Bayesian Segmentation of AM-FM Texture Images

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

Abstract

We present a fully unsupervised parametric modulation domain technique for segmenting textured images. Textured regions in the image are modeled as multicomponent sums of nonstationary AM-FM functions. The dominant modulations at each pixel are estimated using a technique called DCA and used to construct modulation domain feature vectors. The overall feature space is regarded as a mixture of Gaussians, where the modulations within each texture class are modeled by a single multivariate normal distribution. Although this model is somewhat unrealistic, it leads to a robust segmentation algorithm that is able to operate in a fully unsupervised mode. An EM algorithm is used to estimate the parameters of the Gaussian mixture so that approximate maximum-likelihood estimates of the pixel class labels can be obtained. The proposed technique is demonstrated on a variety of images constructed from juxtapositions of Brodatz-like textures.

1. Introduction

Despite intense research efforts, texture remains one of the most poorly understood attributes of visual information. AM-FM models [4, 5, 10, 11, 15, 16] provide a powerful means of characterizing texture in terms of nonstationary joint amplitude and frequency modulations and are also parsimonious with certain plausible models of early processing in biological vision systems [7, 9, 13, 22]. They have been used recently to develop modulation domain solutions to a number of classical problems including shape from texture [19], texture-based stereopsis [6], texture completion [1], and texture segmentation [20, 21, 23].

While the majority of AM-FM modeling techniques previously treated in the literature have been nonparametric and distribution-free, parametric 1D models have been considered in a few cases [8, 14]. In this paper, we investigate multivariate Gaussian mixture models for multidimensional modulations for the first time and use them to perform fully unsupervised maximum-likelihood AM-FM texture segmentation. We obtain high quality segmentations with correct pixel classification rates consistently exceeding 95% for simple juxtapositions of Brodatz-like textures. This fortuitous result was somewhat unexpected since, in the exploratory research presented here, we considered only single multivariate Gaussians for describing the joint modulations within each texture class (resulting in a mixture model for the modulations observed across the entire image). The performance of this approach could almost certainly be improved by employing a Gaussian mixture model for the modulations within each class. As we will discuss in Section 3, however, it is unclear how such a strategy could be made fully unsupervised.

2. Computing the Texture Features

Consider a digital image $t(\mathbf{x}) : \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{C}$. In order for the image modulations to be well defined, we require $t(\mathbf{x})$ to be complex-valued; in applications where real images are of interest, we add an imaginary part equal to the 2D directional Hilbert transform of the image to obtain a complex extension analogous to the 1D analytic signal [12]. We model textured regions in the image using nonlinear AM-FM functions of the form $a(\mathbf{x}) \exp[j\varphi(\mathbf{x})]$. For each such function, the AM term $a(\mathbf{x})$ describes the local texture contrast while the FM term, or *instantaneous frequency vector* $\nabla\varphi(\mathbf{x})$ describes the local texture orientation, coarseness, and granularity. Many nonstationary textures such as a wood grain or a zebra's stripes are well described by a single AM-FM function. However, there are also many textures of practical interest such as cobblestones, burlap, a cloth weave, or a cross-hatch pattern that are inherently multipartite in nature.

Although it is theoretically possible to model arbitrary multipartite textures as single AM-FM functions, doing so requires AM and FM functions that are not locally smooth. This generally makes it impossible to estimate the unknown modulating functions from the image pixel values, since all known AM-FM demodulation algorithms involve approximations that often break down unless the modulating functions possess local smoothness in a sense that can be quan-

0-7803-7147-X/01/$10.00©2001 IEEE

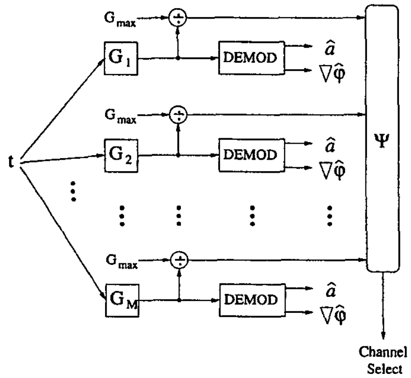

Figure 1. Block diagram of DCA.

tified by certain Sobolev-like derivative norms and smoothness functionals [11]. Therefore, we model textured image regions not as single AM-FM functions, but rather as sums of AM-FM components each having locally smooth modulations. This leads to the image model

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

where each textured image region is modeled by one or more components $t_k(\mathbf{x})$ for which $a_k(\mathbf{x})$ is zero outside the region.

