Nonlinear Operators for Improving Texture Segmentation Based on Features Extracted by Spatial Filtering

MICHAEL UNSER, MEMBER, IEEE, AND MURRAY EDEN, FELLOW, IEEE

Abstract —**An unsupervised texture segmentation system using texture features obtained from a combination of spatial filters and nonlinear operators is described. Local texture features are evaluated in parallel by a succession of four basic operations: a) a convolution for local structure detection (local linear transform), b) a first nonlinearity of the form $f(x)=|x|^{\alpha}$, c) an iterative smoothing operator, and d) a second nonlinearity $g(x)$. The Karhunen–Loève transform is used to reduce the dimensionality of the resulting feature vector and segmentation is achieved by thresholding or clustering in feature space. The combination of nonlinearities $f(x)=|x|^{\alpha}$ (in particular, $\alpha=2$) and $g(x)=\log x$ maximizes texture discrimination and results in a description with variances approximately constant for all feature components and texture regions. This latter property improves the performance of both feature reduction and clustering algorithms significantly.**

I. Introduction

THE CHARACTERIZATION of texture plays two important roles in computer vision. First, texture properties can be used to classify homogeneous image areas as sky, grass, clouds, etc. Second the identification of regions of the same texture character can be an important factor in segmenting images. Many computational methods have been suggested and applied to problems involving texture analysis, classification and image segmentation [1], [2]. Commonly used texture measures include Fourier domain energy [3], densities of local extrema [4], co-occurrence matrix and second-order gray level statistics [5], [6], local orientation and frequency [7], coarseness [8], spatial texture energy [9], [10], as well as parameters of various random field models [11]–[13]. As these methods have had somewhat different motivations from the point of view of application, they tend to be optimal for different properties or statistics.

Among the great variety of texture features, those extracted by spatial filtering [9], [10], [14], [15]–[18] appear to be well suited for texture segmentation based on clustering [19], [20]. In this context, visually distinct image regions can be differentiated effectively by means of differences in the average values of their feature vectors so long as the feature values within the different regions remain as nearly constant as possible.

Spatial texture energy measures are usually computed in three steps [9], [10], [17]. First, the initial image is convolved with a bank of N filters and the output of each convolution mask is stored in a corresponding feature plane or channel. This matching process may be thought of as decomposing the local intensity map into a set of elementary spatial patterns frequently referred to as primary features. At this stage, differences in the variances of these initial features carry the information corresponding to differences between texture regions. The next step is to rectify or square the individual feature values; more generally, to apply a nonlinear transformation. This operation has the effect of converting variance disparities into mean values differences, an essential step if standard clustering and classification procedures are to be employed. Third, these feature channels are smoothed to furnish a set of so-called texture energy measures characterizing certain local texture properties of the neighborhood of a given pixel. This final averaging reduces the feature component variances and results in clusters in feature space that are more compact and easier to distinguish. Another motivation for this operation is that the notion of texture is undefined at the single pixel level but always associated with some set of pixels, which, in our case, corresponds to a neighborhood specified by the range of the averaging window.

This approach to feature extraction combines linear neighborhood functions (filters) and nonlinear point functions (rectifier or squarer). The behavior of the linear filters is well understood as is the optimization of their design [10]. On the other hand, the role of the nonlinear operators is not nearly so transparent and in most instances their selection has been empirically based. The purpose of this paper is to examine this role by providing a detailed analysis of the effects of a class of nonlinearities (including the two cited) and by selecting operators appropriate to the aim of optimizing the discrimination between texture fields that have been determined to be

Manuscript received January 6, 1989; revised August 5, 1989 and February 11, 1990.

M. Unser is on leave from INSERM, Unité U-138, Hôpital Henri-Mondor, F-94010 Créteil CEDEX, France, and is now with Biomedical Engineering and Instrumentation Program, Bldg. 13, Room 3W13, National Institutes of Health, Bethesda, MD 20892.

M. Eden is with Biomedical Engineering and Instrumentation Program, Bldg. 13, Room 3W13, National Institutes of Health, Bethesda, MD 20892.

IEEE Log Number 9035401.

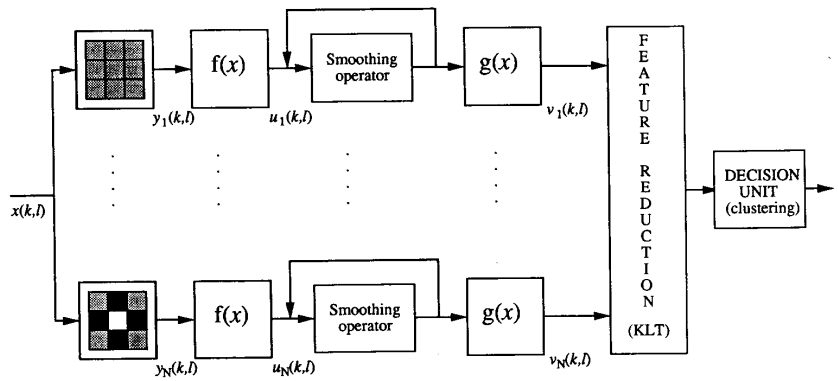

Fig. 1. Block diagram of general segmentation system.

different by other criteria, e.g., visual inspection. As a refinement, we introduce an additional nonlinearity following spatial smoothing. With a view of improving clustering performance, we will also investigate the use of transformations that standardize the variances of the components of the resulting feature vector. The motivation behind this last step is to provide a feature space with an Euclidean metric close to optimal for classification. This last property greatly simplifies classification and should substantially improve the system's performance in an unsupervised mode. The block diagram of the general texture segmentation system to be analyzed as well as the notation that is used throughout the presentation is shown in Fig. 1.

As with most pattern recognition systems, the last two elements of this block diagram are the linear feature reduction component and the decision unit that assigns individual pixels to one or another image region. In this study, we will consider the simplest design for these steps: a Karhunen–Loève transformation (KLT) followed by a simple thresholding or clustering algorithm operating in a reduced feature space to complete the classification.

Ultimately, our goal is to devise an unsupervised texture segmentation system that is easy to implement and whose individual components are carefully selected for optimal performance. The presentation is organized as follows. In Section II, we look at the influence of each processing step and derive the statistical distribution of the individual feature components within a region of homogeneous texture. In Section III, we show how to select nonlinearities that provide texture features with variances very nearly constant in all channels and all texture regions. The statistics of the feature components are then used in Section IV to quantify the system's performance (between-to-within variance ratios) for a two-class texture segmentation problem. With the help of a few simplifying assumptions, these results allow us to compare several combinations of nonlinear operators and to identify the most promising approach. Finally, in Section V, we present some experimental results that support our theoretical findings and illustrate the type of improvement that can be achieved through the proper use of nonlinearities.

II. The System Components

Each of the parallel channels of Fig. 1 is composed of a succession of elementary units. Each component carries out a specific task described below. We step through this chain of operations in order to derive the mean and variance of each feature vector. An important assumption is that the input of the system is zero mean. This condition can be imposed by highpass filtering the image in a preprocessing step. Such a process eliminates the gray tone information which is not essential for texture segmentation and will not be considered further in this study.

A. Local Linear Transform

For a statistical justification of the use of filtering operators to characterize local texture properties and the definition of optimal and sub-optimal operators, we refer to [10]. The underlying principle is to characterize the N-th order probability density function (PDF) of the pixels in a restricted neighborhood by a set of N first-order PDF's (or histograms) estimated along a set of suitably chosen axes. The block diagram in Fig. 1 is comprised of N feature channels that can be processed in parallel.

