EXPERIMENTAL STUDY ON LIP AND SMILE DETECTION

Aryuanto Soetedjo

Department of Electrical Engineering, National Institute of Technology (ITN) Malang, Jl. Bendungan Sigura-gura No. 2, East Java, 65145, Indonesia

E-mail: aryuanto@gmail.com

Abstract

This paper presents a lip and smile detection method based-on the normalized RGB chromaticity diagram. The method employs the popular Viola-Jones detection method to detect the face. To avoid the false positive, the eye detector is introduced in the detection stage. Only the face candidates with the detected eyes are considered as the face. Once the face is detected, the lip region is localized using the simple geometric rule. Further, the the red color thresholding based-on the normalized RGB chromaticity diagram is proposed to extract the lip. The projection technique is employed for detecting the smile state. From the experiment results, the proposed method achieves the lip detection rate of 97% and the smile detection rate of 94%.

Keywords: *face detection, lip detection, normalized RGB chromaticity diagram, smile detection*

Abstrak

Paper ini menyajikan medote pendeteksi bibir dan senyum berdasarkan diagram tingkat kromatis RGB ternormalisasi. Metode ini menggunakan metode Viola-Jones yang populer untuk mendeteksi wajah. Untuk menghindari kesalahan positif, detektor mata diperkenalkan pada tahapan deteksi. Hanya kandidat wajah dengan mata yang telah terdeteksi yang dianggap sebagai wajah. Setelah wajah dideteksi, bagian bibir ditempatkan dengan menggunakan aturan geometris sederhana. Selanjutnya, batasan warna merah berdasarkan pada diagram kromatisitas RGB ternormalisasi digunakan untuk mengekstrak bibir. Teknik proyeksi digunakan untuk mendeteksi keadaan tersenyum. Dari hasil percobaan, metode yang diusulkan mencapai 97% untuk tingkat deteksi bibir dan 94% untuk tingkat deteksi senyum.

Kata Kunci: *deteksi bibir, deteksi senyum, deteksi wajah, normalisasi RGB diagram tingkat kromatis*

1. Introduction

Recently, the applications that employ the image processing techniques increase significantly. Human computer interface is developed in such that the computer behaves as natural like the human. Thus, human facial features recognition becomes a topic that is paid attention by researches. One of the extensively researched areas is the lip detection, which plays an important role in recognizing human expressions and activities [1-8]. In [9-11], the lip detection is used to detect smile expression. While in [12][13] the lip detection is used by the audio-visual speech recognition system.

In [1], a spatial fuzzy clustering is proposed for color lip segmentation. In this method, both the distributions of data in feature space and the spatial interactions between neighboring pixels are considered in the clustering process. A hybrid method that integrating several distinct extraction techniques is proposed to detect mouth features [2]. They employ two stages, i.e. course mouth detection and fine mouth detection. In the course mouth detection stage, a color segmentation based on the HSV color space is employed to determine the location of lip in the facial image. In the fine mouth detection, the curve fitting of the mouth and mouth template matching are used to extract the mouth features: top of the upper lip, bottom of the lower lip, left and right mouth corners.

The rule-based lip detection technique based-on the normalized RGB color space is proposed by [3]. They defined a cresent area on r-g plane using a quadratic polynomial discriminant function for detecting the lip pixels. In [6], a mouth is localized using edge projection technique. It first detects the face region and extracts the intensity valleys to detect the iris candidates. Based on a pair of iris candidates, the projection method is adopted to detect the mouth. In [4], the edge based detection and the segmentation are applied separately. Then

the results of two methods are fused to detect the outer lip contour.

The popular Haar-like classifiers are employed in [7][11][13] to detect the lip region. To overcome the problem of false positive, the geometric relations checking between the detected face features is employed in [11]. The similar method is also adopted in [7].

The smile detection system is proposed in [9]. It first uses Adaboost algorithm to detect the face region in the first image frame and locate the standard facial features position. Then the optical flow is used to track the position of left and right mouth corners. A smile is detected if the distance between the tracked left and right mouth corners is larger than a threshold. In [11], three sets of feature vectors: the lip lengths, the lip angles, and the mean intensities of the cheek areas, are used by a linear discriminant function to detect the smile or non-smile face. In [2], the Artificial Neural Network is employed to classify the mouth movements as smile, neutral, or sad.