Using a technique called DCA [11], we extract the modulations that dominate the local texture structure at each pixel $\mathbf{x} \in \mathcal{D}$. A block diagram of the DCA algorithm is shown in Fig. 1, where $t(\mathbf{x})$ is passed through a multiband Gabor filterbank with channel frequency responses G_m. On a spatially local basis, the filterbank decomposes the image into locally narrowband components such that, at each pixel, the response of each channel is dominated by at most a single component $t_k(\mathbf{x})$. In doing this, the filterbank structure implicitly defines the componentwise image decomposition in (1); it is not required, however, that the filterbank isolate components from one another on a global scale.

For each $\mathbf{x} \in \mathcal{D}$, spatially local discrete nonlinear demodulation algorithms are applied to each channel response to estimate the modulating functions $a_k(\mathbf{x})$, $\nabla\varphi_k(\mathbf{x})$ of the component that dominates the channel at the pixel. The channel containing the modulations from the component that dominates the overall image texture structure at the pixel is then identified using a spatially local channel selection criterion $\Psi(\mathbf{x})$ described in [11]. The dominant modulations $a_D(\mathbf{x})$ and $\nabla\varphi_D(\mathbf{x})$ are extracted on a pointwise basis to construct modulation domain feature vectors

$$Z(\mathbf{x}) = [A(\mathbf{x})\ R(\mathbf{x})\ \Upsilon(\mathbf{x})\ \mathbf{x}]^T, \qquad (2)$$

where $A(\mathbf{x}) = a_D(\mathbf{x})$, $R(\mathbf{x}) = |\nabla\varphi_D(\mathbf{x})|$, and $\Upsilon(\mathbf{x}) = \arg\nabla\varphi_D(\mathbf{x})$.

3. Bayesian Segmentation

In [23], we described a nonparametric technique that obtains texture segmentations by performing k-means clustering in a modulation domain feature space constructed from the histogram of $Z(\mathbf{x})$ in (2). This technique first normalizes the features by scaling and then utilizes novel data dependent similarity measures in the clustering procedure. In this section, we describe an alternative Bayesian approach that performs unsupervised segmentation by using $Z(\mathbf{x})$ to compute maximum-likelihood (ML) estimates of the pixel class labels.

In developing this approach, we made use of an unsupervised Gaussian mixture modeling algorithm called "Cluster" [2] that was designed primarily to facilitate supervised SMAP segmentation of multispectral remote sensing imagery [3]. In normal use, the algorithm addresses the following problem: multispectral images are to be segmented by assigning class labels to each pixel. Vector-valued observations of each class are available with class labels that are known *a priori*. For each class, Cluster uses the labeled samples to compute the parameters of a multivariate Gaussian mixture describing the class. It does this without supervision by using the MDL criterion [18] to determine the number of Gaussian components in each mixture. The class parameters computed by Cluster are then used to supervise the SMAP segmentation algorithm. The advantage of this approach is that it uses mixtures of arbitrary numbers of Gaussians which can closely approximate the unknown true class distributions. The disadvantage is that pre-labeled observations from each class are required.

To adapt the Cluster algorithm into an unsupervised texture segmentation approach, we assume that each texture class is distributed according to a single multivariate Gaussian. $Z(\mathbf{x})$ in (2) may then be regarded as samples from a mixture of an unknown number of these. Estimation of the parameters of such a mixture is precisely the problem that the Cluster algorithm was designed to solve.