Let $x(k,l)$ represent the initial image and $h_i(k,l)$ the finite impulse response of the ith filter. The output of this operator is given by

$$y_i(k,l) = x(k,l) \otimes h_i(k,l) = \boldsymbol{h}_i^T \boldsymbol{x}(k,l) \tag{1}$$

where $\otimes$ denotes the convolution operator. The right hand side of this equation has been expressed as the dot product of a vector $\boldsymbol{h}_i$ characterizing the impulse response of the corresponding filter, and a local neighborhood vector $\boldsymbol{x}(k,l)$ defined for every spatial location (k,l). If all convolution masks are defined on the same support of size N', the equation of the filter bank is given by

$$\boldsymbol{y}(k,l) = \boldsymbol{H}\boldsymbol{x}(k,l) \tag{2}$$

where $\boldsymbol{y}(k,l) = [y_1(k,l) \cdots y_N(k,l)]^T$ and $\boldsymbol{H}$ is an $N \times N'$

matrix defining a local linear transformation. In practice, we will choose $N' = N$ and select orthogonal convolution masks. Therefore, aside from the conventional interpretation of convolution masks as local structure detectors, (2) may be thought of as a rotation in local pixel space, or, equivalently, as a transformation of the original coordinate system. We wish to choose the matrix $\boldsymbol{H}$ so that the channel histograms $\{p(y_i),\ i = 1, \cdots, N\}$ provide the most distinctive texture description. A compact texture characterization is the set of variances of each histogram given by

$$E\{y_i^2\} = \sigma_i^2 = \boldsymbol{h}_i^T \boldsymbol{C}_x \boldsymbol{h}_i \tag{3}$$

where $\boldsymbol{C}_x$ is the $N \times N$ spatial covariance associated with a given texture.

The set of all channel variances provides a complete description of the covariance structure of a given texture field when $\boldsymbol{H}$ diagonalizes $\boldsymbol{C}_x$. A transform having this property is the Karhunen–Loève transform that was first used by Ade for texture analysis [18]. We have shown elsewhere that a similar property is required to discriminate maximally between any two textures and that it is possible to determine optimal sets of convolution masks for any explicit binary texture analysis or classification problem [10]. We have also demonstrated that close-to-optimal performance can be achieved using sub-optimal separable masks associated with the basis vectors of the discrete sine (DST), the cosine (DCT), the Hadamard (DHT), or the real even Fourier (DREFT) transform. The DST usually performs slightly better for odd dimensions of the local neighborhood vector (3×3 and 5×5), while the DHT (or DREFT) gives slightly better results for even neighborhoods (2×2 and 4×4) [10], [21]. These suboptimal operators, structurally very similar to Laws' initial set of convolution masks [9], 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.

B. First Nonlinear Transformation — $f(y)$

At this point, we recall that all filtered channels are zero-mean and that distinct texture regions are presumed to differ mainly in the values of their channel variances. Since most clustering methods are designed to distinguish between classes with non-overlapping centroids, the purpose of the first nonlinear transformation is to translate differences in dispersion characteristics into differences in mean value. To study the effect of such transformations, we define the standardized filtered variable

$$z = y_i / \sigma_1, \qquad i = 1, \cdots, N \tag{4}$$

which has zero mean and unit variance for each of the N local linear features. The normalized density function, $\{p(z)\}$, is characteristic of the shape of the channel histogram, independent of its dispersion. As such, it is representative of a whole class of probability density functions. For example, the normalized Gaussian density can be used to characterize the class of textures for which the joint PDF of $\boldsymbol{x}$ or $\boldsymbol{y}$ are multivariate Gaussian distributions; that is, the so-called Gaussian textures.

A natural choice for $f(y)$ is a member of the class of power functions of the form $|y|^\alpha$ where α is some positive constant, with the most obvious choices being $\alpha = 1$ [9] and $\alpha = 2$ [10]. Let μ_α be the normalized moment defined by

$$\mu_\alpha = E\{|z|^\alpha\} = \int_{-\infty}^{+\infty} |z|^\alpha p(z)\, dz. \tag{5}$$

The expected value of the nonlinear transformed variable $u_i(k,l) = f(y_i(k,l))$ is given by

$$\bar{u}_i = E\{u_i\} = \mu_\alpha \sigma_i^\alpha. \tag{6}$$

Similarly, we find that the variance of $u_i(k,l)$ is

$$\begin{aligned} \sigma_{ui}^2 = \mathrm{Var}\{u_i\} &= (\mu_{2\alpha} - \mu_\alpha^2)\sigma_i^{2\alpha} \\ &= \frac{(\mu_{2\alpha} - \mu_\alpha^2)}{\mu_\alpha^2} \bar{u}_i^2 = \gamma(\alpha)\bar{u}_i^2 \end{aligned} \tag{7}$$

and is proportional to the square of the expected value.

C. Spatial Smoothing

The purpose of the smoothing operator is to decrease the variances of the feature vectors within the various texture regions while preserving the inter-region mean differences insofar as possible. Smoothing is usually achieved by linear filtering. The impulse response of the corresponding lowpass filter may be viewed as an estimation window that is applied to the signal and that provides a local estimate of $E\{f(y_i)\}$, which is given by (6). The choice of a Gaussian filter is particularly attractive because of its effective band rejection and good localization properties, and also because it is the only circularly symmetric filter that is separable in the two principal directions. This operator can be implemented by means of a succession of simple moving average filters, or by using a method initially described by Burt that is based on the cascaded convolution with a separable Gaussian-like kernel that is progressively expanded and filled with zeros to provide a full octave scale progression [22].

A smoothing operator is characterized by an equivalent $n \times n$ window size. For a texture of a given type, it will provide a noise reduction factor $R(n^2) = O(n^2)$. We note that for white noise, $R(n^2)$ is simply equal to n^2. In all cases, the expected value of the feature vectors within a region of uniform texture should remain unchanged.

D. Second Nonlinear Transformation — $g(x)$

A second nonlinearity can be used to counterbalance the effect of $f(x)$ by taking $g(x) = f(|x|)^{-1}$. It will provide a feature value that is in the same units as the input signal independent of the value of α. For $\alpha = 2$, this amounts to using the local standard deviation as a texture feature instead of the variance. The motivation for such a choice

is to make feature sets obtained for different values of α more nearly comparable, thus facilitating the fine tuning of the system. We note that the successive use of $f(x) = |x|^{\alpha}$ and $g(x) = |x|^{1/\alpha}$ is analogous to determining a weighted L_{α} norm in order to measure distance in an n-dimensional space. Another option, which will be considered later on, is to select a transformation that removes the dependency between the feature variance and its mean (cf (7)). Before analyzing the effect that this may have on texture discrimination, we will derive the approximate distribution of the resulting feature vector.

Due to the noise reducing effect of the smoothing operation, the signal at the input of this operator is the superposition of a constant (the expected value of $f(y)$) and a residual variation Δu of the order of $1/n$. We use the Taylor series expansion

$$g(\bar{u}_i + \Delta u) = g(\bar{u}_i) + g'(\bar{u}_i)\Delta u + O(1/n^2) \quad (8)$$

where

$$g'(x) = \frac{\partial g(x)}{\partial x}.$$

For n sufficiently large, the terms in $O(1/n^2)$ are negligible and the expected value of the feature vector is simply

$$E\{v_i\} \cong g(\bar{u}_i) + g'(\bar{u}_i)E\{\Delta u\} = g(\bar{u}_i). \quad (9)$$

Similarly, we find that the variance is given by

$$\mathrm{Var}\{v_i\} \cong g'(\bar{u}_i)^2 \mathrm{Var}\{\Delta u\} = \frac{g'(\bar{u}_i)^2 \sigma_{ui}^2}{R(n^2)}. \quad (10)$$

As a consequence of the central limit theorem the output of the smoothing operator tends to be Gaussian distributed. When n is sufficiently large (8) is essentially a linear transformation. In such a circumstance, the PDF of the feature component $v_i(k,l)$ can be approximated by a Gaussian distribution with mean and variance given by (9) and (10), respectively.

