An Automatic Tongue Detection and Segmentation Framework for Computer-Aided Tongue Image Analysis

Wentao Xu[†] *, Ratchadaporn Kanawong[†], Dong Xu[†], Shao Li[‡], Tao Ma[‡], Guixu Zhang*, Ye Duan[†]

[†] Computer Science Department and Informatics Institute, University of Missouri-Columbia, MO, 65211, USA
‡ MOE Key Laboratory of Bioinformatics and Bioinformatics Division, TNLIST and Department of Automation, Tsinghua University, Beijing 100084, China.
*Computer Science Department, East China Normal University, Shanghai, 200062, China

***Abstract*—Traditional Chinese Medicine (TCM) has a long history and has been recognized as a popular alternative medicine in western countries. Tongue diagnosis is a significant procedure in computer-aided TCM, where tongue image analysis plays a dominant role. In this paper, we proposed a fully automatic tongue detection and tongue segmentation framework, which is an essential step in computer-aided tongue image analysis. Comparing with other existing methods, our method is fully automatic without any need of adjusting parameters for different images and do not need any initialization.**

***Index Terms*—Tongue analysis, TCM, computer-aide diagnosis.**

I. INTRODUCTION

Traditional Chinese Medicine (TCM) has a long history in China and it has been recognized as a popular complementary and alternative medicine in western countries [1]. TCM diagnosis is based on the information obtained from four diagnostic processes, i.e, inspection, listening, smelling, inquiry and palpation . The inspection is to examine the patient by observing his or her shape, expression and tongue, etc. The listening and smelling is to collect information for diagnosis by smelling the odor and listening to the voice of the patient. The inquiry is to diagnose through conversation with the patient. The palpation process is to examine the pulses of the patient's wrist at some important locations. Then the possible causes of diseases can be concluded based on the information summarized from these processes and the treatment is implemented hereafter [2]. Since tongue diagnosis is one of the most important and widely used diagnostic methods in TCM [2], recently researchers have started to develop computer-aided tongue analysis algorithms based on the new advancement in digital photogrammetry, image analysis and pattern recognition technologies. In general computer-aided tongue analysis algorithm consists of the following main steps: tongue detection, tongue segmentation, tongue feature extraction and tongue analysis [3].

For tongue detection, Wang *et al*. developed a color-texture detection algorithm in [4]. Firstly, a number of homogenous regions are obtained by an improved JSEG [5]. Then these regions are classified into different categories of colors of substances and coatings based on the Earth Mover's distance. Locally linear embedding techniques are employed to eliminate the discrepancies among samples. In the area of tongue segmentation, Pang *et al.* proposed a bi-elliptical deformable contour (BEDC) method [6], which is based on a combination of a bi-elliptical deformable template (BEDT) and an active contour model (ACM) or Snakes. The BEDT obtains coarse shape features through utilizing the steepest decent method on its energy function in the parameter space. The BEDC is evolved from the BEDT by replacing classical internal forces with template forces, and can be deformed to fit local details. The algorithm can automatically segment tongue images but need initializing. Liu *et al.* present a novel tongue segmentation method that uses hyperspectral images and the support vector machine [7]. This method combines spatial and spectral information to analyze the medical tongue image and can provide better tongue segmentation results. Hyperspectral imaging devices however are not always easily accessible.

Despite considerable progresses in computer-aided tongue detection and segmentation, there are still some problems with the existing approaches. For example, the BEDC method of [6] is very useful in segmentation, but is not robust, especially when the surface color of the tongue is similar to its ambient tissue. Contour initialization and template are indispensable in this method, which weakens the automaticity of segmentation. Furthermore the bi-elliptical tongue template is not always suitable for every tongue. In [7], the device named pushbroom hyperspectral imager (PHI) is required, which is not always available. Good results can be obtained through JSEG [5-6], but it is required to interactively adjust the parameters for different images.

In this paper, we propose a fully automatic tongue image detection and segmentation framework that can overcome the aforementioned limitations of the existing methods. More specifically, we developed a Principal Component Analysis (PCA) based tongue detection algorithm as well as a new hybrid tongue segmentation algorithm that fully exploits the robustness of the Mean Shift algorithm and the efficiency of the Canny Edge Detection algorithm, together with the outlier removal capability of the Tensor Voting algorithm. Experimental results show

Ratchadaporn Kanawong, Ye Duan (Corresponding author email: duanye@missouri.edu), Dong Xu are with the Computer Science Department and Informatics Institute, University of Missouri-Columbia, MO, 65211, USA.

Wentao Xu is a visiting scholar of the Computer Science Department, Univsersity of Missouri-Columbia, MO, 65211,USA and with East China Normal University, Shanghai, 200062 China. Guixu Zhang is with the Computer Science Department, East China Normal University, Shanghai, 200062 China.

