Adaptive active contour model based automatic tongue image segmentation

Jingwei Guo   Yikang Yang   Qingwei Wu   Jionglong Su   Fei Ma*

Mathematical Sciences

Xi'an Jiaotong-Liverpool University, Suzhou, China

*Telephone: (86) 512 8816 1633, Email: fei.ma@xjtlu.edu.cn

Abstract—For about 1800 years, tongue inspection has been one of the four major diagnostic methods in Traditional Chinese Medicine (TCM). The tongue is believed to be able to reflect the health status of the human body. However, making an accurate diagnose with the tongue is not a trivial task. It usually requires enormous training on the TCM doctor before he can make a reasonable diagnosis. Recently, image processing methods have been proposed to automatically process the tongue images and make diagnosis. This study proposes a k-means clustering and adaptive active contour model based automatic tongue region segmentation algorithm. This study is the first step towards the automatic tongue diagnosis. The method was applied on a set of real tongue images. To quantitatively evaluate the segmentation results, the automatically extracted boundaries were compared to the tongue boundaries drawn by experts. An average coverage ratio of 92% was found, indicating the accuracy of the proposed algorithm.

I. Introduction

Traditional Chinese medicine (TCM) has existed in China for several thousands years, and has been proved to be a scientific and effective treatment by study, research and practice. There are four major methods of diagnosis in TCM, including observation, auscultation, interrogation and palpation. Tongue inspection, as a method of observation diagnosis, has about 1800 years history. An ancient book named Treatise on Febrile and Miscellaneous Diseases, written in the 3rd century by Zhang Zhongjing [1], recorded the study of tongue coating. As an important organ of the human body, the tongue is believed to be like a mirror, it reflects the conditions of human body. By carefully inspecting the coating, the color as well as the tip of the tongue, a veteran TCM doctor can identify the physical conditions of a human body, diagnose the illness of a patient and determine the treatment. However, tongue diagnosis is not a trivial task, it requires enormous training before a TCM doctor can make a correct diagnosis. In the past, such kind of training greatly relies on the supervision of another experienced TCM doctor. The difficulty of tongue inspection has hampered the development and improvement of TCM [2].

With the development of computer technology, digital image processing technique has been applied to many areas, such as medical science, military, aerospace and so on [3]. Many attempts have been made in using image processing techniques to automatically process tongue images, and to automatically make diagnosis on the tongue images [4][5][6]. These researches could greatly promote the development of TCM.

As one of the key steps towards computer aided automatic tongue diagnosis, the tongue image segmentation has been explored in many studies. Zhuo et al. [7] proposed a content-based image retrieval (CBIR) method to analyze colors of tongue and coating. They applied their method on 288 images and reported classification accuracy of 87.85% for the color of tongue and 88.54% for the color of tongue coating. Bai et al. [8] also presented a tongue coating segmentation method based on threshold and splitting-merging methods. In their experiment on 49 images they found the accuracy of their segmentation method is 95.9%. Wang et al. analyzed the statistical distribution characteristics of human tongue color and used the tongue color to analyze patients' physical condition [4]. Fuzzy C-means Algorithm (FCMA) has also been applied to process the tongue image based on artificial intelligence. In the study of Ghaleh and Behrad [9], they use RGB color space and FCMA to isolate lip contour. In [10], a hue threshold control function was used to determine a threshold to obtain an initial tongue region. The obtained region was then enhanced to improve the luminance to facilitate the segmentation of tongue region in L*A*B color space. The paper claimed that over 70% images yield satisfactory results.

