Tongue Recognition From Images

Ryszard S. Choraś

Institute of Telecommunications and Computer Science
UTP University of Sciences and Technology
85-796 Bydgoszcz, Poland
Email: choras@utp.edu.pl

***Abstract*—This paper proposes a method of personal identification based on tongue image. Tongue images have many advantage for personal identification and verification. In this paper a texture tongue features are extracted based on Gabor filters, and *WLD* (Weber Law Descriptors) transform. These features can be used in forensic applications and with other robust biometrics features can be combined in multi modal biometric system.**

Keywords*—*Tongue image, Gabor filters, Weber Law Descriptors.

I. Introduction

Tongue recognition is attracting a great deal of attention because of its usefulness in many applications. Traditional, tongue recognition are often classified into two groups:

- Tongue recognition and analysis for the patient disease diagnosis. Tongue recognition for diagnosis has played an important role in traditional Chinese medicine (*TCM*) and in this area most investigation has been focused on extraction of chromatic features [11] [16], shape and textural features [7] [8] [12].
- Tongue recognition for biometric personal identification.

Our work concerns the biometric applications of the tongue recognition and efficient feature extraction.
Biometrics human identifications uses automated methods of recognizing a person based on a physiological or behavioral characteristics [6]. A biometric system is a pattern recognition system that recognizes a person on the basis of a feature vector derived from a specific physiological or behavioral characteristics that the person possesses. Physiological Biometrics - also known as static biometrics - is based on the data derived from the measurement of a part of a person's anatomy [9].

Tongue image analysis have received much attention in image analysis and computer vision. Tongue texture has many advantages for human identification and verification [10] [18]. The identification of people can be based on the texture features. As a biometric identifier, tongue image has the following properties:

- Tongue images are unique to every person. Texture features of the tongue are distinctive to each person,
- Texture features of an individual tongue are stable and unchangeable during the life of a person,
- The human tongue is well protected in mouth and is difficult to forge.

Tongue recognition system is presented in Figure 1 and it involves five major modules: tongue image acquisition, preprocessing, tongue feature extraction, visual features and classification.

In this paper, a tongue image feature extraction method is proposed, which utilizes Gabor filters and local features such as *WLD*, because these features are robust against some types of geometric modifications. Weber local descriptors (*WLD*) is a simple but powerful local descriptor, which simulates the human visual perception.

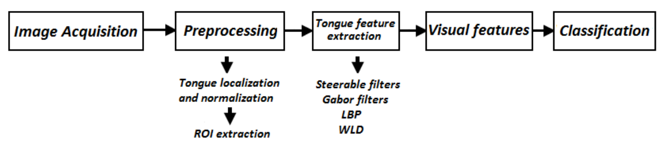

Fig. 1. Tongue recognition system

Images which are considered in this paper are displayed in Figure 2.

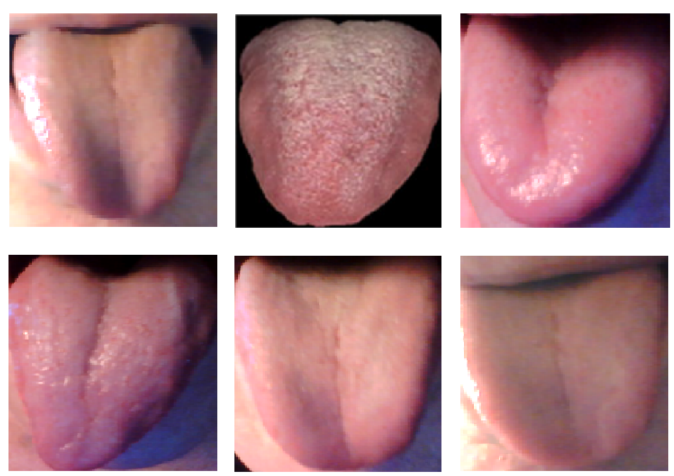

Fig. 2. Tongue images

The remainder of this paper is organized as follows. Section 2 briefly describes the preprocessing operations. In Section 3, we describe Gabor filters, propose a local descriptor called *WLD* and briefly describes the definition of Weber magnitude and orientation. Conclusions and references are given thereafter.

