Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 27 (2006) 386–392
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Unsupervised color–texture segmentation based on soft criterion with adaptive mean-shift clustering

Yuzhong Wang *, Jie Yang, Ningsong Peng

Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, P.O. Box 104, No. 1954 Hua Shan Road, Shanghai 200030, PR China

Received 4 April 2004; received in revised form 17 July 2005
Available online 25 October 2005

Communicated by S. Ablameyko

Abstract

An improved approach for *J* value segmentation (JSEG) is presented for unsupervised color–texture segmentation. Instead of the color quantization algorithm used in JSEG, an automatic classification method using adaptive mean-shift (AMS) clustering is applied for nonparametric clustering of image data set. The clustering results are used to construct Gaussian mixture modelling for the calculation of soft *J* value. The region growing algorithm used in JSEG is then applied in segmenting the image based on the multiscale soft *J*-images. Experiments show that the improved method overcomes the limitations of JSEG successfully and is more robust.

Keywords: Color–texture segmentation; JSEG; Adaptive mean-shift clustering; Gaussian mixture modeling; Soft *J* value

1. Introduction

Color image segmentation is useful in many applications. From the segmentation results, it is possible to identify regions of interest and objects in the scene, so it is widely used for content-based image retrieval, object detection and recognition, video coding and so on. However, the problem is difficult because natural scenes are rich in color and texture and it is difficult to identify image regions containing color–texture patterns. A variety of techniques have been proposed, for example: direct clustering method in color space (Comaniciu and Meer, 1997), stochastic model based approaches (Belongie, 1998; Delignon, 1997; Wang, 1998), morphological watershed based region growing (Shafarenko et al., 1997), energy diffusion (Ma and Manjunath, 1997), and graph partitioning (Shi and Malik, 1997). In these approaches, (Comaniciu and Meer, 1997) will work well on homogeneous color regions while (Belongie, 1998; Delignon, 1997; Wang, 1998; Shafarenko et al., 1997; Ma and Manjunath, 1997; Shi and Malik, 1997) can be used for color or texture segmentation.

However, these algorithms cannot work well on a large variety of data. The reason is that the methods such as (Comaniciu and Meer, 1997) cannot segment texture regions and the methods such as (Belongie, 1998; Delignon, 1997; Wang, 1998; Shafarenko et al., 1997; Ma and Manjunath, 1997; Shi and Malik, 1997) require the estimation of texture model parameters or the special methods of texture features extraction, which are also very difficult problems. As for this problem, Yining Deng proposed a new approach called JSEG which can be used to segment images into homogeneous color–texture regions (Deng and Manjunath, 2001). JSEG is computationally more feasible than model parameter estimation or general texture features extraction, so it is a more general algorithm for color–texture segmentation.

* Corresponding author. Tel.: +86 2162934115; fax: +86 2162933739.
E-mail address: oliverwang@sjtu.edu.cn (Y. Wang).

0167-8655/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2005.09.014

However, JSEG has two limitations which affect the segmentation results. One is caused by color quantization parameter which determines the minimum distance between two quantized colors, and the quantization results directly influences the segmentation results. A good parameter value yields the minimum number of colors necessary to separate two regions. However, it is very difficult to select a good parameter. Facing an unfamiliar image, the user will have difficulty to select a suitable quantization parameter. In addition, it is inappropriate to process different kinds of images using the same parameter. Therefore, the existence of this parameter degrades the flexibility of JSEG. The other limitation is caused by the varying shades due to the illumination. The problem is difficult to handle because, in many cases, not only the illuminant component, but also the color components of a pixel, change their values due to the spatially varying illumination (Deng and Manjunath, 2001) and this problem usually cause oversegmentation.

In this paper, a new approach is presented to improve JSEG. First, we use adaptive mean-shift clustering to finish color classification instead of using original color quantization algorithm. By this classification method, image data can be divided into appropriate clusters automatically, so the adaptability of JSEG without quantization parameter is improved. Second, motivated by segmentations based on fuzzy theories, we make an assumption that colors distributions in the image obey Gaussian mixture modeling, and calculate soft J values to construct soft J-image using Gaussian mixture modeling. This can effectively restrain the oversegmentation in regions with smooth color transition. Experiments show that our new method is successful and more robust than JSEG.