Shao Li, Tao Ma are with the MOE Key Laboratory of Bioinformatics and Bioinformatics Division, TNLIST and Department of Automation, Tsinghua University, Beijing, 100084, China.

that the new framework is robust and efficient. Based on our tests our new algorithm clearly outperforms other existing methods.

II. ALGORITHM

Our algorithm can be divided into three primary phases:

1) Mean Shift-based clustering;
2) Tongue detection;
3) Tongue segmentation.

Given a tongue image, we will first conduct Mean Shift based clustering to divide the image into a number of clusters based on the color and spatial similarity. Next we will apply a PCA-based tongue detection method to detect the image cluster that contains the tongue. Finally we will employ the Tensor Voting based image segmentation method to extract the boundary contour from the tongue cluster.

Mean Shift-based Clustering: The initial clustering is done based on the Mean Shift algorithm [8]. Mean Shift is a powerful technique for clustering scattered data. It is an iterative approach to cluster image by searching a number of density modes on color domain. Firstly, an initial mean x is estimated. Secondly, the kernel density function K(x) is computed. Thirdly, x is replaced by m(x):

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)} \qquad (1)$$

where N(x) is the neighborhood of x. These three steps will iterate until it converges.

PCA-based Tongue Detection: In this paper, we developed a Principal Component Analysis (PCA) based tongue detection algorithm. Since tongue has very similar shape of an ellipse [6], we can utilize ellipse as a template to detect the tongue. More specifically, for each of the individual cluster obtained from the Mean Shift algorithm, we will employ PCA to fit an ellipse into the cluster. We will then compute the similarity measurements between the cluster and the fitting ellipse. If the similarity is greater than a threshold, then the cluster is detected as a cluster that contains the tongue.

PCA computation consists of the following two major steps: 1) compute the covariance matrix of a cluster, and 2) compute the eigenvectors (v_1, v_2) and eigenvalues (l_1, l_2) of the covariance matrix. Once the PCA is conducted, an ellipse equation $f(\hat{x}, \hat{y}) = 0$ can then be derived from the eigenvectors and eigenvalues by using (2), (3).

$$\frac{(x-c_x)^2}{a^2} + \frac{(x-c_y)^2}{b^2} = 1 \qquad (2)$$

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad (3)$$

Where $a = 2\sqrt{l_1}$, $b = 2\sqrt{l_2}$. $\theta = \tan^{-1}\frac{y_{v_1}}{x_{v_1}}$ is the angle of the major axis of the ellipse to the x-axis, (c_x, c_y) is the center of the ellipse, which is also the mean of the cluster.

We define three similarity measurements (M_1, M_2, M_3) based on the intersection region between the cluster and its fitted ellipse (4). These measurements give similarity scores ranging from 0 (most dissimilar) to 1 (most similar).

$$M_1 = \frac{A_I}{A_C}, M_2 = \frac{A_I}{A_E}, \; M_3 = M_1 * M_2 \qquad (4)$$

Where A_I, A_C, A_E are the area of the intersection region, the area of the cluster and the area of the fitted ellipse, respectively.

We implement our method on a group of tongue images. In our experiments, we use 0.9, 0.9, 0.8 as the thresholds of M_1, M_2, M_3 respectively. Negative results (rejected clusters) of detection are shown in Table 1. Positive results (detected clusters) of detection are shown in Table 2. From the comparison of Table 1 and Table 2, we can see that M_3 is more effective than M_1 and M_2.

Table 1. Negative results of tongue detection.

	Image1	Image2	Image3	Image4
Cluster				
M_1	0.87	0.77	0.70	0.55
M_2	0.64	0.90	0.89	0.86
M_3	0.56	0.69	0.62	0.47

Table 2. Positive results of tongue detection.

	Image1	Image2	Image3	Image4
Cluster				
M_1	0.95	0.92	0.90	0.92
M_2	0.94	0.94	0.94	0.95
M_3	0.89	0.86	0.85	0.88

Tensor Voting-based Tongue Segmentation: Tensor voting [9] is a method of information propagation where tokens convey their orientation preferences to their neighbors in the form of votes. Each vote is an estimate of orientation or termination of a perceptual structure consisting of just two tokens: the voter and the receiver. Smooth perceptual structures are fitted between the two locations to generate the orientation estimates at the receiver. The strength of the votes attenuates both with distance and with increased curvature of the hypothesized structure, making straight continuations preferable to curved ones following the principles of smooth continuation and simplicity.

