Automatic Tongue Diagnosis Using a Smart Phone

Min-Chun Hu
National Cheng Kung University, Taiwan
anita_hu@mail.ncku.edu.tw

Guang-Yu Zheng
National Cheng Kung University, Taiwan
d84991066@mail.ncku.edu.tw

Yian-Ting Chen
kla@csie.ncku.edu.tw

ABSTRACT
An automatic tongue diagnosis framework is proposed to analyzing tongue images taken by smart phones. Different from conventional tongue diagnosis systems, our input tongue images are usually in low resolution and taken under unknown lighting conditions. Consequently, existing tongue diagnosis methods cannot be directly applied to give accurate results. We propose a lighting condition estimation method based on the SVM classifier to predict the color correction matrix according to color difference of images taken with and without flashlight. We also modify the state of the art work of fur and fissure detection and successfully improve the detection accuracy by taking hue information into consideration and adding a de-noising step.

General Terms
Algorithms, Measurement.

Keywords
Tongue diagnosis, tongue segmentation, color correction, lighting condition estimation

1. INTRODUCTION
Tongue diagnosis [1] is one of mostly widely used diagnostic methods among the four diagnostic processes [2] in traditional Chinese medicine (TCM). The usefulness of tongue diagnosis lies in its simplicity and immediacy: inspection of the tongue can instantly clarify one’s pathological problem so that people seeking health care can have their tongues routinely examined. However, the medical application traditional tongue diagnosis is limited by the fact that the clinical competence of tongue diagnosis is decided by the experience and knowledge of the practitioners. The diagnostic results based on the subjective analysis of the examiners may be unreliable and inconsistent. Therefore, it is important to have an objective and quantitative diagnostic process for tongue diagnosis.

To circumvent the subjective and qualitative problems of traditional tongue diagnosis, several computer aided tongue diagnosis systems have been proposed [3-6]. For example, Zhang et al. [3] proposed a system based on Bayesian networks that can identify five different diseases with an accuracy about 75%. Kanawong [6] proposed a hybrid image segmentation algorithm that integrates the region-based method with the boundary-based method for automatic tongue classification. However, to the best of our knowledge, these tongue diagnostic systems generally assume that the tongue images are taken from a well-controlled environment (e.g. with controlled light conditions, built-in color palettes for tongue color calibration, and fixed tongue position, etc) and used only by the TCM doctors [7].

Recently, the need for continuous monitoring of health conditions have attracted increasing attention in both academia and industry. Such an extensive monitoring capability has many direct benefits to the quality of patient care. Continuous monitoring reduces the time to react to sudden changes in patient condition. The large amount of collected data from each patient also allow doctors to predict future hazards of each patient more accurately, and to apply pre-emptive care accordingly. Given that mobile phone penetration in some countries is almost 100% and most people carry their phones with them everywhere, smart-phones provide a socially appropriate means of displaying timely information to the user and enable the physiological sensors to transmit measurements directly to health care providers. Over the past decade, smartphones have become prevalent, e.g., more than 50% of U.S. mobile phone users had a smartphone and 500 million smartphones were being sold worldwide in 2012. In addition, an increasing availability of smartphone built-in sensors (e.g., magnetometer, accelerometer, camera) enable a development of new sensor systems for measurements of a state of a phone user and his/her surrounding environment. These sensors can measure user’s psychophysiology and environmental conditions, together with associated mobile applications, they can be used as a tool for gathering quality data for medical research, or regular healthcare practice, as data can be gathered from the subjects unobtrusively for long periods of time, in a laboratory, as well as in a subject’s natural environments. In this paper, we discuss an architecture that allows a person to monitor his health by performing the tongue diagnosis on the smart phone. At the same time, his tongue images are stored in a cloud server for continuous analysis.

2. SYSTEM OVERVIEW
Figure 1 shows the overview of the proposed automatic tongue diagnosis system, which is composed of four main components, i.e. tongue photo taking guide, tongue image color correction, tongue region segmentation, and tongue image diagnosis. Conventional tongue devices require the user to pose his/her head at a specific position such that the device can easily capture tongue images under a constrained lighting environment. Instead of using a fixed tongue device, we propose to let users take tongue images with their own smart phones no matter where they are. Two main challenges arise in our application scenario: 1) Usually, photos taken by the back camera have better image quality than the ones taken by the front cameras of the smart phone. However, if the back camera is used to obtain photos with better quality, the user cannot see the current visual content captured by the camera and consequently has difficulty in locating the tongue within the proper sensing region. Therefore, we design a tongue photo taking guide to facilitate the tongue image capturing process and ensure that the size and position of the tongue region is appropriate for
further analyses. 2) Even though we take photos of the same object in the same view, images may look quite different in color due to various sensing devices and diverse lighting conditions. Hence, before making diagnosis on the tongue image, we have to correct the color information first. In this work, we propose a light estimation method based on analyzing the color of photos taken with/without flashlight. Knowing the estimated light information helps us to correct the image color based on the color correction matrix, which can be trained in advance by the method introduced in [8]. Considering the real-time and power consumption issues on the user side, only the tongue photo taking guide task is left to the smart phone processor, while other components are done in the cloud server. All tongue data, including tongue images and diagnosis reports, will be stored in the server database for future examination. We will detail each component in the following sections.

