AM-FM IMAGE SEGMENTATION

Tanachit Tangsukson and Joseph P. Havlicek

School of Electrical and Computer Engineering
University of Oklahoma, Norman, OK 73019-1023 USA

ABSTRACT

We introduce a new modulation domain texture segmentation algorithm. The approach begins by constructing a dominant component AM-FM image model, where the dominant amplitude and frequency modulations are used as segmentation features. Statistical clustering is applied in this feature space to compute an initial segmentation which is then refined by morphological filtering and connected components labeling. The algorithm, which consistently delivers correct pixel classification rates exceeding 94%, is only partially unsupervised at present since the desired number of regions must be known *a priori*. Our future work is focused on developing strategies to make the approach fully unsupervised.

1. INTRODUCTION

In this paper, we present a novel modulation domain feature-based approach for segmenting textured images. Segmentation is a classical image processing problem that involves partitioning an image into two or more disjoint regions that are each homogeneous with respect to some specified properties. Often, the processing goal is to obtain a segmentation that is consistent with human visual perception. The segmentation problem is critical in a multitude of diverse applications including target identification and tracking, remote sensing, autonomous vehicle navigation, automated manufacturing and inspection, robot control, and automated or computer aided diagnostic medicine. Indeed, segmentation plays a fundamental early processing role supporting object recognition and identification in virtually all computer vision systems. Moreover, the general segmentation problem is notoriously difficult, as evidenced by the vast body of technical literature that has been devoted to it [1–3].

The new approach we present here begins by formulating a multicomponent AM-FM model for the image and then extracting estimates of the dominant amplitude and frequency modulations at each pixel for use as features. A preliminary segmentation is obtained by applying a statistical clustering algorithm in the modulation domain feature space, where we utilize two new objective functions based on feature space entropy and average local feature deviations. Cluster validation is performed using the squared-error criterion and invariants of the within- and between-cluster scatter matrices. The preliminary segmentation is then mapped back into the pixel domain where it is refined by morphological filtering and connected components labeling with minor region removal to obtain the final segmentation.

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 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. At present, our algorithm is only partially unsupervised, because the number of textured regions in an image must be known *a priori*. Our ongoing research is focused on using the modified Hubert index to automatically determine the number of regions so that the algorithm will be fully unsupervised [4].

2. IMAGE MODEL AND FEATURE SPACE

Our approach is based on the complex-valued multicomponent AM-FM image model

$$t(m,n) = \sum_{k=1}^{K} a_k(m,n) \exp[j\varphi_k(m,n)], \tag{1}$$

where $a_k(m,n)$ is the amplitude modulation of component k and $\nabla\varphi_k(m,n) = [U(m,n)\ \ V(m,n)]^T$, which contains the samples of the corresponding continuous-domain phase gradient field, is the frequency modulation function of component k. For a real-valued image $s(m,n)$, the complex image $t(m,n) = s(m,n) + jq(m,n)$ is constructed by adding an imaginary part equal to the (linear) directional multidimensional Hilbert transform of $s(m,n)$ [5, 6]. In this case $t(m,n)$ admits several important properties analogous to those of the corresponding 1D analytic signal; we call $t(m,n)$ the *analytic image* associated with $s(m,n)$.

We analyze $t(m,n)$ with a 42-channel multiband bank of octave-band unity-aspect-ratio Gabor filters similar to the one described in [7]. The purpose of this processing is to isolate the multiple image components in the model (1) from one another on a local basis in both space and spatial frequency prior to performing demodulation. We assume that each image component may generally lie within different filterbank channels at different pixels, but that each channel is dominated by *only one* component in any 3 × 3-pixel neighborhood. We then estimate the modulating functions of *all* components by applying the approximate discrete demodulation algorithm [8]

$$\widehat{a}(m,n) = |t(m,n)|, \tag{2}$$

$$|\widehat{U}(m,n)| = \arccos\left[\frac{t(m+1,n)+t(m-1,n)}{2t(m,n)}\right], \tag{3}$$

$$\operatorname{sgn}\widehat{U}(m,n) = \operatorname{sgn}\arcsin\left[\frac{t(m+1,n)-t(m-1,n)}{2jt(m,n)}\right], \tag{4}$$

