Texture Segmentation Using Local Energy in Wavelet Scale Space

Zhi-Yan Xie and Michael Brady

Robotics Research Group, Department of Engineering Science
University of Oxford, Oxford OX1 3PJ, U.K.

Abstract. Wavelet transforms are attracting increasing interest in computer vision because they provide a mathematical tool for multiscale image analysis. In this paper, we show that i) the subsampled wavelet multiresolution representation is translationally variant; and ii) a wavelet transform of a signal generally confounds the phase component of the analysing wavelet associated with that scale and orientation. The importance of this observation is that commonly used features in texture analysis also depend on this phase component. This not only causes unnecessary spatial variation of features at each scale but also makes it more difficult to match features across scales.
In this paper, we propose a complete 2D decoupled local energy and phase representation of a wavelet transform. As a texture feature, local energy is not only immune to spatial variations caused by the phase component of the analysing wavelet, but facilitates the analysis of similarity of across scales. The success of the approach is demonstrated by experimental results for aerial Infrared Line Scan (IRLS), satellite, and Brodatz images.

1 Introduction

In a multiresolution approach, texture segmentation typically consists of several basic stages: 1) application of a set of wavelet-like filters tuned to different frequency band and orientations; 2) one or more nonlinear operations to the linear filtering outputs; 3) average of the resultant responses to derive "texton" density measure; 4) finally, the segmentation. To choose appropriate filters (wavelets) and adequate nonlinear operations in the first two stages are critical for the subsequent processing.
For a continuous wavelet transform (CWT), it is defined by

$$W_\psi f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x)\overline{\psi}(\frac{x-b}{a})\,dx$$

where $a \in \mathbf{R}^+, b \in \mathbf{R}$ are scale and translation parameters and the wavelet $\psi(x)$ can be any bandpass function. If the scale and space parameters of CWT are sampled at $\{a = 2^{-j};\ b = 2^{-j}k;\ j,k \in \mathbf{Z}\}$, and the wavelet $\psi(x)$ is biorthogonal, then the most compact and complete wavelet transform is defined

by

$$W_\psi f(j,n) = 2^{\frac{j}{2}} \sum_k f(k)\overline{\psi}(2^j k - n) \triangleq < f(x), \psi_{j,n}(x) > \quad (1)$$

where $\psi_{j,n}(x) = 2^{\frac{j}{2}}\psi(2^j k - n)$. Although this representation has been shown useful in image compression because of its completeness and compactness, $W_\psi f(j,n)$ is translationally variant by noticing

$$\begin{aligned} W_\psi f(j,n) &= 2^{\frac{j}{2}} < f(x - x_0), \psi_{j,n}(x) > \\ &= 2^{\frac{j}{2}} < f(x), \psi_{j,n-2^j x_0}(x) > \\ &\neq 2^{\frac{j}{2}} < f(x), \psi_{j,n-m}(x) > \end{aligned}$$

For an arbitrary translation x_0, $x_0 = 2^{-j}m$; $m \in \mathbf{Z}$ won't in general be satisfied. In other words, if two identical signals were to appear in different positions, their representations $W_\psi f(j,n)$ can be quite different. Such performance may not cause problem for image compression but it badly affects texture segmentation.

The need for non-linear operation after linear filtering has been recognised [5, 1], but there are still no principles determining what kind of operation should be used. Overwhelmingly, the nonlinear operations used to generate features are nonlinear smoothing [3, 5, 7] and energy measure [8, 4]. Although smoothing can remove weak variations, it is also liable to destroy important details. The most commonly used energy measures are full- and half-wave rectification and the square power. However, we show that, in general the features generated by applying these energy measures to the output of the wavelet transform are coupled with the local phase component that depends not only on the analysed image but also on the analysing wavelet at that scale. This is because the output of a wavelet transform oscillate in space depending on the shape of the wavelets. Consequently, their moduli are also affected by the oscillation as illustrated in Figure 1 (top row). Such performance is unacceptable for texture analysis because one wants a uniform feature response in those regions of the image which have uniform texture, while a wavelet is typically an oscillating, wave-like function. Hence, some other nonlinear operation must be found to derive features which can be invariant to *the phase at each scale.*

In this paper, we first define a translationally invariant wavelet transform and propose a definition of 2D local energy and phase in wavelet scale-space. Based on this definition, a complete, decoupled local energy and phase representation of 2D wavelet transforms is presented. As a result of this theory, the local energy is phase independent and used as feature for texture segmentation.