2.1 Tongue Photo Taking Guide

When the user faces the back camera, the system automatically captures one image every 0.2 second and detects if there are enough tongue color pixels in the image. The tongue color distribution is trained by the kernel density estimation technique [9], 20 tongue images of different people are captured under various lighting conditions, and the hue values of all pixels inside the tongue region are collected to form a tongue color distribution. For a given input image, each pixel is determined to be a tongue color pixel if its hue value has high probability in the pre-trained distribution. The mean position of all tongue color pixels is calculated and considered to give the user instructions of how to move the camera to capture a complete tongue with an appropriate size. The tongue photo taking guide will be terminated only when there are more than 50% tongue color pixels inside the current scene.

![Figure 1. System Flowchart.](image)

2.2 Tongue Image Color Correction

Tongue image color correction is an important issue in the field of automatic tongue diagnosis. Existing research literatures can be classified into four categories, i.e. methods based on simple image statistics, color temperature curve calibration, supervised learning, and double exposure theory [11]. The first kind of methods fail when the given scene contains a large object having the same color, which is exactly our case since the tongue is usually of similar color and occupies a large portion of the whole image. The other kinds of methods require a standard color checker to be a reference for color correction. However, ordinary smart phone users usually do not have a standard color checker. Hence, we proposed to estimate the current lighting condition \( L' \) without a color checker and apply the color correction matrix \( M' \) trained under the corresponding lighting condition \( L' \) to the captured tongue photo.

Lighting condition estimation is a challenging problem and we simplify it into finding the color temperature of the current light. We collect 150 tongue image pairs \( \{L_i, L'_i\} \) of different people taken under three kinds of lighting temperature, where \( L_i \) and \( L'_i \) means the image of the \( i \)th tongue is captured with and without flashlight under the lighting temperature of \( L_i \), respectively. We transform each tongue images from sRGB to CIE xyY color space and get the \((x, y)\) color values for each pixel. For each image, the mean coordinate \((x_{mean}, y_{mean})\) of all the pixels inside the tongue region is calculated and regarded as the color center of the image. We plot color centers of all images in the \((x, y)\) color coordinate. Figure 2 shows part of them to more clearly present our idea. Each cross and circle sign represents the mean coordinate for an image captured with and without flashlight, respectively. Moreover, different colors indicate that the images are taken in different \( L_i \).

![Figure 2. (x,y) color coordinates of tongue images captured under three different lighting conditions.](image)

Form Figure 2, we observe that under the same lighting condition \( L_i \), even though the tongue color varies along with different people, the image pairs \( \{L_i, L'_i\} \) have similar distance vector in terms of the mean coordinate \((x_{mean}, y_{mean})\). The arrows in Figure 2 indicate the mean distance vectors for each lighting condition \( L_i \). Therefore, we train SVM classifiers to predict the current lighting condition according to the distance vector of each tongue image pair. When the user uses the smart phone to capture the tongue image, our system will automatically capture two images (one with flashlight and one without flashlight) and estimate the lighting condition based on the trained SVM classifiers. Given a known lighting condition, we approximate the corresponding color correction matrix based on [8] and apply it to obtain the corrected tongue color image. Please refer to Section 3 for more details.

2.3 Tongue Region Segmentation

In previous subsection, we roughly extract tongue color pixels of the current image. To more precisely segment the tongue region, Otsu’s method [10] is applied to the hue values of all tongue color pixels in the image. The Otsu’s threshold can further classify pixels into two clusters, i.e. the real tongue pixels and the noisy
pixels (usually caused by poor lighting conditions that make the face skin color looks like the tongue color). Note that the cluster having smaller standard deviation in terms of pixel position is assigned to be the real tongue color cluster. Moreover, the hole filling and contour detection techniques are utilized to obtain the complete tongue region.