0-7803-6297-7/00/$10.00 © 2000 IEEE

$$|\widehat{V}(m,n)| = \arccos\left[\frac{t(m,n+1)+t(m,n-1)}{2t(m,n)}\right], \quad (5)$$

$$\operatorname{sgn}\widehat{V}(m,n) = \operatorname{sgn}\arcsin\left[\frac{t(m,n+1)-t(m,n-1)}{2jt(m,n)}\right] \quad (6)$$

to every filterbank channel response at all pixels.

For segmentation features, we use the estimated modulations (2)-(6) corresponding to the image component that locally dominates the image spectrum at each pixel. Let $y_i(m,n)$ be the response of channel i and $G_i(U,V)$ be the channel frequency response. The dominant component at pixel (m,n) is the one that maximizes the channel selection criterion [8]

$$\Gamma_i(m,n) = \frac{|y_i(m,n)|}{\max_{U,V}|G_i(U,V)|}. \quad (7)$$

Modulation domain texture features $A(m,n) = \widehat{a}(m,n)$, $R(m,n) = [\widehat{U}^2(m,n) + \widehat{V}^2(m,n)]^{\frac{1}{2}}$ and $\theta(m,n) = \arctan[\widehat{V}(m,n)/\widehat{U}(m,n)]$ are selected from the channel that maximizes (7) on a pixel-by-pixel basis.

3. STATISTICAL CLUSTERING

To obtain a preliminary image segmentation, we apply the well-known k-means clustering algorithm [9] in the A-R-θ feature space. While we are currently developing a technique based on the modified Hubert index [4] for automatically determining the number of clusters, our algorithm presently requires *a priori* knowledge of k,

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 $\widetilde{\theta}$. This scaling ensures that a feature with relatively large numerical values will not be permitted to dominate the clustering algorithm. The similarity measure between pixels (i,j) and (m,n) is then given by

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

where the weights α, β, and γ, which seek to emphasize the feature that provides the best class separability, are chosen by one of two methods as described in the following two sections.

3.1. Entropy-based similarity measure

Intuitively, we expect that a feature with a flat histogram supplies relatively little class separability information, whereas one with a histogram that is tightly concentrated about several distinct modes is powerful in its ability to discriminate between textures. This admittedly imprecise notion suggests that the weights α, β, and γ in (8) should be chosen based on modulation domain entropy: a feature with a histogram that is relatively more localized will have lower entropy and should therefore be assigned a larger weight since it is expected to provide a relatively greater amount of class separability information. When this line of reasoning is applied to calculate the weights, we refer to (8) as the *entropy-based similarity measure*.

Let $p_{\widetilde{A}}(q)$ be the normalized histogram of $\widetilde{A}(m,n)$ and define $p_{\widetilde{R}}(q)$ and $p_{\widetilde{\theta}}(q)$ similarly. As usual, the entropy of each feature is defined by, *e.g.*,

$$E_{\widetilde{A}} = -\sum_q p_{\widetilde{A}}(q)\log_2 p_{\widetilde{A}}(q). \quad (9)$$

The weight α is then calculated according to

$$\alpha = \frac{(\mathcal{E}_T - \mathcal{E}_{\widetilde{A}})^2}{\mathcal{E}_{\widetilde{A}}}, \quad (10)$$

where

$$\mathcal{E}_{\widetilde{A}} = \frac{E_{\widetilde{A}}}{\max_q p_{\widetilde{A}}(q) - \min_q p_{\widetilde{A}}(q)} \quad (11)$$

and

$$\mathcal{E}_T = \mathcal{E}_{\widetilde{A}} + \mathcal{E}_{\widetilde{R}} + \mathcal{E}_{\widetilde{\theta}}. \quad (12)$$