 ISBN: 978-1-61208-487-9

II. Preprocessing

Before performing feature extraction, the original tongue images are subjected to some image processing operations, such as:

1) Image stretching. The contrast level is stretched according to

$$f_{out}(x,y) = 255 \times \left(\frac{f_{in}(x,y) - f_{in_{min}}(x,y)}{f_{in_{max}}(x,y) - f_{in_{min}}(x,y)} \right)^{\gamma} \quad (1)$$

$f_{out}(x,y)$ is the color level for the output pixel (x,y) after the contrast stretching process. $f_{in}(x,y)$ is the color level input for data the pixel (x,y). $f_{in_{max}}(x,y)$ - is the maximum value for color level in the input image. $f_{in_{min}}(x,y)$ - is the minimum value for color level in the input image, γ - constant that defines the shape of the stretching curve.

2) Extraction of region of interest *(ROI)* from original tongue images. The tongue images are normalized with respect to position, orientation, scale, reflection, as follows.

The new invariant coordinates (x,y) of image pixels and the old coordinates (x',y') are related by

$$[x,y,1] = [x',y'1] \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -i_0 & -j_0 & 1 \end{bmatrix} \times$$
$$\times \begin{bmatrix} \frac{1}{\delta_x} & 0 & 0 \\ 0 & \frac{1}{\delta_y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where x_0, y_0 is the centroid of image; δ_x and δ_y represent standard deviation relative to variable x, y; and β is an angle between the major axis of an object and the vertical line

$$\tan 2\beta = \frac{2\sum_x \sum_y (x-x_0)(y-y_0)}{\sum_x (x-x_0^2) - \sum_y (y-y_0^2)} \quad (3)$$

Next, the *ROI*'s tongue blocks are automatically selected on the centroid of tongue normalized images. The size of whole *ROI* is $w_x \times w_y$ where $w_x = (x_0 + \frac{K}{2}) - (x_0 - \frac{K}{2})$, $w_y = (y_0 + \frac{K}{2}) - (y_0 - \frac{K}{2})$ where $K = 128$ pixels (Figure 3). Next, the *ROI* image is divided into the four sub-blocks. The size of sub-block is $\frac{K}{2} \times \frac{K}{2}$ pixels (Figure 4).

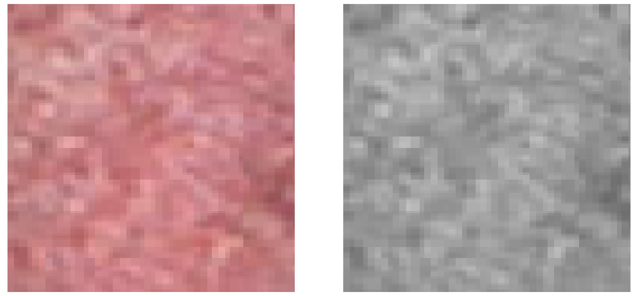

Fig. 3. Tongue *ROI*

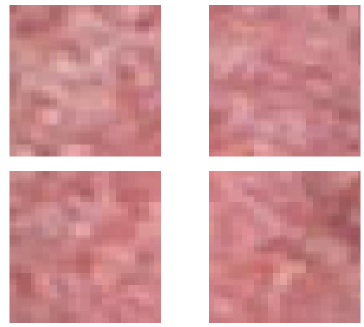

Fig. 4. Four *ROI*'s sub-blocks

III. Feature extraction

A. Gabor filters for feature extraction

Gabor filters are a powerful tool to extract texture features and in the spatial domain is a complex exponential modulated by a Gaussian function. In the most general the Gabor filters are defined as follows [3] [14] [15].