As a first step towards a long term target of automatic tongue diagnosis, this study focuses on segmentation of the tongue from the tongue image using active contour model. The active contour model, which is also known as the snake model, is an effective algorithm for image segmentation. It was put forward by Kass, Witkin and Terzopoulos in 1987 [11]. As its name suggests, its process looks like a snake keeping moving in an image in order to acquire the object. Under the action of internal energy, external energy and image forces [12], an active contour model usually requires many iterations to finally obtain the object. There have been a few studies on the tongue segmentation with active contour model. Shen et al. [6] found the accuracy rate of using an active contour model on tongue image segmentation is 82% after their experiment on 441 pictures. Jia et al. [13] combined the watershed transform and active contour model to extract the contour of tongue. They applied their method on a set of 678 images and claimed that 96.9% of them were successfully segmented. Zhai et al. [14] proposed a double snake model to improve the segmentation accuracy. In their experiment, they reported that the accuracy of the single-snake is 81.63% while the double snake is 92.89%. Pang [15] combined a bi-
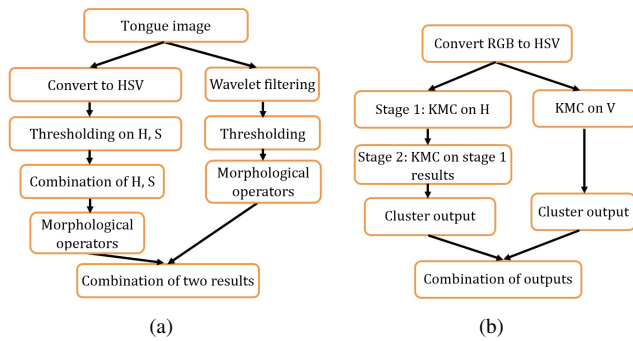
Fig. 1. Procedures for initial boundary extraction. (a) is the procedure to obtain an initial tongue boundary from the tongue image. (b) lists the steps of initial tongue boundary extraction based on k-means clustering.

elliptical deformable template (BEDT) and an active contour model, namely the bi-elliptical deformable contour (BEDC), and applied it to automated tongue segmentation via template force. They tested their method on a set of 3572 images, out of which 3405 image were reported as successfully extracted.

This paper proposes an adaptive active contour model based tongue image segmentation method. The paper is organized as follows. Section II describes the proposed method, including a procedure for initial boundary extraction in section II-A, a two stage k-means clustering method in section II-B, and the adaptive active contour model in section II-C. Section III describes the experiments of applying the proposed method on a set of tongue images and the obtained results. Finally discussion and conclusion are made in section IV.

II. METHODOLOGY

A. Initial boundary extraction

An initial boundary was firstly obtained using the procedure listed in figure 1(a). In the procedure, the tongue image was firstly converted to HSV color space. The Otsu's threshold method [16] was then used on H and S channel to convert them into binary images (step ' Thresholding on H, S'). The two binary images obtained on H and S channel were then combined (step 'Combination of H, S') by performing the logical "AND" on the two results. In the step of "morphological operators", image dilation, flood-fill operation and image erosion were applied to the binary image.

On the other hand, another binary image was obtained by "Wavelet filtering", "Thresholding", and "Morphological operators". For the "wavelet filtering", a 2-D image filter based on a class of rational orthogonal wavelets (ROWs)[17] is used to pre-filter the image. The same Otsu's threshold method was then used to convert the filtered image into a binary image. The same set of morphological operators was applied on the binary image.

In the final step, the logical "AND" was performed on the two binary images to create a final binary image containing only the tongue region, from which an initial boundary can be easily extracted.

B. Two stage k-means clustering

In most cases the initial boundary obtained with the previous procedure is close to the true boundary. In our experiments, however, the procedure picked the entirely wrong boundary in several cases. To correct the wrong results, a further procedure based on two stage k-means clustering (KMC) is developed.

Visually the tongue in a tongue image is a distinct region. The HSV color space is less correlated within its three channels Hue, Saturation and Value, comparing to RGB color space, and is capable in differentiating colors. In this study the two stage KMC method is applied on both H and V channels to form a better binary image for boundary extraction.

Figure 1(b) describes the procedure of this procedure. In the first stage, the KMC is applied on the H channel with 12 centers. In the second stage, the KMC is applied again on the set of cosine value of 12 cluster means with seven centers. The cluster with greatest mean in the second stage KMC was kept as the result. The clusters of the first KMC, belonging to the cluster of second KMC with greatest mean, were retained as output. From these two stage KMC, the tongue region can normally be isolated. However, the final binary image may also contain large blocks belonging to the background. To eliminate these noisy parts, the KMC was also applied to the V channel with four centers. The output is a binary image with cluster having largest cluster mean retained as 1. The final output of the procedure is the logic "AND" between two binary outputs.