2. Background

The basic idea of the JSEG method is to separate the segmentation process into two stages, color quantization and spatial segmentation. In the first stage, colors in the image are quantized to several representative classes that can be used to differentiate regions in the image. Then, the image pixel values are replaced by their corresponding color class labels, thus forming a class-map of the image. The class-map can be viewed as *a special kind of texture composition*. In the second stage, spatial segmentation is performed directly on this class-map without considering the corresponding pixel color similarity.

A criterion for "good" segmentation using spatial data points in class-map is proposed. Let Z be the set of all N data points in a class-map. Let $z=(x,y)$, $z\in Z$ and μ be the mean,

$$\mu=\frac{1}{N}\sum_{z\in Z} z. \tag{1}$$

Suppose Z is classified into C classes, Z_i, $i=1,\ldots,C$. Let μ_i be the mean of the N_i data points of class Z_i

$$\mu_i=\frac{1}{N_i}\sum_{z\in Z_i} z. \tag{2}$$

Let

$$S_T=\sum_{z\in Z}\|z-\mu\|^2 \tag{3}$$

and

$$S_W=\sum_{i=1}^{C} S_i=\sum_{i=1}^{C}\sum_{z\in Z_i}\|z-\mu_i\|^2, \tag{4}$$

S_W is the total variance of points belonging to the same class. Define

$$J=(S_T-S_W)/S_W. \tag{5}$$

The definition of J comes from the Fisher's multiclass linear discriminant. J applied to a local area (sampling window) of the class-map is called local J value. The basic window at the smallest scale is a 9×9 round window and the smallest scale is denoted as scale 1. From scale 1, the window size is doubled each time to obtain the next larger scale. Image whose pixel values correspond to those local J values calculated over sampling windows centered at the pixels is called J-image. The higher the local J value is, the more likely that the corresponding pixel is near a region boundary. Contrarily, a small local J value indicates corresponding pixel be in a region interior. The characteristics of the J-image allow us to use a region-growing method to segment the image. Consider the original image as one initial region. The algorithm starts the segmentation of the image at a coarse initial scale. Then, it repeats the same process on the newly segmented regions at the next finer scale. The initial result often has oversegmented regions, so a region merging algorithm must be employed to eliminate oversegmentation. Two neighboring regions are merged based on their color similarity denoted by the distance between two color histograms. A maximum threshold for the distance is needed in region merging algorithm. For more detail information, please refer to (Deng and Manjunath, 2001).

In fact, an accurate construction of class-map can result in an outstanding segmentation; furthermore, an exact quantization parameter can result in an accurate class-map. Therefore, the existence of quantization parameter, as mentioned above, degrades the flexibility of JSEG. In addition, the second limitation is caused by hard classification for colors. Because hard classification often divides colors with smooth transition into several classes, therefore makes several obvious visually different regions in class-map which originally belong to the same region.

A synthetic image, its J-image at scale 2 and segmentation result are shown in Fig. 1 (quantization parameter is 200, the number of scales is 2 and region merge threshold is 0.4). In the synthetic image, there is phenomenon of colors smooth transition in yellow and blue regions. Therefore, it is impossible to avoid oversegmentation no matter what the quantization parameter is selected. This simple example well shows the limitations of JSEG.

Fig. 1. (a) A synthetic image, (b) J-image at scale 2 and (c) segmentation result at scale 2.

3. The improved method based on Gaussian mixture modeling

To overcome the first limitation of JSEG, a nonparametric clustering based on adaptive mean-shift is used for colors quantization. While to overcome the second limitation, a soft criterion based on Gaussian mixture modeling replaces hard classification to calculate soft J value for labeling every pixel.

3.1. Adaptive mean-shift clustering

Adaptive mean-shift clustering method (Comaniciu, 2003) is an excellent unsupervised kernel density estimation method. It can automatically select bandwidth of every sample point to finish mode seeking.

