Signal Processing 29 (1992) 1–16
Elsevier

Frequency and orientation sensitive texture measures using linear symmetry

Josef Bigün
Swiss Federal Institute of Technology, Signal Processing Laboratory, Department of Electrical Engineering, CH-1015 Lausanne, Switzerland

Received 7 January 1991
Revised 2 July 1991 and 12 March 1992

Abstract. An efficient method for computing texture features based on dominant local orientation is introduced. The features are computed as a Laplacian pyramid is built. At each level of the Laplacian pyramid, the linear symmetry feature is computed. This feature is anisotropic and estimates the optimal local orientation in the Least Square Error (LSE) sense. It is complex valued and hence consists of two components, the local orientation estimate and its confidence measure based on the error. The algorithm is based on convolutions with simple separable filters and pixel-wise non-linear arithmetic operations. These properties allow highly parallel implementation, for example on a pyramid machine, yielding real time applications. Comparative experimental results are presented using the feature for unsupervised segmentation on test images of natural aerial image textures.

Zusammenfassung. Es wird eine effiziente Methode zur Berechnung von Textur-Merkmalen eingeführt, die auf einer lokalen Orientierung basiert. Die Merkmale werden entsprechend dem Bau einer Laplace-Pyramide berechnet. Auf jeder Stufe der Laplace-Pyramide werden die linearen Symmetrie-Merkmale ermittelt. Solche Merkmale sind anisotrop und schätzen die optimale lokale Orientierung im Sinne des kleinsten quadratischen Fehlers (LSE). Sie sind komplexwertig und bestehen daher aus zwei Komponenten, dem Schätzwert der lokalen Orientierung und seinem auf dem Fehler basierenden Vertrauensmaß. Der Algorithmus basiert auf Faltungen mit Hilfe separierbarer Filter und pixelweiser nichlinearer arithmetischer Operationen. Diese Eigenschaften erlauben ein hohes Maß paralleler Implementierung, zum Beispiel auf einer Pyramide-Maschine, und eröffnen die Möglichkeit der Echtzeit-Verarbeitung. Vergleichende experimentelle Resultate werden wiedergegeben, die von den Eigenschaften der 'unsupervised' Segmentierung von Testbildern und natürlichen Texturen von Luftaufnahmen Gebrauch machen.

Résumé. Une méthode efficace de calcul d'attributs de texture basée sur l'orientation locale dominante est introduite. Les attributs sont calculés pendant q'une pyramide laplacienne est construite. A chaque niveau de la pyramide laplacienne, l'attribut de symétrie linéaire est calculé. Cet attribut est anisotrope et estime l'orientation locale optimale au sens de la plus petite erreur quadratique (LSE). C'est une valeur complexe qui comprend donc deux composantes, l'estimée d'orientation locale et sa mesure de fiabilité basée sur l'erreur. L'algorithme est basé sur des convolutions avec des filtres simples séparables et sur des opérations arithmétiques non-linéaires au niveau des pixels. Ces propriétés permettent une implantation hautement parallèle, par exemple sur une machine de type pyramidal, engendrant des applications temps réel. Des résultats expérimentaux comparatifs sont présentés sur des attributs utilisés pour la segmentation non supervisée d'images de test réalisées à partir d'images aériennes texturées.

Keywords. Local symmetry; linear symmetry; segmentation; textures; texture features; aerial images; local image description; local orientation; frequency decomposition; Laplacian pyramid; resolution pyramid.

Correspondence to: Dr. Josef Bigün, Swiss Federal Institute of Technology, Signal Processing Laboratory, Department of Electrical Engineering, CH-1015 Lausanne, Switzerland.

1. Introduction

Texture segmentation has been an important part of image processing activities. One of the reasons is certainly due to its applicability in classification tasks. A vital step in this process is extraction of texture features, allowing description and segmentation of regions.

It is now an established fact that mammalia are equipped with cells performing local processing at an early stage of vision. In particular it is known that the human visual system is capable of discriminating textures differing in their spatial statistics higher than the first order [20]. That is, not only the gray value occurrence frequency but also the spatial arrangement of these occurrences are of high importance for vision. Since there exist segmentation methods which perform well on features differing by their first order statistics, the problem has been to find features representing higher order statistics. This has been done by utilizing autocorrelation functions, partitioning of the Fourier transform energy into radial and angular bins [1], interpretation of co-occurrence matrices through their moments [17], identification of Markov random processes parameters [14], energy of real valued special texture masks [22, 26], etc.

Although the size of a filter mask is just a parameter which can be changed, it is not a trivial task to determine an optimal strategy for the size choice since real images contain textures from a very broad range of scales. Also the demands of computational resources may increase dramatically when adapting the methods to coarse textures if special care is not taken. The patterns which can be studied within the size of a local image are restricted to that size. A constant size can be useful in characterizing a texture well in one scale but perform poorly in another. Psycho-physical experiments indicate [7, 19] that there exist frequency and orientation selectivity in the human visual system, giving a hint of how this *scaling problem* can be avoided. The difficult questions of how this selectivity is accomplished and how this is further used have been the cause of accumulating experimental and theoretical investigations, none offering a through insight yet.

The scaling problem has, in image processing, been attacked by applying Laplacians of different sizes [23, 10] or using different size sensitive decompositions like Gabor [13], Wavelet [16] and Wigner [29] basis functions. In [11] the moments of Laplacian pyramid are proposed providing anisotropic information about the local image. In [12] orientation selectivity within a frequency band is done by filtering a Laplacian image with a directional cosine filter. In principle this technique can be utilized for 2-D orientation selectivity. But such an approach results in spatial phase dependency, that is, the response to a sinusoid is also a sinusoid. This property is not desirable in texture analysis because one would like to have a uniform response throughout regions of the image which have a dominant orientation.

To achieve phase independence, the magnitudes of Gabor filter responses or quadrature filter responses can be used. The latter is proposed in [15, 21], where interpolation on the magnitudes of responses, obtained from the directional filters, is utilized. However, the filters used in that case are non-separable and require dedicated hardware for efficient computation. Our main contribution is to propose the dominant orientation as a texture feature in spatial frequency bands separated through *octave* progression and to compute these features by using computationally efficient, very few, *separable filters*. We also present experimental results using real, *aerial textures* with comparisons to other feature extraction methods. Although the filters used are not the Gabor filters, the result of the method will be a good estimate of the orientation obtained by these. In addition we develop a nonlinear logarithmic transform, CLOG, to improve the discrimination power of the texture features in general.

To increase the compuational efficiency further, we will decompose the local image into its components at different resolutions by means of a Laplacian pyramid [10]. At each level in the Laplacian pyramid, i.e. a frequency channel, we will apply

linear symmetry [4, 6] measuring the orientation or lack thereof. Linear symmetry is the simplest type in a series of symmetries which can be modeled within the same mathematical concept [2]. Symmetry is defined in terms of harmonic functions which model isogray value curves of local images. Moreover, we will use the local energy of each frequency channel to measure the activity in that channel. This is done to characterize isotropic textures.