The initial boundary obtained with the two stage KMC method is generally not as accurate as the boundary obtained with the previous procedure. However, in our experiment, on the cases where the first procedure extracted the wrong initial boundary, the two stage KMC worked well. The decision of using the results of two stage k-means clustering rather than the result of previous procedure is based on the criteria that the center of the image is not inside the region obtained with the previous procedure. This is based on the fact that the image acquiring facility ensures that the tongue is in the center of the image. In figure 2 the results after each step of two stage KMC are given.

C. Adaptive active contour model

Although the obtained initial tongue boundary generally contains the target region, visually the initial boundary does not fit the actual boundary of tongue very well. Figure 4(a) shows an example of an initial boundary obtained with procedures described in the previous sections. In order to improve the accuracy of the tongue boundary, in this study, an version of active contour model developed by Kass et al. [18] is employed to move the initial boundary to the true boundary of the tongue. The core principle of the active contour model is to minimize the total energy below.

$$E_{\text{whole}} = \alpha * E_{\text{internal}} + \beta * E_{\text{external}} \tag{1}$$

In this formula, α and β denote the contribution of the energy terms to the whole energy; E_{internal} and E_{extenal} represent the
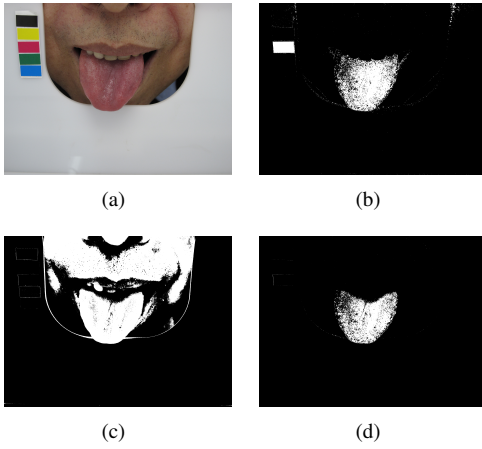
(a)　　　　(b)

(c)　　　　(d)

Fig. 2. Results from two stage KMC procedure. (a) is the original tongue image. (b) is the binary image resulted from applying two stage KMC on H channel. (c) is the result of KMC on V channel, and (d) is the final result after combining (b) and (c).

Fig. 3. The figure shows how the search area of a point p_i is defined. The p_{i-1} and p_{i+1} are the neighboring points of p_i. D is in the middle of p_{i-1} and p_i and C is in the middle of p_i and p_{i+1}. A and E are on the line perpendicular to CD and satisfy $DA = DE = Dp_i$. Similarly, BF is perpendicular to CD and satisfies $CB = CF = Cp_i$. The region ABCD is denoted by R_1 and the region CDEF is denoted by R_2. The whole search region is ABFE.

internal and external energy respectively which are defined as

$$E_{\text{internal}} = a_1 * V'^2 + a_2 * V''^2 \qquad (2)$$
$$E_{\text{external}} = b_1 * E_{\text{edge}} + b_2 * E_{\text{term}} + b_3 * E_{\text{line}} \qquad (3)$$

In (2) the term V' and V'' represent the first and second derivatives of the boundary, with a_1 and a_2 their weighting coefficients. In this study, empirically a_1 is set as 0.4 and a_2 is set as 0.2 for better controlling the continuity and smoothing of the contour.

In (3) the term E_{edge} is defined as

$$E_{\text{edge}} = -|\nabla I(x, y)|^2 \qquad (4)$$

where $I(x, y)$ is the gray scale of the image. This external energy term focuses on driving the contour to the place where the largest gray scale change happens, which normally is the edge of an image. The term E_{term} captures the curvature of the boundary and is defined as

$$E_{\text{term}} = \frac{I_{yy}I_x^2 - 2I_{xy}I_xI_y + I_{xx}I_y^2}{\sqrt{(I_x^2 + I_y^2)^3}} \qquad (5)$$

Furthermore the term E_{line} in (3) is simply defined as

$$E_{\text{line}} = I(x, y). \qquad (6)$$

By using the image intensity, this term aligns the boundary to the lightest or darkest nearby contour.