Assume that each data point $x_i \in R^d$, $i = 1, \ldots, n$ is associated with a bandwidth value $h_i > 0$. The sample point estimator

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^d}k\left(\left\|\frac{x - x_i}{h_i}\right\|^2\right) \tag{6}$$

based on a spherically symmetric kernel K is an adaptive nonparametric estimator of the density at location x in the feature space. It can be proven that at location x the weighted mean of the data points selected with kernel G is proportional to the normalized density gradient estimate obtained with kernel K. The kernel G is the shadow kernel of K (Comaniciu, 2003). The implication of the mean shift is a hill climbing technique to the nearest stationary point of the density, i.e., a point in which the density gradient vanishes. There are numerous methods described in the statistical literature to define h_i, and for reducing the computational complexity a simple method is used to decide adaptive bandwidth in next section.

3.2. The kernel selection and the method of deciding adaptive bandwidth

Adaptive mean-shift clustering is employed to classify color image data. Images are usually stored and displayed in the RGB space. However, to ensure the isotropy of the feature space, a uniform color space with the perceived color differences measured by Euclidean distances should be used. We have chosen the $L^*U^*V^*$ space, whose coordinates are related to RGB values by nonlinear transformations, thus allowing the use of spherical windows (Comaniciu and Meer, 1997). We assume image data obey Gaussian mixture modeling in $L^*U^*V^*$ space, so we employ the multivariate normal kernel

$$K(x) = (2\pi)^{-d/2}\exp\left(-\frac{1}{2}\|x\|^2\right) \tag{7}$$

in adaptive mean-shift procedure.

We know that if there are large quantities of sampling points the influence of bandwidth h can be weakened extremely (Bian and Zhang, 2001). Making an assumption that we use enough sampling points with fixed number for sample point estimator, we can see an attractive behavior of the adaptive estimator: the data points lying in large density regions affect a narrower neighborhood since the kernel bandwidth h_i is smaller, but given a larger importance due to the weight $\frac{1}{h_i^d}$. By contrast, the points that correspond to the tails of underlying density are smoothed more and receive a smaller weight. The extreme points (outliers) receive very small weights, being thus automatically discarded. For computational reasons, the simplest way to obtain the density estimate is by nearest neighbors of fixed number. Let $x_{i,k}$ be the k-nearest neighbor of the point x_i. Then, we take $h_i = \|x_i - x_{i,k}\|_1$. This method is also used in (Georgescu et al., 2003). In order to get the appreciate parameter k, five images of different sizes, which are the same as Fig. 1a, are used to test adaptive mean-shift clustering. Because the pixel number of the real image is about 10^5 or so, the sizes of the artificial images mentioned above are 100×100, 200×200, 300×300, 400×400 and 500×500. Since the number of neighbors used for density estimation does not have a strong influence, let the parameter k is a multiple of 100 and let $k = 100$–1200. An approximation technique (Georgescu et al., 2003), locality-sensitive hashing (LSH), was employed to reduce the computational complexity of adaptive mean-shift procedure and convergence is declared when the magnitude of the shift becomes less than 0.1. Fig. 2 shows the mean classifying accuracy for different values of k and colors classification result with $k = 200$. Visually the synthetic image should be classified into three color classes and it is actually decomposed into three clusters. The results show that the classifications