Texture patches from Brodatz album [8] are utilized by many researchers to demonstrate the discrimination power of texture features. These texture images are easily available and represent a large number of homogeneous textures and as such have great value. However, they suffer from not being very representative in many respects. They have a very high signal to noise ratio and are formed under highly controlled different imaging conditions such as illumination and distance. An important application domain for texture feature usage is segmentation of aerial images. Such textures are formed under approximately the same imaging conditions, are noisy and frequently lack a high degree of homogeneity. For this reason we will use real aerial textures in our experiments to evaluate the proposed texture features. The emphasis of this paper is on feature extraction, although experimental results based on unsupervised segmentation techniques will be presented.

The ideas underlying the concept of linear symmetry, essential to this work will be developed in Section 2. In Section 3, the local energy feature is discussed. In Section 4, the implementation of the algorithm in a pyramid structure will be presented. In Section 5, the experimental results which include signal theoretical tests and experiments using aerial textures will be presented. Finally, in Section 6 the conclusions are given.

2. Local linear symmetry feature

Consider a two-dimensional function $f(x, y)$ representing the local image around a certain point of an image. We assume that the Fourier transform of f is concentrated to a certain frequency band. This is not a loss of generality, since it can be thought that the image is obtained by means of band-pass filtering the original image and multiplying the result with a smooth and compact function such as a Gaussian. To be able to infer whether the local image f has a dominant orientation one should define this more strictly so that it is meaningful to any local image. The way it is done in this paper is illustrated in Fig. 1. It illustrates F, the Fourier transform of f, which has a dominant orientation. It can be shown that if a function $f(x, y)$ can be expressed as

$$f(x, y) = g(k_{0x}x + k_{0y}y) = g(\bar{k}_0^{\mathrm{T}}\bar{r}),$$
$$g: E_1 \to E_1, \qquad (1)$$

where $\bar{k}_0$ is a real constant vector, then F is concentrated to a line passing through the origin. E_1 represents the Euclidean space of dimension 1, the axis of real numbers. Such a local image f has isogray values consisting of parallel lines as expressed by $g(\bar{k}_0^{\mathrm{T}}\bar{r})$ and will be called linearly symmetric. As f deviates continuously from g, that is, it is no longer possible to represent f with g without error, so does F from being perfectly concentrated to a line. This

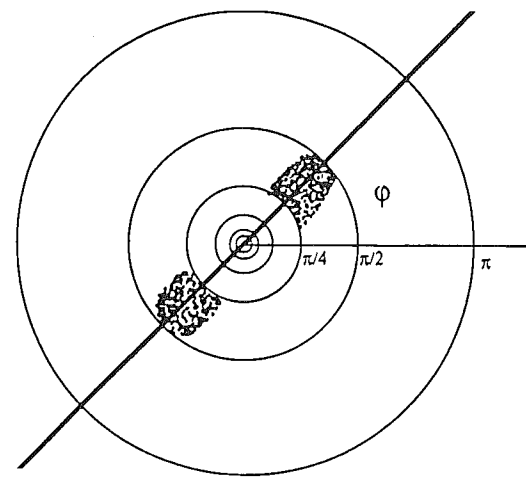

Fig. 1. A stylistic illustration representing the Fourier transform of a local image whose energy concentration is confined to a certain frequency band. The linear symmetry algorithm fits an optimal line, in the LSE sense, to the frequency domain of the local image.

deviation from a line is small if the deviation of the local image from being one-dimensional, i.e. dominance of an orientation, is small since the Fourier transform is continuous. For an arbitrary local image the linear symmetry algorithm fits a line in the Fourier domain such that the following error measure is minimized:

$$\min_{\|\bar{k}\|=1} e(\bar{k}) = \int_{E_2} \|\bar{\omega} - (\bar{\omega}^{\mathrm{T}}\bar{k})\bar{k}\|^2 \times |F(\bar{\omega})|^2 \, \mathrm{d}\bar{\omega}. \quad (2)$$

This is the well known line fitting problem or the principal axis decomposition problem formulated for the local frequency domain. The minimum always exists for some $\bar{k}_{\min}$. When $\bar{k}_{\min}$ and its corresponding error $e(\bar{k}_{\min})$ are found, the question of whether this error is large or small needs to be answered. We propose that this is done by comparing the smallest error with the largest error:

$$e(\bar{k}_{\max}) - e(\bar{k}_{\min}). \quad (3)$$

By estimating this nonnegative quantity and computing $\bar{k}_{\min}$, it is possible to know whether there exists a dominant orientation and if so its corresponding direction. Conventional differential calculus can be utilized in order to minimize/maximize (2) and compute (3). The closed form result depends only on the second order moments of the energy function, $|F|^2$. Such a term is translated to the spatial domain by using the Parsevall relationship:

$$\int_{E_2} \omega_1\omega_2|F|^2 \, \mathrm{d}\bar{\omega} = \int_{E_2} \frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \, \mathrm{d}\bar{r}. \quad (4)$$

Since f is the local image function it contains a smooth window function implicitly. Having the theory of the band limited functions in mind, we discretize the right-hand side of (4). As a result, the moments of the local Fourier transforms can be obtained for *all* local neighborhoods of an image by means of convolutions:

$$\sum_j m_j \mathbf{f}_x^{i-j}\mathbf{f}_y^{i-j}. \quad (5)$$

Here $\mathbf{f}_x^i$ represents the partial derivative of the original, (large) image with respect to x at the image position i. m_j represents the coefficients of a low-pass filter, which is defined by the interpolation function used in the discretization and the smooth neighborhood function contained in f. When Gaussians are used as both of these functions the set $\{m_j\}$ represents a discrete Gaussian function controlled by one parameter, the standard deviation:

$$m_j = \exp\left(-\frac{x_j^2 + y_j^2}{2\sigma^2}\right).$$

By using the theory of the complex number fields the optimal solution can be expressed compactly as [4, 6]

$$z_1 = (\nabla \mathbf{f})^2 * m, \quad (6)$$

where $\mathbf{f}$ represents the original image and $\nabla\mathbf{f}$ represents the complex image $\mathbf{f}_x + \mathrm{i}\mathbf{f}_y$ and $*m$ represents convolution with an averaging filter. We note here that the $\nabla\mathbf{f}$ is complex valued and can be exponentiated. Ultimately the solution of the minimization problem will be dependent on the neighborhood size defining the local image, f, since this is formally defined as a product of a smooth window function and the original image function $\mathbf{f}$. However, the choice of m will not be critical since we will apply the method to a collection of $\mathbf{f}$'s each representing one frequency band of the original. Thus the standard deviation parameter of this filter will be dependent of the frequency bandwidth of the applied image. We have determined m empirically through experiments, the full account of which will be given in a subsequent section. The complex number, z_1, obtained at every pixel represents

$$|z_1| = e(\bar{k}_{\max}) - e(\bar{k}_{\min}),$$
$$\arg(z_1) = 2\arg(\bar{k}_{\min}). \quad (7)$$

Thus, representation of $\bar{k}_{\min}$ through an angle which is twice the inclination angle of the vector $\bar{k}_{\min}$ is inherent to the method. This representation of the orientation of an axis is convenient since two different representations always correspond

uniquely to two different axes [15]. That is, $-\bar{k}_{\min}$ and $\bar{k}_{\min}$ have different arguments (the angles φ and $\varphi+\pi$) although they represent the same axis. Using $2\arg(\bar{k}_{\min})$, instead of $\arg(\bar{k}_{\min})$, eliminates this problem.