E. Linear Feature Reduction

A detailed discussion of feature reduction techniques in the context of texture segmentation may be found in [20]. The representation that discriminates maximally between different texture regions is obtained by using multiple or generalized Fisher's linear discriminant functions [23], [24]. However, the use of this technique requires a priori knowledge of the mean vectors and the covariance matrices of the feature vectors in all texture regions. It is applicable only for supervised image segmentation.

For unsupervised segmentation, a natural choice is the KLT, the most commonly used feature reduction technique [24]. One drawback of the KLT is that features are weighted according to their overall energy contribution rather than their discriminatory efficiency. Its performance in the task of texture segmentation can sometimes be quite poor. A more satisfactory approach is the multiple-resolution feature reduction technique proposed by us earlier, which provides a much closer approximation to Fisher's linear discriminants [20]. Alternatively, we may want to keep the system as simple as possible and try to improve the performance of the KLT through an appropriate choice of nonlinearities, which is the approach taken here.

III. Variance Equalization

As mentioned in the introduction, the texture segmentation system shown in Fig. 1 can be refined by selecting a sequence of transformations with the property that the variances of the feature components are nearly constant for all textures and all feature channels. The search for a class of texture features with an equalized variance is motivated by a desire for Euclidean metric suitability in feature space, which not only facilitates further processing but should also improve the system's performance in an unsupervised mode. This goal can be achieved by adjusting either the first or the second nonlinearity.

A. $f(x) = \log|x|$

Instead of using a first nonlinearity of the form $|x|^{\alpha}$, we may use the logarithmic transformation $f(x) = \log|x|$. It is straightforward to see that

$$\bar{u}_i = \log \sigma_i + \mu_{\log} \quad (11)$$

$$\sigma_{ui}^2 = \sigma_{\log}^2 = \text{const.} \quad (12)$$

where $\mu_{\log}$ and $\sigma_{\log}^2$ are the mean and variance of $\log|z|$ where z is the standardized variable defined by (4). Hence, feature vectors have a constant variance, apart from some normalized shape factor. In Appendix A, we derive the statistical distribution and the moment generating function of $\log|z|$ when z is Gaussian distributed. In particular, we show that $\mu_{\log} = -0.6351\cdots$ and $\sigma_{\log}^2 = \pi^2/8$, which is a result that holds for the class of all Gaussian textures.

B. Variance Stabilization for Power Functions

An important property shared by the power functions considered in Section II-B is that the variance of a given texture feature u_i is proportional to the square of the expected value (cf (7)). We can choose $g(x)$ so that the variances of the resulting texture features are approximately constant. This requirement may be rewritten as

$$\sigma_{\mathrm{vik}} \cong \frac{g'(\bar{u}_i)\sigma_{uik}}{\sqrt{R(n^2)}} = g'(\bar{u}_i)\bar{u}_{ik}\sqrt{\frac{\gamma(\alpha)}{R(n^2)}} = \text{const.} \quad (13)$$

which, assuming some constant values of $R(n^2)$ and $\gamma(\alpha)$, is equivalent to the differential equation:

$$dg(\bar{u}_i) = \frac{c_1}{\bar{u}_i} d\bar{u}_i$$

where c_1 is some fixed constant. By integrating this ex-

pression, we obtain

$$g(\bar{u}_i) = \int \frac{c_1}{u_i} d\bar{u}_i = c_1 \log(\bar{u}_i) + c_0 \tag{14}$$

which, once again, points towards the choice of the logarithmic transformation.

IV. Performance Analysis

In order to compare the performance of various combinations of nonlinear operators for texture segmentation, we will consider the simplified problem of distinguishing between two regions with distinct texture types: ω_1 and ω_2. Obviously, the discriminatory power of a given texture feature will depend on the ratio:

$$\lambda_i^2 = E\{y_i^2|\omega_1\}/E\{y_i^2|\omega_2\} = (\sigma_{i1}/\sigma_{i2})^2 \tag{15}$$

where σ_{i1}^2 and σ_{i2}^2 are the channel variances of the two textures. We recall that these quantities are entirely determined by the covariance structure of the underlying texture fields (cf (3)).

A. Between-to-Within Variance Ratio

A commonly used measure of separability is the between-to-within variance ratio that for a given feature v_i and a K-class classification problem is given [24] by

$$\beta_i = \frac{\sum_{k=1}^{K} P_k[\bar{v}_{ik} - \bar{v}_{io}]^2}{\sum_{k=1}^{K} P_k \sigma_{vik}^2} \tag{16}$$

where $P_k = \text{Prob}(\omega_k)$, $\bar{v}_{ik} = E\{v_i|\omega_k\}$, $\sigma_{vik}^2 = \text{Var}\{v_i|\omega_k\}$, and $\bar{v}_{io}$ is the overall average feature value:

$$\bar{v}_{io} = \sum_{k=1}^{K} P_k \bar{v}_{ik}.$$

For our two class classification problem, we will assume for simplicity that $P_1 = P_2 = 1/2$ and that the noise reduction factors for both image regions are the same. We also denote by $\bar{u}_{ik}$ and σ_{uik}^2, $(k = 1,2)$, the mean and variance of u_i for both texture fields. By substituting the expressions for the mean and variance given by (9) and (10) in (16), we find that

$$\beta_i \cong \frac{1}{2R(n^2)} \frac{[g(\bar{u}_{i1}) - g(\bar{u}_{i2})]^2}{g'(\bar{u}_{i1})^2 \sigma_{ui1}^2 + g'(\bar{u}_{i2})^2 \sigma_{ui2}^2}. \tag{17}$$

This equation will be used for comparing various combinations of nonlinear operators.

B. Comparison of Different Combinations of Nonlinear Operators

Using (17), we can now attempt a theoretical comparison of the performance of the various combinations of nonlinear operators that have been considered so far.

TABLE I
Normalized Moments μ_α (5) and Associated Shape Factors $\kappa(\alpha)$ (18) as Function of Exponent of α for Gaussian and Laplacian Distribution

	Gauss		Laplace	
α	μ_α	$\kappa(\alpha)$	μ_α	$\kappa(\alpha)$
1/2	0.822	1.38	0.886	0.915
1	0.798	1.75	1	1
2	1	2.	2	0.8
3	1.596	1.84	6	0.474
4	3	1.5	24	—
6	15	—	720	—
8	105	—	—	—

This comparison is based on the two following assumptions.

1) the standardized (unit variance and zero mean) PDFs of the primary features for both textures under consideration are approximately the same, at least to the extent that $\mu_\alpha = E\{|z|^\alpha|\omega_1\} \cong E\{|z|^\alpha|\omega_2\}$. In particular, we note that this condition is satisfied for the entire class of Gaussian textures and that it also holds across all features.

2) the noise reduction factor of the smoothing filter $R(n^2)$ is the same for both textures under consideration and for all nonlinear operators that will be considered. Obviously, this condition is only satisfied exactly for white noise. Otherwise, approximate equality can be considered as an acceptable hypothesis whenever the equivalent size of the averaging kernel is greater than the maximum distance over which pixels are significantly correlated.

At this stage, it is convenient to introduce the distribution parameter

$$\kappa(\alpha) = \frac{\alpha^2 \mu_\alpha^2}{\mu_{2\alpha} - \mu_\alpha^2} \tag{18}$$

which, as it will turn out later on, is a direct measure of the discriminatory power of the function $f(y) = |y|^\alpha$ for a given class of distributions. The values of μ_α and $\kappa(\alpha)$ for the standardized Gaussian and Laplacian (or exponential) distributions are given in Table I.