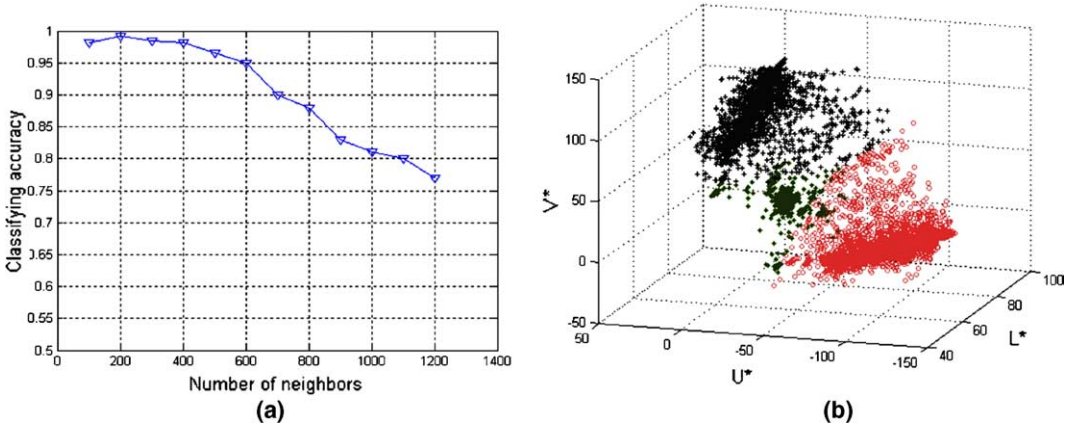

Fig. 2. (a) The mean classifying accuracy for different values of k and (b) clustering result with $k = 200$.

are basically correct and acceptable with $k = 100$–500, however considering a tradeoff between computation speed and accuracy, we select the parameter $k = 200$ in our method.

3.3. Soft J value based on Gaussian mixture modeling

Suppose $\{I_k\}$, $k = 1, \ldots, N$ is the set of all pixels of the color image $I(x, y)$, and I_k obey Gaussian mixture distribution of C classifications. Mark sub-Gaussian distribution as ω_i, $i = 1, \ldots, C$. Then, the statistical distribution $p(I_k)$ of I_k can be approximately expressed with Gaussian mixture modelling of C classes, and the probability density function of every subsidiary Gaussian distribution ω_i can be expressed as follows:

$$p(I_k|\omega_i, \theta_i) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(I_k - \mu_i)^T \Sigma_i^{-1}(I_k - \mu_i)\right\} \quad i = 1, \ldots, C, \tag{8}$$

$\theta_i = (\mu_i, \Sigma_i)$ denotes the parameters of Gaussian mixture modelling, and μ_i is the mean and Σ_i is the covariance matrix; the prior probability of ω_i is $P(\omega_i)$. μ_i and Σ_i can be calculated with the data belonged to the ith class and $P(\omega_i)$ is the ratio of the number of pixels of the ith class to total number of pixels.

Then we can calculate every pixel's membership $(m_{I_k,j}(k = 1, \ldots, N; i = 1, \ldots, C))$ of every class with Bayesian equation

$$m_{I_k,j} = \frac{P(\omega_i)p(I_k|\omega_i, \theta_i)}{\sum_{j=1}^{C} P(\omega_j)p(I_k|\omega_j, \theta_j)} \quad k = 1, \ldots, N; \ i = 1, \ldots, C. \tag{9}$$

After finishing calculation of pixel's membership, we redefine the calculation of J value, still letting Z be the set of all N data points in a class-map and $z = (x, y)$, $z \in Z$. Suppose image data set is classified into C classes. Eqs. (1), (3) and (5) need not to be changed. Modify Eq. (2) as follows:

$$\mu_i = \frac{\sum_{z \in Z} z \cdot m_{z,i}}{\sum_{z \in Z} m_{z,i}} \quad i = 1, \ldots, C \tag{10}$$

and modify Eq. (4) as follows:

$$S_W = \sum_{i=1}^{C} S_i = \sum_{i=1}^{C} \sum_{z \in Z} \left(m_{z,i} \cdot \|z - \mu_i\|^2\right). \tag{11}$$

Then, the J value calculated with new rules is called soft J value, and the new J-image constructed by soft J values is called soft J-image. The second limitation can be overcome by using region growing in soft J-image.

Soft J-image of the synthetic image and corresponding segmentation result are shown in Fig. 3. The experimental results prove that the improved method overcomes the limitations of JSEG successfully.

4. Experimental results

Taking 200 human labeled images from Berkeley segmentation dataset as ground truth, we test the improved algorithm in order to compare with edge flow algorithm (Ma and Manjunath, 1997) and normalized cuts algorithm (Shi and Malik, 1997) as well as JSEG. Using the relative area of region coincidence as an evaluation criterion, we get statistics on segmentation accuracy. The results show that the new method is more robust than these algorithm mentioned above.