In this paper, researcher proposes a lip color segmentation based on the normalized rgb chromaticity diagram. The proposed lip color segmentation extracts the lip region from the face region detected using the Viola-Jones face detector. The eye detection is added to overcome the problem of false face detection. Further, the projection technique is employed to detect the inner area of the lip to determine the smile or non-smile face.

The rest of paper is organized as follows. Section 1 presents the fundamental theory used in this research. Section 2 discusses the proposed method. Section 3 presents the experimental results. Conclusion is covered in section 4.
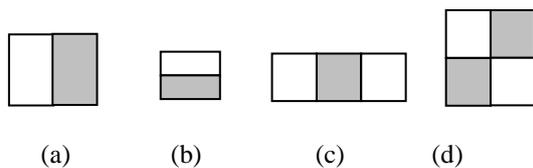

Figure 1. Three kinds of rectangle features.

Haar-like features use the change in contrast values between adjacent rectangular groups of pixels [7]. Figures 1(a) and 1(b) show the two-rectangle feature which is calculated as the difference of the sum of the pixels within two rectangular regions. Figure 1(c) shows three-rectangle feature which is calculated as the subtraction of the sum of the pixels within two outer rectangles from the sum of the pixels in the center rectangle. Figure 1(d) shows the four-

rectangle feature which is calculated as the difference between diagonal pairs of rectangles.
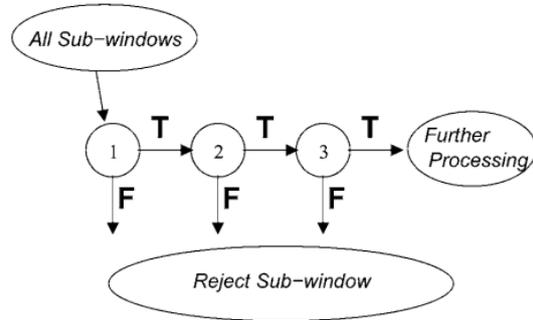

Figure 2. Schematic of cascade classifier [14].

Rectangle features which are described above could be computed rapidly if researcher use the integral image [14]. Integral image is an intermediate representation for the image which is expressed as:

$$int\ i(x,y) = \sum_{x'\le x, y'\le y} i(x',y') \qquad (1)$$

where $int\_i(x,y)$ is integral image at location x,y and $i(x,y)$ is the original image. From Eq. (1) researcher could say that integral image at location x,y is the sum of the pixels in the left and above of x,y. The integral image in equation (1) could be computed using the following formulas

$$row\ sum(x,y) = row\ sum(x,y-1) + i(x,y) \qquad (2)$$

$$int\ i(x,y) = int(x-1,y) + row\ sum(x,y) \quad (3)$$

where $row\_sum(x,y)$ is the cumulative row sum.

The Adaboost learning algorithm is used to select a small set of features and train the classifier [15]. Basically, the Adaboost is used to boost the performance of a weak/simple classifier. Hence the weak classifier is designed to select the single rectangle feature which best separates the positive and negative samples. Mathematically, a weak classifier $h_j(x)$ is defined as

$$h_j(x) = \begin{cases} 1 \ if \ p_j f_j(x) < p_j \theta_j \\ 0 \ otherwise \end{cases} \qquad (4)$$

where x is a 24×24 pixel sub-window, f_j is a feature, θ_j is a threshold and p_j is a parity that defines whether x should be classified as positive sample or negative sample.