2 Translation Invariant Wavelet Transform

To overcome the translation variant problem of the most compact wavelet transform, we change scale-dependent sampling, i.e. $\{a = 2^{-j};\ b = 2^{-j}k;\ j,k \in \mathbf{Z}\}$,

to a uniform spatial sampling given by $\{a = 2^{-j};\ b = k;\ j, k \in \mathbf{Z}\}$. Hence, the discrete wavelet transform is redefined by

$$WW_\psi f(j,n) = 2^j \sum_k f(k)\overline{\psi}(2^j(k-n))$$

Clearly, $WW_\psi f(j,n)$ is an oversampled version of $W_\psi f(j,n)$ given in Eqn: 1 and it becomes translationally invariant. Although $WW_\psi f(j,n)$ sacrifices the compactness of the representation compared with $W_\psi f(j,n)$, it provides translational invariance which is essential for texture segmentation, and many other image processing tasks.
Similarly, a translationally invariant 2D separable wavelet transform can be defined by

$$\begin{aligned} DD_x^j(m,n) &= < f(x,y), 2^j\phi(2^j(x-m))2^j\psi(2^j(y-n)) > \\ DD_y^j(m,n) &= < f(x,y), 2^j\psi(2^j(x-m))2^j\phi(2^j(y-n)) > \\ DD_d^j(m,n) &= < f(x,y), 2^j\psi(2^j(x-m))2^j\psi(2^j(y-n)) > \end{aligned} \tag{2}$$

It can be shown that DD_x, DD_y, DD_d have same orientation emphasis as their subsampled version as given in [6]. More precisely, they give strong response to spatial structures in the horizontal, vertical and diagonal directions[1], and so they are called the horizontal, vertical and diagonal channels, respectively.

3 2D Decoupled Local Energy and Phase

A 1D decoupled local energy and phase representation of a real-valued wavelet transform has been developed using the Hilbert transform [10]. The principal theoretical difficulty in extending the local energy and phase representation of the 1D wavelet transform to 2D is that there does not exist a universal 2D Hilbert transform.
Nevertheless, we propose a definition of the local energy and local phase of a 2D wavelet transform which not only provides a complete representation of a 2D wavelet transform in scale-space, but also facilitates the local energy to be independent from the phase components of the analysing wavelets. Moreover, the relationship between a 2D wavelet transform and its local energy is established both in *scale-space* and in *frequency* space.

3.1 Horizontal and vertical channels

Recall the mother wavelets associated with the horizontal and vertical channels:

$$\Psi^1(x,y) = \phi(x)\psi(y) \qquad \Psi^2(x,y) = \psi(x)\phi(y)$$

[1] The orientation of the spatial structure is defined as perpendicular to the direction of maximum gradient.

Since the scaling function $\phi(x)$ is low-pass, $DD_x^j(m,n)$ can be considered as a 1D wavelet transform with respect to $\psi(y)$ for each column (y axis) after first smoothing each row (x axis). Similarly, $DD_y^j(m,n)$ can be considered as a 1D wavelet transform for each row after first smoothing each column. Hence, the local energy and phase can be defined as given in 1D [10]. More precisely,