The two-dimensional Gabor filter is defined as

$$Gab(x,y,W,\theta,\sigma_x,\sigma_y) =$$
$$= \frac{1}{2\pi\sigma_x\sigma_y} e^{\left[-\frac{1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2 \right) + jW(x\cos\theta + y\sin\theta) \right]} \quad (4)$$

where $j = \sqrt{-1}$ and σ_x and σ_y are the scaling parameters of the filter, W is the radial frequency of the sinusoid and $\theta \in [0, \pi]$ specifies the orientation of the Gabor filters.

Figure 5 presents the real and imaginary parts of Gabor filters.

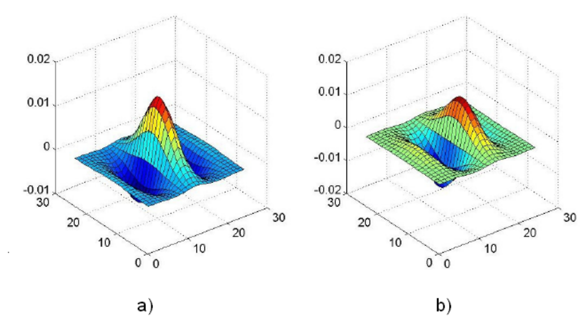

Fig. 5. The real and imaginary parts of Gabor filters

In our work we use a bank of filters built from the real part of Gabor expression called as even symmetric Gabor filter. Gabor filtered output of the image is obtained by the convolution of the image with Gabor even function for each of the orientation/spatial frequency (scale) orientation (Figure 6).

Given an image $F(x, y)$, we filter this image with $Gab(x, y, W, \theta, \sigma_x, \sigma_y)$

$$FGab(x, y, W, \theta, \sigma_x, \sigma_y) = \\ = \sum_k \sum_l F(x-k, y-l) * Gab(x, y, W, \theta, \sigma_x, \sigma_y) \qquad (5)$$

The magnitudes of the Gabor filters responses are represented by three moments

$$\mu(W, \theta, \sigma_x, \sigma_y) = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} FGab(x, y, W, \theta, \sigma_x, \sigma_y) \qquad (6)$$

$$std(W, \theta, \sigma_x, \sigma_y) = \\ = \sqrt{\sum_{x=1}^{X} \sum_{y=1}^{Y} ||FGab(x, y, W, \theta, \sigma_x, \sigma_y)| - \mu(W, \theta, \sigma_x, \sigma_y)|^2} \qquad (7)$$

$$Energy = \sum_{x=1}^{X} \sum_{y=1}^{Y} [FGab(x, y, W, \theta, \sigma_x, \sigma_y)]^2 \qquad (8)$$

By selecting different center frequencies and orientations, we can obtain a family of Gabor kernels, which can then be used to extract features from an image. The feature vector is constructed using *mean* - $\mu(W, \theta, \sigma_x, \sigma_y)$, *standard deviation* - $std(W, \theta, \sigma_x, \sigma_y)$ and *energy* as feature components (Table I).

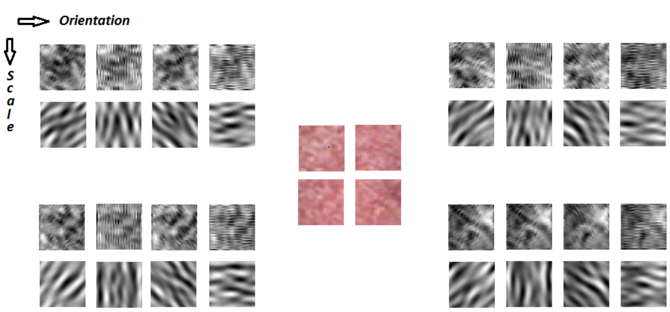

Fig. 6. Gabor images of tongue ROI's

To reduce dimension of feature vector [13], [17], we use the Principle Component Analysis (*PCA*) algorithm to keep the most useful Gabor features.

Let $X = [x_1, x_2, \ldots, x_n]$ denote an n-dimensional feature vector. The mean of the vector X and the total scatter covariance matrix of the vector X are defined as: $\overline{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $S_X = \sum_{i=1}^{n} (x_i - \overline{\mu}) \cdot (x_i - \overline{\mu})^t$.

