Modulation Domain Image Segmentation

T. Tangsukson and J.P. Havlicek
School of Electrical & Computer Engineering
The University of Oklahoma, Norman, OK 73019-1023

Abstract

We present a modulation domain technique for segmenting textured images modeled as 2-D multicomponent AM-FM functions. A Gabor filterbank is used to derive estimates of the dominant modulations at each image point and statistical clustering is then performed with respect to a similarity measure based on modulation domain entropy. Post processing is applied in the image domain to regularize the solution and enforce spatial correspondence. Several synthetic and natural examples are presented.

1. Introduction

The unsupervised segmentation of images into disjoint, connected regions corresponding to objects in the imaged scene has long been recognized as a problem of fundamental importance in computer vision and image understanding [9, 24]. It is a difficult problem, and no general solution has yet been formulated [19]. However, many good solutions have been proposed for specific classes of segmentation problems. Among such classes, the problem of segmenting textured images is particularly challenging. A primary reason for this is that a well-posed definition of texture has only begun to emerge. Many recent successful techniques have been based on multiband filtering [7, 15, 25, 26] or on purely stochastic models [1, 4, 17, 19, 21, 27].

In this paper, we take a distinct approach by characterizing texture in terms of nonstationary amplitude and frequency *modulations*. We model textured images as sums of nonstationary AM-FM functions, or *components*, of the form $t(\mathbf{x}) = a(\mathbf{x})\exp[j\varphi(\mathbf{x})]$. By definition, the instantaneous amplitude $a(\mathbf{x})$ is the AM function of $t(\mathbf{x})$ and the instantaneous frequency $\nabla\varphi(\mathbf{x})$ is the FM function of $t(\mathbf{x})$. Using a recently developed theory of AM-FM image modeling [12], we compute estimates of the modulating functions of the multiple components in an image to arrive at a modulation domain representation. We then address the segmentation problem in modulation space.

Our solution is a partially unsupervised feature-based segmentation algorithm — the number of image regions must be known, although no *a priori* information about the region characteristics is required. While feature-based segmentation techniques have been investigated extensively in the literature, our approach is significant for two reasons. First, we believe that this is the first case where an image processing problem of substantial practical interest has been formulated and systematically solved directly in the modulation domain. Second, the quality of the results we obtain is competitive with the best results that have been reported in the literature to date. Our AM-FM image model is introduced in Section 2 and feature extraction techniques for estimating the dominant image modulations on a spatially local basis are described in Section 3. In Section 4 we detail our segmentation approach, which utilizes a statistical clustering algorithm in the modulation domain followed by image domain post processing. Examples and conclusions are presented in Section 5.

2. Image Model

Modulation domain representation of a real-valued image $s(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, derives from considering that $s(\mathbf{x})$ is a sum of locally coherent (*i.e., locally smooth*) nonstationary AM-FM functions $s_k(\mathbf{x})$:

$$s(\mathbf{x}) = \sum_{k=1}^{K} s_k(\mathbf{x}). \tag{1}$$

The reason for considering a multicomponent model such as the sum (1) is that it is generally impossible to compute accurate modulating function estimates unless the individual AM-FM functions $s_k(\mathbf{x})$ are locally coherent almost everywhere [12]. Most images of practical interest contain complicated, nonstationary texture structures and may also be inherently multipartite. Thus single component AM-FM representations with locally smooth modulations simply do not exist, and it becomes a computational necessity to seek multicomponent models of the form (1) wherein each individual AM-FM component admits modulations that *are* locally smooth.

The AM-FM components $s_k(\mathbf{x})$ in (1) are real-valued; their instantaneous amplitude envelopes (AM functions) and instantaneous frequency vectors (FM functions) are therefore not unique. By way of analogy to the well known 1-D analytic signal, we use the directional 2-D Hilbert transform to disambiguate the demodulation problem by associating with $s(\mathbf{x})$ a complex-valued image $t(\mathbf{x}) = s(\mathbf{x}) + jq(\mathbf{x})$ where, with $\mathbf{x} = [x\ y]^T$,

$$q(\mathbf{x}) = \text{p.v.}\frac{1}{\pi}\int_{\mathbb{R}} s(x-\xi, y)\ \frac{d\xi}{\xi}. \tag{2}$$