Similarly, $e(\bar{k}_{\max})+(k_{\min})$ can be obtained as a result of a convolution [2]:

$$e(\bar{k}_{\max})+e(\bar{k}_{\min})=|\nabla \mathbf{f}|^2 * m, \tag{8}$$

$|z_1|$ is an energy dependent certainty measure, that is, it decreases as the contrast of the neighborhood decreases. Normalizing (6) with the sum of the errors we obtain an energy independent complex valued texture measure.

$$z=\frac{(\nabla \mathbf{f})^2 * m}{|\nabla \mathbf{f}|^2 * m}. \tag{9}$$

We note that $|z|$ is a certainty measure just like $|z_1|$:

$$|z|=\frac{e(\bar{k}_{\max})-e(\bar{k}_{\min})}{e(\bar{k}_{\max})+e(\bar{k}_{\min})}.$$

$|z|$ attains the maximum value 1 if and only if $e(\bar{k}_{\min})=0$, since both $e(\bar{k}_{\min})$ and $e(\bar{k}_{\max})$ are non-negative. It decreases when the difference between the errors decreases. z is defined as 0 when $e(\bar{k}_{\max})+e(\bar{k}_{\min})=0$. This happens only when we have a constant image in the neighborhood and thus no unique orientation is present. The scheme is illustrated in Fig. 2.

The meaning of the line fitting process for the spatial domain can be investigated easily by interpreting the error function $e(\bar{k})$ given by (2) through the Parsevall relationship. By inspection and identifying the partial derivatives as Lie translation operators one arrives at the following conclusion: The axis $\bar{k}_{\min}$ found represents the direction along which an infinitesimal translation of the image changes the considered local gray image minimally and the error found represents the error made in this infinitesimal translation. Thus, the direction found has a physical meaning for the local gray image.

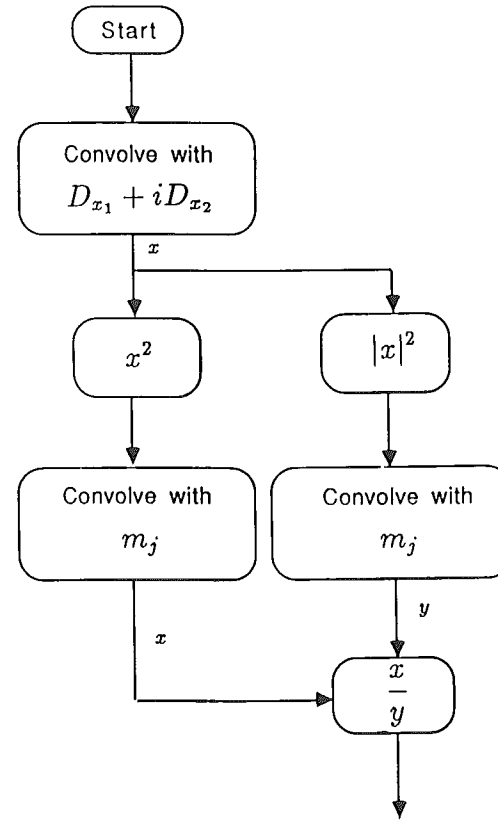

Fig. 2. The flow chart of the linear symmetry algorithm computing the complex image which represents the local orientation and its corresponding certainty. m_j represents the coefficients of a low-pass filter (Gaussian).

3. Local energy feature

When a local neighborhood has zero response to the linear symmetry measure it can be shown that this corresponds to an isotropic image with no particular orientation dominance. We mean isotropic in the sense that translation of the local image in any direction by a small amount yields the same error. Hence, the linear symmetry measure is 'quiet' when the local image is isotropic. In order to capture the isotropic information we propose the use of a local energy measure.

Consider again the local image f, and its Fourier transform, F. Assume that f is confined to a frequency band as before. Moreover, assume that it is isotropic. Then F will have its energy smeared

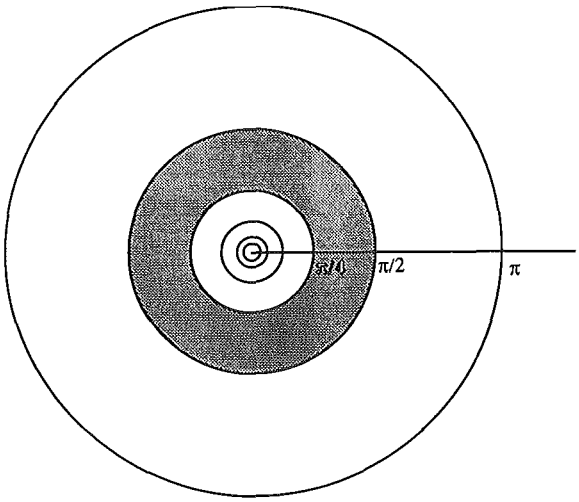

Fig. 3. The local energy measure picks up the energy of the local image in a rotationally symmetric frequency band.

out along its passband, as in Fig. 3, resulting in a bad characterization if an attempt to fit a line is made. For this reason, as a feature, we propose to measure the local energy

$$\int_{E_2} |F(\bar{\omega})|^2 \, \mathrm{d}\bar{\omega}. \tag{10}$$

This is implementable in the spatial domain by using Parsevall relation yielding

$$\mathbf{f}^2 * m, \tag{11}$$

which is a convolution of the square of the entire image with an averaging filter [11].

4. Implementation using pyramids

To implement the linear symmetry algorithm for the purpose of texture segmentation, we use the approach in [23, 10]. We decompose the image into a sequence of essentially non-overlapping band-pass channels by approximating the band-pass filters as differences of two isotropic low pass filters with different cut off frequencies. An efficient implementation of this uses pyramids and is illustrated by the Laplacian pyramid in Fig. 4.

The Laplacian pyramid consists of a series of images decreasing in size and approximates non-overlapping frequency channels. As the images in the pyramid decrease in size, so do the represented frequencies. Thus, a signal consisting of frequencies corresponding to a specific channel will be observable, essentially in that channel.

Since the linear symmetry algorithm itself is hierarchical, the implemenation yields an intermediate pyramid before the orientation pyramid is obtained. This pyramid, like the orientation pyramid, consists of the complex images

$$(\nabla \mathbf{f})^2 = \left(\frac{\partial \mathbf{f}}{\partial x} + \mathrm{i}\frac{\partial \mathbf{f}}{\partial y}\right)^2 = (\mathbf{f}_x^2 + \mathbf{f}_y^2)\exp(2\varphi(\mathbf{f})), \tag{12}$$

where $\varphi(\mathbf{f}) = \arg(\mathbf{f}_x + i\mathbf{f}_y)$. Thus, every level of the Laplacian pyramid is filtered linearly and then squared resulting in an image of $(\nabla \mathbf{f})^2$. This complex image and its real valued magnitude image are in turn filtered separately with a Gaussian yielding an orientation image (9). The magnitude represents the degree of orientation dominance and the argument represents the orientation. This process is illustrated in the upper branch of Fig. 4.

Similarly, the local energy pyramid is built up from the Laplacian pyramid. The real valued pixels in the Laplacian pyramid are first squared pixelwise and then averaged by means of the same Gaussian as before (11).

The resulting features can be fed to a general vector image segmentation algorithm after the necessary size equalization of the images through interpolations [10]. In that case, the complex images can be treated as two separate features representing the real and imaginary parts.