The tensor field generated by a voting procedure allows effective and robust communications among the data to extract both coherent geometry (through the tensor structure) and likelihood information (through the saliency field). Suppose that there exists a smooth curve connecting the origin O and a point P, and the normal to the curve at O is known. Then what is the most likely normal direction at P? It is argued in [9] that the osculating circle connecting O and P is the most likely connection since it keeps the curvature constant along the hypothesized circular arc. So the most likely normal is given by the normal to the circular arc at P. This normal at P is oriented such that its inner product with the normal at O is nonnegative. The length of this normal, which represents the voting strength, is inversely proportional to the arc length s and curvature k. So the decay function of vote strength is defined as (5):

$$DF(s,\kappa,\sigma) = e^{-\left(\frac{s^2+c\kappa^2}{\sigma^2}\right)} \quad (5)$$

where σ controls smoothness, which determines the effective neighborhood size , and constant c controls the decay with high curvature.

Votes are cast from token to token and accumulated from their neighboring points. After votes are collected at every location, stick and ball saliency maps are built, local structures such as curves, junctions, region boundaries can be detected based on their high region saliency and high polarity.

In our study, the tensor matrix is calculated from the gradient $[d_x, d_y]$ of the extracted boundary contour:

$$S = \begin{bmatrix} d_x^2 & d_x d_y \\ d_x d_y & d_y^2 \end{bmatrix} \quad (6)$$

The gradient can be computed using finite difference:

$$\begin{aligned} dx &= x(n+1) - x(n-1) \\ dy &= y(n+1) - y(n-1) \end{aligned} \quad (7)$$

The boundary contour extracted from the Mean Shift algorithm is not always very smooth which can result in unreliable gradient estimation if computed directly by the above finite difference method (Fig. 1(a)). In this paper we propose a local average gradient (LAG) algorithm, which can attenuate the influence of the noise (Fig. 1(b)). In particular, we exploit local line fitting technique to fit a straight line in the local neighborhood . The slope of locally fitted line is then utilized as the smoothed gradient estimation at the point.

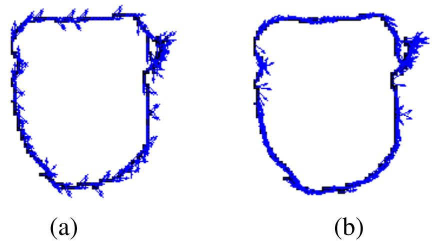

Figure 1. (a) Gradient directions of the tongue boundary contour estimated using equation (7); (b) Smoothed gradient directions by the local average gradient (LAG) algorithm.

Since the Mean Shift algorithm is a region-based clustering algorithm, it might not be very sensitive to the discontinuity of the image gradient and thus can miss important image edges. To overcome this issue, after an initial boundary contour is extracted from the Mean Shift algorithm, we will apply a Localized version of the Canny Edge Detection algorithm next to provide complement information, which will be fused in the following Tensor Voting step. The Localized Edge Detection (LED) will apply only to the local neighborhood around the region of interest, which is obtained by applying the mathematical morphology dilation operation with a disc element onto the boundary contour extracted from the above initial cluster step.

Finally, we will combine the boundary contour extracted by the Mean Shift algorithm (Fig. 2(a)) with the edges extracted by the Localized Canny edge detection algorithm (Fig. 2(b)) into one binary image (Fig. 2(c)). We will then compute the gradient vector and the tensor matrix at the edge and contour pixels of the binary image, which will then be casting votes on all the pixels of the image to generate a global tensor field. Figure 2(d) shows the stick saliency map constructed from the global tensor field. In the end, we will extract the refined object contour based on the stick saliency map (Fig. 2(e)).

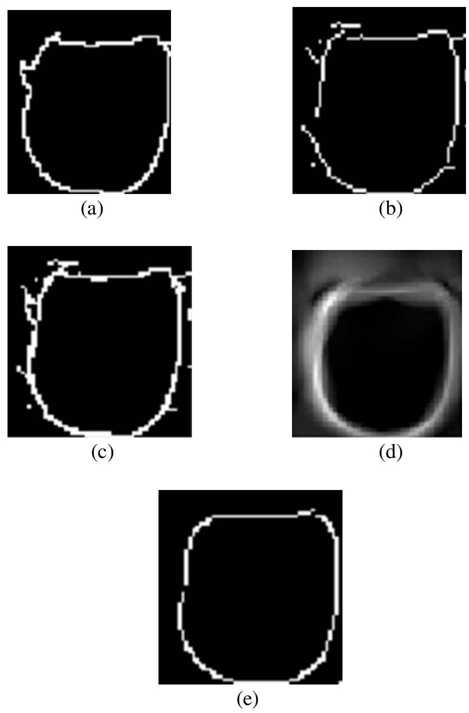

Figure 2. Tongue image segmentation by Tensor Voting. (a) Object boundary contour extracted by Mean Shift algorithm; (b) Edges extracted by Localized Canny Edge detection; (c) Boundary contour extracted from (a) is overlaid with edges extracted from (b); (d) Stick saliency map generated by the Tensor Voting; (e) Refined object boundary contour extracted from the stick saliency map of (d).