The coefficients b_1, b_2 and b_3 determine the contribution of E_{edge}, E_{term} and E_{line}. In our study, these values are empirically set as $b_1 = 1.75, b_2 = 0.27, b_3 = 0.22$.

Traditionally minimization of the energy defined in equation (1) is transformed into solving the equation below according to Lagrange's theorem [19].

$$\partial E_{\text{internal}} + \partial E_{\text{external}} = 0 \qquad (7)$$

In our implementation of active contour model, we apply the model on each point in the current boundary one by one. We choose a nearby point to replace the current point if it yields a lower energy. Let $P = \{p_1, p_2,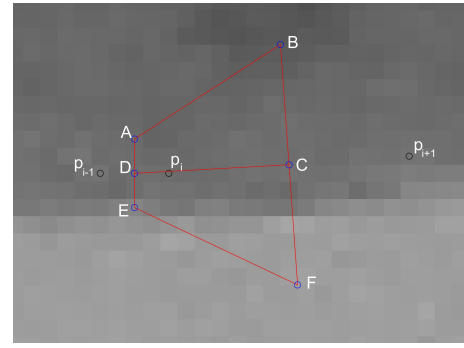 \ldots, p_N\}$ represents the set of points in the initial boundary. For each of these boundary point, a search area is first defined. The definition of the search area of a point p_i is shown in figure 3.

Rather than choosing a fixed weight for the internal and external energies, in this study, the weights α and β are adjusted automatically. For a point p_i in the boundary, let $\varphi_i = \{n_1, n_2, \ldots, n_i, \ldots, n_N, p_i\}$ denotes the set of points making up the search area of p_i. Let R_1^i and R_2^i be the two point sets consisting the search area φ_i (see figure 3). They satisfy $R_1^i \bigcup R_2^i \bigcup p_i = \varphi_i$ and $R_1^i \bigcap R_2^i = \emptyset$. To define the formula for α, a similarity measure of color μ is first defined based on the RGB value of pixels within R_1^i and R_2^i [20].

$$r_t = \frac{\sum_{n_j \in R_1^i} r(n_j)}{|R_1^i|} \bigg/ \frac{\sum_{n_j \in R_2^i} r(n_j)}{|R_2^i|}$$

$$g_t = \frac{\sum_{n_j \in R_1^i} g(n_j)}{|R_1^i|} \bigg/ \frac{\sum_{n_j \in R_2^i} g(n_j)}{|R_2^i|}$$

$$b_t = \frac{\sum_{n_j \in R_1^i} b(n_j)}{|R_1^i|} \bigg/ \frac{\sum_{n_j \in R_2^i} b(n_j)}{|R_2^i|}$$

Here $r(n_j), g(n_j), b(n_j)$ gives respectively the red, green and blue value of the original RGB image on n_j point. $|R|$ gives the number of points inside region R. The similarity measure μ and the weight of internal energy α are then defined as

$$\mu = \frac{3^2}{(r_t + g_t + b_t) * (\frac{1}{r_t} + \frac{1}{g_t} + \frac{1}{b_t})} \qquad (8)$$

$$\alpha = \exp(\mu) * \frac{a_1 * p_i'^2 + a_2 * p_i''^2}{\|a_1 * p_i'^2 + a_2 * p_i''^2\|} \qquad (9)$$

Here $\|x\| = 10^{\lfloor \log_{10}(x) \rfloor}$ gives the magnitude of x, where $\lfloor \cdot \rfloor$ rounds the value to the nearest integer less than or equal to that value. By dividing $\| \cdot \|$ a value is turned into a number of 1 to 10. μ gives the degree of similarity between two sub-area of the search area decided by p_i. It ranges from 0 to 1, with 0 being the lowest in similarity and 1 being the highest in similarity.