5. Experiments

In the previous sections, we have described how to obtain two feature pyramids and then proposed the images in these pyramids to be texture features.

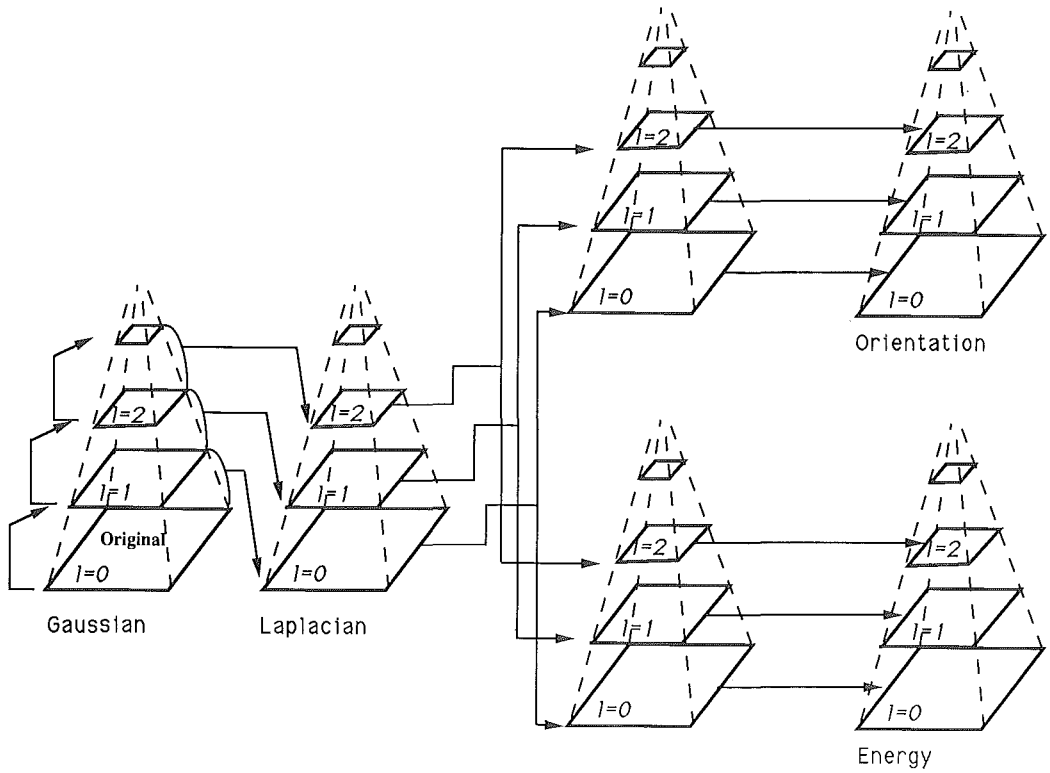

Fig. 4. The data flow in the pyramid structure proposed. The arrows represent dependencies. The first two pyramids on the left represent the Gaussian and the Laplacian pyramids, respectively. The upper branch on the right computes the linear symmetry according to Fig. 2. The first pyramid in this branch represents the $(\nabla f)^2$, while the second corresponds to the local orientation estimation. The images in this pyramid are obtained by dividing the smoothed $(\nabla f)^2$ with the smoothed $|(\nabla f)^2|$. The lower branch computes the local energy. The first pyramid is obtained by squaring the Laplacian pyramid inputs. The second pyramid corresponds to the results obtained by smoothing the preceding pyramid.

In the experiments presented below, the octave low-pass filter used to build the Laplacian pyramid consisted of an 11×11 Gaussian filter. It was truncated at a radius when the value of the filter reached 0.1% of its maximum value. This choice was made because a Gaussian is isotropic and separable, although the 5×5 filter proposed in [10] could be used with somewhat poorer performance. This degradation was noticeable at the highest frequency channel and can be explained by the fact that the cascades of the 5×5 filter approximate a rotationally symmetric filter first at medium and low frequencies. To build the local energy pyramid a 17×17 Gaussian filter, averaging the squares of the Laplacian pyramid images was utilized. The filter used to obtain the partial derivative in order to compute orientation is given by

$$(x + iy) \exp(-\beta(x^2 + y^2)). \tag{13}$$

It had a size of 9×9 and was composed of separable components. That is, (13) can be rewritten as

$$\exp(-\beta y^2)[x \exp(-\beta x^2)] + i \exp(-\beta x^2)[y \exp(-\beta y^2)], \tag{14}$$

resulting in two sets of separable (in x and y) convolutions.

The truncation of the filter was done in the same way as it is done for the Gaussian filters, that is, the complex filter was truncated at a radius when the absolute value of the filter reached 0.1% of its maximum absolute value. The orientation pyramid was built by averaging the squares of the complex derivatives with a Gaussian filter of the size 17×17, the same filter used for building the local energy pyramid. Moreover, the image pyramids in the following experiments were built to level 3 starting with level 0 with the largest image having the size of 256×256. Thus, the images at the highest levels of the pyramids had the size of 32×32. It should be remembered that the total amount of data in any octave pyramid is less than one and one half times the size necessary to represent the largest image in a pyramid even if one goes up to the size of 1×1.

5.1. *Results on the frequency modulated test image*

The frequency modulated (FM) test image, 256×256 in size, Fig. 5, is designed for the purpose of understanding the signal theoretic behavior of the features proposed. It includes spatial frequencies with all directions and all frequencies up to the half Nyquist frequency. The rate of increase of the spatial frequencies in the radial direction is exponential because this forces the center frequencies of a set of band pass filters (decreasing in octaves) to be separated in an equidistant manner. Figure 6 consists of the profiles of the local energy responses along a line cutting through the center. It illustrates the equidistant separation of the frequency channels, confirming the octave decrease of the channels as obtained by means of the Laplacian pyramid, second in Fig. 4.

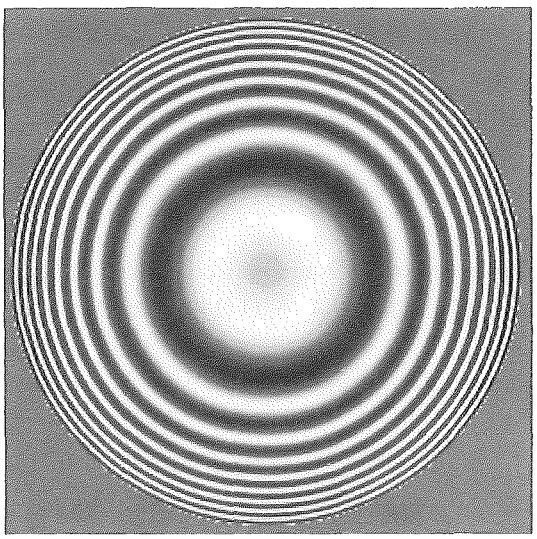

Fig. 5. The frequency modulated (FM) test image.

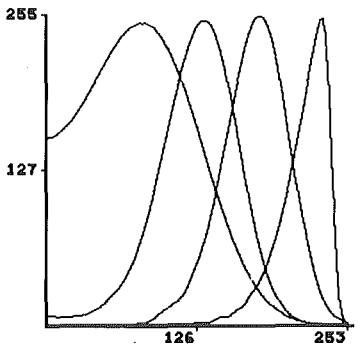

Fig. 6. Cross sections of the energy of Laplacian pyramid of the FM test image.