The *PCA* projection matrix S can be obtained by eigenanalysis of the covariance matrix S_X. We compute the eigenvalues of $S_X : \lambda_1 > \lambda_2 > \cdots > \lambda_n$ and the eigenvectors of $S_X : s_1, s_2, \ldots, s_n$. Thus $S_X s_i = \lambda_i s_i, \;\; i = 1, 2, \ldots, m$. s_i is the ith largest eigenvector of S_X, $m \ll n$ and $S = [s_1, s_2, \ldots, s_m]$.

TABLE I. FEATURES OF TONGUE ROI'S.

Scale	Orientation	Energy	Mean	Std
		ROI 1		
2	45°	1968755.0	480.65308	35.064735
2	90°	1999251.1	488.09842	40.68314
2	135°	1968758.9	480.65402	34.23929
2	180°	1999252.2	488.0987	42.19372
8	45°	3.181323E7	7766.903	669.66876
8	90°	3.1827126E7	7770.2944	657.5264
8	135°	3.181321E7	7766.897	762.5107
8	180°	3.1827142E7	7770.2983	728.3237
		ROI 2		
Scale	*Orientation*	*Energy*	*Mean*	*Std*
2	45°	1968755.0	480.65308	35.064735
2	90°	1999251.1	488.09842	40.68314
2	135°	1968758.9	480.65402	34.23929
2	180°	1999252.2	488.0987	42.19372
8	45°	3.181323E7	7766.903	669.66876
8	90°	3.1827126E7	7770.2944	657.5264
8	135°	3.181321E7	7766.897	762.5107
8	180°	3.1827142E7	7770.2983	728.3237
		ROI 3		
Scale	*Orientation*	*Energy*	*Mean*	*Std*
2	45°	1867638.8	455.9665	36.981506
2	90°	1896568.8	463.02948	44.25503
2	135°	1867638.6	455.96646	36.304142
2	180°	1896567.2	463.0291	41.741203
8	45°	3.0179278E7	7367.988	707.88116
8	90°	3.0192456E7	7371.205	717.7458
8	135°	3.0179264E7	7367.984	774.104
8	180°	3.0192524E7	7371.221	614.612
		ROI 4		
Scale	*Orientation*	*Energy*	*Mean*	*Std*
2	45°	1906920.6	465.5568	27.945246
2	90°	1936458.5	472.7682	35.911766
2	135°	1906919.8	465.55658	28.149256
2	180°	1936459.8	472.7685	35.161488
8	45°	3.0813984E7	7522.9453	786.99506
8	90°	3.08275E7	7526.245	728.3674
8	135°	3.0814002E7	7522.9497	562.7714
8	180°	3.082751E7	7526.2476	521.4142

Any vector x can be written as a linear combination of the eigenvectors (S is symmetric, $s_1, s_2, \ldots, s_n$form a basis), i.e. $x = \sum_{i=1}^{n} b_i u_i$. For dimensionality reduction we choose only m largest eigen values, i.e. $x = \sum_{i=1}^{m} b_i u_i$. m is choose as follows: $\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} > t$ where t is threshold.

By removing the principal components that contribute little to the variance, we project the entire feature vector to a lower dimensional space, but retain most of the information.

B. Weber Law Descriptors

In 1834 Ernst Weber stated that "the ratio between the smallest perceptual change in a stimulus $\triangle f_{min}$ and the background level of the stimulus f is constant e.g. $\frac{\triangle f_{min}}{f} = k$" [6]. Inspired by Weber's Law, a robust and powerful Weber Local Descriptor (WLD) is a recently developed for local feature extraction. For each pixel of the input image, we compute two joint descriptors: a differential excitation DE operator and a gradient orientation GO descriptor. The DE is a function of the ratio between two terms: one is the relative intensity differences of a current pixel against its neighbors (e.g., 3×3 square region) and the other is the intensity of the current pixel. The orientation component is the GO of the current pixel. Figure 7 show how the DE and GO are calculated [1].