III. Experiment

In order to demonstrate the robustness and the efficacy of the proposed algorithm, we conducted some experiments to compare our algorithm with other popular image segmentation algorithms such as Mean Shift, JSEG, as well as Canny edge detection on some testing tongue images. The preliminary testing results (Fig.3) show that our algorithm clearly outperforms the above three algorithms.

IV. Conclusion

In this paper, we proposed a novel tongue image detection and segmentation framework. Comparing with other existing methods, our method is fully automatic without any need of adjusting parameters for the different images and does not need any initialization. In the future we would like to continue this endeavor and apply the developed framework in areas such as automatic tongue image classification and computer-aided diagnosis.

References

[1] D. Xu, "Mutual understanding between Traditional Chinese Medicine and systems biology: gaps, challenges and opportunities," *International Journal of e Functional Informatics and Personalised Medicine*, vol. 2, no. 3, pp. 248-260, 2009.

[2] S. Lukman,Y.L. He, and S.C. Hui, "Computational methods for Traditional Chinese Medicine:A survey," *Computer Methods and Program in Biomedicine,* vol. 88, no. 3, pp. 283-294, Dec. 2007.

[3] C.C. Chiu, "A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue", Computer Methods and Programs in Biomedical, vol. 61, pp. 77-89, 2000.

[4] Y.G. Wang, J. Yang,Y. Zhou, and Y.Z. Wang,"Region partition and feature matching based color recognition of tongue image",*Pattern Recognition Letters, vol.* 28, no. 1, pp. 11-19, Jan. 2007.

[5] Yining Deng, and B.S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Analysis and Machine Intelligence.* vol. 23, pp. 800-810. 2001.

[6] B. Pang, D. Zhang, and K.Q. Wang,"The Bi-Elliptical Deformable Contourand Its Application to Automated Tongue Segmentation in Chinese Medicine", *IEEE Trans. On Med. Imaging,* vol. 24, no. 8, pp. 946-956, Aug. 2005.

[7] Z. Liu, J.Q Yan,D. Zhang, and Q.L Li,"Automated tongue segmentation in hyperspectral images for medicine," *Applied Optic , vol. 46, no. 34, pp.* 8328-8334, Nov. 2007.

[8] D. Comaniciu, P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans.Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 603–619, May. 2002.

[9] G. Medioni and S.B. Kang, *Emerging topics in computer vision*. Prentice Hall, 2004.

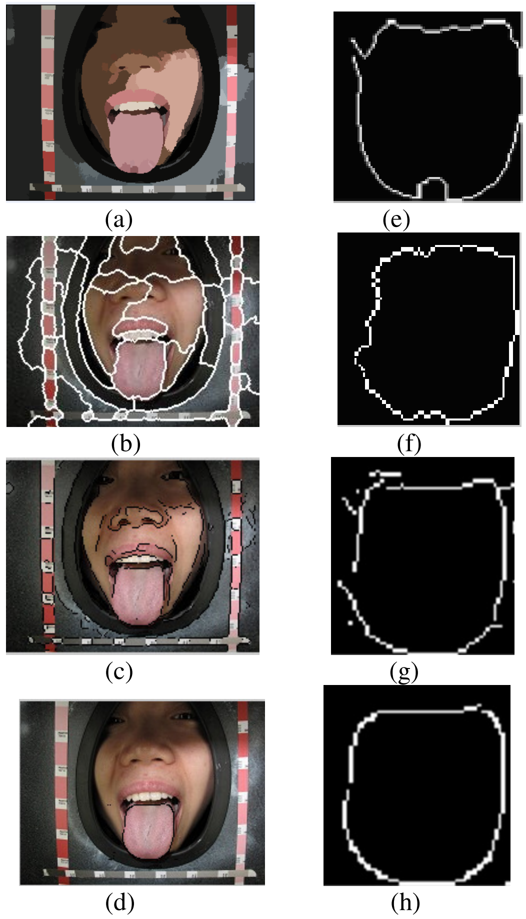

Figure 3. Comparison of our algorithm with other image segmentation algorithms such as Mean Shift, JSEG and Canny edge detection. (a) Segmentation result of the Mean Shift algorithm; (b) Segmentation result of the JSEG algorithm; (c) Canny Edge Detection result; (d) Segmentation result of the proposed approach; (e) extracted boundary contour of (a); (f) extracted boundary contour of (b); (g) Edges extracted of (c); (h) extracted boundary contour of our algorithm.