Denote the number of classes by K and let π_k be the probability that a pixel belongs to class k. Each class is described by its mean μ_k and covariance matrix R_k. The mixture is then defined by parameter $\theta = (\pi, \mu, R)$, where $\pi = [\pi_1\ \pi_2\ \ldots\ \pi_k]^T$, $\mu = [\mu_1\ \mu_2\ \ldots\ \mu_k]^T$, and $R = [R_1\ R_2\ \ldots\ R_k]^T$. We induce an arbitrary 1D ordering on the image domain $\mathcal{D}$ to obtain a set $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ of pixel sites where $N = |\mathcal{D}|$ and $s_n \in \mathcal{D}$. $\mathcal{S}$ is used to map the observed feature vectors (2) to a set $z = \{z_n : n \in \mathcal{S}\}$ regarded as a sample of a random vector $Z = \{Z_n : n \in \mathcal{S}\}$.

Let the class label of pixel s_n, which takes values in $[0, K-1]$, be given by a random variable Λ_n and let

$\Lambda = \{\Lambda_n : n \in \mathbb{S}\}$. The conditional density of Z_n given that $\Lambda_n = k$ is then

$$f(z_n|k,\theta) = (2\pi)^{-\frac{5}{2}}|R_k|^{-\frac{1}{2}}e^{-\frac{1}{2}(z_n-\mu_k)^T R_k^{-1}(z_n-\mu_k)}. \tag{3}$$

Assuming that the variables Z_n are mutually independent given K, θ, and Λ yields the log-likelihood function

$$L(\theta) = \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k f(z_n|k,\theta)\right) \tag{4}$$

with ML estimate $\widehat{\theta} = \max_\theta L(\theta)$ for the parameter. However, the joint ML estimate for K and θ is not useful because (4) is a monotonically increasing function of the number of classes K.

In [2], an EM algorithm is given that approximates the ML estimates $\widehat{K}$, $\widehat{\theta}$ by minimizing a minimum description length criterion MDL(K,θ) that is equal to the negative of the log-likelihood function (4) with an added term to penalize solutions where K is large. The strategy is to initialize K with a value K_0 that is much larger than the actual number of classes that are suspected to exist. The MDL criterion is then iteratively minimized with respect to θ, where $\theta^{(K,i)}$ denotes the estimate of θ at iteration i. To obtain an initial value $\theta^{(K_0,1)}$, the class probabilities π_k are all set equal to one another, the covariance matrices R_k are all set equal to the sample covariance computed globally over all N samples in z, and K_0 samples z_n are selected randomly to initialize μ.

For each $n \in \mathbb{S}$ and each $k \in [0, K-1]$, a soft membership function giving the probability that z_n belongs to class k is computed according to

$$P(\Lambda_n = k|z_n, K, \theta^{(K,i)}) = \frac{\pi_k f(z_n|k,\theta^{(K,i)})}{\sum_{m=1}^{K} \pi_m f(z_n|m,\theta^{(K,i)})}. \tag{5}$$

The parameter estimate $\theta^{(K,i)}$ and soft membership functions are then alternately updated in an iteration until the MDL is minimized.

The number of classes K is then reduced by one by merging the two classes that are closest to one another with respect to a distance metric given in [2] and, for this new value K, the MDL is again minimized with respect to θ. Normally, this procedure is repeated until $K = 1$, whereupon the pair K,$\theta^{(K,i)}$ that produced overall the lowest value of the criterion MDL(K,θ) is used to approximate the ML estimates $\widehat{K}$, $\widehat{\theta}$. Labels $\arg\max_k P(\Lambda_n = k|z_n, K, \theta^{(K,i)})$ can then be assigned to each pixel to obtain an unsupervised segmentation.

There is a practical difficulty with this approach, however. Single Gaussians provide only rough approximations to the true unknown class distributions. Thus, the MDL criterion tends to significantly overestimate the number of classes K that are present. This problem can be overcome by using a robust alternative method for estimating the number of classes that was given in [17] and applied in the modulation domain in [23]. With this alternative method, low pass Gaussian filtering is applied to the histogram of $Z(\mathbf{x})$ to estimate the density of pixels about each point in modulation space. A gradient ascent technique is then used to form clusters in the filtered result, where the number of clusters K^* accurately estimates the number of texture classes present in the image. Our approach is to apply this alternative method prior to running the Cluster algorithm and then terminate the Cluster algorithm after the MDL has been minimized with respect to θ for $K = K^*$.