Furthermore the weight β is the weight of external energy which drives the points to the real boundary of tongue. In this study, β is automatically adjusted based on

$$\beta = \exp(1-\mu) * \frac{\max E^i_{\text{external}} - \min E^{\varphi_i}_{\text{external}}}{\text{mean}E^{\varphi_i}_{\text{external}}} \quad (10)$$

Here E^i_{external} is the external energy calculated on the point n_i, and $\max E^i_{\text{external}}$ is the maximum external energy ever calculated on this point. $\min E^{\varphi_i}_{\text{external}}$ and $\text{mean}E^{\varphi_i}_{\text{external}}$ are the minimum and mean value of external energy calculated on the points within the area φ_i.

During the searching process, the growth of α allows the internal energy to dominate the whole energy, straightening and shrinking the contours gradually. Similarly, when the value of β become large, the contour tends to get closer to the real boundary. Under the control of those two adaptive coefficients, the contour can gradually move to the real boundary of tongue.

To be able to compare the internal and external energies among all points within the search area, for each point both external and internal energies are normalized according to the following.

$$EE^j_{\text{internal}} = \frac{E^j_{\text{internal}}}{\|E^j_{\text{internal}}\|} \quad (11)$$

$$EE^j_{\text{external}} = \frac{E^j_{\text{external}}}{\|E^j_{\text{external}}\|} \quad (12)$$

Here $\|\cdot\|$ is the same as defined in (9).

Within each iteration, the whole energy is calculated for each point n_j of φ_i of point p_i

$$EE^j_{\text{whole}} = \alpha * EE^j_{\text{internal}} + \beta * EE^j_{external}. \quad (13)$$

The point within φ_i having the minimum energy is then selected to replace point p_i, to become the ith point of the boundary. In one iteration, every point will be updated. In our implementation, we fixed the iteration number to be 100. We noticed that for most images, the active contour model becomes stable after 40 to 50 iterations. Figure 4(b) to 4(d) gives the result of our active contour model after 30, 70 and 100 iterations.

III. EXPERIMENTS AND RESULTS

To evaluate our proposed method, in this study, 16 images randomly selected from a tongue image database of Medical School of Xiamen University were used. The images are in 3456x2592 pixels, 24 bit depth and JPG format, and were obtained using a Canon PowerShot G10 camera. To be able to quantitatively evaluate the proposed method, for each image, a tongue boundary was manually drawn by two experts using a public software ImageJ. The manually drawn boundaries were severed as "true boundaries", and the boundaries automatically extracted through our active contour model were then compared with the "true boundaries". For evaluation, a measure, κ, also called the area coverage ratio [21], was used to assess the efficiency of our adaptive active contour. It is defined by

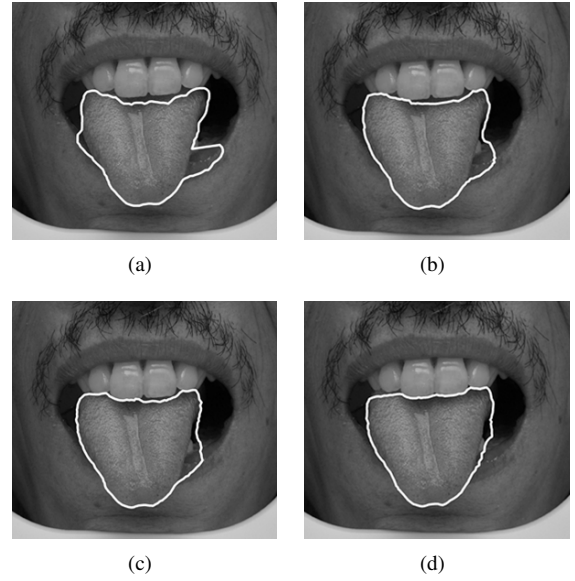$$\kappa = \frac{R_{\text{snk}} \bigcap R_{\text{man}}}{R_{\text{snk}} \bigcup R_{\text{man}}} \quad (14)$$

Fig. 4. Results of active contour model. (a) is the initial boundary of tongue. (b)(c) and (d) are the boundaries obtained after 30, 70 and 100 iterations respectively.
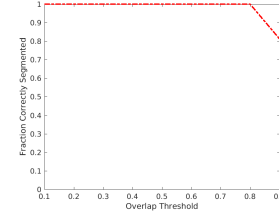

Fig. 5. The performance of the proposed method.

where R_{snk} is the region enclosed by the boundary obtained by the proposed adaptive active contour model, and R_{man} is the region enclosed by the manually drawn boundary.