The explicit evaluation of (15) for various combinations of nonlinear operators using the mean and variance values given by (6), (7) or (11), (12) and performing the appropriate substitutions yields the following results.

- *Case 1:* $f(x) = |x|^\alpha$ and $g(x) = x$:

$$\beta_i \cong \frac{1}{2R(n^2)} \kappa(\alpha) \eta_1(\alpha, \lambda_i)$$

$$\text{where} \quad \eta_1(\alpha, \lambda_i) = \frac{(1 - \lambda_i^\alpha)^2}{\alpha^2(1 + \lambda_i^{2\alpha})} \tag{19}$$

- *Case 2:* $f(x) = \log x$ and $g(x) = x$:

$$\beta_i \cong \frac{1}{2R(n^2)} \frac{1}{\sigma_{\log}^2} \eta_2(\lambda_i) \quad \text{where} \quad \eta_2(\lambda_i) = \frac{(\log \lambda_i)^2}{2} \tag{20}$$

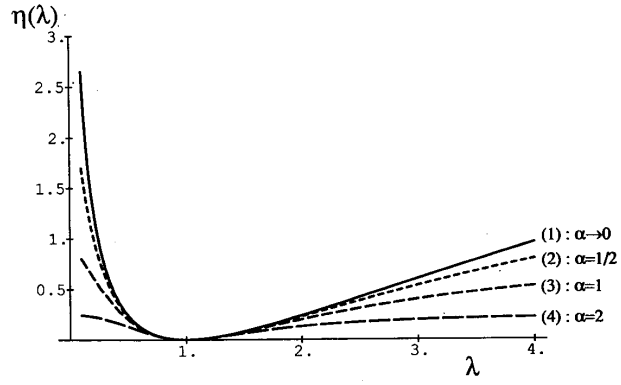

Fig. 2. Texture separation measures as function of λ (ratio of standard deviations) for various combinations of nonlinear operators. (1) $f(x)=|x|^{\alpha}$ and $g(x)=\log x$. (2) $f(x)=|x|^{1/2}$ and $g(x)=x$. (3) $f(x)=|x|$ and $g(x)=x$ (or $f(x)=|x|^{\alpha}$ and $g(x)=|x|^{1/\alpha}$). (4) $f(x)=|x|^2$ and $g(x)=x$.

- *Case 3:* $f(x)=|x|^{\alpha}$ and $g(x)=|x|^{1/\alpha}$:

$$\beta_i \cong \frac{1}{2R(n^2)}\kappa(\alpha)\eta_3(\lambda_i) \quad \text{where} \quad \eta_3(\lambda_i)=\frac{(1-\lambda_i)^2}{(1+\lambda_i^2)} \tag{21}$$

- *Case 4:* $f(x)=|x|^{\alpha}$ and $g(x)=\log x$:

$$\beta_i \cong \frac{1}{2R(n^2)}\kappa(\alpha)\eta_4(\lambda_i) \quad \text{where} \quad \eta_4(\lambda_i)=\frac{(\log\lambda_i)^2}{2}. \tag{22}$$

All these expressions have been expressed as the product of three basic terms.

The first one, $1/R(n^2)$, is the noise reduction factor of the smoothing filter intended to decrease the intraclass variance of the texture features and which, to a first approximation, can be assumed to be the same in all cases.

The second term, $\kappa(\alpha)$ (or $1/\sigma_{\log}^2$ for case 2), defined by (5) and (18), characterizes the basic shape of the family of underlying distributions and is independent of the dispersion parameters that allow the separation of different texture fields. For a Gaussian distribution, $\kappa(\alpha)$ is maximized when $\alpha=2$ (cf Table I) and is significantly greater than $1/\sigma_{\log}^2$ that is equal to $8/\pi^2=0.81$ (cf Appendix A). The choice of $\alpha=1$ is most appropriate for the exponential family and can still be considered as acceptable for Gaussian textures.

The third term, which satisfies the property $\eta(\lambda_i)=\eta(1/\lambda_i)$, is a texture separation measure that depends only on the ratio of the standard deviations in the corresponding channel, irrespective of the family of underlying distributions. In particular, we note that $\eta_3(\lambda_i)=\eta_1(\alpha=1,\lambda_i)$ and $\eta_2(\lambda_i)=\eta_4(\lambda_i)$. In the two last cases, it is interesting to note that the initial dependence of $\eta_1(\alpha,\lambda_i)$ on α has been removed through the appropriate selection of the second nonlinearity.

In Cases 2, 3 and 4, the three basic factors of β_i are fully decoupled and can be optimized separately. The comparison between Cases 1, 3, and 4 is simplified as the two first terms in (19), (21) and (22) are identical in all cases. The essential result, illustrated by the graph in Fig. 2, is that for any value of λ_i, we have

$$\eta_4(\lambda_i)=\lim_{\alpha\to 0}(\eta_1(\alpha,\lambda_i)) \geqslant \eta_1(\alpha_1,\lambda_i) \geqslant \eta_1(\alpha_2,\lambda_i), \quad \text{for} \quad 0\geqslant\alpha_1\geqslant\alpha_2. \tag{23}$$

The intuitively pleasing result that $\eta_1(\alpha,\lambda_i)$ tends to $\eta_4(\lambda_i)$ in the limit is demonstrated in Appendix B.

Therefore, provided that the assumptions underlying the derivation of (19)–(22) hold, the most favorable combination of nonlinear operators is Case 4 and is independent of α. It has the additional property of providing features with variances approximately equal in all feature channels and all texture regions (cf Section III-B). For $\alpha\geqslant 1$, the combination studied in Case 3 surpasses that of Case 1, while for $\alpha<1$ the reverse is true. We note, however, that for λ_i close to one all curves in Fig. 2 are almost indistinguishable and that methods 1, 3, and 4 are essentially equivalent. This result points to the importance of choosing α appropriately so that equivalent improvements in performance can be expected for the last three of these methods independent of the value of λ_i. Clearly, Case 2 appears to be the least favorable because we usually have $\kappa(\alpha)>1/\sigma_{\log}^2$. This observation suggests that the approach to variance equalization considered in Section III-B is the most effective.

V. Experimental Results

The complete texture segmentation system has been implemented on an Apple Macintosh II personal computer. We present here some processing examples and provide an experimental evaluation of the performance of various combinations of nonlinear operators.

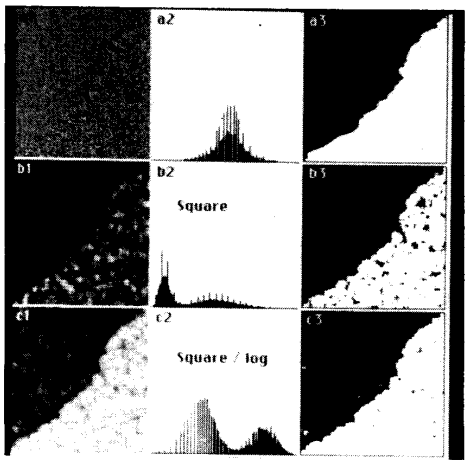

Fig. 3. Segmentation of generated white Gaussian noise with $\lambda = 2$. (a1) 128×128 test image. (a2) histogram. (a3) definition of image regions. (b1) extracted Feature channel with $f(x) = |x|^2$ and $n = 5$. (b2) histogram. (b3) binary result of segmentation. (c1) extracted feature channel with $f(x) = |x|^2$, $g(x) = \log x$ and $n = 5$. (c2) histogram. (c3) binary result of segmentation.