Postprocessing is required to refine the segmentations delivered by the Bayesian segmentation approach described in this Section since incorporation of the image domain position term $\mathbf{x} \in \mathcal{D}$ in the feature vectors $Z(\mathbf{x})$ is not sufficient to enforce a spatial correspondence constraint on the resulting segmentation. As described in [20, 21], connected components labeling with minor region removal is first applied to remove all but the K^* largest components from the image of labels in the segmentation. A morphological majority filter is then applied to the label image to smooth the boundaries of the segmented regions.

4. Examples

Figure 2 illustrates several examples where the unsupervised texture segmentation algorithm described in Section 3 is applied to juxtapositions of Brodatz or Brodatz-like textures. The original images are shown in Fig. 1(a), (c), (h), (k), and (n). For the image in Fig. 1(a), the un-post processed segmentation delivered by the modulation domain k-means clustering technique described in [23], while the result obtained by the technique described in Section 3 without post processing is shown in Fig. 2(c). The final segmentation result of the Bayesian approach presented in this paper with post processing is shown in Fig. 2(d), where the correct pixel classification rate is 98%. For comparison, the final segmentation resulting from the k-means result in Fig. 2(b) had a correct pixel classification rate of 98.23%.

For the images of Fig. 2(e) and (h), the un-post processed results obtained from the Bayesian technique of Section 3 are shown in Fig. 2(f) and (i), respectively. Final segmentation results after post processing are given in Fig. 2(g) and (j), where the correct pixel classification rates were 98.4% and 97%, respectively.

Fig. 2(k) shows a natural scene from the MIT VisTex database. The result obtained from the approach of Section 3 prior to post processing is shown in Fig. 2(l), while the final segmentation after post processing is shown in Fig. 2(m). While ground truth is not available for this image, the Segmentation in Fig. 2(m) is in excellent agreement with our visual perception.

Finally, the synthetic image in Fig. 2(n) is a juxtaposition of five Brodatz textures. The un-post processed segmentation is given in Fig. 2(o), while the final result obtained from the approach described in Section 3 is shown in Fig. 2(p). In this case, the correct pixel classification rate is only 91%. With a large number of textures present in the image, we believe that the fact that single multivariate Gaussians only roughly approximate the true texture class distributions leads to an increased probability of misclassifications in this case.

5. Conclusion

A fully unsupervised modulation domain parametric approach for texture segmentation was presented. The EM algorithm used to approximate the maximum likelihood parameter estimates was adapted from the "Cluster" algorithm given in [2], most notably through the addition of a density-based method for determining the number of texture classes present. In both performance and computational complexity, the proposed approach is comparable to the one given in [23], where segmentations were obtained by applying k-means clustering in a related modulation domain feature space.

The approach presented in this paper demonstrates that modulation domain parametric techniques can be both highly effective and robust against modeling errors. The main issue that needs to be addressed by future research is the development of a method permitting Gaussian mixtures or other sophisticated distributions to be used for modeling the individual texture class distributions that are present in the image.

References

[1] S. T. Acton, D. P. Mukherjee, J. P. Havlicek, and A. C. Bovik. Oriented texture completion by AM-FM reaction-diffusion. *IEEE Trans. Image Proc.*, 10(6):885–896, June 2001.

[2] C. A. Bouman. CLUSTER: an unsupervised algorithm for modeling Gaussian mixtures. December 1999, `http://www.ece.purdue.edu/~bouman`.

[3] C. A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. Image Proc.*, 3(2):162–177, March 1994.

[4] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(1):55–73, January 1990.

[5] 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.

[6] T. Y. Chen, A. C. Bovik, and L. K. Cormack. Stereoscopic ranging by matching image modulations. *IEEE Trans. Image Proc.*, 8(6):785–797, June 1999.

