation," in *Proc. Int. Conf. Comput. Vision*, London, England, June 8–11, 1987, pp. 439–443.
[19] P. J. Burt, "Fast algorithms for estimating local image properties," *Comput. Graph. Image Processing*, vol. 21, pp. 368–382, 1983.
[20] S. Watanabe, "Karhunen-Loève expansion and factor analysis," *Trans. Fourth Prague Conf. Inform. Theory*, Prague, 1965, pp. 635–660; also in *Pattern Recognition: Introduction and Foundation*, J. Sklansky, Ed. Dowden, Hutchinson & Ross, 1973, pp. 146–171.

[21] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, part II, pp. 179–188, 1936; also in *Contributions to Mathematical Statistics*. New York: Wiley, 1950.
[22] J. G. Bryan, "The generalized discriminant function: Mathematical foundation and computational routines," *Havard Educ. Rev.*, vol. 21, pp. 90–95, 1951.
[23] O. D. Faugeras, "Texture analysis and classification using a human visual model," *Proc. Int. Joint Conf. Pattern Recog.*, Kyoto, Japan, Nov. 1978, pp. 549–559.
[24] G. H. Granlund, "Description of texture using the general operator approach," in *Proc. 5th Int. Conf. Pattern Recog.*, 1980, pp. 776–779.
[25] D. Wermser and C.-E. Liedtke, "Texture analysis using a model of the visual system," in *6th Int. Conf. Pattern Recog.*, Munich, Germany, Oct. 19–22, 1982, pp. 1078–1081.
[26] F. Ade, "Characterization of Texture by 'eigenfilter,'" *Sig. Processing*, vol. 5, no. 5, pp. 451–457, Sept. 1983.
[27] R. M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-8, pp. 610–621, Nov. 1973.
[28] M. Kendall and A. Stuart, *The Advanced Theory of Statistics—Volume 1: Distribution Theory*. New York: MacMillan, 1977.
[29] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pp. 429–459, 1946.
[30] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[31] D. W. Peterson and R. L. Mattson, "A method of finding linear discrimant functions for a class of performance criteria," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 380–387, July 1966.
[32] P. Brodatz, Textures—A photographic album for artists and designers. New York: Dover, 1966.

Michael Unser was born in Zug, Switzerland, on April 9, 1958. He received the M.S. (with honors) and Ph.D. degrees in electrical engineering in 1981 and 1984, respectively, from the Swiss Federal Institute of Technology in Lausanne, Switzerland.

He is currently a Visiting Associate at the Biomedical Engineering and Instrumentation Branch, National Institutes of Health, Bethesda, which he joined in 1985. His main area of research is the application of image processing and pattern recognition techniques to various biomedical problems such as the restoration and quantitative analysis of high resolution electron micrographs of biological macromolecules, and computer-assisted analysis of echocardiograms.

Dr. Unser is a member of the European Association for Signal Processing. He was awarded the Brown-Bowery Price in 1984 for his work on texture analysis and automated inspection.

Murray Eden (M'60–F'73) was born in Brooklyn, NY, on August 17, 1920. He received the B.S. degree from City College of New York in 1939 and the Ph.D. degree from the University of Maryland in 1951.

He is currently Chief, Biomedical Engineering and Instrumentation Branch, Division of Research Services, National Institutes of Health, and Professor of Electrical Engineering, Emeritus, Massachusetts Institute of Technology. His research interests include pattern recognition, analytical uses of image processing, and models for perception.