Definition 1. For a real valued ψ and $f(x,y) \in \mathbf{L^2(Z^2)}$ the local energy ρ_x, ρ_y and the local phase φ_x, φ_y of $DD_x^j(m,n)$ and $DD_y^j(m,n)$ are given by

$$\rho_x^j(m,n) = \sqrt{[DD_x^j(m,n)]^2 + [H_y\{DD_x^j(m,n)\}]^2}$$
$$\varphi_x^j(m,n) = Atan2\frac{H_y\{DD_x^j(m,n)\}}{DD_x^j(m,n)}$$
$$\rho_y^j(m,n) = \sqrt{[DD_y^j(m,n)]^2 + [H_x\{DD_y^j(m,n)\}]^2}$$
$$\varphi_y^j(m,n) = Atan2\frac{H_x\{DD_y^j(m,n)\}}{DD_y^j(m,n)}$$

where $H_x\{.\}$ ($H_y\{.\}$) denotes the Hilbert transform of the 1D function $DD(m,n)$ when n (m) is fixed.

3.2 Diagonal channel

The diagonal channel of the wavelet transform of a function $f(x,y)$ is given by

$$DD_d^j(m,n) =< f(x,y), 2^{2j}\Psi^3(2^j(x-m), 2^j(y-n)) > \tag{3}$$

The mother wavelet associated with $DD_d^j(m,n)$ is

$$\Psi^3(x,y) = \psi(x)\psi(y) \tag{4}$$

We construct four complex functions as follows

$$\begin{aligned} G_1(x,y) &= [\psi(x) + i\psi_H(x)][\psi(y) + i\psi_H(y)] \\ G_2(x,y) &= [\psi(x) - i\psi_H(x)][\psi(y) - i\psi_H(y)] \\ G_3(x,y) &= [\psi(x) + i\psi_H(x)][\psi(y) - i\psi_H(y)] \\ G_4(x,y) &= [\psi(x) - i\psi_H(x)][\psi(y) + i\psi_H(y)] \end{aligned} \tag{5}$$

Noting the following conjugacy relationships,

$$\begin{aligned} DG_1^j(m,n) &= \overline{DG_2^j}(m,n) \\ DG_3^j(m,n) &= \overline{DG_4^j}(m,n) \end{aligned} \tag{6}$$

only one pair $\{DG_k^j(m,n); k = 1,3\}$ or $\{DG_k^j(m,n); k = 2,4\}$ needs to be considered. In the following discussion, we take the first pair. Substituting $\Psi^3(x,y)$ with $\{G_k(x,y); k = 1,3\}$ in Eqn. 3, we generate two complex images by:

$$DG_k^j(m,n) \stackrel{\text{def}}{=} < f(x,y), 2^{2j}G_k(2^j(x-m), 2^j(y-n)) > \; ; \; k = 1,3 \tag{7}$$
$$= (f(x,y) * 2^{2j}\overline{G}_k(-2^j x, -2^j y) \tag{8}$$

The properties of functions $\{DG_k^j(m,n); k = 1,3\}$ are essential for deriving the decoupled local energy and local phase representation of the diagonal channel. We present them in the following lemma.

Lemma 2. *For each scale j*

1. *The functions $DG_1^j(m,n)$ can be represented by $DD_d^j(m,n)$ as*

$$DG_1^j(m,n) = DD_d^j(m,n) - H_y\{H_x\{DD_d^j(m,n)\}\} + i(H_x\{DD_d^j(m,n)\} + H_y\{DD_d^j(m,n)\}) \quad (9)$$

   and gives a strong response to spatial structures at or close to $\frac{\pi}{4}$.

2. *The functions $DG_3^j(m,n)$ can be represented by $DD_d^j(m,n)$ as*

$$DG_3^j(m,n) = DD_d^j(m,n) + H_y\{H_x\{DD_d^j(m,n)\}\} + i(H_x\{DD_d^j(m,n)\} - H_y\{DD_d^j(m,n)\}) \quad (10)$$

   and gives a strong response to spatial structures at, or close to $\frac{3\pi}{4}$.

Proof. See [9].