The selection of the number of scales relates to images resolution. Generally speaking, the higher the resolution is, the larger the number of scales is. Because the size of all images from Berkeley segmentation dataset is 321×481, three scales are used in this test. With JSEG, it has been very difficult for us to identify a combination of a quantization parameter and a region merge threshold which will make the algorithm more efficient. The reason lies in that when quantization parameter is small, oversegmentation is very obvious and a large region merge threshold is often needed to eliminate oversegmentation as much as possible. Contrarily, when quantization parameter is large, region merge threshold should be small. Since

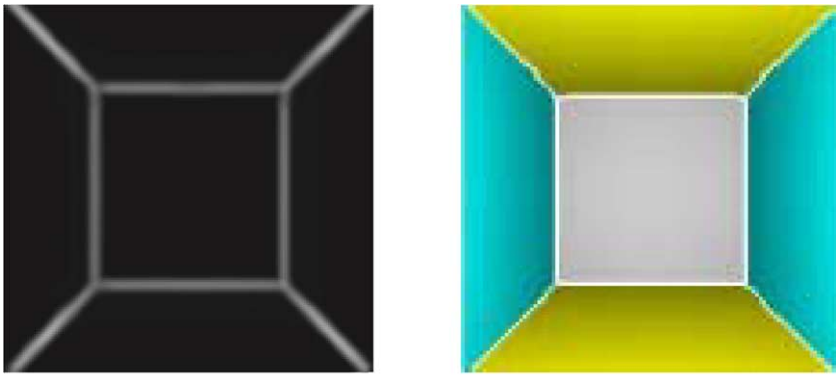

Fig. 3. (a) Soft J-image at scale 2 and (b) corresponding segmentation result.

quantization parameter is not needed in our method, the selection of region merge threshold is relatively simple. Fig. 4 shows the comparison of mean accuracy between our method and JSEG (quantization parameter, QT, respectively equals 100, 200, 300, 400, 500 and 600) respectively with different region merge thresholds $(0, 0.1, 0.2, \ldots, 0.9, 1)$. From Fig. 4, we know an appropriate region merge threshold should be selected from 3.5 to 5 in our method. When the region merge threshold equals 0.4, the both methods achieved the best segmentation results. The highest accuracy of our method is 75.6% and that of JSEG is 58.2%. Since edge flow algorithm and normalized cuts algorithm (DOOG filters are used in the two methods to extract texture features) are affected by some necessary critical parameters, their adaptability and flexibility are not so good. The best mean segmentation accuracy of edge flow and normalized cuts are 68.3% and 64.5%, respectively. Fig. 5 shows the segmentation results of four images (flower, leopard, stone, eagle) respectively from manual label, our method (the region merge threshold is 0.4), JSEG (quantization parameter is 200 and the region merge threshold is 0.4), edge flow and normalized cuts.

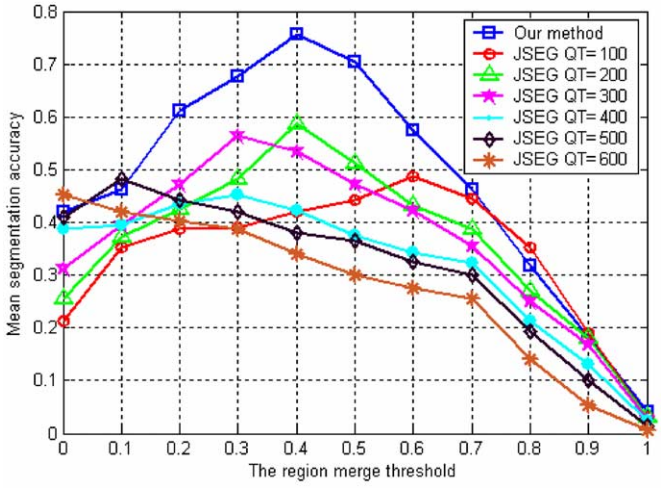

Fig. 4. The comparison of segmentation accuracy between our method and JSEG respectively with different region merge thresholds.