In (11), note that the modulation domain entropy $E_{\widetilde{A}}$ is divided by the range of the normalized histogram, a quantity that tends to grow inversely with entropy. While it is true that lower entropy generally implies increased class discrimination power in a feature, it is also true that a feature exhibiting a perfectly unimodal histogram will have the lowest possible entropy and yet provide no class separability information whatsoever. The numerator and denominator of (11) carefully balance these considerations. The value assigned to the weight α in (10) is inversely proportional to the fraction of the total feature space entropy $\mathcal{E}_T$ that is contributed by the feature $\widetilde{A}$. Calculations completely analogous to those just described are used to calculate the weights β and γ for the normalized features $\widetilde{R}$ and $\widetilde{\theta}$.

3.2. Local feature deviations-based similarity measure

A deficiency of the entropy-based similarity measure is that the normalized histogram entropy indicates only localization of the feature histogram and does not consider spatial relationships between pixels in the image. In this section we describe a second method for calculating the weights in (8) that does not suffer from this deficiency. When this method is used we refer to (8) as the *local feature deviations-based similarity measure.* It is based on the idea that we seek to segment the image into texturally homogeneous disjoint regions that are substantially different from one another. Thus, for a feature that provides good class separability information, the local deviations of the feature should be small within regions. Moreover, since the number of pixels that lie on boundaries between regions is small compared to the number that lie in the interior of the regions, the average of the local deviations computed across the image should also be small.

With $K = N-1$, we define the average local deviation of the $N \times N$ feature image $\widetilde{A}(m,n)$ by

$$D(\widetilde{A}) = \frac{1}{K^2}\sum_{i,j=1}^{N-2}\left\{\sum_{k=-1}^{1}\left[\widetilde{A}(i,j) - \widetilde{A}(i-k,j)\right]^2 + \left[\widetilde{A}(i,j) - \widetilde{A}(i,j-k)\right]^2\right\}^{\frac{1}{2}}. \quad (13)$$

The weight α in (8) is then given by

$$\alpha = \frac{\left(\partial_{\text{Total}} - D(\widetilde{A})\right)^2}{D(\widetilde{A})}, \quad (14)$$

where $\partial_{\text{Total}} = D(\widetilde{A}) + D(\widetilde{R}) + D(\widetilde{\theta})$ is the total feature space deviation. The weights β and γ in (8) are calculated in an entirely analogous fashion.

3.3. Clustering algorithm

The k-means algorithm is run with respect to the similarity measure (8) for ten iterations starting from random initial cluster seeds. In cases where the entropy-based method is used for determining the weights α, β, and γ, cluster validation is performed with respect to the usual squared-error criterion [9].When the method based on feature deviations is used, the invariant $\text{tr}S_W^{-1}S_B$ is used for cluster validation, where S_W and S_B are the within- and between-cluster scatter matrices, respectively. In either case, the final clustering that optimizes the validation criterion is retained for the preliminary segmentation.

4. 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. To ameliorate these effects and arrive at the final segmentation, we apply two pixel domain post processing operations to the image of region labels obtained from clustering. First, an isotropic morphological majority filter is applied to smooth the region boundaries. For images of size 256×256, we use a 9×9 filter kernel. Second, connected components labeling and minor region removal are applied; only the k largest connected components are retained, where k is equal to the number of clusters delivered by the k-means algorithm. This final step enforces a spatial correspondence constraint on the final segmentation in the image domain.

5. EXAMPLES

Eight segmentation examples using the technique described in this paper are presented in Fig. 1. The original images shown in Fig. 1(a), (e), (g), (j), (l), and (n) are juxtapositions of Brodatz-like textures, whereas the natural scenes in Fig. 1(q) and (s) are from the MIT *VisTex* database. The images in Fig. 1(a), (e), and (g) each contain two textured regions. Those in Fig. 1(j) and (l) contain three textured regions, while the image in Fig. 1(n) contains four textures. Our perception suggests that the images in Fig. 1(q) and (s) each contain two primary textured regions, and this value was supplied to the k-means algorithm. The feature deviation-based similarity measure was used for all images except the two natural scenes, which utilized the entropy-based measure.