[7] M. Clark and A. C. Bovik. Texture discrimination using a model of visual cortex. In *Proc. IEEE Int'l. Conf. Syst., Man, Cybern.*, Atlanta, GA, 1986.

[8] B. Friedlander and J. M. Francos. Estimation of amplitude and phase parameters of multicomponent signals. *IEEE Trans. Signal Proc.*, 43(4):917–926, April 1995.

[9] J. P. Havlicek. The evolution of modern texture processing. *Elektrik, Turkish Journal of Electrical Engineering and Computer Sciences*, 5(1):1–28, 1997.

[10] 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.

[11] J. P. Havlicek, D. S. Harding, and A. C. Bovik. Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models. *IEEE Trans. Image Proc.*, 9(2):227–242, February 2000.

[12] J. P. Havlicek, J. W. Havlicek, N. D. Mamuya, and A. C. Bovik. Skewed 2D Hilbert transforms and computed AM-FM models. In *Proc. IEEE Int'l. Conf. Image Proc.*, pages 602–606, Chicago, IL, October 4-7 1998.

[13] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional Gabor model of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58(6):1233–1258, 1987.

[14] S. Lu and P. C. Doerschuk. Nonlinear modeling and processing of speech based on sums of AM-FM formant models. *IEEE Trans. Signal Proc.*, 44(4):773–782, April 1996.

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

[16] P. Maragos, J. F. Kaiser, and T. F. Quatieri. Energy separation in signal modulations with applications to speech analysis. *IEEE Trans. Signal Proc.*, 41(10):3024–3051, October 1993.

[17] E. J. Pauwels and G. Frederix. Finding salient regions in images. *Comput. Vision, Image Understand.*, 75(1/2):73–85, July/August 1999.

[18] J. Rissanen. A universal prior for integers and estimation by minimum description length. *Annals of Stat.*, 11(2):417–431, 1983.

[19] B. J. Super and A. C. Bovik. Shape from texture using local spectral moments. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(4):333–343, 1995.

[20] T. Tangsukson and J. P. Havlicek. AM-FM image segementation. In *Proc. IEEE Int'l. Conf. Image Proc.*, pages 104–107, Vancouver, BC, Canada, September 10-13 2000.

[21] T. Tangsukson and J. P. Havlicek. Modulation domain image segmentation. In *Proc. IEEE Southwest Symp. Image Anal., Interp.*, pages 46–50, Austin, TX, April 2-4, 2000.

[22] M. A. Webster and R. L. D. Valois. Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *J. Opt. Soc. Am. A*, 2(7):1124–1132, July 1985.

[23] T. B. Yap, T. Tangsukson, P. C. Tay, N. D. Mamuya, and J. P. Havlicek. Unsupervised texture segmentation using dominant image modulations. In *Proc. 34th IEEE Asilomar Conf. Signals, Syst., Comput.*, pages 911–915, Pacific Grove, CA, October 29-31, 2000.