The cascade scheme is constructed to increase detection performance while reducing computation time [14]. The scheme is to use simpler classifiers which reject the majority of

sub-windows before more complex classifiers to achieve low false positive rates. Figure 2 shows the schematic of cascade classifier. A positive result from the first classifier triggers the evaluation of a second classifier. A positive result from the second classifier triggers a third classifier, and so on. A negative result at any point leads to the immediate rejection of the sub-window. The cascade scheme attempts to reject as many negatives as possible at the earliest stage possible.
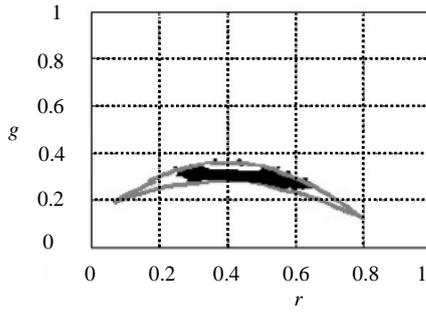

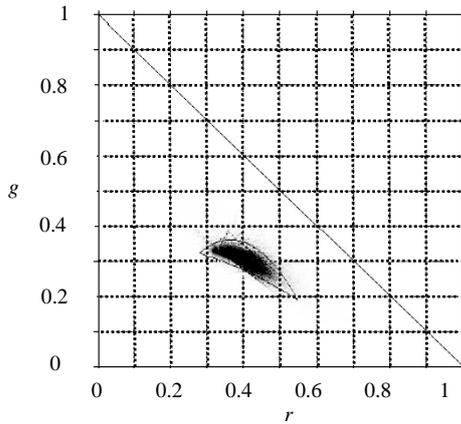Figure 3. Skin locus proposed by [3][17].

Figure 4. Skin locus proposed by [18].

In the normalized RGB color space, the distribution of human skin is a shell-shaped area, which is called as skin locus [16]. Figures 3, 4, and 5 show the skin color locu proposed by [3][17], [18], [19] respectively. In figure 3, the upper bound quadratic function of skin locus is $g = -1.3767r2 + 1.0743r + 0.1452$, while the lower bond quadratic function is $g = -0.776r2 + 0.5601r + 0.1766$. In figure 4, the upper bound quadratic function is $g = -5.05r2 + 3.71r - 0.32$, while the lower bound quadratic function is $g = -0.65r2 + 0.05r + 0.36$. In figure 5, the boundary of skin locus is expressed as

$$\text{line-G} : g = r \tag{5}$$

$$\text{line-R} : g = r - 0.4 \tag{6}$$

$$\text{line-B} : g = -r + 0.6 \tag{7}$$

$$\text{line-up} : g = 0.4 \tag{8}$$

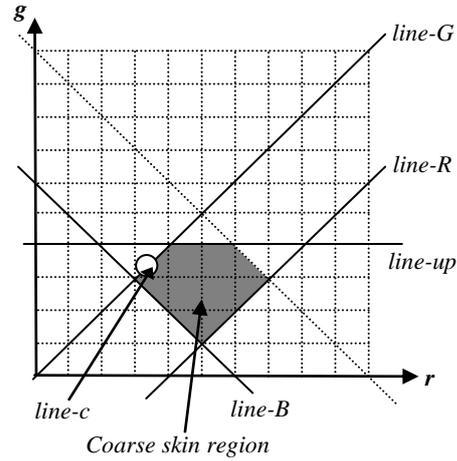$$\text{line-c} : (g - 0.33)2 + (r - 0.33)2 = 0.0004 \tag{9}$$

Figure 5. Skin locus proposed by [19].

2. Methodology

Figure 6 shows the overview of proposed system. It starts with the Viola-Jones face detector [14] to detect face in the image. The Viola-Jones face detector has three important aspects: a) image feature called integral image; b) learning algorithm based-on Adaboost; c) combining classifiers using a cascade scheme. The Viola-Jones face detector is very fast method to detect face with high detection rate. However, the algorithm has the drawback, i.e. it produces the false positive.To reduce the false positive, the eye detector is introduced in the second stage. Researcher apply the eye detector to each face candidate obtained by the face detector. If the eye is detected, then the face candidate is considered as the face, otherwise researcher reject the face candidate. It is noted here that even the Viola-Jones approach is applied to detect the eye, since the image region is limited to the face candidate's region only, the false eye detection (false positive) is avoided.