Preliminary segmentations obtained using the k-means clustering algorithm alone are given in Fig. 1(h) and (o). Fig. 1(d), (f), (i), (k), (m), and (p) show final segmentations obtained after post processing. In all cases, the percentage of pixels correctly classified exceeds 94%. In Fig. 1(r) and (t), the final segmentations are shown overlayed on the original images. Ground truth was not available for these images.

In selecting these examples, we have endeavored to choose some for which the algorithm performed very well and some that are more interesting because they are either extremely difficult or contain region boundaries that seem visually ambiguous to us. For example, the region boundary near the center of the right edge of the image in Fig. 1(l) is quite difficult to detect, as is the lower right portion of the boundary between the two regions in Fig. 1(e).

6. CONCLUSION

AM-FM image modeling is an important emerging area and the texture segmentation technique presented in this paper is significant because it represents the first time that an image processing problem of substantial practical interest has been formulated and solved directly in the modulation domain. In view of the fact that many modern descriptions of texture are formulated in terms of nonstationary amplitude modulations (local contrast) and frequency modulations (local orientation and granularity), it is not entirely unexpected that the results shown in Fig. 1 are of such high quality. For a suite of 30 test images not unlike those depicted in Fig. 1 we obtained correct pixel classification rates ranging from 99.53% to 94.18% using both the entropy-based and feature deviations-based similarity measures, which is competitive with the best reported techniques. While we have observed that the feature deviations-based measure often works best for synthetic images and the entropy-based measure often works best for natural images, we have not developed a characterization that would enable us to predict their relative performance *a priori*. Our future work in this area is focused on the development of improved similarity measures and cluster validation criteria that will enable the algorithm to run in a fully unsupervised mode.

7. REFERENCES

[1] R. M. Haralick, "Image segmentation survey," in *Fundamentals in Computer Vision*, O. D. Faugeras, Ed., pp. 209–224. Cambridge Univ. Press, Cambridge, 1983.

[2] T. R. Reed and J. M. H. DuBuf, "A review of recent texture segmenation and feature extraction techniques," *CVGIP: Image Understanding*, vol. 57, no. 3, pp. 359–372, May 1993.

[3] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision, et al.* C. H. Chen, Ed., pp. 235–276. World Scientific, Singapore, 1993.

[4] R. C. Dubes, "How many clusters are best? an experiment," *Pattern Recognit.*, vol. 20, pp. 645–663, 1987.

[5] 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.

[6] J. P. Havlicek and A. C. Bovik, "Image modulation models," in *The Image and Video Processing Handbook*, A.C. Bovik, Ed. Academic Press, San Diego, 2000, *to appear*.

[7] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo, "Localized measurement of emergent image frequencies by Gabor wavelets," *IEEE. Trans. Info. Theory*, vol. 38, no. 2, pp. 691–712, March 1992.

[8] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models," *IEEE Trans. Image Proc.*, vol. 9, no. 2, pp. 227–242, February 2000.