The result, illustrating z_1 in pyramid, is presented here as a gray image with lines, Fig. 7(b). But in reality, every pixel is complex valued and the advantages of color screens can be utilized to inspect local orientation images. The frequency bands of the corresponding input images decrease clockwise from top-left. The lines in this figure represent different orientations with the lengths being certainties. The center frequencies as well as the band widths of the inputs change in octave progression, Fig. 6. The local energy pyramid is illustrated as a gray scale image on the left in the same figure. In full size, it is illustrated as a set of gray scale images in a similar way on the left. It can be observed that both the certainty response of the linear symmetry and the response of the local energy have no ripples, that is, they are phase independent.

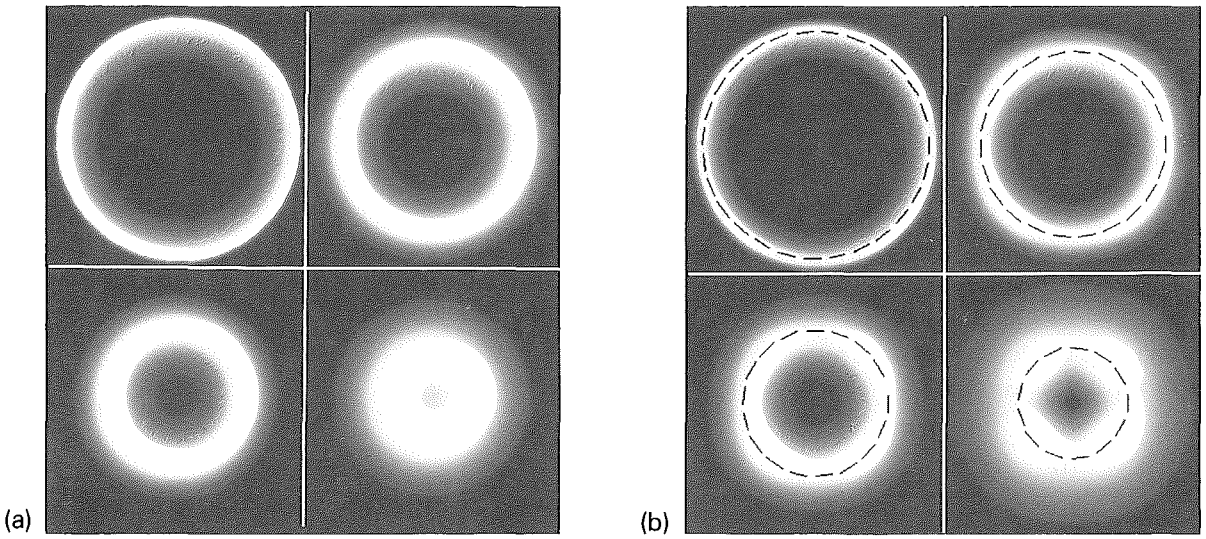

Fig. 7. (a) The local energy pyramid and (b) the local orientation pyramid of the FM test image. The lines represent the orientations at the respective positions while gray values represent certainties.

5.2. CLOG mapping

For many reasons, for example the speed and simplicity of the relevant hardware, the integer number representation is by far the most common technique for storing images, in particular texture feature images. However, special care should be taken in order not to destroy the discrimination power of the features. In image classification, the absolute values of the feature pixels are not important, but their relative values compared to the values of the other pixels are crucial. Since we intend to have integer representation at the final stage, we cannot have the luxury of having good relative quantization error representations at all parts of the unit interval. In the following, we propose an adaptive mapping of the feature values followed by a uniform quantization. Many image processing algorithms require that images, including feature images, should be represented by means of integer numbers per pixel (typically 8 bits) mainly because of storage requirements. But also other more fundamental reasons, like the need of histogram based image processing may be the cause of integer representation. To utilize the available dynamic range better, the feature images should be re-mapped non-linearly in such a way that the outliers, the few pixels which have high values, should be suppressed. In our case this is particularly important since it is known that the histograms of the derivatives of natural images have the form of an exponential function. For this reason, we have applied the centroid adaptive logarithmic mapping, CLOG, described in the following, to our features. That is, the magnitude of the complex pixels are mapped through the CLOG mapping. Similarly, the local energy features were mapped through CLOG mapping before they were quantized. These integer feature images (12 in total) were processed further by the feature reduction and segmentation methods.

We assume that the images are in the floating point representation and are scaled so that they have the dynamic range of [0, 1]. Here 1 represents the maximum and 0 represents the minimum magnitude in the feature images. Utilizing uniform discretization to represent the feature value is neither economical nor accurate for the subsequent arithmetic operations. If n is the available number of bits per pixel we have $s = 2^n$ different labels for the numbers to be represented. Quantizing

uniformly the unit interval, the representable real numbers are given by $q_1(k)=k/s$, where k is an integer and runs between 0 and $s-1$. Differentiating q_1 with respect to k gives us

$$dq_1=\frac{1}{s}dk. \tag{15}$$

This equation shows that the quantization error is independent of k and is constant throughout the interval. But on the contrary the relative quantization error is not constant:

$$\frac{dq_1}{q_1}=\frac{1}{sq_1}dk=\frac{1}{k}dk. \tag{16}$$

Thus, the closer the integer k is to 0, the larger the relative quantization error gets. The image and its histogram in Fig. 8 (left part of the figure) illustrate the magnitude of the Laplacian of the test image with uniform discretization. It can clearly be seen that the majority of the pixels are at the lower end of the interval. Thus, the relative quantization errors made in this representation of Laplacian are very large. The wish to represent very few large numbers with high precision caused the majority of the pixel values represented very poorly.

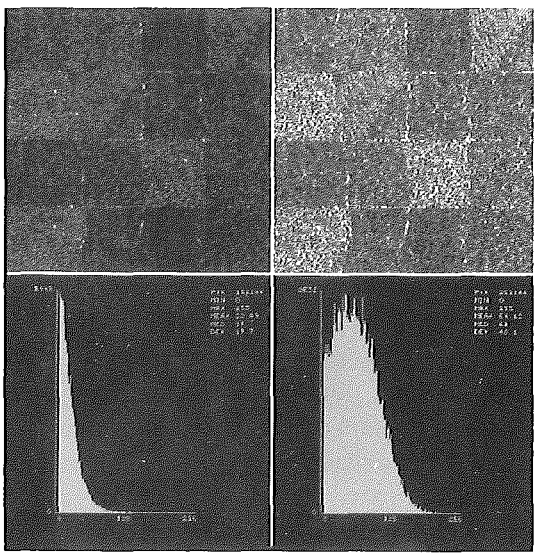

Fig. 8. The magnitude of Laplacian of a test image and its histogram. Results of linear mapping (left) and the CLOG mapping (right).