Once face is detected, the simple geometric rule is adopted to localize lip region. The lip region is localized by considering the lower part of the face region defined empirically as follows:

$$Lip\_left = Face\_left + (Face\_width/4) \tag{10}$$

$Lip\_top=Face\_top+(2*(Face\_height)/3)$ (11)

$Lip\_width=Face\_width/2$ (12)

$Lip\_height=Face\_height/3$ (13)

where

Lip_left: *x*-coordinate of the left border of lip region
Lip_top: *y*-coordinate of the top border of lip region
Lip_width: the width of lip region
Lip_height: the height of lip region
Face_left: *x*-coordinate of the left border of face region
Face_top: *y*-coordinate of the top border of face region
Face_width: the width of face region
Face_height: the height of face region.

After lip region is localized, the next step is the lip detection using color segmentation based on the normalized RGB chromaticity diagram as discussed in the following section. Further, the smile detection based on the openness of the mouth/lip is applied to the lip area to detect smile or non smile expressions.

Similar to the skin color locus, a lip detection method based on the *r - g* color space is proposed by [3]. They defined the two discriminant functions for extracting lip pixels as

$$l(r) = -0.766r^2 + 0.5601r + 0.2123 \quad (14)$$
$$f_{lower}(r) = -0.776r^2 + 0.5601r + 0.1766$$

then the lip pixels are detected by the following rule

$$L = \begin{cases} 1 \; if \; f_{lower}(r) \leq g \leq l(r) \\ and \; R \geq 20 \; and \; G \geq 20 \; and \; B \geq 20 \\ 0 \; otherwise \end{cases} (15)$$

The lip region described above is defined empirically, therefore in some cases it might fail to detect the lip as observed from our experiments.

To overcome the drawback, researcher propose a new lip detection method which is the extension of our previous research [19,20]. In [20], researcher utilize the normalized RGB chromaticity diagram for extracting the red color of traffic signs. Since the colors of lips range from dark red to purple [3], researcher might extract the lip color using the similar way.
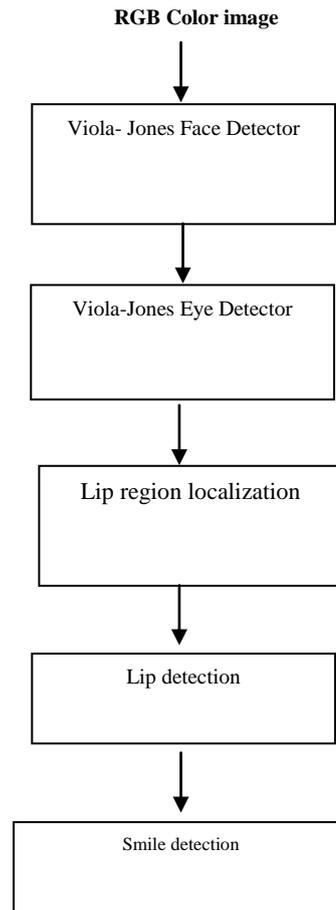
RGB Color image

Viola- Jones Face Detector

Viola-Jones Eye Detector

Lip region localization

Lip detection

Smile detection

Figure 6. Overview of proposed system.

In the method, to extract the lip pixels researcher employ the following rule:

If g-r < TR, then assign pixel as lip (16)

where *TR* is a threshold. Rather than defining the threshold *TR* fixedly, researcher find the *TR* automatically by developing a new transformed image called *Igr* using the following equations:

$$Igr = \big(100 + (100 \times g) - (100 \times r)\big)/200 (17)$$

using this transformed image, the optimal threshold *TR* could be found by employing the Otsu's thresholding method [21].