In our experiment, the average κ over the whole image set is 92.2%. Figure 5 shows a plot of the fraction of correctly extracted tongue boundaries at various area coverage ratio levels for all images. Some results are shown in figure 6.
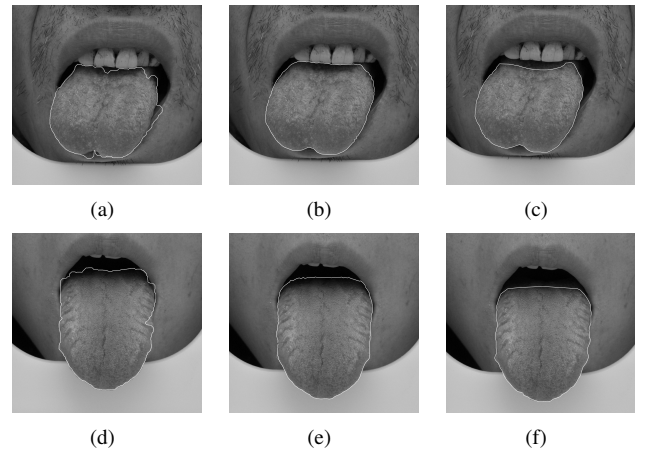

Fig. 6. Results of active contour model. (a) and (d) are the original tongue image with initial boundaries in white color. (b) and (e) are the results obtained by the active contour model. (c) and (f) are manually drawn boundaries imposed on the original images.
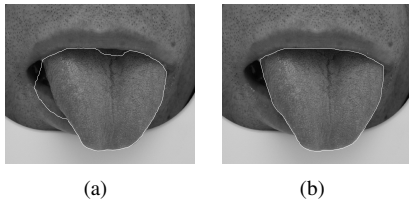
Fig. 7. Results of active contour model. (a) is the boundary obtained by our active contour model, (b) is the manually drawn boundaries imposed on the original images.

IV. Discussion and Conclusion

The initial boundaries, obtained by the combination of the simple procedure and the procedure based on two stage KMC, normally contains the tongue region. However, the boundaries are normally unsmooth, and generally do not meet the true boundaries well. This can be seen from the examples shown in figure 4 and 6. After our proposed adaptive active contour model, the boundaries can generally be improved. The average area coverage ratio of 92% indicates that the proposed method is capable in extracting the tongue boundary.

In many cases, the extracted initial boundary enclosed part of the lip. Both figure 4 and 7 are examples of this case. This happens especially when the color of the lip is very close to the color of the tongue. In some cases, the adaptive active contour model can finally pull the boundary to the true boundary, and hence corrected the error, like the case shown in figure 4. In some cases, the adaptive active contour model failed to do so, like the case shown in figure 7. Basically, when the initial boundary is close to the true boundary, we have found that the adaptive active contour model can normally drive the boundary to the correct place.

In this paper we proposed a tongue image segmentation method to automatically extract the contour of the tongue. In the method, the initial tongue boundary was firstly extracted based on a simple initial boundary extraction procedure and a two stage k-means clustering method. Based on the extracted initial boundary, an adaptive active contour model was used to obtain a more precise tongue boundary. The experiment of applying the proposed method on a set of tongue images, resulted in an average area coverage ratio of 92%, showing that the method can be used to isolate the tongue region from the tongue image. The experimental data set is relatively small. This is partly due to the restriction of the resource. However the objective of this paper was primarily to propose an original methodology for tongue boundary extraction, and the experiment has shown that the proposed method can work on the used images.

References

[1] L. Rong, "The exploration of development history of chinese medicine before 1949," *Studies in the history of natural science*, vol. 23, no. 3, pp. 257–273, 2004.