Now we are in the position to define the local energy and local phase of the diagonal channel:

Definition 3. The function $DG_1^j(m,n)$ is a complex function and can be written as

$$DG_1^j(m,n) = \rho_{x+y}^j(m,n)e^{i\varphi_{x+y}^j(m,n)}$$

where

$$\rho_{x+y}^j(m,n) = \sqrt{[g - H_{xy}\{g\}]^2 + [H_x\{g\} + H_y\{g\}]^2} \quad (11)$$

$$\varphi_{x+y}^j(m,n) = Atan2[\frac{g - H_{xy}\{g\}}{H_x\{g\} + H_y\{g\}}] \quad (12)$$

where $g = DD_d^j(m,n)$ and $H_{xy}\{.\}$ denotes the Hilbert transform along x, followed by along y. The $[\rho_{x+y}^j(m,n)]^2$ and $\varphi_{x+y}^j(m,n)$ are called the local energy and the local phase of $DD_d^j(m,n)$,respectively, at or close to $\frac{\pi}{4}$.

Definition 4. The function $DG_3^j(m,n)$ is a complex function and can be written as

$$DG_3^j(m,n) = \rho_{x-y}^j(m,n)e^{i\varphi_{x-y}^j(m,n)}$$

where

$$\rho_{x-y}^j(m,n) = \sqrt{[g + H_{xy}\{g\}]^2 + [H_x\{g\} - H_y\{g\}]^2} \quad (13)$$

$$\varphi_{x-y}^j(m,n) = Atan2[\frac{g + H_{xy}\{g\}}{H_x\{g\} - H_y\{g\}}] \quad (14)$$

where $g = DD_d^j(m,n)$ and $[\rho_{x-y}^j(m,n)]^2$ and $\varphi_{x-y}^j(m,n)$ are called the local energy and the local phase of $DD_d^j(m,n)$,respectively, at or close to $\frac{3\pi}{4}$.

Now we have defined four local energy channels for each scale j, denoted by $\rho_x^j(m,n)$, $\rho_y^j(m,n)$, $\rho_{x+y}^j(m,n)$, $\rho_{x-y}^j(m,n)$, which are oriented in the horizontal, vertical and $\frac{\pi}{4}$, $\frac{3\pi}{4}$ directions, respectively. Comparing a patch of local energy and moduli surfaces shown in Figure 1, the phase dependency embedded in the moduli has been removed in the local energy representations. Moreover, the local energy images at different scales become comparable in terms of shape similarity as indicated by the correlation of the images between adjacent scales shown in Figure 1 (bottom row).

3.3 Properties of the local energy and local phase

The following theorem shows that the local energy and local phase defined above provide a complete representation of 2D wavelet transform. The local energy and the wavelet transform are equivalent in frequency domain (conserve energy), but they are very different in scale-space.

Theorem 5. *For a real valued $\psi(x)$ and $f(x,y) \in \mathbf{L^2(Z^2)}$,*

1. *The wavelet transform $DD_x^j(m,n)$, $DD_y^j(m,n)$ and $DD_d^j(m,n)$ can be represented completely by the local energies and local phases and are given by*

$$\begin{aligned} DD_x^j(m,n) &= \rho_x^j(m,n)\cos\varphi_x^j(m,n) \\ DD_y^j(m,n) &= \rho_y^j(m,n)\cos\varphi_y^j(m,n) \\ DD_d^j(m,n) &= \frac{1}{2}(\rho_{x+y}^j(m,n)\cos\varphi_{x+y}^j(m,n) + \rho_{x-y}^j(m,n)\cos\varphi_{x-y}^j(m,n)) \end{aligned} \tag{15}$$

2. *For each scale $j < 0$,*

$$\begin{aligned} \sum_m\sum_n [DD_x^j(m,n)]^2 &= \frac{1}{2}\sum_m\sum_n [\rho_x^j(m,n)]^2 \\ \sum_m\sum_n [DD_y^j(m,n)]^2 &= \frac{1}{2}\sum_m\sum_n [\rho_y^j(m,n)]^2 \\ \sum_m\sum_n [DD_d^j(m,n)]^2 &= \frac{1}{8}\sum_m\sum_n ([\rho_{x+y}^j(m,n)]^2 + [\rho_{x-y}^j(m,n)]^2) \end{aligned} \tag{16}$$

Proof. See [9].