TABLE II
EXPERIMENTAL BETWEEN-TO-WITHIN VARIANCE RATIOS FOR TWO TEXTURES IN FIG. 2 USING VARIOUS FEATURE EXTRACTION METHODS APPLIED TO NONFILTERED SIGNAL ($\lambda = 2$ AND $n = 5$)

	Case 1	Case 2	Case 3	Case 4
$f(x)$	$\lvert x\rvert^\alpha$	$\log x$	$\lvert x\rvert^\alpha$	$\lvert x\rvert^\alpha$
$g(x)$	x	x	$\lvert x\rvert^{1/\alpha}$	$\log x$
$\alpha = 1/2$	3.505	2.342	3.235	3.453
$\alpha = 1$	4.160	2.342	4.160	4.379
$\alpha = 2$	3.454	2.342	4.695	4.78

A. White Noise

In a first experiment, we generated a test image composed of two regions with zero-mean Gaussian white noise, shown in Fig. 3(a1). The ratio between the noise variances was chosen to be 4 ($\lambda = 2$). As these random fields have no spatial correlation, no initial spatial filtering was used. Segmentation was based on a single local texture feature evaluated using various combinations of nonlinear operators and a 5×5 moving average filter. In effect, this approach is equivalent to computing some form of local estimate of the noise variance over a 5×5 window. The corresponding features obtained by using $f(x) = x^2$ alone and in combination with $g(x) = \log x$ are displayed in Figs. 3(b1) and 3(c1), respectively. The efficiency of various feature extraction methods was compared by computing the between-to-within ratio defined by (17) using the mean and variance values estimated in the reference regions shown in Fig. 3(a3). These values are given in Table II. Not surprisingly, the combination $f(x) = |x|^2$ and $g(x) = \log x$ is the most efficient by this measure and is more than twice as good as the least favorable one ($f(x) = \log x$). It can be verified that all results are in good agreement with the predictions that can be made from the equations in Section IV-B using the theoretical values of $\kappa(\alpha)$ for a Gaussian distribution which are given in Table I.

For examples of this type with only one-component feature vector, final segmentation can be obtained by simple thresholding. In the present context, it makes sense to choose the threshold value leading to the partition with maximal between-to-within variance ratio. It is the method described in [25]. The corresponding binary segmentation obtained for the feature channels in Fig. 3(b1) and 3(c1) are shown in Fig. 3(b3) and 3(c3), respectively. Obviously, this threshold selection technique performs best with the combination $f(x) = x^2$ and $g(x) = \log x$, in which case the classification errors are fewest and the most nearly symmetric in distribution. An explanation lies in the fact that our unsupervised segmentation method relies implicitly on the Euclidean metric for classifying individual pixels. According to the optimal Bayesian decision rule under class conditional Gaussian assumption, the use of a common metric is only optimal when all classes have the same covariance matrix. This is precisely the case in Fig. 2(c) because of the variance stabilizing property of the logarithmic transformation. This property is illustrated by comparing the histograms in Fig. 2(b2) and 2(c2). We have performed the same experiment using various levels of noise and always obtained experimental results in excellent agreement with the theory. In accordance with our predictions, we also verified that the log transform (Case 2 and 4) equalizes the variance of the feature component independent of the amount of noise initially present in the image regions.

B. Texture Segmentation Examples

To investigate how well our theoretical results were applicable to "natural" textures, we created test images with predefined texture regions created by combining textures taken from [26]. The original images were digitized and preprocessed to compensate for transducer nonuniformities; their mean and variance were set to prescribed constants: $\bar{x} = 0$ and $\sigma_x = 30$, respectively. The use of synthetic texture composites has the advantage of allowing a precise performance evaluation. Texture properties were extracted using a bank of four filters associated to the 2×2 Hadamard transform: the first operator $\left(\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix}\right)$ is a lowpass filter while the three others are vertical $\left(\begin{vmatrix} 1 & -1 \\ 1 & -1 \end{vmatrix}\right)$, horizontal $\left(\begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix}\right)$ and diagonal $\left(\begin{vmatrix} 1 & -1 \\ -1 & 1 \end{vmatrix}\right)$ edge detectors, respectively. The filtered channels were then processed using the sequence of operations outlined in Section II.

A first test image (a) created by combining two Brodatz textures (bark and grass) is displayed in Fig. 4(a); the two reference regions are shown in Fig. 4(a3). For this particular example, smoothing was performed using two iterations of a 7×7 moving average filter that is roughly equivalent to a 10×10 window. The processing results for

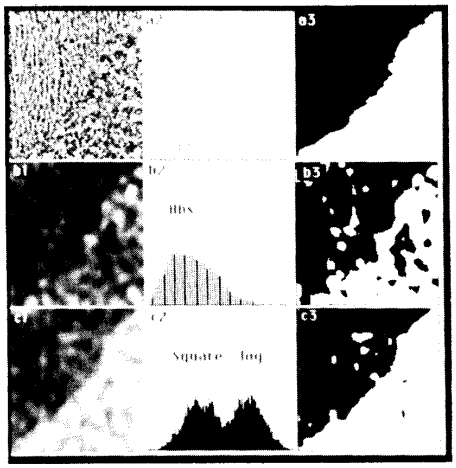

Fig. 4. Segmentation of test image (a) with two texture regions (bark and grass). (a1) 128×128 test image. (a2) histogram. (a3) definition of image regions. (b1) first component of KLT for features computed with $f(x)=|x|$ and $n=10$ (two iterations of a 7×7 moving average filter). (b2) histogram. (b3) binary result of segmentation. (c1) first component of KLT for features computed with $f(x)=|x|^2$, $g(x)=\log x$ and $n=10$. (c2) histogram. (c3) binary result of segmentation.

TABLE III
EXPERIMENTAL BETWEEN-TO-WITHIN VARIANCE RATIOS FOR TEST IMAGE IN FIG. 4(a1) FOR VARIOUS FEATURES DERIVED FROM THE FOUR CHANNELS OF THE 2×2 HADAMARD TRANSFORM

Features	$f(x)$ / $g(x)$	$\lvert x\rvert^{\alpha}$ / x	$\lvert x\rvert^{\alpha}$ / $\lvert x\rvert^{1/\alpha}$	$\lvert x\rvert^{\alpha}$ / $\log x$	$\log x$ / x
#1	$\alpha=1/2$	0.37	0.37	0.37	0.29
#2		0.75	0.74	0.75	0.58
#3		1.63	1.60	1.62	1.32
#4		0.13	0.14	0.13	0.10
KLT 1		2.12	0.82	2.40	2.06
Fisher 1		2.64	2.60	2.62	2.17
#1	$\alpha=1$	0.41	—	0.42	—
#2		0.87	—	0.87	—
#3		1.76	—	1.78	—
#4		0.16	—	0.16	—
KLT 1		0.97	—	2.58	—
Fisher 1		2.87	—	2.92	—
#1:	$\alpha=2$	0.44	0.46	0.46	—
#2		0.95	0.98	0.98	—
#3		1.66	1.79	1.81	—
#4		0.16	0.17	0.17	—
KLT 1		0.55	1.20	2.71	—
Fisher 1		2.86	3.07	3.11	—