[2] G. Rui, W. Yiqin, Y. Jianjun, L. Fufeng, Y. Xiahai, and Zhaoxia X., "Objective study of tongue diagnosis of chinese medicine," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 29, no. 7, pp. 642–645, 2009.

[3] Zuo Fei, *Digital Image Processing: Principles and Practice (MATLAB)*, Beijing: Publishing House of Electronics Industry, 2014.

[4] X. Wang, B. Zhang, Z. Yang, H. Wang, and D. Zhang, "Statistical analysis of tongue images for feature extraction and diagnostics," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5336–5347, Dec 2013.

[5] B. Zhang, B. V. K. Vijaya Kumar, and D. Zhang, "Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 491–501, Feb 2014.

[6] S. Lansun, W. Aimin, W. Baoguo, W. Yonggang, and Z. Zhongxun, "Image analysis technology in the application of objectivity of tongue diagnosis," *Acta Electronica Sinica*, vol. 29, no. z1, pp. 1762–1765, 2001.

[7] L. Zhuo, P. Zhang, B. Cheng, X. Li, and J. Zhang, "Automatic tongue color analysis of traditional chinese medicine based on image retrieval," in *Control Automation Robotics Vision (ICARCV), 2014 13th International Conference on*, Dec 2014, pp. 637–641.

[8] L. Y. Bai, Y. Shi, J. Wu, Y. Zhang, W. Wong, Y. Wu, and J. Bai, "Automatic extraction of tongue coatings from digital images: A traditional chinese medicine diagnostic tool," *Tsinghua Science and Technology*, vol. 14, no. 2, pp. 170–175, April 2009.

[9] V. E. C. Ghaleh and A. Behrad, "Lip contour extraction using rgb color space and fuzzy c-means clustering," in *Cybernetic Intelligent Systems (CIS), 2010 IEEE 9th International Conference on*, Sept 2010, pp. 1–4.

[10] Li Chen, Dongyi Wang, Yiqin Liu, Xiaohang Gao, and Huiliang Shang, "A novel automatic tongue image segmentation algorithm: Color enhancement method based on l*a*b* color space," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*, Nov 2015, pp. 990–993.

[11] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1987.

[12] L. Peihua and Zhang Tianwen, "The summary of active contour models (snake)," *Journal of Software*, vol. 11, no. 6, pp. 751–757, 2000.

[13] W. Jia, Z. Yonghong, B. Jing, W. Weiliang, C. Yu, H. Yao, and L. Jinghua, "The picture segmentation of tongue contour on the basis of watershed transform and active contour models," *Tsinghua Science and Technology*, vol. 48, no. 6, pp. 32–38, 2008.

[14] X. Zhai, H. d. Lu, and L. Zhang, "Application of image segmentation technique in tongue diagnosis," in *Information Technology and Applications, 2009. IFITA '09. International Forum on*, May 2009, vol. 2, pp. 768–771.

[15] Bo Pang, D. Zhang, and Kuanquan Wang, "The bi-elliptical deformable contour and its application to automated tongue segmentation in chinese medicine," *IEEE Transactions on Medical Imaging*, vol. 24, no. 8, pp. 946–956, Aug 2005.

[16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[17] L. Yu and L. B. White, "Optimum receiver design for broadband doppler compensation in multipath/doppler channels with rational orthogonal wavelet signalling," *IEEE Trans. Sig. Proc.*, vol. 55, pp. 4091–4103, Aug 2007.

[18] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, "Snake: Active contour models," *Internaltional Journal of Computer Vision*, pp. 321–331, 1988.

[19] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. on Image Processing*, vol. 7, no. 3, pp. 359–369, Mar. 1998.

[20] YANG Kang-Ye and WU Chun-Xue, "Color image segmentation based on color similarity of rgb model," *Computer Systems and Applications*, vol. 22, no. 3, pp. 128–131, 2013.

[21] Fei Ma, Mariusz Bajger, and Murk J. Bottema, "Automatic mass segmentation based on adaptive pyramid and sublevel set analysis," *DICTA*, pp. 236–241, 2009.