The wavelet detail images and their associated local energy images are very different in scale-space: the former confounds the phase component, the latter does not. Further, from Eqns. 15, it is clear that full-, half-wave rectification or squaring of the wavelet transform also confound the phase component.

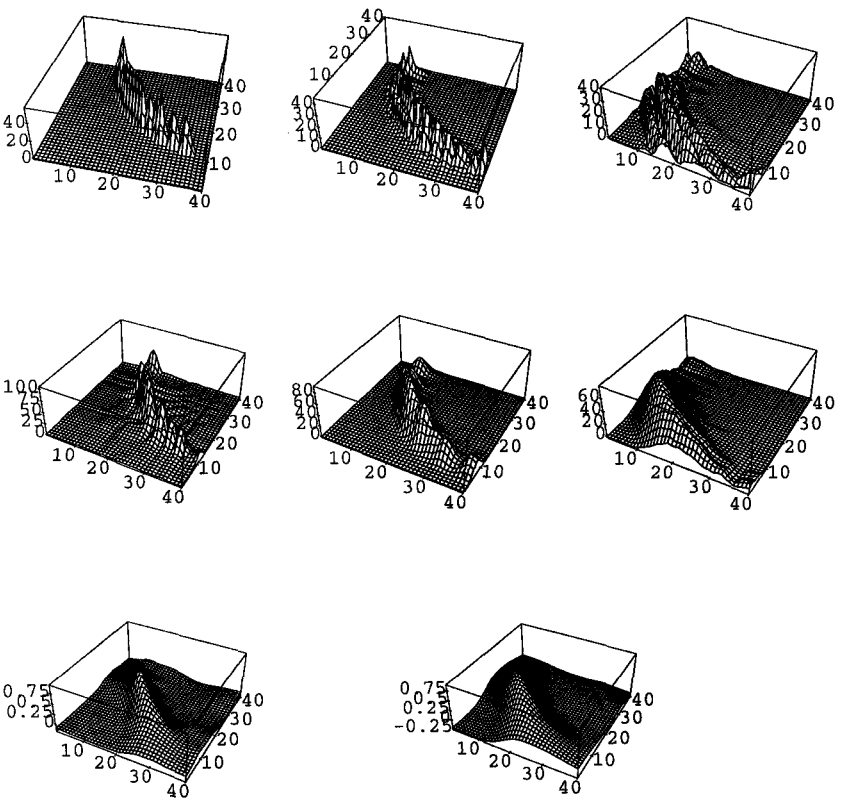

Fig. 1. Comparison between the modulus and the local energy of the wavelet transform. Top row (left to right): the plot of patches of $|DD_d^j|$ of an circular impluse edge, at scale $j = -1, -2, -3$, respectively; middle row: the plot of the local energy ρ_{x-y}^j for the same patch at scale $j = -1, -2, -3$, respectively; bottom row: the plot of the linear correlation coefficients of two local energy images at adjacent scales, $j = -1, -2$ and $j = -2, -3$ respectively.

4 Application to texture segmentation

As shown in the last section, the outputs of real-valued wavelet transform at each scale are coupled with the phase of the wavelet associated with that scale and orientation. As a consequence, squaring, half- and full-wave rectification of the outputs are also phase dependent. Furthermore, it is known that the outputs of wavelet transform at a given location and orientation fail to match across scales according to the local image structures giving rise to the responses. In order to overcome the phase dependency and spatial localisation problems, we propose a four level computation scheme for texture segmentation.

At the **first level**, the 2D **oversampled wavelet transform** is applied to an image. This transform decomposes an image into a stack of images denoted by $DD(\theta, j, x, y)$ at sampled orientation $\theta = \{\theta_1, \cdots, \theta_n\}$ and sampled scale $a = \{2^{-j};\ j = -1, -2, \cdots, -J\}$. For a 2D separable wavelet transform, an image is decomposed into a pile of images $\{DD_x^j(x, y), DD_y^j(x, y), DD_d^j(x, y); j = -1, -2, \cdots, -J\}$.