We assume that the relative quantization error in the representation is constant:

$$\frac{dq}{q}=\frac{C}{s}dk, \tag{17}$$

where C is a positive real constant and s is a positive integer. The solution of this differential equation (with $q(0)=1$) is given by

$$q(k)=\exp\left(\frac{kC}{s}\right). \tag{18}$$

However, when k and s are positive integers then $q(k)\geq 1$. Thus, to represent a value $x\in[0, 1]$ by means of $q(k)$ we apply translation and scaling (translation would suffice but we will explain the reason of scaling with a constant a, soon):

$$q(k)=1+ax. \tag{19}$$

Thus

$$\frac{k}{s}\approx\frac{\ln(1+ax)}{C}. \tag{20}$$

Here the $\approx$ is used since the right-hand side is in general irrational and k/s can only be its rational approximation if k and s are restricted to integers. Remembering that $x\in[0, 1]$ and $k\in\{0, \ldots, s-1\}$ forces C to be $\ln(1+a)$. Equation (20) suggests that we should discretize the image uniformly after having transformed it through the mapping

$$h(x)=\frac{\ln(1+ax)}{\ln(1+a)}. \tag{21}$$

Here x represents the feature image. h is well defined if $-1<a$ for $x\in[0, 1]$. When a approaches zero the mapping approaches the linear mapping, $h(x)=x$, which can be verified by Taylor expansion.

Figure 9 illustrates $h(x)$ for 3 different values of a. The limit functions when a approaches -1 and ∞ are simply unit impulses at the positions 0 and 1, respectively.

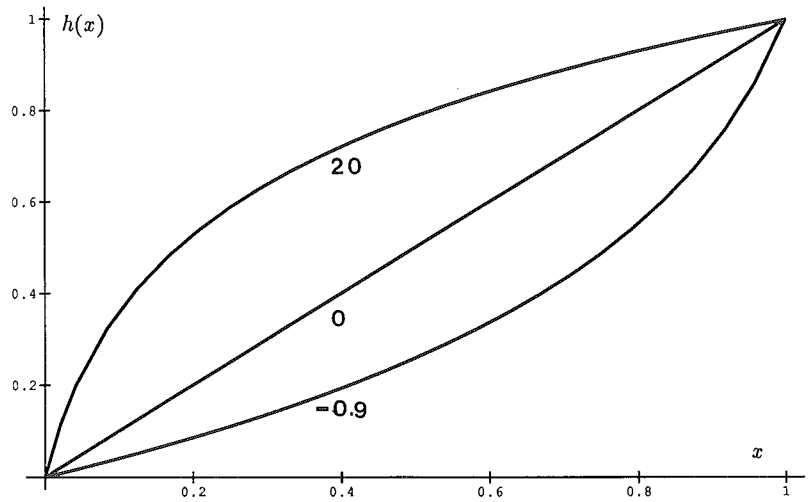

Fig. 9. The function utilized in CLOG mapping for 3 different values of a: 20, 0 and −0·9.

The constant a is proposed to depend on the centroid of the historgram. Let μ be the abscissa (the coordinate in the bin axis) of the centroid of the continuous image histogram (frequency function). We approximate μ as the mean value of the feature image. We choose a in such a way that μ is mapped to m_c, $m_c = g(\mu)$, e.g. by the Newton–Raphson method:

$$m_c = \frac{\ln(1+a\mu)}{\ln(1+a)}, \tag{22}$$

where m_c is an empirically chosen constant in the central one third of the unit interval $[\frac{1}{3}, \frac{2}{3}]$. Thus, for every feature image we have a new function h which is adapted to the histogram of the image in such a way that the histogram of the resulting image has a mass shift towards the center of symbol range. For this reason we call this non-linear mapping a centroid adaptive logarithmic mapping, CLOG. It is worth noting that the mass center of the new histogram does not become m_c in general. Similarly, the mass of the new histogram moves towards m_c from above if the original histogram has a centroid with its x coordinate larger than m_c. This bit-allocation defavors the extreme values which are obtained after a derivative filtration.

In clustering methods the problem of outliers has been known as robust discrimination [9, 24]. In particular, we note that an important class of these methods cluster the rank of the samples instead of the samples themselves. Clustering in this space has been observed to be more robust in comparison with the original space. One reason is that the feature space remains unchanged even if an extreme value is very far away from the rest of the samples. Another reason is that the distribution of the rank data is expected to be more normal like even if the original data is drawn from an unknown distribution [9]. In our particular application, ranking means that the pixel values (in floating point representation) within a feature image are sorted and replaced by their rank in the sorted set. The ranking is carried out in such a way that the smallest value is assigned the rank value 1 and the next smallest value is assigned the value 2, etc. This procedure can be seen as a nonlinear mapping followed by a discretization where the mapping is similar to the CLOG mapping when the image histogram is exponential like.

Figure 8 (right) illustrates the result of the CLOG mapping when $m_c = 0.4$ followed by a uniform discretization. We have tested the CLOG mapping only in connection with our features.

These tests were carried out on hardware using an integer arithmetic convolver, GOP-302. We have utilized the CLOG mapping not only on the magnitudes of the output images coming from the GOP-302 integer arithmetic convolver, but also at the intermediate results between the cascade of the three required convolutions: Laplacian filtering, partial derivative filtering and averaging. Our test results indicate that the texture discrimination property is not very sensitive to different choices of m_c as long as m_c is approximately in the central one third of the unit interval. We have therefore used $m_c = 0.4$ throughout our experiments presented in this paper.

5.3. Results on real texture images

The original image, with the true borders superimposed for the purpose of elucidation, which we used in the subsequent tests, is presented in Fig. 10(a). It consists of different types of field and forest textures and a texture which represents an urban area. The textures were selected by a group of 3 people, independent of the author, with experience in aerial images to decrease the unconscious bias in the choice of interesting textures. It consists of 16 patches and has size 256×256. Among the patches there are only 7 distinct textures which are arranged in such a way that each of these has all the other textures as a neighbor at least once. Table 1 illustrates the spatial arrangement of the textures corresponding to Fig. 10. The horizontal borders of the patches are straight, while the vertical borders are different realizations of a Markovian process. Moreover each texture is brightness (mean) and contrast (variance) normalized individually, before it is incorporated to the test image.

Table 1
The special arrangement of the textures in the test image

T4	T7	T6	T1
T2	T1	T4	T5
T4	T3	T2	T7
T2	T5	T6	T3

Figure 10(a) illustrates the result obtained by utilizing the proposed features, the automatic feature sub-selection method of [3, 5] and the unsupervised segmentation method of [25]. The feature images were blown up to the size of the original using the interpolation technique