From the previous lip detection stage, a blob image (black and white) is obtained as shown in figures 7(b) and (d). In this work, researcher assume that a smile is detected when the mouth is open, i.e. appearing the teeth as shown in figure 7(c). Therefore our smile detection method searches the black area inside the lip region by projecting the blob image into *y*-coordinate.
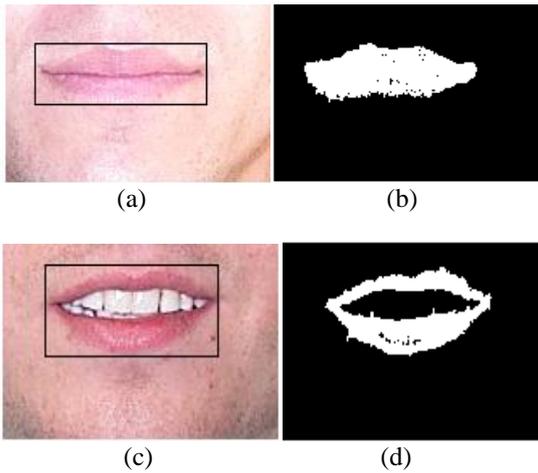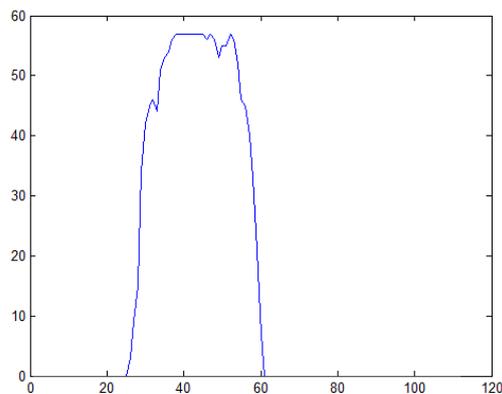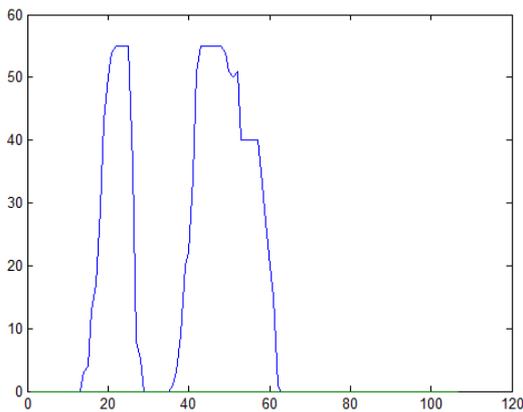
Figure 7. (a), (c): Bounding box of the detected lip; (b), (d): Blob image.

(a)

(b)

Figure 8. (a) *y*-projection of Fig. 7(b); (b) *y*-projection of Fig. 7(d).

Figures 8(a) and (b) show the *y*-projections of the blob image in figures 7(b) and (d) respectively. From the figures, it is clear that the opened mouth (smile state) could be detected if there are two peaks in the *y*-projection, otherwise the mouth is close (non smile state).

3. Result and Analysis

Our algorithm is implemented using MATLAB. Researcher tested the algorithm using one hundred (100) face images from Caltech Database [22]. The database contains front face color images under different lighting, expressions, and backgrounds. Each face image only contains one face. There are thirty six (36) smile face images.

Researcher make comparison to other methods, and conduct four different schemes as follows: a) M1: proposed method, b) M2: proposed method with Viola-Jones mouth detector to localize lip region, c) M3: lip detection method proposed by [3], d) M4: lip detection method proposed by [3] with Viola-Jones mouth detector to localize lip region. In M1 and M3, to localize the lip region, researcher uses our proposed method as defined by Eqs. (10)-(13).

TABLE I
DETECTION RESULTS

	M1	M2	M3	M4
Lip detection rate (%)	97	92	65	67
Smile detection rate (%)	94	78	17	17

Table I shows the detection results. From the table, researcher obtain that our proposed method shows the best detection rate, both the lip detection rate and the smile detection rate. From the table, researcher could see that the detection rate of M2 is lower than M1. It is caused by the fact that some of the lip regions obtained by Viola-Jones mouth detector are too small (not covering all the mouth/lip regions). Thus the smile could not be detected properly. In M3 and M4, the lip segmentation method proposed by [3] could not extract the white color of the teeth from the lip color properly. Therefore the smile detection rates of those methods are very low.

Figure 9 show some of the detection results, where images in the first column are the original images, images in the second column are the localized lip regions and the bounding boxes of the detected lips, images in the third column are the blob images representing the extracted lips.