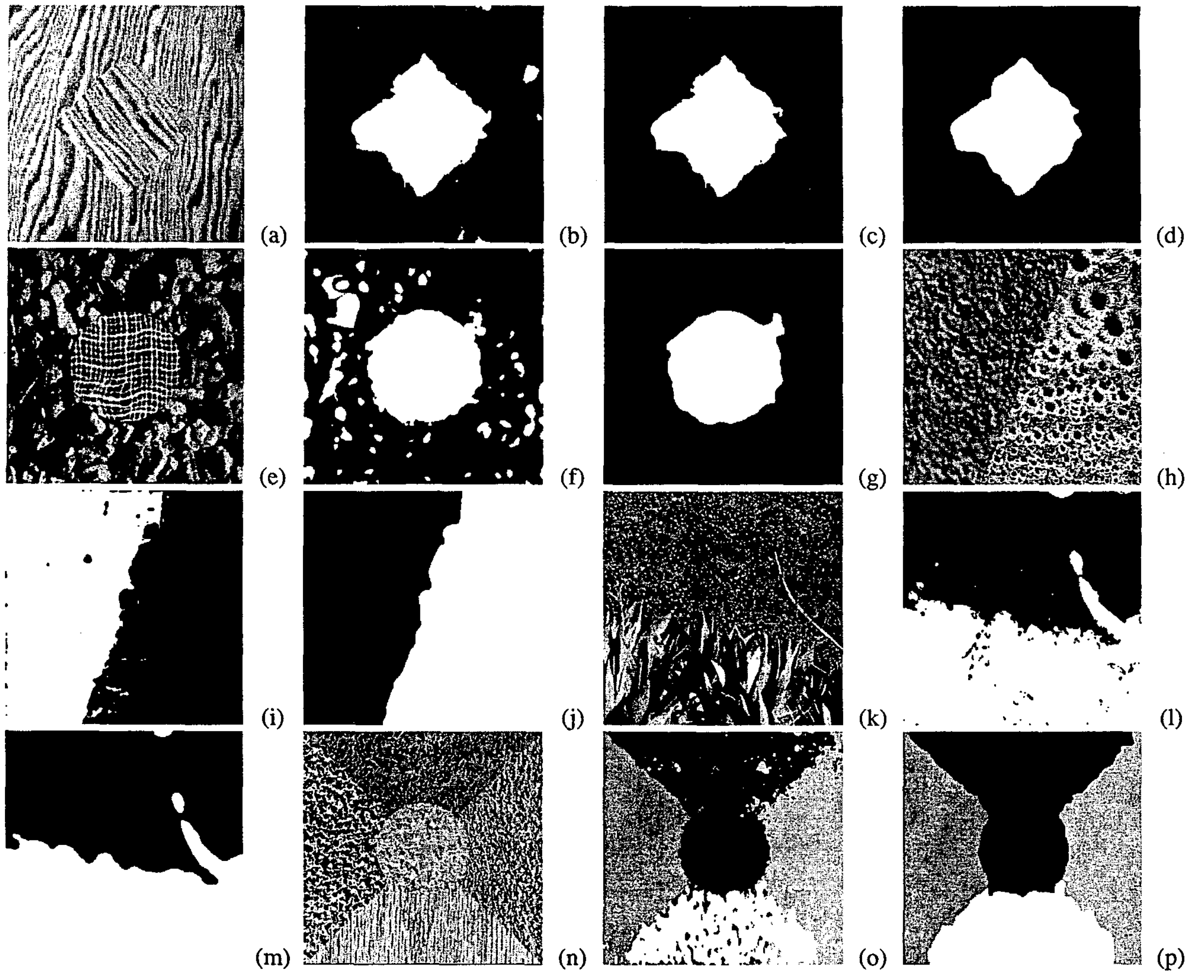

Figure 2. (a) Original two-texture image *WoodWood*. (b) Initial segmentation obtained by applying the k-means clustering approach of [23] without post processing. (c) Initial segmentation obtained by applying the Bayesian approach of of Section 3 without post processing. (d) Final segmentation obtained using the Bayesian approach of Section three with post processing. Correct pixel classification rate is 98%. (e) Original two-texture image *MicaBurlap*. (f) Initial segmentation obtained by applying the Bayesian approach of Section 3 without post processing. (g) Final segmentation result after post processing; correct pixel classification rate is 98.4%. (h) Original two-texture image *CorkFlowers*. (i) Initial segmentation result obtained by applying the Bayesian approach of Section 3 without post processing. (j) Final segmentation result after post processing; correct pixel classification rate is 97%. (k) Original *GrassPlantsSky.0005* image from the MIT VisTex database. (l) Initial segmentation result obtained by applying the Bayesian approach of Section 3 without post processing. (m) Final segmentation result after post processing; ground truth is not available for this image. (n) Original five-texture image. (o) Initial segmentation obtained by applying the Bayesian approach of Section 3 without post processing. (p) Final segmentation result after post processing; correct pixel classification rate is 91%.