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

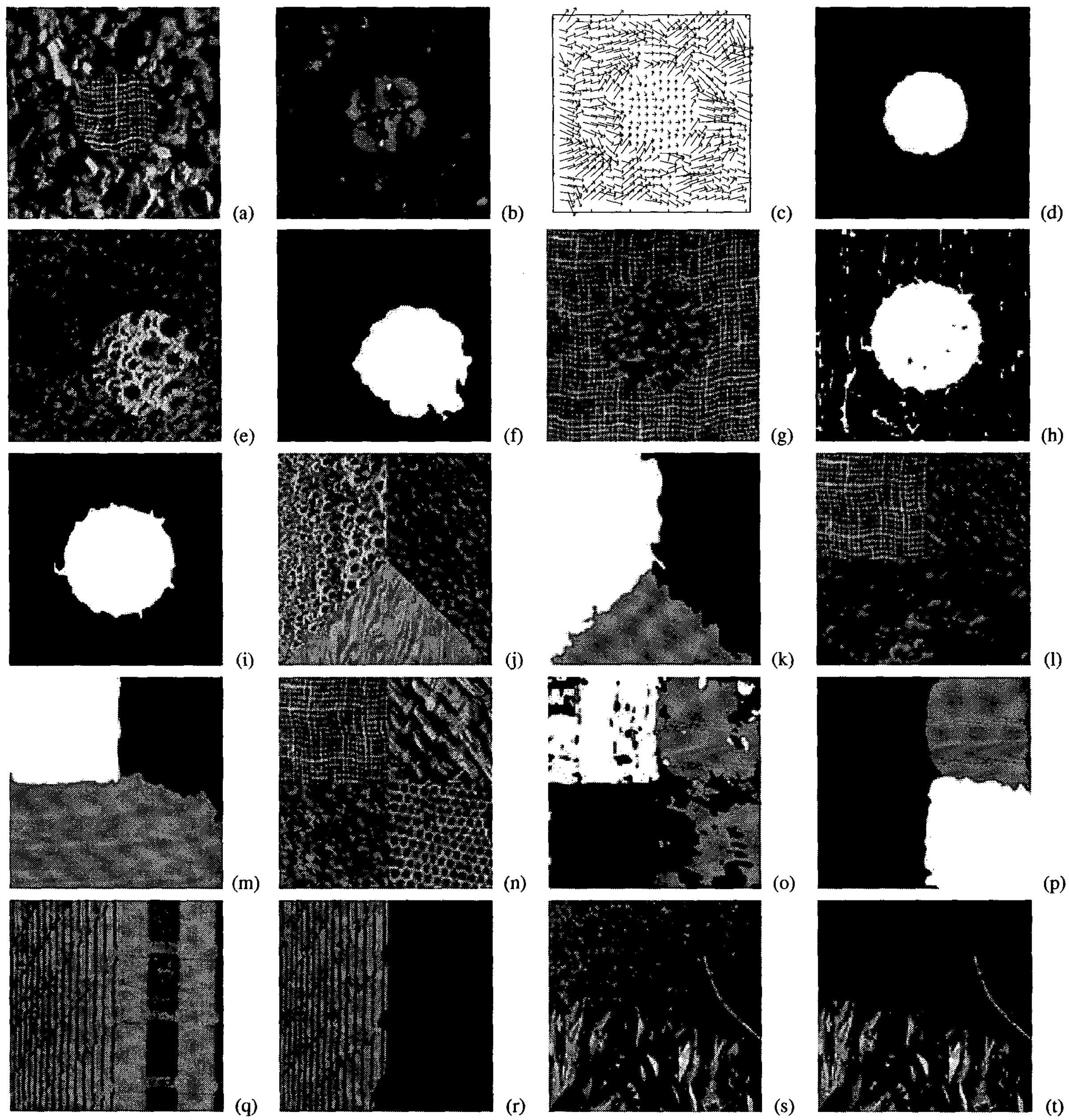

Fig. 1. Examples. (a) *Mica-Burlap* image. (b) Computed dominant AM function. (c) Computed dominant FM function; $\theta(m,n)$ is given by needle orientation and needle length is inversely proportional to $R(m,n)$. (d) Final segmentation delivered by the proposed algorithm. The correct classification rate is 99.53%. (e) *Grass-Flowers* image. (f) Final segmentation; correct classification rate is 97.43%. (g) *Burlap-Pigskin* image. (h) Preliminary segmentation delivered by the k-means algorithm. (i) Final segmentation; correct classification rate is 99.02%. (j) *Flowers-Wood-Cork* image. (k) Final segmentation; correct classification rate is 94.53%. (l) *Burlap-Cork-Fieldstone* image. (m) Final segmentation; correct classification rate is 96.84%. (n) Four-texture example. (o) Preliminary segmentation delivered by the k-means algorithm. (p) Final segmentation; correct classification rate is 97.33%. (q) *Building.0010* image. (r) Overlay of final segmentation 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. (s) *GrassPlantsSky.0005* image. (t) Overlay of final segmentation 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.