4. Conclusion

The lip and smile detection method based on the normalized RGB chromaticity diagram is presented. The method works efficiently in extracting the lip from the face image.
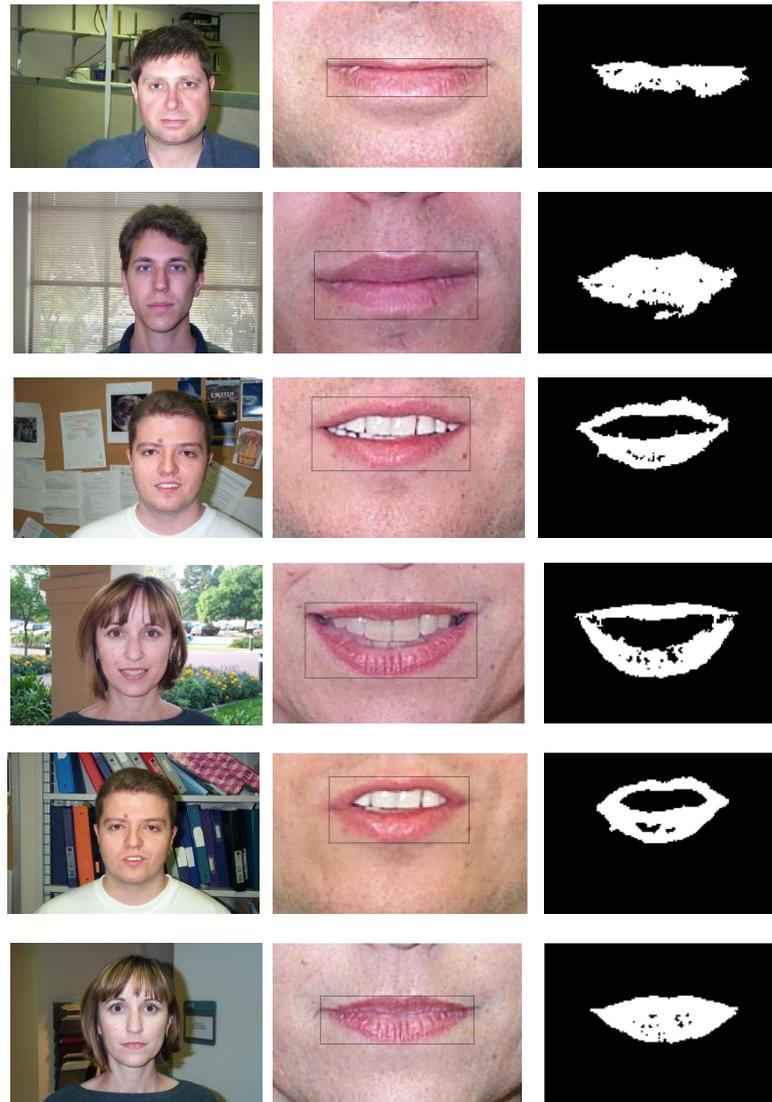
Figure 9. Some of the detection results.

Further, a simple projection technique is employed to detect the smile or non-smile state by considering the openness of the mouth. In future researcher will extend our approach to detect the smile from other features, not only openness mouth's features, and also to recognize the other human expressions. Further, the real-time implementation will be conducted.

Reference

[1] A.W.C. Liew, S.H. Leung,& W.H. Lau, "Segmentation of Color Lip Images by Spatial Fuzzy Clustering," _IEEE Transactions on Fuzzy Systems_, vol. 11, pp. 542-549, 2003.

[2] M. Pantic, M. Tomc, &L.J.M. Rothkrantz, "A Hybrid Approach to Mouth Features Detection"_InProceeding of the 2001 IEEE Systems, Man, and Cybernetics Conference_, pp. 1188-1193, 2001.

[3] C.C. Chiang, W.K. Tai, M.T. Yang, Y.T. Huang, &C.J. Huang, "A Novel Method for Detecting Lips, Eyes and Faces in Real Time," _Real-Time Imaging_, vol.9, pp. 277-287, 2003.