different combinations of nonlinear operators are summarized in Table III in terms of the between-to-within variance ratios evaluated for each feature component, as well as for the first component of the KLT. We have also included the β associated to the linear Fisher discriminant function, which can be considered an upper bound on performance. We observe first that for either method there are slight differences in performance depending on the value of α. The tendency is more or less the same as in Table II but the differences in performance are not nearly as important, which may be explained by the fact that the underlying random fields are not exactly Gaussian. The use of the logarithmic function alone is still the least favorable approach. The results obtained in case 1, 3 and 4 are approximately equivalent although there is usually a slight advantage in using the latter. This similarity in performance is explained by observing that the initial variance ratios in all channels are not very different from one and correspond to values for which all curves in Fig. 2 are very close. Overall, the best performance is achieved with the combination of operators: $f(x)=x^2$ and $g(x)=\log x$. We have also found that for Case 4, the KLT performs much the best, and that it approximates the optimal linear discriminant function closely. This behavior is illustrated in Figs. 4(b) and 4(c), which show the first component of the KLT and the corresponding segmentation results using the absolute value alone, which is the conventional approach and the combination of square and logarithmic transformations. These two cases correspond to the underlined values of β in Table III. The weights for the first component of the KLT in Fig. 4(b1) and 4(c1), given in the order we gave the 2×2 normal basis functions, are $(-0.91, 0.25, -0.33, -0.004)$ and $(-0.36, 0.46, -0.76, -0.25)$, respectively. It is particularly striking to compare the histogram in Fig. 4(b2) in which both classes overlap substantially, and the one in Fig. 4(c2) that is clearly bimodal.

It may be argued that the joint PDFs of textures such as the ones used in the preceding example are close to Gaussian distributions, in which case the assumptions underlying our derivations are well satisfied. We therefore devised a series of experiments using texture composites created by combining Brodatz textures that had their grayscale histograms equalized to 32 levels and that were therefore clearly non-Gaussian. The segmentation results obtained with our second test image (b) using the same parameters as before are shown in Fig. 5. A more detailed performance assessment is provided by Table IV. These results are consistent with the previous ones and the same qualitative observations can be made. The two texture regions are now differentiable primarily by their proportion of vertical edge components (second feature component). In order to check the reproducibility of these findings, we processed 30 test images created by combining six equalized Brodatz textures (D17, D19, D24, D29, D68, and D84) in all possible pairs. The segmentation parameters were identical to those used in Fig. 5 and the overall performance was assessed from the magnitude of the within-to-between variance ratio of the first component of the KLT. The results were unambiguous. In all cases, the combination $|x|^2-\log x$ significantly outperformed the squarer alone. The same pattern was also observed in the comparison of combinations 1 and 4 for $f(x)=|x|$, except for three cases in which the two methods essentially performed the same. For all test images, the KLT approximation of Fisher linear discriminants obtained in Case 4 was excellent (within a few percent)

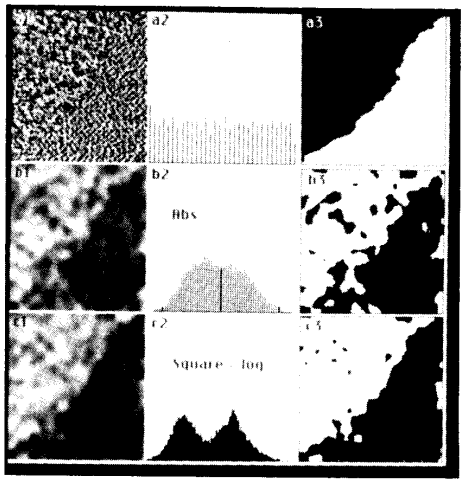

Fig. 5. Segmentation of test image (b) with two texture regions (paper and sand). (a1) 128×128 test image. (a2) histogram. (a3) definition of image regions. (b1) first component of KLT for features computed with $f(x)=|x|$ and $n=10$ (two iterations of 7×7 moving average filter). (b2) histogram. (b3) binary result of segmentation. (c1) first component of KLT for features computed with $f(x)=|x|^2$, $g(x)=\log x$ and $n=10$. (c2) histogram. (c3) binary result of segmentation.

while the results obtained in Case 1 (or 3) were highly variable.

We also conducted experiments using larger sets of texture features. Table V summarizes the results obtained for the segmentation of test images (a) and (b) using nine texture features derived from the 3×3 DST [10]. The β values given for the optimal Fisher discriminant function serve as a compact characterization of the performance of the entire feature set. When compared with Tables III and IV, the performance is somewhat improved due to the use of a larger number of features. As before, the performance obtained by use of sequence 4 are the best in almost all cases. The combination $f(x)=|x|^2$ and $g(x)=|x|^{1/2}$ is also consistently better than the use of a rectifier or a squarer alone. Surprisingly, the performance of $f(x)=\log|x|$ is not as bad as it was in the previous experiments. As in all other examples, the first component of the KLT provides a better approximation of the optimal discriminant function when the variances of the feature components are equalized through the use of a logarithmic transformation.

The general pattern that is apparent in Tables III–V and that is qualitatively consistent with our theoretical findings has been observed with a variety of different textures. It is also virtually independent of the average kernel size. We have also used test images with as many as four different textures and verified that the KLT consistently performed the best for Case 4.

C. Discussion

The conventional approach to extracting texture features from a bank of convolution masks is to apply a rectifier followed by some smoothing operator on each filtered channel. This concept is initially due to Laws who used the so-called texture energy measures [9]. This approach has been investigated and refined by other researchers, especially that part which deals with the choice of proper convolution operators [10], [18]. With this study, we have attempted to show from both a theoretical and experimental point of view how further to improve Laws' procedure through the careful use of additional nonlinearities. From our experimental results, the most promising approach seems to be the combination of a squarer and a logarithmic transformation. The system that we are proposing is flexible and can be tuned by changing the exponent of the first nonlinearity to provide the best performance for a particular application. For the class of all Gaussian textures, the value $\alpha=2$ is optimal, an unsurprising result given that the sample variance of a zero-mean Gaussian random variable is a sufficient statistic.

Overall, there is good agreement between the theoretical and experimental results. For images with nonsynthetic textures, however, the results agree only qualitatively. Again, this is not surprising because we had made several approximations and assumptions in deriving the formulas in Section IV-B. In particular, the distribution of texture images is rarely precisely Gaussian and different texture fields may have slightly different values of $\kappa(\alpha)$. Moreover, the noise reduction factor $R(n^2)$ is not precisely the same in all image regions. In our derivation we have also assumed that n was sufficiently large to justify the use of a local linear approximation to the second nonlinearity.

Even though our theoretical results were derived for a simple two-class classification problem, we have reason to believe that our conclusions should apply as well to the processing of images with multiple texture regions. From

TABLE IV
EXPERIMENTAL BETWEEN-TO-WITHIN VARIANCE RATIOS FOR TEST IMAGE IN FIG. 5(a1) FOR VARIOUS FEATURES DERIVED FROM THE FOUR CHANNELS OF THE 2×2 HADAMARD TRANSFORM

Features	$f(x)$ $g(x)$	$\lvert x\rvert^\alpha$ x	$\lvert x\rvert^\alpha$ $\lvert x\rvert^{1/\alpha}$	$\lvert x\rvert^\alpha$ $\log x$	$\log x$ x
#1	$\alpha=1/2$	0.15	0.15	0.15	0.14
#2		1.42	1.42	1.41	1.11
#3		0.03	0.03	0.03	0.01
#4		0.72	0.72	0.54	0.54
KLT 1		1.33	0.18	1.59	1.26
Fisher 1		1.75	1.76	1.74	1.36
#1	$\alpha=1$	0.15	—	0.15	
#2		1.56	—	1.56	
#3		0.03	—	0.04	
#4		0.85	—	0.86	
KLT 1		0.89	—	1.82	
Fisher 1		1.94	—	1.95	
#1	$\alpha=2$	0.15	0.16	0.16	
#2		1.53	1.57	1.56	
#3		0.04	0.04	0.04	
#4		0.93	0.95	0.96	
KLT 1		0.33	0.18	2.05	
Fisher 1		2.20	2.22	2.26	