Since the Hilbert transform (2) is linear, for any given decomposition of $s(\mathbf{x})$ into components we then have the complex model

$$t(\mathbf{x}) = \sum_{k=1}^{K} t_k(\mathbf{x}) = \sum_{k=1}^{K} a_k(\mathbf{x}) \exp[j\varphi_k(\mathbf{x})] \tag{3}$$

wherein the AM functions $a_k(\mathbf{x})$ and FM functions $\nabla\varphi_k(\mathbf{x})$ are unique [14, 22]. A method for discretizing (2) and (3) was given in [10].

3. Feature Extraction

Feature extraction involves computing dense, pointwise estimates of the unknown modulating functions $a_k(\mathbf{x})$ and $\nabla\varphi_k(\mathbf{x})$ in (3). Our approach for doing this is to first isolate the components from one another locally in space and frequency by applying a multiband linear filterbank and then subsequently to apply a spatially local nonlinear demodulation algorithm to the filterbank channel responses for estimating each $a_k(\mathbf{x})$ and $\nabla\varphi_k(\mathbf{x})$. While it is not necessary for the filterbank to isolate components on a global scale, *i.e.*, a single channel response may be dominated by different components in different spatial regions, what is required is that each channel response be dominated by at most one component at each point in the spatial domain. This requirement is satisfied automatically in practice, because the structure of the filterbank will determine the multicomponent decomposition of the image as indicated in (1) and (3).

Let $g_i(\mathbf{x})$ and $G_i(\mathbf{\Omega})$ be the impulse response and frequency response of a particular one of the filterbank channels, and suppose that, at a particular point $\mathbf{x}$, the channel response $y_i(\mathbf{x})$ is dominated by image component $t_k(\mathbf{x})$, so that

$$y_i(\mathbf{x}) = t(\mathbf{x}) * y_i(\mathbf{x}) \approx t_k(\mathbf{x}) * y_i(\mathbf{x}). \tag{4}$$

We estimate the AM and FM functions of component $t_k(\mathbf{x})$ at the point $\mathbf{x}$ using the local nonlinear algorithms [13]

$$\nabla\varphi_k(\mathbf{x}) \approx \text{Re}\left[\frac{\nabla y_i(\mathbf{x})}{j y_i(\mathbf{x})}\right], \tag{5}$$

$$a_k(\mathbf{x}) \approx \left|\frac{y_i(\mathbf{x})}{G[\nabla\varphi_k(\mathbf{x})]}\right|. \tag{6}$$

Rigorous derivations of (5), (6) and their discrete equivalents are given in [12]. It should be noted that the multidimensional Teager-Kaiser operator [20] could also be considered as an alternative to (5) and (6).

The features of interest for modulation domain image segmentation are those modulations $a_D(\mathbf{x})$, $\nabla\varphi_D(\mathbf{x})$ that dominate the image spectrum on a spatially local basis. Since different components in (3) are expected to be dominant in different image regions, the dominant modulations will come from different components at different points. Thus $a_D(\mathbf{x})$ and $\nabla\varphi_D(\mathbf{x})$ are expected to be neither homogeneous nor locally coherent, and this fact is the impetus behind applying statistical clustering in the modulation domain to perform segmentation.

For each point $\mathbf{x}$, we define the dominant component as the one that maximizes the channel selection criterion

$$\Psi_i(\mathbf{x}) = \frac{|y_i(\mathbf{x})|}{\max_{\mathbf{\Omega}} |G_i(\mathbf{\Omega})|}. \tag{7}$$

As explained in [12], this criterion tends to select high-amplitude components with frequency vectors lying near the maximum transmission frequency of the channel, thereby reducing demodulation errors due to noise and cross-component interference. Computed estimates of $a_D(\mathbf{x})$ and $\nabla\varphi_D(\mathbf{x})$ are extracted by applying (5) and (6) to the response of the channel that maximizes (7) on a pointwise basis.