Table 1 shows the comparison of the colors classification results of the four images showed in Fig. 5, which are respectively from color quantization method in JSEG and adaptive mean-shift clustering. The results show that adaptive mean-shift clustering algorithm is very robust, while the original color quantization algorithm is not flexible. It indicates that it is inappropriate to process different kinds of images using the same parameters and that the existence of quantization parameter degrades the flexibility of JSEG.

Since our method is a time-consuming approach, it is suitable for offline processing, not for real-time application. For example, for a 321×481 image, the running time is about 43 s on a Pentium IV 1.8 GHz processor.

5. Application

The improved algorithm is successfully applied in our tongue characterization system and is mainly used to segment homogenous regions of substance and coat for colors identification. In (Shen et al., 2001) a tongue characterization system is introduced and segmentation of homogenous regions of substance and coat as well as colors identification are finished by recognizing pixels through the standard color samples. However, it does not coincide with human perception and it only contains statistical information, disregarding spatio-temporal information that is very important to doctors. In fact, there is already enough information in a single tongue that can be used for segmenting homogenous regions of substance and coat. To obtain results that coincide with human perception, we should employ a fine-to-coarse then coarse-to-fine method, that is, substance and coat should be segmented into different homogenous regions at first and then every pixel in different regions is recognized by using standard color samples.

Therefore, there is no doubt that it is a correct choice to segment homogenous regions in tongue by using our improved algorithm of JSEG through which we have

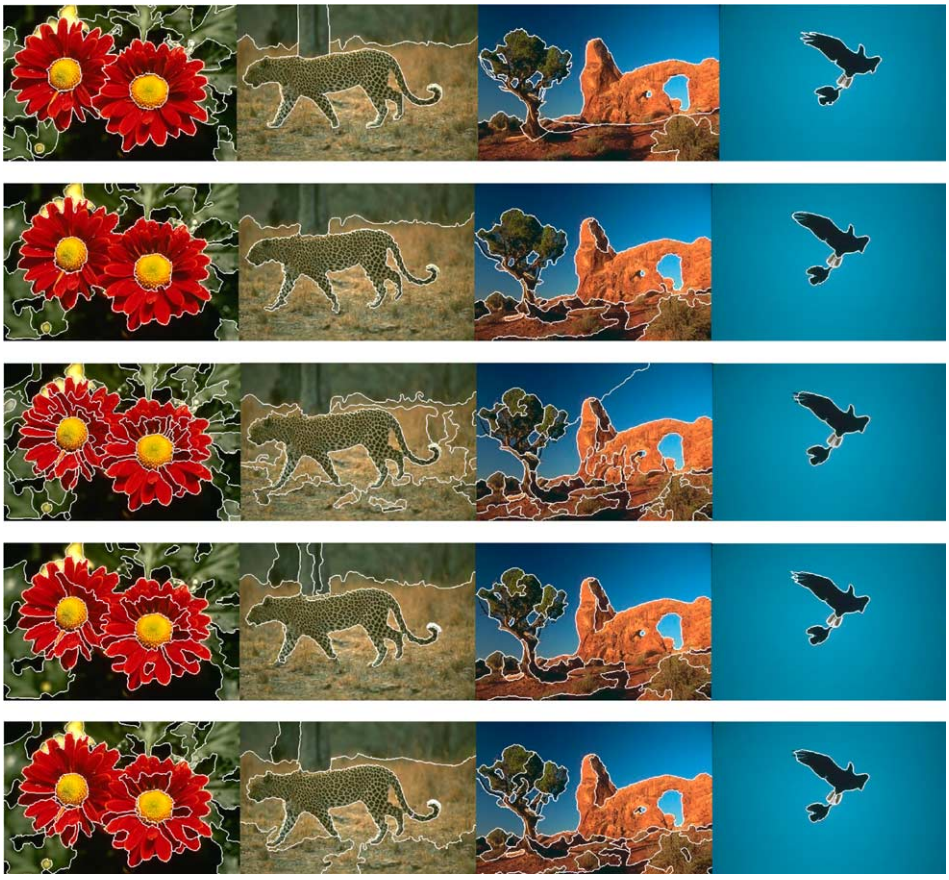

Fig. 5. The manual segmentation results and the results from our method, JSEG, edge flow and normalized cuts are shown form the first row to the fifth row in turn.