### 2.4 Tongue Diagnosis

In this work, we focus on measuring the color/ratio of fur and the quantity of fissures in the tongue region, which are important bases for tongue diagnosis. Tongue fur is a substance covering the tongue and usually appear in white, yellow, dark purple or black. Different color/ratio of tongue fur implies various symptoms of diseases. Hsu et al. [12] extracted fur pixels from the tongue image according to the saturation value. Unfortunately, their method only works on white fur detection. To more effectively distinguish fur and non-fur regions, we utilize both hue (H) and saturation (S) values in a given tongue region to construct the corresponding 2-D color distribution, i.e. $P(H, S)$. If the entropy of $P(H, S)$ is larger than $Th_2$, it implies the color diversity in the tongue region is large, and we apply the 2D Otsu’s method to find the proper thresholds ($Th_h, Th_s$) for separating the tongue pixels into two clusters. The separating process repeats in each cluster until the color entropy is small enough. Finally, the color distribution of each cluster is classified into one of the forementioned four fur colors or non-fur color by comparing with each fur color distribution model, which is pre-trained and stored on the server. Our system will present all fur color pixels and the ratio of fur pixels to the user.

A normal tongue is relatively flat across the length of the organ. A fissured tongue is marked by a deep groove or fissure in the middle or other areas of the tongue. People may have one or more fissures of varying sizes and depths. Fissures on the tongue can be intuitively detected by edge detection algorithms, such as Canny edge detection [13]. Hsu et al. [12] also proposed a statistic-based method which considers pixels with relatively dark intensity values among pixels in the same row/column as the fissure pixels. However, Hsu’s method results in noisy fissures around flat areas due to the difficulty in selecting proper stand deviation for thresholding. We further improve Hsu’s method by applying a 10x10 denoising window to the tongue region. Only pixels with intensity values smaller than the mean intensity inside the window will be kept. Example results of the proposed fur/fissure detection methods are shown in Table 1 and Table 2.

### 3. Experimental Results

To evaluate the proposed automatic tongue diagnosis framework, we develop our system on a Samsung smart phone (Galaxy S2 / GSsmart Rio R1) with android 4.0.4 OS and a PC server with Intel(R) Core(TM)2 CPU (1.86GHz, 32bit). To collect the training image pairs introduced in Section 2.2, 15 people are invited to take tongue photos under four kinds of lighting conditions, including the CIE standard illuminant D50, fluorescent illuminant, incandescent illuminant, and halogen illuminant. We use the tongue images taken under the standard illuminant D50 as the ground truth to evaluate whether the tongue color is properly corrected. Table 3 shows the color correction results. In each grid (i, j), the left side is the i-th original tongue image captured under fluorescent illuminant, and the right side is the corrected image using the color correction matrix obtained by the method introduced in [8] using the j-th tongue image, which is also taken under the fluorescent illuminant. Compared with the ground truth images, we can observe that as long as the Tongue, and Tongue, are taken under the same fluorescent illuminant, the color of Tongue, can be corrected by using the correction matrix obtained from Tongue. This phenomenon also appears in the incandescent illuminant and halogen illuminant. Therefore, we can store corresponding color correction matrices of common lighting conditions on the server, and the color of a new tongue image can be corrected after the lighting condition is estimated by the SVM classifier introduced in Section 2.2. Table 1 and Table 2 shows some example results of the proposed fur detection and fissure detection methods applying on the color-corrected tongue images.

| Table 1. Fur detection results applying different methods. |
| --- | --- | --- |
| **Input Tongue** | **Hsu [12]** | **Ours** |
| ![Input Tongue Image](image1) | ![Hsu Image](image2) | ![Ours Image](image3) |

| Table 2. Fissure detection results applying different methods. |
| --- | --- | --- |
| **Input Tongue** | **Canny Edge** | **Hsu [12]** | **Ours** |
| ![Input Tongue Image](image4) | ![Canny Edge Image](image5) | ![Hsu Image](image6) | ![Ours Image](image7) |

### 4. CONCLUSIONS

We proposed an automatic tongue diagnosis framework which can be applied to smart phones. Different from conventional tongue diagnosis systems, our input tongue images are usually in low resolution and taken under unknown lighting conditions. Consequently, existing tongue diagnosis methods cannot be directly applied to give accurate results. Hence, we propose a lighting condition estimation method based on the SVM classifier to predict the color correction matrix according to color difference of images taken with and without flashlight. Compared to the state of the art work of fur and fissure detection, we also successfully improve the detection accuracies by taking hue information into consideration and adding a denoising step, respectively. In the future, we will conduct experiments on more diverse lighting environment, such as different lighting distances and angles, to verify the robustness of the proposed system.
Table 3. Color correction results using different color correction matrix calculated by [8]. (The left side is the original image captured under fluorescent illuminant and the right side is the corrected image. The ground truth is the same tongue captured under the standard illuminant D50.)

<table>
<thead>
<tr>
<th>Fluorescent</th>
<th>Ground Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-M₁</td>
<td>F-M₂</td>
</tr>
<tr>
<td>F-M₃</td>
<td>F-M₄</td>
</tr>
<tr>
<td>F-M₅</td>
<td></td>
</tr>
</tbody>
</table>

5. REFERENCES