Although our segmentation algorithm is independent of the particular filterbank that is used, the filterbank design is nevertheless a critically important issue: it should be designed to yield components $t_k(\mathbf{x})$ in (3) that admit locally coherent modulating functions so that approximation errors in the estimators (5) and (6) will be as small as possible. Furthermore, the filterbank must cover the right frequency half-plane where the complex image (3) has support with a tessellation of filters that are spatially localized to capture nonstationary spatial structures in the components $t_k(\mathbf{x})$ and simultaneously spectrally localized to resolve the multiple components from one another.

The efficacy of using Gabor channel filters, which in the continuous-domain case uniquely realize the uncertainty principle lower bound on joint spatio-spectral resolution, has been rigorously studied and well established [2, 3, 15, 23]. We typically employ a bank of one-octave isotropic unity L^2-norm Gabor filters similar to the filterbanks described in [3, 11]. Such filterbanks generally provide good performance across a wide variety of images and image types. In situations where *a priori* samples of the expected region types are available, it should be noted that improved performance can be obtained by using techniques similar to those described in [7, 26] to design an optimal or near optimal set of filters.

4. Segmentation

We have found that improved segmentation performance is obtained by transforming the Cartesian dominant frequency vector $\nabla\varphi_D(\mathbf{x})$ to a polar representation. Thus, switching now to an explicitly discrete notation, we consider a modulation domain feature vector with entries $A(m,n) = a_D(m,n)$, $R(m,n) = |\nabla\varphi_D(m,n)|$ and $\theta(m,n) = \arg\nabla\varphi_D(m,n)$ at each image pixel. While spatial coordinates could also be included in the feature vector to encourage formation of connected regions in the segmentation, there is a significant computational advantage to keeping the dimension of the feature space as small as possible [6, 23]. We therefore obtain a preliminary segmentation by performing statistical clustering in the 3-D modulation space and subsequently impose a spatial correspondence constraint by refining the segmentation with post processing as described in Section 4.2.

4.1. Modulation domain clustering

To obtain the preliminary segmentation, we apply the well-known k-means clustering algorithm in the A-R-θ feature space [16]. While we are currently developing a technique based on the modified Hubert index [5] for automatically determining the number of clusters, our algorithm presently requires *a priori* knowledge of the number of regions in the image to specify the value of k in the k-means algorithm.

Prior to clustering, we scale each feature by the reciprocal of the sample standard deviation computed for the feature to obtain the scaled features $\widetilde{A}$, $\widetilde{R}$, and $\widetilde{\theta}$. The similarity measure between pixels (i,j) and (m,n) is then given by

$$\begin{aligned}\mathcal{S}(i,j,m,n) = \Big\{ &\alpha\left[\widetilde{A}(i,j) - \widetilde{A}(m,n)\right]^2 \\ &+ \beta\left[\widetilde{R}(i,j) - \widetilde{R}(m,n)\right]^2 + \gamma\left[\widetilde{\theta}(i,j) - \widetilde{\theta}(m,n)\right]^2 \Big\}^{\frac{1}{2}},\end{aligned} \tag{8}$$

where the weights α, β, and γ, which seek to emphasize the feature that provides the best class separability, are chosen based on modulation domain entropy. Let $p_{\tilde{A}}(q)$ be the normalized histogram of $\widetilde{A}(m,n)$ and define $p_{\tilde{R}}(q)$ and $p_{\tilde{\theta}}(q)$ similarly. As usual, the entropy of each feature is defined by, *e.g.*, $E_{\tilde{A}} = -\sum_q p_{\tilde{A}}(q)\log_2 p_{\tilde{A}}(q)$. The weight α is then calculated according to $\alpha = (\mathcal{E}_T - \mathcal{E}_{\tilde{A}})^2/\mathcal{E}_{\tilde{A}}$, where

$$\mathcal{E}_{\tilde{A}} = \frac{E_{\tilde{A}}}{\max_q p_{\tilde{A}}(q) - \min_q p_{\tilde{A}}(q)} \tag{9}$$

and $\mathcal{E}_T = \mathcal{E}_{\tilde{A}} + \mathcal{E}_{\tilde{R}} + \mathcal{E}_{\tilde{\theta}}$. The calculations for β and γ are completely analogous.