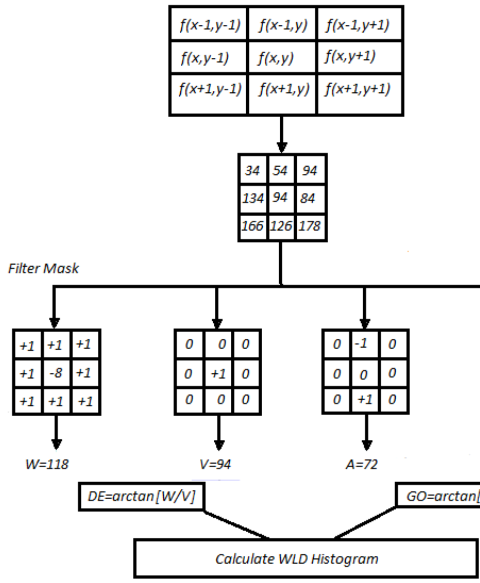

Fig. 7. Illustration of the WLD algorithm

If $f(x,y)$ is the center pixel of a 3×3 window, and $f(x+i, y+j)$; $i=-1,0,1$ $j=-1,0,1$ are the neighbors of the center pixel, DE is calculated as

$$DE = \arctan\left[\frac{\sum_{i=-1}^{1}\sum_{j=-1}^{1} f(x+i,y+j) - 8f(x,y)}{f(x,y)}\right] = \arctan\left[\frac{W}{V}\right] \tag{9}$$

where: $f(x+i,y+j)$ $i=-1,0,1;$ $j=-1,0,1$ is the gray level intensity of the corresponding pixel. The positive value of DE indicates that the current pixel is darker than the neighboring pixel,while the negative value represents the opposite.

The main purpose of the DE component is to extract the local salient patterns from the image.

The GO of the center pixel $f(x,y)$ is calculated as

$$GO = \arctan\left[\frac{f(i,j-1)-f(i,j+1)}{f(i+1,j)-f(i-1,j)}\right] = \arctan\left[\frac{A}{B}\right] \tag{10}$$

where the numerator is the intensity difference between the left and the right of $f(x,y)$, while the denominator is the intensity difference between the below and the above of $f(x,y)$.

Next, the GO are quantized into dominant orientations as follows:

$$GO' = \arctan 2\left[\frac{A}{B} + \pi\right] \tag{11}$$

$$\arctan 2\left[\frac{A}{B}\right] = \tag{12}$$
$$= \begin{cases} GO & A > 0 \;\; and \;\; B > 0 \\ \pi - GO & A > 0 \;\; and \;\; B < 0 \\ GO - \pi & A < 0 \;\; and \;\; B < 0 \\ -GO & A < 0 \;\; and \;\; B > 0 \end{cases}$$

The GO is then quantized into T dominant orientations. For each dominant orientation, histogram, H, is calculated using the DE (Figure 8) [1].

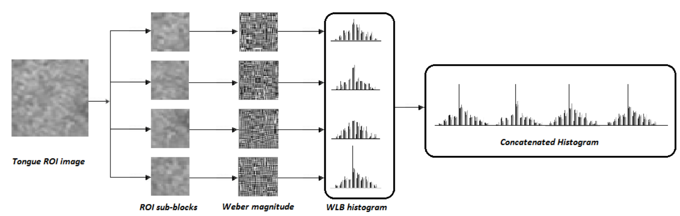

Fig. 8. Tongue feature extraction

Because not all the features are equally important, the feature selection technique is used.

To compute the distance of two histograms chi-square χ^2 distance is used

$$\chi^2(H^1,H^2) = \sum_i \frac{(h_i^1 - h_i^2)^2}{h_i^1 + h_i^2} \tag{13}$$

where H^1 and H^2 are two histograms and h_i^1 , h_i^2 are the ith bin of the histograms.