Table 1
Comparison between color quantization method in JSEG and adaptive mean-shift clustering

	Color quantization algorithm in JSEG								AMS clustering
	QT	50	100	200	250	300	400	500	600	$k = 200$
Flower	Number of color classes	15	13	11	11	10	10	9	7	7
Leopard		10	4	3	3	2	2	2	1	5
Stone		16	15	12	10	8	8	6	5	8
Eagle		4	3	3	3	3	3	2	2	4

Fig. 6. Four segmentation results of homogeneous regions of substance and coat.

achieved excellent results. Fig. 6 shows four examples of regions segmentation of substances and coats. These results are quite coinciding with human perception and are appreciated by experts on traditional Chinese medicine. Since the segmentation of homogenous regions of substance and coat is a monotone application, the algorithm work very

well after we specify the numbers of scale and the region merge threshold. We have tested the algorithm with 1500 tongue images and the mean accuracy is 92.6%.

6. Conclusions

In this work, an improved approach for JSEG is presented for the fully unsupervised segmentation color–texture regions in color images. An automatic classification method based on adaptive mean-shift clustering is used for nonparametric clustering of image data set. Gaussian mixture modeling of image data constructed with classifications achieved by adaptive mean-shift procedure is applied in the calculation of soft J value.

If we want to get good results by JSEG, the parameters used in JSEG must be adjusted repeatedly. Fortunately, the JSEG is relatively insensitive to scale threshold and region merging threshold. However, improper quantization threshold badly affects the result of color classification leading to wrong segmentation. Therefore, repeated selecting quantization threshold will exhaust users and is forbidden in automatic systems. In the traditional clustering techniques, we know, the feature space is usually modeled as a mixture of multivariate normal distributions, which can introduce severe artifacts due to the elliptical shape imposed over the clusters or due to an error in determining their number. However, the adaptive mean-shift based nonparametric feature space analysis eliminates these artifacts. Therefore, Gaussian mixture modeling constructed from the results obtained by adaptive mean-shift clustering method is consequentially more exact.

Experiments show the new method overcomes the limitations of JSEG successfully and is more robust. Excellent adaptability and flexibility of the improved method make it more applicable in practical systems.

References

Belongie, S., 1998. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. In: Proc. of ICCV, pp. 675–682.

Bian, Z.Q., Zhang, X.G., 2001. Pattern Recognition. Tsinghua Press, Beijing, China.

Comaniciu, D., 2003. An algorithm for data-driven bandwidth selection. IEEE Trans. PAMI 2, 281–288.

Comaniciu, D., Meer, P., 1997. Robust analysis of feature spaces: Color image segmentation. IEEE Proc. CVPR, pp. 750–755.

Delignon, Y. et al., 1997. Estimation of generalized mixtures and its application in image segmentation. IEEE Trans. Image Process. 6, 1364–1376.

Deng, Y., Manjunath, B.S., 2001. Unsupervised segmentation of color–texture regions in images and video. IEEE Trans. PAMI 8, 800–810.

Georgescu, B., Shimshoni, I., Meer, P., 2003. Mean shift based clustering in high dimensions: A texture classification example. In: Proc. 9th Internat. Conf. on Computer Vision, pp. 456–463.

Ma, W.Y., Manjunath, B.S., 1997. Edge flow: A framework of boundary detection and image segmentation. In: Proc. of CVPR, pp. 744–749.

Shafarenko, L., Petrou, M., Kittler, J., 1997. Automatic watershed segmentation of randomly textured color images. IEEE Trans. Image Process. 11, 1530–1544.

Shen, L.S., Wang, A.M., Wei, B.G., 2001. Image analysis for tongue characterization. Acta Electron. Sinica 12, 1762–1765.

Shi, J., Malik, J., 1997. Normalized cuts and image segmentation. In: Proc. of CVPR, pp. 731–737.

Wang, J.P., 1998. Stochastic relaxation on partitions with connected components and its application to image segmentation. IEEE Trans. PAMI 6, 619–636.