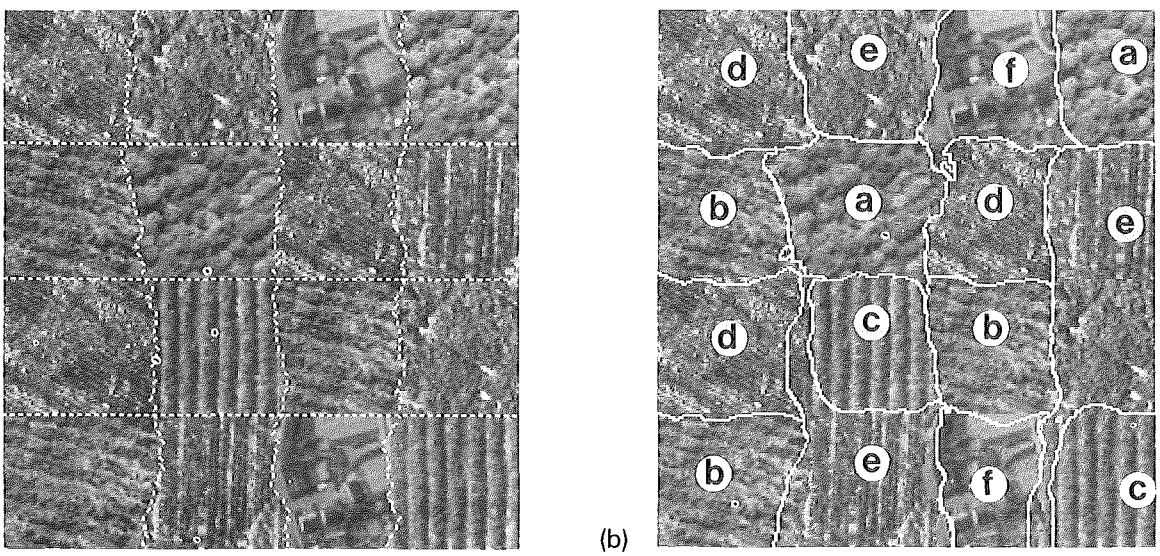

Fig. 10. (a) The used aerial texture image containing 7, different textures corresponding to fields, forests and an urban area. (b) The unsupervised segmentation result of the proposed features.

described by [10] before the selection method was applied. The selection method is based on the use of the KL transform on local neighborhoods as well as on the global image. It applies the KL transform to the vectors which are the best representations of the original feature vectors in a neighborhood. The result is a set of images which are ranked according to a significance vector. Finally, the feature images are selected by thresholding the significance vector. The segmentation method performs clustering in the feature space which now has a reduced dimensionality. In order to increase the computational efficiency and to reduce the noise further, this method performs a size reduction of the feature images. Each feature space point is compared to the other points in order to see whether they lie within a small sphere around the current point. A new feature space is computed, which corresponds to the centroids of the points within that sphere, and the old space is replaced by the new one. This process is repeated until a convergence of the feature space has been obtained. The points in the converged space are then taken as the class centers.

The result predicts 6 classes, of which 5 correspond to true classes and the remaining class corresponds to a merging of two true classes. Almost all borders are found with good accuracy. It should be noted that the classification as well as feature reduction is unsupervised. Thus the border quality and the classes found should be compared with other results, like the subsequent ones to be presented, using unsupervised procedures.

Figure 11(a) illustrates the result when the texture energy measures proposed in [22] are utilized as features along with the same feature reduction and clustering technique as before. The texture energy planes are obtained by first convolving the original by 7 filters (E5L5, L5E5, R5R5, E5S5, S5E5, L5S5, S5L5 as labeled in [22]) and then applying the standard deviation filtering using a 7×7 window. We have also tried larger sizes which resulted in a poorer class and border performance compared to Fig. 11(a). The feature dimensionality reduction process of [3, 5], tested also on many other different feature sets, seems to be an adequate means to compress the information and increase the discrimination power of the texture features. The texture feature dimensionality reduction has also been dealt with recently in [28]. The result suggests that 4 classes exist of which two more or less represent two classes while the two others represent the merged classes. Many borders are identified with good accuracy.

Figure 11(b) illustrates the segmentation result when the features proposed by [27] are utilized. These features, totaling 4, are obtained by first computing the discrete Hadamard transform within a neighborhood of 2×2 and then applying the variance filtering to those in a window of 8×8. In the result 4 classes can be distinguished and of these one corresponds to a true class, while the remaining are merged classes. Some of the borders can be identified with reasonable accuracy.

We have also tried to obtain a segmentation result using the 13 features proposed in [17]. The features used represented certain, in [17] well defined, properties (angular second moment, contrast, etc.) of the co-occurrence matrix with 64 quantization levels computed in 9×9 neighborhoods using a distance vector of $(0, 1)^T$. The result was a single class for the entire image and naturally no borders.

Here we stress the fact that no attempt to choose the best features manually from the originally proposed total set was made. The used automatic selection method is based on local homogeneity and cannot compete with a human, who subconsciously also uses her expertise in classification, when making her choice. Moreover, the clustering methods are in general quite bad in finding different class modes as the dimensionality of the feature space increases. Thus, we do not exclude that the result stemming from this feature set can be improved if the conditions of the entire segmentation process are changed. However, in order to limit the scope of this paper, we have refrained from such an attempt.

An explanation for why the proposed features have performed well is that the textures in the test

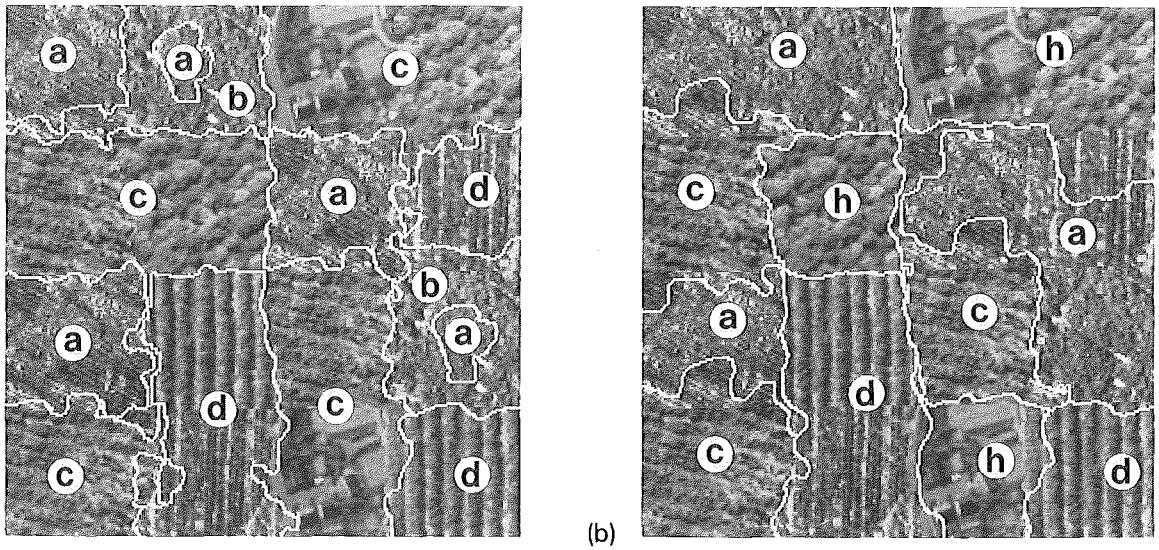

Fig. 11. The segmentation result using the features of (a) [22] and (b) [26].

image differ in their 'repetition frequency', e.g. scale, considerably while the used neighborhoods in the 3 compared methods were constant. On the other hand, a straight-forward application of these techniques to every level of a multiresolution pyramid would result in too many texture features for further unsupervised processing. The proposed features are directly measuring the dominant orientation and frequency information in each channel. This is a property which is in line with the current understanding of the human visual system. The results of these unsupervised experiments and