The k-means algorithm is run with respect to the similarity measure (8) for ten iterations starting from random initial cluster seeds, and the best final configuration is selected using the usual squared-error criterion for cluster validation [6, 16].

4.2. Post processing

The preliminary segmentation delivered by the k-means algorithm is generally unsatisfying and is rarely in good agreement with visual perception. Many small regions of misclassified pixels are typically present and we have also observed long, narrow "streaks" of misclassified pixels. In addition, numerous irregularities frequently appear along the boundaries of regions that were smooth in the original image.

We ameliorate these effects by applying two image domain post processing stages to arrive at the final segmentation. First, similar to the post processing operator that was used in [26] to regularize region boundaries, we apply an isotropic morphological majority filter to smooth the boundaries of the segmented regions. For images of size 256×256, we use a 9×9 structuring element. Second, connected components labeling and minor region removal are applied to the segmentation in the image domain. Only the k largest connected components are retained, where k is equal to the number of clusters delivered by the k-means algorithm. As mentioned above, this final step enforces an image domain spatial correspondence constraint on the final segmentation.

5. Examples and Conclusion

Several two-texture segmentation examples are presented in Fig. 1. The center portion of the image in Fig. 1(a) has been rotated counterclockwise by 45°. The computed dominant AM function is shown as a gray scale image in Fig. 1(b) and a needle diagram depicting the dominant FM function appears in Fig. 1(c), where needle length is inversely proportional to the dominant frequency magnitude $R(m,n)$. The segmentation obtained using our proposed algorithm appears in Fig 1(d), where 98.25% of the pixels are classified correctly. The image in Fig. 1(e) is a juxtaposition of two Brodatz textures. The computed dominant modulations are shown in Fig. 1(f) and (g), and the obtained segmentation is given in Fig. 1(h). In this case, 99.19% of the pixels are correctly classified. Finally, Fig. 1(i) and (k) show two natural scenes from the MIT Media Laboratory VisTex database. The segmentation results appear in Fig. 1(j) and (l), where pixel intensities in one of the two segmented regions are equal to those of the original images and pixel intensities in the second region have been divided by two. While precise ground truth is not available for these images,

the segmentation results appear to be in excellent qualitative agreement with human visual perception.

As demonstrated by these examples, the textured image segmentation approach proposed in this paper is competitive with the best existing techniques, many of which are described in the references below. We have run the algorithm against numerous synthetic two- and three-texture images and observed correct pixel classification rates ranging from 94% to 99%. It also seems to work well on natural scenes such as the ones in Fig. 1(i) and (k). The approach is significant because it represents the first time that an image processing problem of substantial practical interest has been formulated and solved in the modulation domain. Our ongoing research is focused on developing a technique for automatically determining the number of regions in an image so that the algorithm will be fully unsupervised. Several of the new cluster validation techniques that have emerged recently [8, 18] may prove useful in this regard.

References

[1] C. Bouman and B. Liu. Multiple resolution segmentation of textured images. *IEEE. Trans. Pattern Anal. Machine Intell.*, 13(2):99–113, February 1991.

[2] A. C. Bovik. Analysis of multichannel narrow-band filters for image texture segmentation. *IEEE. Trans. Signal Proc.*, 39(9):2025–2043, September 1991.

[3] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo. Localized measurement of emergent image frequencies by Gabor wavelets. *IEEE. Trans. Info. Theory*, 38(2):691–712, March 1992.

[4] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE. Trans. Pattern Anal. Machine Intell.*, 9(1):39–55, January 1987.

[5] R. C. Dubes. How many clusters are best? an experiment. *Pattern Recognit.*, 20:645–663, 1987.

[6] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, New York, 1973.

[7] D. Dunn and W. E. Higgins. Optimal Gabor filters for texture segmentation. *IEEE Trans. Image Proc.*, 4(7):947–964, July 1995.

[8] H. Frigui and R. Krishnapuram. A robust competitive clustering algorithm with applications in computer vision. *IEEE. Trans. Pattern Anal. Machine Intell.*, 21(5):450–464, May 1999.

[9] R. M. Haralick. Image segmentation survey. In O. D. Faugeras, editor, *Fundamentals in Computer Vision*, pages 209–224. Cambridge Univ. Press, Cambridge, 1983.