TABLE V
EXPERIMENTAL BETWEEN-TO-WITHIN VARIANCE RATIOS FOR TEST IMAGES (a) AND (b) FOR FIRST COMPONENT OF KLT AND FISHER'S OPTIMAL LINEAR DISCRIMINANT FUNCTION USING VARIOUS FEATURES SETS DERIVED FROM THE NINE CHANNELS OF THE 3×3 DST TRANSFORM

Features	$f(x)$ / $g(x)$	$\lvert x\rvert^{\alpha}$ / x	$\lvert x\rvert^{\alpha}$ / $\lvert x\rvert^{1/\alpha}$	$\lvert x\rvert^{\alpha}$ / $\log x$	$\log x$ / x
$\alpha = 1$					
	test image (a):				
KLT 1		2.12	—	2.89	2.73
Fisher 1		3.06	—	3.11	2.87
	test image (b):				
KLT 1		1.82	—	2.27	2.09
Fisher 1		3.04	—	3.00	2.47
$\alpha = 2$					
	test image (a):				
KLT 1		0.87	2.19	2.88	
Fisher 1		2.85	3.07	3.11	
	test image (b):				
KLT 1		1.14	2.00	2.30	
Fisher 1		3.15	3.17	3.21	

our experiments with four distinct texture regions, we have found that the combination of operators $|x|^{\alpha} - \log x$ performed somewhat better overall than the other schemes and that the improvement of KLT in this particular case was even more striking as can be seen in Tables III–V. For such examples, however, the number of significant components grows (typically, $K-1$ in a K-class problem), and segmentation requires the use of clustering algorithms that are more sophisticated than simple thresholding.

The combination $f(x) = |x|^{\alpha}$ and $g(x) = \log x$ is particularly attractive because it stabilizes the variance of the resulting feature components. Unlike Case 2, this is achieved without sacrifice in performance. In this respect, our theoretical results are quite general and not restricted to the class of Gaussian textures alone. The sole requirement is that the normalized shape factor $\gamma(\alpha)$ in (7) be close to a constant across all channels and all texture regions. Even if this condition is not precisely met, we should not expect variance fluctuations greater than the variation of this shape factor (cf (13)). This variance equalization property has some very useful implications, particularly in the final stages of processing. Following logarithmic transformation, the feature vectors, apart from differences in their mean values, are roughly identically distributed across all feature channels and all texture regions. This also implies that the simple (unweighted) Euclidean distance between feature vectors is appropriate for measuring similarity between patterns. The data are therefore in a form suitable for the great majority of unsupervised clustering techniques usually based on this metric. As a consequence, these methods will tend to work better on the logarithm-transformed feature vector as illustrated in our first example (Figs. 3(b) and 3(c)). Another benefit is that energy-based feature reduction techniques such as the KLT will perform extremely well (cf Tables III–V) because the percentage of energy representing within-region variation is about the same for each feature vector component. It follows that the features with the greatest energy contribution are also those that are the most discriminating. The two last properties are particularly relevant in an unsupervised setting and will almost certainly improve the performance of the complete system. These considerations are obviously less important for supervised problems for which it is possible to use *a priori* knowledge of the region statistics to design optimal decision rules. In the latter case, the use of the second nonlinearity is not necessary as it does not intrinsically change the amount of available information but it does no harm.

Finally, there are some practical considerations. First of all, a prerequisite for the system to perform as described is that the initial outputs of the filter bank are all zero-mean. Usually, there is only one operator (the first one) that is not a highpass filter and that has to be treated carefully by removing the dc component before or after filtering. Incidentally, this dc (or lower frequency) component corresponds to the gray level part of the visual information; it is not directly related to texture and as such should be treated separately. There may also be some roundoff errors when applying nonlinear transformations to images stored in byte or integer arrays. A partial solution to this problem is to use scaling factors to make best use of the available dynamic range.

VI. CONCLUSION

In this paper, we have considered a general texture segmentation system using features extracted by spatial filtering and two levels of nonlinearity. By making certain plausible assumptions, we have been able to analyze the effect of each stage of processing and quantify the system's performance for a simplified two class classification problem. The global performance of this system (between-to-within variance ratio) has been expressed as the product of three basic terms: 1) a distribution factor, $\kappa(\alpha)$, charac-

terizing the response to a particular class of normalized probability density functions and which is a function of the exponent of the first nonlinearity; 2) a noise-reduction factor, $1/R(n^2)$, describing the ability of the spatial smoothing operator to decrease the within-region variances of the feature vector; and 3) a texture separation term, $\eta(\lambda)$, which is a function of the texture energy ratio in a particular feature channel (λ^2) and depends on the combination of nonlinear operators used for feature extraction. This decomposition allows the separate optimization of each term. The most favorable computational scheme combines a squarer and a logarithmic transformation. These results have been supported experimentally, at least qualitatively.

In the context of unsupervised texture segmentation, there are two separate but equally important issues related to feature extraction. The first one is obviously to find a set of features that is maximally discriminative. The second is to simplify subsequent processing (feature reduction and decision making). The latter can be achieved by making sure insofar as possible that the variances are the same for all components of the feature vector and over every texture region. When this last requirement is fulfilled, one has the guarantee that basic unsupervised pattern recognition techniques such as the KLT or clustering algorithms based on the Euclidean metric will perform at their best. The system that we propose has both of these properties and should therefore be useful in many image processing applications.

Appendix A

Logarithmic Transformation of a Gaussian Random Variable with Zero Mean

Let x be a Gaussian distributed random variable with zero mean and variance σ_x^2. The probability density function (PDF) of the normalized variable $z = x/\sigma_x$ is

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2).$$

The logarithmic transformation of x is defined by

$$y = \log(|x|) = \log(\sigma) + \log(|z|) = \log(\sigma) + u.$$

The PDF of $u = \log(|z|)$ and $\log(|y|)$ are identical apart from a constant difference in the mean. By applying the principle of equal probability, which in this case implies that $p(u)\,du = 2p(z)\,dz$, it is straightforward to show that

$$p(u) = \frac{2}{\sqrt{2\pi}} e^n \exp\left(-\frac{e^{2u}}{2}\right). \tag{24}$$

In order to determine the moments of this distribution, we propose to compute the characteristic or moment generating function, which is the Fourier transform of the PDF:

$$\phi_u(j\omega) = E\{e^{j\omega u}\} = \int_0^{+\infty} \frac{2}{\sqrt{2\pi}} \exp(-z^2/2) \exp\{j\omega \log|z|\}\, dz.$$

By making the change of variable $v = z^2/2$ and using the fact the $\exp\{j\omega \log v\} = v^{j\omega}$, we find that

$$\phi_u(j\omega) = \frac{2^{j\omega/2}}{\sqrt{\pi}} \int_0^{+\infty} v^{-1/2} v^{j\omega/2} e^{-v}\, dv = 2^{j\omega/2} \frac{\Gamma\left(\frac{1+j\omega}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}$$

where the gamma function $\Gamma(x)$ is defined by, in [27, p. 255, eq. 6.1.1]:

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-x}\, dt$$

with the property that $\Gamma(1/2) = \sqrt{\pi}$. Since the derivatives of $\log \Gamma(x)$ have been tabulated and define the so-called Psi function [27], it is more convenient to use the cumulant generating function:

$$\log \phi_u(j\omega) = \frac{j\omega}{2} \log 2 + \log \Gamma\left(\frac{1+j\omega}{2}\right) - \log \Gamma\left(\frac{1}{2}\right).$$