[4] U. Saeed & J.L. Dugelay, "Combining Edge Detection and Region Segmentation for Lip Contour Extraction"_InProcs. of 6th Conference on Articulated Motion and Deformable Objects_, pp. 11-20, 2010.

[5] T. Gernoth, R. Kricke, & R.R. Grigat, "Mouth Localization for Appearance-based Lip Motion Analysis," _WSEAS Transactions on Signal Processing_, vol. 3, pp. 275-281, 2007.

[6] M. Rizon, "Automatic Mouth Localization Using Edge Projection," _Journal of_

Computer Science, vol. 6, pp. 679-683, 2010.

[7] P.I. Wilson & J. Fernandez, "Facial Feature Detection Using Haar Classifiers," *Journal of Computing Science in Colleges*, vol. 21, pp.127-133, 2006.

[8] P. Dalka& A. Czyzewski, "Human-Computer Interface Based-on Visual Lip Movement and Gesture Recognition," *International Journal of Computer Science and Applications*, vol. 7, pp. 124-139, 2010.

[9] Y.H. Huang& C.S. Fuh, "Face Detection and Smile Detection"*InProceeding of IPPR Conference on Computer Vision, Graphics and Image Processing*, pp. 108, 2009.

[10] A. Ito, X. Wang, M. Suzuki, &S. Makino, "Smile and Laughter Recognition using Speech Processing and Face Recognition from Conversation Video"*InProceeding of the 2005 International Conference on Cyberworlds*, pp. 437-444, 2005.

[11] Y. Jang&W. Woo, "Adaptive Lip Feature Point Detection Algorithm for Real-Time Computer Vision-Based Smile Training System"*In Proceedingof the 4th International Conference on E-Learning and Games: Learning by Playing. Game-based Education System Design and Development*, pp. 379-389, 2009.

[12] S. Werda, W. Mahdi,& A.B. Hamadou, "Lip Localization and Viseme Classification for Visual Speech Recognition," *International Journal of Computing and Information Sciences*, vol. 5, pp. 62-75, 2007.

[13] R. Navarathna, P. Lucey, D. Dean, C. Fookes, &S. Sridharan, "Lip Detection for Audio-Visual Speech Recognition in Car Environment" *InProceedingsof 10th International Conference on Information science, Signal Processing and their Applications*, pp. 598-601, 2010.

[14] P. Viola& M. Jones, "Robust Real-Time Face Detection,"*International Journal of Computer Vision*, vol. 57, pp. 137-154, 2004.

[15] Y. Freund& R. Schapire, "A Decision-theoritic Generalization of On-line Learning and An Application to Boosting,"*Journal of Computer System Science*,vol. 55, pp. 119-139, 1997.

[16] M. Storring, H. Andersen,& E. Granum, "Skin Colour Detection Under Changing Lighting Condition"*InProceeding of 7th Symposium on Intelligent Robotics Systems*, pp. 187-195, 1999.

[17] M. Soriano, B. Martinkauppi, S. Huovinen,& M. Laaksonen, "Using The Skin Locus To Cope With Changing Illumination Conditions in Color-based Face Tracking"*InProceeding of IEEE Nordic Signal Processing Symposium*, pp. 383-386, 2000.

[18] J. Fritsch, S. Lang, M. Kleinehagenbrock, G.A. Fink, &G. Sagerer, "Improving Adaptive Skin Color Segmentation by Incorporating Results From Face Detection"*InProceeding of IEEE International Workshop on Robot and Human Interactive Communication*, pp. 337-343, 2002.

[19] A. Soetedjo& K. Yamada, "Skin Color Segmentation Using Coarse-to-Fine Region on Normalized RGB Chromaticity Diagram for Face Detection," *IEICE Trans. On Information and Systems*, vol. E91-D, pp. 2493-2502, 2008.

[20] A. Soetedjo& K. Yamada, "A new approach on red color thresholding for traffic sign recognition system," *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 19, pp. 457-465, 2007.

[21] N. Otsu, "A thresholding selection method from gray-level histogram," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62-66, 1979.

[22] California Institute of Technology, http://www.vision.caltech.edu/Image_Datasets/faces/faces.tar, 2003, retrieved March 3, 2011.