[10] J. P. Havlicek and A. C. Bovik. Image modulation models. In A. Bovik, editor, *The Image and Video Processing Handbook*. Academic Press, San Diego, 1999. *to appear.*

[11] J. P. Havlicek, A. C. Bovik, and D. Chen. AM-FM image modeling and Gabor analysis. In C. W. Chen and Y. Zhang, editors, *Visual Information Representation, Communication, and Image Processing*, pages 343–385. Marcel Dekker, New York, 1999.

[12] J. P. Havlicek, D. S. Harding, and A. C. Bovik. Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models. *IEEE Trans. Image Proc. to appear Feb. 2000.*

[13] J. P. Havlicek, D. S. Harding, and A. C. Bovik. The multicomponent AM-FM image representation. *IEEE Trans. Image Proc.*, 5(6):1094–1100, June 1996.

[14] J. P. Havlicek, J. W. Havlicek, and A. C. Bovik. The analytic image. In *Proc. IEEE Int'l. Conf. Image Proc.*, Santa Barbara, CA, October 26-29 1997.

[15] T. Hoffmann, J. Puzicha, and J. M. Buhmann. Unsupervised texture segmentation in a deterministic annealing framework. *IEEE. Trans. Pattern Anal. Machine Intell.*, 20(8):803–818, August 1998.

[16] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[17] C. Kervrann and F. Heitz. A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics. *IEEE Trans. Image Proc.*, 4(6):856–862, June 1995.

[18] D. A. Langan, J. W. Modestino, and J. Zhang. Cluster validation for unsupervised stochastic model-based image segmentation. *IEEE Trans. Image Proc.*, 7(2):180–195, February 1998.

[19] B. S. Manjunath and R. Chellappa. Unsupervised texture segmentation using Markov radom field models. *IEEE. Trans. Pattern Anal. Machine Intell.*, 13(5):478–482, May 1991.

[20] P. Maragos and A. C. Bovik. Image demodulation using multidimensional energy separation. *J. Opt. Soc. Amer. A*, 12(9):1867–1876, September 1995.

[21] H. H. Nguyen and P. Cohen. Gibbs random fields, fuzzy clustering, and the unsupervised segmentation of textured images. *CVGIP: Graph. Models and Image Proc.*, 55(1):1–19, January 1993.

[22] F. Peyrin, Y. M. Zhu, and R. Goutte. Extension of the notion of analytic signal for multidimensional signals. Application to images. In I. T. Young, *et al.*, editor, *Signal Processing III: Theories and Applications*, pages 677–680. Elsevier Science Publishers, Amsterdam, B. V. (North-Holland), 1986.

[23] O. Pichler, A. Teuner, and B. J. Hosticka. An unsupervised texture segmentation algorithm with feature space reduction and knowledge feedback. *IEEE Trans. Image Proc.*, 7(1):53–61, January 1998.

[24] T. R. Reed and J. M. H. DuBuf. A review of recent texture segmenation and feature extraction techniques. *CVGIP: Image Understanding*, 57(3):359–372, May 1993.

[25] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Proc.*, 4(11):1549–1560, November 1995.

[26] T. P. Weldon and W. E. Higgins. An algorithm for designing multiple Gabor filters for segmenting multi-textured images. In *Proc. IEEE Int'l. Conf. Image Proc.*, Chicago, IL, October 4-7 1998.

[27] C. S. Won and H. Derin. Unsupervised segmentation of noisy and textured images using Markov random fields. *CVGIP: Graph. Models and Image Proc.*, 54(4):308–328, July 1992.