We now use the property that the mean and variance correspond to the two first cumulants which are the first and second derivatives of the cumulant generating function evaluated at $\omega = 0$, respectively [28]. We therefore have that

$$\mu_u = \left[\frac{\partial \log \phi_u(j\omega)}{\partial j\omega}\right]_{\omega=0} = \frac{\log 2}{2} + \frac{1}{2}\left[\frac{\partial \log \Gamma(x)}{\partial x}\right]_{x=1/2} = \frac{1}{2}\left[\log 2 + \Psi\left(\frac{1}{2}\right)\right] = -\frac{1}{2}[\gamma + \log 2] = -0.6351$$

where we use the change of variable $x = (1 + j\omega)/2$. $\Psi(x)$ is the Psi or Digamma function and it is known that $\Psi(1/2) = -(\gamma + 2\log 2)$, where $\gamma = 0.5772$ is the Euler constant [27, p. 258, eq. 6.3.3]. The mean of $y = \log x$ therefore simply

$$\mu_y = E\{\log x\} = \log(\sigma_x) - \frac{1}{2}[\gamma + \log 2]. \tag{25}$$

Similarity, we find that

$$\sigma_u^2 = \sigma_y^2 = \left[\frac{\partial^2 \log \phi_u(j\omega)}{\partial (j\omega)^2}\right]_{\omega=0} = \left(\frac{1}{2}\right)^2 \left[\frac{\partial^2 \log \Gamma(x)}{\partial x^2}\right]_{x=1/2} = \frac{1}{4}\Psi'\left(\frac{1}{2}\right) = \frac{\pi^2}{8} = 1.22370\cdots \tag{26}$$

where $\Psi'(x)$ is the tri-gamma function. From [27, p. 260, eq. 6.4.5], we know that $\Psi'(1/2) = \pi^2/2$.

Appendix B

The task is to determine the limit:

$$\lim_{\alpha \to 0} \left\{ \frac{(1-\lambda_i^\alpha)^2}{\alpha^2(1+\lambda_i^{2\alpha})} \right\} = \lim_{\alpha \to 0} \left\{ \frac{(1-\lambda_i^\alpha)^2}{2\alpha^2} \right\} = \lim_{\alpha \to 0} \left\{ \frac{A(\alpha)}{B(\alpha)} \right\}.$$

We clearly have an indeterminate form 0/0. We can therefore use l'Hôpital's rule and differentiate twice until the indetermination is resolved:

$$\begin{aligned} \lim_{\alpha \to 0} \left\{ \frac{A(\alpha)}{B(\alpha)} \right\} &= \lim_{\alpha \to 0} \left\{ \frac{A'(\alpha)}{B'(\alpha)} \right\} \\ &= \lim_{\alpha \to 0} \left\{ \frac{-\log \lambda_i [\lambda_i^\alpha - \lambda_i^{2\alpha}]}{2\alpha} \right\} \\ &= \lim_{\alpha \to 0} \left\{ \frac{A''(\alpha)}{B''(\alpha)} \right\} \\ &= \lim_{\alpha \to 0} \left\{ \frac{-(\log \lambda_i)^2 [\lambda_i^\alpha - 2\lambda_i^{2\alpha}]}{2} \right\} \\ &= \frac{(\log \lambda_i)^2}{2}. \end{aligned} \tag{27}$$

Acknowledgment

We would like to thank Dr. Akram Aldroubi for helpful suggestions and discussions.

References

[1] R. M. Haralick, "Statistical and structural approaches to texture," in *Proc. IEEE*, vol. 67, pp. 786–804, 1979.
[2] L. Van Gool, P. Dewaele, and A. Oosterlink, "Texture analysis Anno 1983," *Comput. Vision, Graph., Image Processing*, vol. 29, pp. 336–357, 1983.
[3] R. Bajcsy and L. Lieberman, "Texture gradient as a depth cue," *Computer Graphics and Image Processing*, vol. 5, pp. 52–67, 1976.
[4] O. R. Mitchell, C. R. Myers, and W. Boyne, "A max-min measure for texture analysis," *IEEE Trans. Comput.*, vol. C-26, pp. 408–414, 1977.
[5] 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.
[6] M. Unser, "Sum and Difference Histograms for Texture Classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 118–125, January 1986.
[7] H. Knutson and G. H. Granlund, "Texture analysis using two-dimensional quadrature filters," in *Proc. IEEE Conf. on Acoust., Speech, Signal Processing*, 1983.
[8] S. W. Zucker, A. Rosenfeld, and L. S. Davies, "Picture segmentation by texture discrimination," *IEEE Trans. Comput.*, vol. C-24, pp. 1228–1233, 1975.
[9] K. I. Laws, "Textured image segmentation," Ph.D. Thesis, Rept. 940, Image Processing Institute, University of Southern California, Jan. 1980.
[10] M. Unser, "Local linear transforms for texture measurements," *Signal Processing*, vol. 11, pp. 61–79, July 1986.
[11] R. L. Kashyap, R. Chellappa, and A. Khotanzad, "Texture classification using features derived from random field models," *Pattern Recog. Lett.*, vol. 1, pp. 43–50, 1982.
[12] C. W. Therrien, "An estimation-theoretic approach to terrain image segmentation," *Comput. Vision, Graph., Image Processing*, vol. 22, pp. 313–326, 1983.
[13] H. Derin and W. S. Cole, "Segmentation of textured images using gibbs random fields," *Comput. Vision, Graph., Image Processing*, vol. 35, pp. 72–98, 1986.
[14] O. Faugeras, "Texture analysis and classification using a human visual model," in *Proc. Int. Joint Conf. Pattern Recog.*, Kyoto, Japan, Nov. 1978, pp. 549–559.
[15] G. H. Granlund, "Description of texture using the general operator approach," in *Proc. 5th Int. Conf. Pattern Recog.*, 1980, pp. 776–779.
[16] D. Wermser and C.-E. Liedtke, "Texture analysis using a model of the visual system," in *Proc. 6th Int. Conf. Pattern Recog.*, Munich, West Germany, Oct. 19–22, 1982, pp. 1078–1081.
[17] 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.
[18] F. Ade, "Characterization of Texture by 'eigenfilter'," *Signal Processing*, vol. 5, pp. 451–457, Sept. 1983.
[19] G. B. Coleman and H. C. Andrews, "Image segmentation by clustering," *Proc. IEEE*, vol. 67, pp. 773–785, May 1979.
[20] M. Unser and M. Eden, "Multi-resolution feature extraction and selection for texture segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 717–728, July 1989.
[21] M. Unser, "Description statistique de textures: application à l'inspection automatique," Ph.D. Thesis No. 534, Swiss Federal Institute of Technology, Lausanne, Switzerland, 1984.
[22] P. J. Burt, "Fast algorithms for estimating local image properties," *Comput. Graphics and Image Processing*, vol. 21, pp. 368–382, 1983.
[23] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, part II, pp. 179–188, 1936.
[24] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[25] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-9, pp. 62–66, Jan. 1978.
[26] P. Brodatz, *Textures — A Photographic Album for Artists and Designers*. New York: Dover, 1966.
[27] M. Abramowitz and I. A. Stegun, "*Handbook of mathematical functions*," National Bureau of Standards, 1972.
[28] M. Kendall and A. Stuart, *The Advanced Theory of Statistics — Volume 1: Distribution Theory*. New York: MacMillan Publishing Co., 1977.

Michael Unser (M'89) 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 Scientist at the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, which he joined in 1985. He has also been affiliated with INSERM (French National Institutes of Health and Biomedical Research) since April 1988. 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.

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

He is currently Chief, Biomedical Engineering and Instrumentation Program, National Center for Research Resources, 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.