The **second level** is a nonlinear operation to remove the phase dependency from each image $DD(\theta, j, x, y)$, to obtain a pile of local energy images $\rho(\theta, j, x, y) = \{\rho_x^j(x, y), \rho_y^j(x, y), \rho_{x+y}^j(x, y), \rho_{x-y}^j(x, y)\}$. This level operates only within a single scale, hence it is also called **intra-scale nonlinear fusion**.

The **third level** derives two texture features in wavelet scale-space, i.e. a multiscale orientational measure $\alpha(j, x, y)$ and an energy measure $F(j, x, y)$. This level is composed of two sub-processes, namely **inter-scale clustering** and **inter-orientation fusion**. The inter-scale clustering designed to associate the local energy descriptors $\rho(\theta, j, x, y)$ across scales such that, for the resultant new feature image $\rho'(\theta, j, x, y)$, the spatial localisation problem is minimised globally. Unlike the other levels given above, the inter-orientation fusion is not universal. It is specific to each application and to the meaning of different orientation channels. Currently, a simple formula is used to combine four oriented local energy images into quantitative and orientational measures of local energy given by

$$F(j, x, y) = \sqrt{[\rho'(\theta, j, x, y)|_{\theta=0}]^2 + [\rho'(\theta, j, x, y)|_{\theta=\frac{\pi}{2}}]^2} + c * \sqrt{[\rho'(\theta, j, x, y)|_{\theta=\frac{\pi}{4}}]^2 + [\rho'(\theta, j, x, y)]^2|_{\theta=\frac{3\pi}{4}}} \quad (17)$$

$$\alpha(j, x, y) = \arg\left(\frac{\rho'(\theta, j, x, y)|_{\theta=\frac{\pi}{2}}}{\rho'(\theta, j, x, y)|_{\theta=0}}\right) \quad (18)$$

The **fourth level** is the segmentation, which is carried out by Gaussian smoothing, clustering and post-processing.

The texture segmentation scheme given above is implemented and has been tested on more than 30 real aerial and satellite images. Typical results are shown in Figure 2. Figure 2 (a), (b) show typical IRLS images taken from a low flying aircraft. The goal (part of a system under development for matching images on successive fly-pasts and matching/constructing a map) is to segment rural and urban areas. The patches in Figure 2 (b) correspond to parks within the surrounding urban area. Figure 2 (c) shows the segmentation of a satellite image taken over Plymouth area, the segmentation result is matched quite well with the map over same area. Finally Figure 2 (d) shows the segmentation of two Brodatz textures [2] (cotton canvas and woolen cloth).

5 Conclusions

In this paper, we developed a complete, decoupled local energy and phase representation of a 2D oversampled wavelet transform. This representation provides