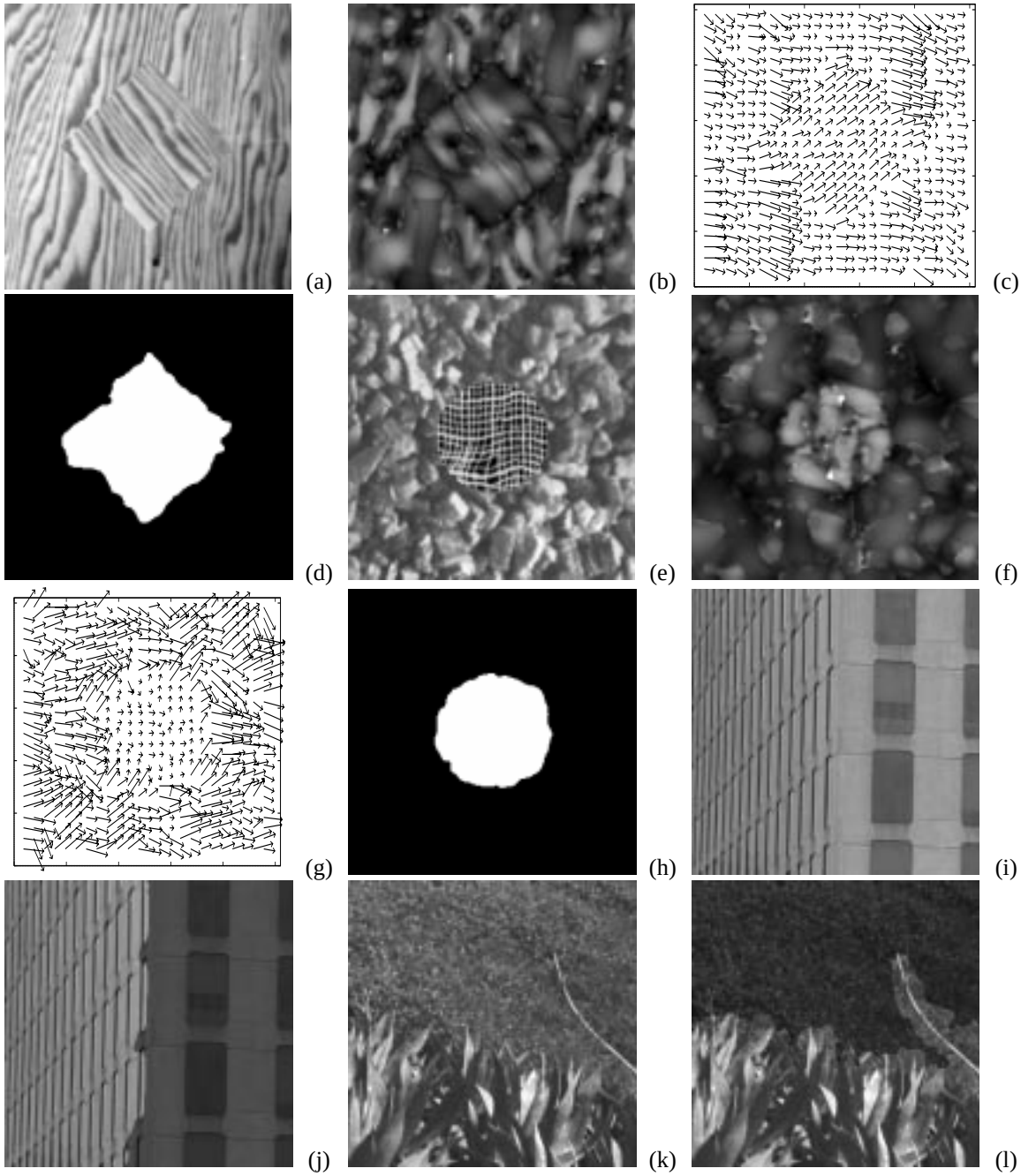

Figure 1. (a) Original *WoodWood* image. (b) Computed dominant AM function. (c) Computed dominant FM function. (d) Segmentation result delivered by the proposed algorithm. The correct classification rate is 98.25%. (e) Original *MicaBurlap* image. (f) Computed dominant AM function. (g) Computed dominant FM function. (h) Segmentation result. The correct classification rate is 99.19%. (i) Original *Building.0010* image. (j) Overlay of segmentation result on original image. Pixel values in the left-hand region are unaltered from the original; pixel values in the right-hand region are divided by two. (k) Original *GrassPlantsSky.0005* image. (l) Overlay of segmentation result on original image. Pixel values in the lower region are equal to those in the original; pixel values in the upper region are divided by two.