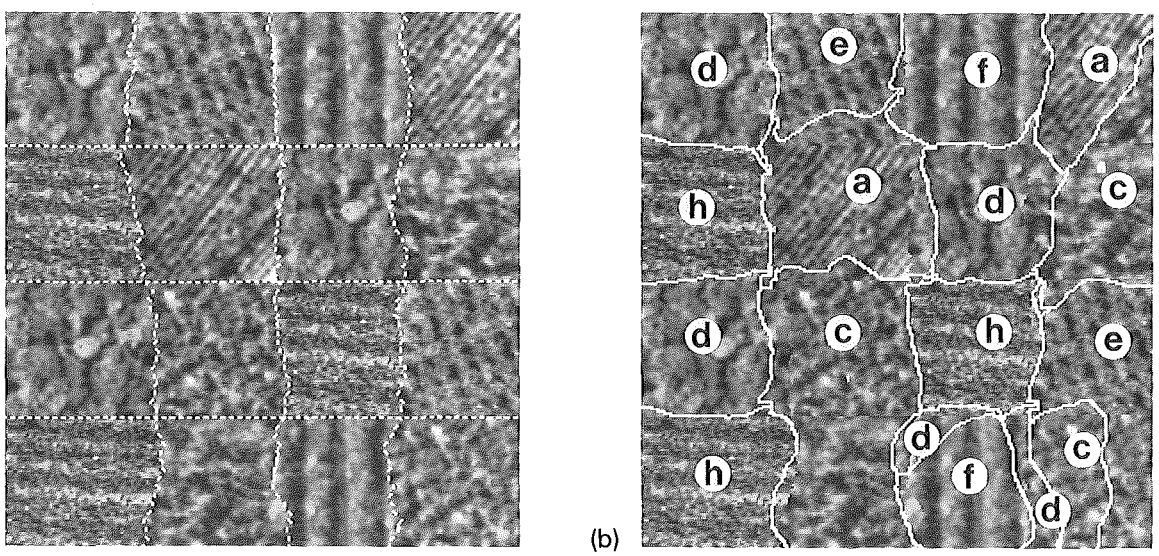

Fig. 12. The used test image containing another set of aerial textures and the segmentation result obtained by using the proposed features.

others, like those of Figs. 10 and 12, we have performed, using aerial images suggest that the linear symmetry features can be used successfully in texture analysis.

6. Conclusion

A conceptually common feature in the local energy and local orientation pyramid is that these results are obtained as averaging the result of a linear filtering followed by the same non-linear operation, squaring. The averaging after this non-linear operation has an optimization effect as indicated in Section 2.

If a Gaussian pyramid based segmentation algorithm is utilized, averaging will be done inherently by the segmentation algorithm. For the images of a high level in the local energy and the local orientation pyramid this means that the segmentation algorithm can start from a small image, offering computational advantages. The experiments done during this work indicate that the linear symmetry pyramid together with the local energy pyramid are efficient texture features. The pyramid structure facilitates the modeling of textures at different sizes with respect to isotropy and anisotropy allowing the proposed features to be incorporated into real time recognition systems. Moreover, the linear symmetry applied to a certain frequency channel, i.e. a level in the orientation pyramid, has a similar response to the orientation estimation through interpolation of the Gabor or quadrature mirror filter responses. But the computational cost per pixel is much less.

Acknowledgments

Credits should be given to Roland Bapst for his work during the implementation, to Thompson CSF and to the Swiss Federal Institute of Technology in Lausanne for the financial support.

References

[1] R. Bajcsy and L.I. Lieberman, "Texture gradient as a depth cue," *Computer Graph Image Process.*, Vol. 5, 1976, pp. 52–67.

[2] J. Bigün, Pattern recognition by detection of local symmetries, in: E.S. Gelsema and L.N. Kanal, eds., *Pattern Recognition and Artificial Intelligence*, North-Holland, Amsterdam, 1988, pp. 75–90.

[3] J. Bigün, "Feature reduction in image segmentation by local Karhunen–Loève transform", *Proc. of ICPR-11*, The Hague, The Netherlands, 31 August–3 September 1992.

[4] J. Bigün and G.H. Granlund, "Optimal orientation detection of linear symmetry", *First Internat. Conf. Computer Vision*, London, June 1987, pp. 433–438.

[5] J. Bigün and C. Horne, "Feature space dimensionality reduction by using local representability", *Internat. Conf. on Signal Process.*, Beijing, 22–26 October 1990, pp. 555–558.

[6] J. Bigün, G.H. Granlund and J. Wiklund, "Multidimensional orientation estimation with applications to texture analysis and optical flow", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No. 8, 1991, pp. 775–790.

[7] C. Blakemore and F.W. Campbell, "On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images", *J. Physiology*, Vol. 203, 1969, pp. 237–260.

[8] P. Brodatz, *Textures*, Dover, New York, 1966.

[9] J.D. Broffitt, "Nonparametric classification", in: P.R. Krishnaiah and L.N. Kanal, eds., *Handbook of Statistics*, Vol. 2, North-Holland, Amsterdam, 1982, pp. 139–168.

[10] P. Burt, "Fast filter transforms for image processing", *Comput. Graph. Image Process.*, Vol. 16, 1981, pp. 20–51.

[11] P. Burt, "Smart sensing within a pyramid vision machine", *IEEE Proc.*, Vol. 76, No. 8, 1988, pp. 1006–1015.

[12] D.J. Fleet and A.D. Jepson, "Hierarchical construction of orientation and velocity selective filters", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No. 3, 1989, pp. 315–325.

[13] D. Gabor, "Theory of communication", *J. Inst. Elec. Engrg.*, Vol. 93, 1946, pp. 429–459.

[14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian resaturation of images", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 11, 1984, pp. 721–741.

[15] G.H. Granlund, "In search of a general picture processing operator", *Comput. Graph. Image Process.*, Vol. 8, 1978, pp. 155–173.

[16] A. Grossman and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", *SIAM J. Math.*, Vol. 15, 1984, pp. 723–736.

[17] R.M. Haralick and I. Dinstein, "Textural features for image classification", *IEEE Trans. System Man Cybernet.*, Vol. 3, No. 11, 1973, pp. 610–621.

[18] M.K. Hu, "Visual pattern recognition by moment invariants", *IRE Trans. Information Theory*, 1962, pp. 179–187.

[19] D.H. Hubel and T.N. Wiesel, "Brain mechanisms of vision", *Scientific American*, Vol. 241, No. 3, September 1979, pp. 150–162.

[20] B. Julesz, "Visual pattern discrimination", *IRE Trans. Information Theory*, Vol. IT-8, 1962, pp. 84–92.

[21] H. Knutsson, Filtering and reconstruction in image processing, Linköping studies in science and technology Dissertations No. 88, 1982.

[22] K.I. Laws, Textured image segmentation, Ph.D. Dissertation, Department of Electrical Engineering, University of Southern California, 1980.

[23] D. Marr, *Vision*, Freeman, San Fransisco, 1982.

[24] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.

[25] M. Spann and R. Wilson, "A quad-tree approach to image segmentation which combines statistical and spatial information", *Pattern Recognition*, Vol. 18, Nos. 3/4, 1989, pp. 257–269.

[26] M. Unser, "Sum and difference histograms for texture classification", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, No. 1, January 1986, pp. 118–125.

[27] M. Unser, "Local linear transforms for texture measurements", *Signal Processing*, Vol. 11, No. 1, July 1986, pp. 61–79.

[28] M. Unser and M. Elden, "Multiresolution feature extraction and selection for texture feature segmentation", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, July 1989, pp. 717–727.

[29] E. Wigner, "On the quantum correction for thermodynamic equilibrium", *Phys. Rev.*, Vol. 40, 1932, pp. 749–759.