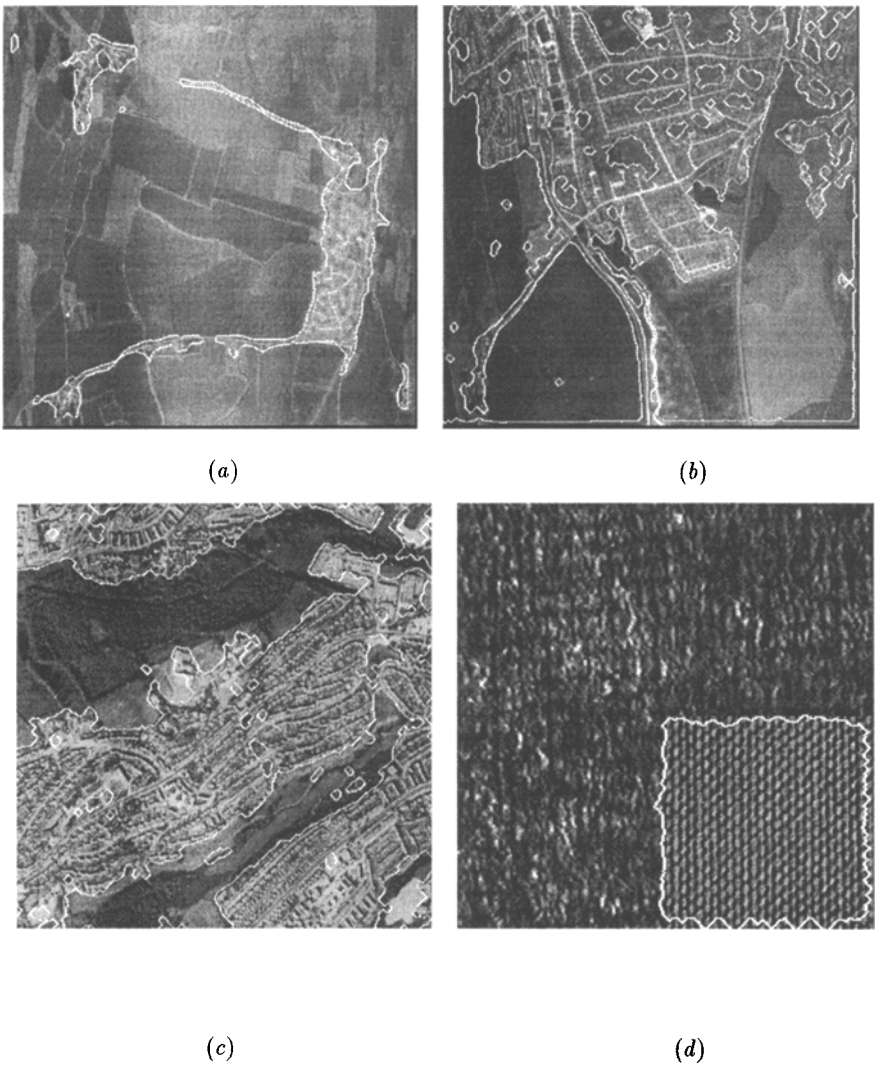

Fig. 2. Examples of texture segmentation results. Texture boundaries are extracted and superimposed on their original images. Top row: Urban regions have been extracted for real IRLS aerial images; bottom row (left to right): Urban regions have been extracted for a satellite image, cotton canvas (in right bottom corner) have been picked up from woolen cloth background for a Brodatz montage image.

a guide to choose appropriate wavelet by revealing the phase dependency problem associated with widely used real-valued wavelet transforms.
- an approach to construct a complex-valued wavelet from a real-valued wavelet function such that the phase dependency problem can be overcome;
- a method to derive local energy in wavelet scale-space. As a local feature, the local energy is not only immune to spatial variations caused by the phase component of the analysing wavelet, but facilitates the analysis of similarity of across scales;
- a way to formulate 2D Hilbert transform which is still an open problem.

The usefulness of this decoupled local energy and phase representation is demonstrated by its application to segment textures in several classes of natural images.

References

1. A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, 1987.
2. Phil Brodatz. *Textures.* Dover Publications, Mineola NY, 1966.
3. A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24:1167–1186, 1991.
4. A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE PAMI*, 15(11):1186–1190, 1993.
5. Jitendra Malik and Pietro Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7(5):923–932, 1990.
6. Stephane G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE PAMI*, 11:671–693, December 1989.
7. T. Randen and J. Husoy. Multiscale filtering for image texture segmentation. *Optical Engineering*, 33:2617–2625, August 1994.
8. A. Sutter, J. Beck, and N. Graham. Contrast and spatial variables in texture segregation:testing a simple spatial-frequency channels model. *Pecept. Psychophys.*, 46:312–332, 1989.
9. Z. Xie. *Multi-scale Analysis and Texture Segmentation.* PhD thesis, University of Oxford, 1994.
10. Zhi-Yan Xie and Michael Brady. A decoupled local energy and phase representation of a wavelet transform. In *VCIP'95 (Visual Communications and Image Processing)*, 1995.