IV. Conclusion

In the paper, are presented some approaches for tongue recognition from images. To evaluate the performance of tongue recognition methods we use own tongue database that consists 30 images. We proposed a method which combines the recognition results of Gabor filters and *WLD* features to tongue recognition. The *WLD* texture features are robust against rotation and noise. The proposed system will be evaluated on other tongue databases in the future study.

Acknowledgment

The research was supported by the UTP University of Sciences and Technology by the Grant BS01/2014.

References

[1] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikinen, X. Chen, W. Gao, ”WLD: A robust local image descriptor”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), pp. 1705 - 1720, 2010.

[2] R. S. Choras, ”Iris-based person identification using Gabor wavelets and moments”, Proc. 2009 Int. Conf. on Biometrics and Kansei Engineering ICBAKE, CPS IEEE Computer Society, pp. 55-59, 2009.

[3] R. S. Choras, ”Lip-prints feature extraction and recognition”, Image Processing & Communications Challenges 3, AISC 102, pp.33-41 , Springer-Verlag Berlin Heidelberg, 2011.

[4] R. S. Choras, ”Vascular Biometry”, Image Processing & Communications Challenges 6, AISC 313, pp. 21 -28, Springer-Verlag Berlin Heidelberg 2015.

[5] R. S. Choras, ”Thermal face recognition”, Image Processing & Communications Challenges 7, AISC 389, pp. 37-46, Springer-Verlag Berlin Heidelberg 2016.

[6] A. K. Jain, Fundamentals of a Digital Signal Processing, Prentice-Hall, Englewood Clifts, NJ., 1989.

[7] W. Huang, Z. Yan, J. Xu and L. Zhang, ”Analysis of the tongue fur and tongue features by naive bayesian classifier”, Proceedings of the Int. Conf. Computer Application and System Modeling, Taiyuan, Oct. 2010, vol. 4, pp. 304-308.

[8] B. Huang, D. Zhang, Y. Li, H. Zhang and N. Li, ”Tongue coating image retrieval”, Proceedings of the 3rd Int. Conf. Advanced Computer Control, Harbin, Jan. 2011, pp. 292-296.

[9] A. K. Jain, R. Bolle and S. Pankanti, Biometrics: Personal Identification in Networked Society, Kulwer Academic, 1999.

[10] S. Lahmiri, ”Recognition of tongueprint textures for personal aAuthentication: A wavelet approach”, Journal of Advances in Information Technology, vol.3 ,no.3, pp. 168 - 175, 2012.

[11] C. Li and P. Yuen, ”Tongue image matching using color content”, Pattern Recogn. 35(2), pp. 407-419, 2002.

[12] W. Li, S. Hu, J. Yao and H. Song, ”The separation framework of tongue coating and proper in traditional Chinese medicine”, Proceedings of the 7th Int. Conf. Information, Communications and Signal Processing, Macau, Dec. 2009, pp. 1-4.

[13] C. J. Liu and H. Wechsler, ”Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition”, IEEE Transactions on Image Processing, vol. 11, no. 4, pp. 467- 476, 2002.

[14] D. H. Liu, K. M. Lam, and L. S. Shen, ”Optimal sampling of Gabor features for face recognition”, Pattern Recognition Letters, vol. 25, no. 2, pp. 267-276, 2004.

[15] L. Shen and L. Bai, ”Face recognition based on Gabor features using kernel methods”, Proc. 6th IEEE Conf. on Face and Gesture Recognition Korea, pp. 170-175, 2004.

[16] Y. Wang, J. Yang and Y. Zhou, ”Region partition and feature matching based color recognition of tongue image”, Pattern Recogn. Lett., 28(1), pp. 11-19, 2007.

[17] B.C. Zhang, S.G. Shan, X.L. Chen and W. Gao, ”Histogram of Gabor phase patterns(HGPP): a novel object representation approach for face recogniton”, IEEE Trans. Image Processing, 16, pp. 57-68, 2007.

[18] D. Zhang, Z. Liu, J. Yan and P. Shi, ”Tongue-print: A novel biometrics pattern”, ICB 2007, Springer-Verlag LNCS 4642, pp. 1174 - 1183, 2007.