An Efficient Mouth Detection Based on Face Localization and Edge Projection

Hui-Yu Huang and Yan-Ching Lin

Abstract—**In this paper, we propose a fast mouth detection scheme. For mouth detection, there are many related techniques which could not exactly locate the correct mouth position, especially profile face or light influence. In order to solve this problem, we propose an approach to address this problem based on the effective skin-color segmentation filter and edge projection to detect the mouth feature. The main procedures of this approach consist of face extraction, mouth feature extraction, and mouth detection. For face extraction, we use the Viola and Jones detector to extract the face region. In order to more effectively and correctly detect the mouth position, we adopt two phases which are the correct facial features extraction and the projection computation for the specified area in skin region to achieve the purpose. The results show that our approach can obtain a high accuracy and a fast executed time with face extraction and mouth feature localization.**

Index Terms—**Adaboost algorithm, skin color transformation, Canny detector, projection.**

I. INTRODUCTION

As the rapid development of science and technology, it makes human life more functional and convenient. In recent years, there are many researches which focus on human attributes, such as face recognition, or face detection, etc. As for applications, one of important schemes is eyes detection which can avoid or alarm a fatigue situation for driving a car. In addition, for mouth detection, it can provide a lip-read recognition for help the blind persons. Based on the interesting scheme, in this paper, we will focus on locate the correct mouth position based on our proposed method. However, based on this topic, firstly, the face region in an image must be clearly labeled out. For face detection, there are many approaches which have been presented [1]-[4].

Zhang *et. al.* [1] presented an efficient method of face detection based on skin-color segmentation and principal components analysis (PCA). It consisted of three procedures. The first is segmenting image which using GLHS color model to filter out non-face on candidates roughly; Secondly, Eye-analogue segments at a suitable scale are discovered by finding regions which their neighborhoods to filter non-face father; Finally, PCA method is used to extract the relevant information in human faces. Yang *et. al.* [2] put forward an adaptive face detection algorithm. Firstly, face is chosen by the adaptive skin color model and eyebrow and eyes are

located by integral projection. Then, the coordinates of eye center are confirmed by the method which makes use of mixed round and elliptic based on Hough transform. Finally, face is detected by the human eye center mapping.

Viola and Jones [3], [4] presented a robustness algorithm for face detection. It is combined of Haar-like feature, Adaboost algorithm and Cascade method to filter and find correct face region. This system is robust and efficient.

For face detection, one interesting topic is mouth detection. Hsu, *et. al.*, [5] proposed a mouth method by color transformation. The color of mouth region contains more red component and less blue component than other facial regions. Hence, authors used the *Cb* and *Cr* components in *YCbCr* space to find mouth region. He and Zhang [6] presented a new real time lip detection based on the discrepancy of skin and lip on chrominance and R/G.

In this paper, we propose a mouth method based on face detection, skin color segmentation, canny detector and projection. The skin color segmentation is to find face mask which purposes to filter out non-skin color regions and raises accuracy before mouth localization process.

The rest of this paper is organized as follows: Section II presents the proposed method. Experimental results and performance evaluation are presented in Sections III. Finally, Section IV concludes this paper.

II. PROPOSED APPROACH

In the section, we will describe our approach. This approach concludes preprocess, segment the face region, edge detection, and projection. Fig. 1 shows the system flowchart. Details of procedures are described in the following subsections.
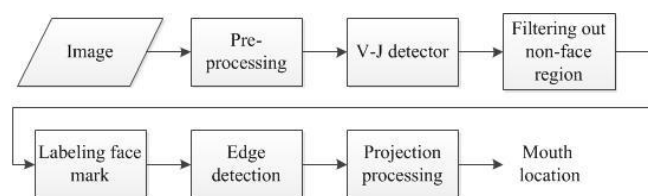

Fig. 1. System flowchart.

A. Preprocessing

In order to decrease the computational complexity and fast obtain the face features, we take the normalization process for all test data. Here, we use a bi-cubic interpolation to perform the normal size because the bi-cubic interpolation has better image quality than other interpolation methods. The normal size is defined as 500×350, 350×500, and 500×500.

B. Viola and Jones Detector

For face detection, one of many researchers, Viola and Jones (V-J) method [3], is usually adopted to work this field that possesses a high efficiency and accuracy to locate the face region in an image. The V-J method consists of three phases. Firstly, the Haar-like features is rectangular type that is obtained by integral image; Secondly, The Adaboost algorithm is a learning process that is a weak classification and then uses the weight value to learn and construct as a strong classification. Finally, we can obtain the non-face region and face region after cascading each of strong classifiers. Details of V-J detector can study Ref. [4]. Here, we don't further describe it.

C. Filtering Out Non-Faces

Although the face detection by used V-J detector has a higher performance, the threshold may affect the location result. Hence, we employ a ratio of skin region filtered candidates to further decide which the correct face region is, so that it can improve the face location accuracy after V-J detector. In other words, the better location of skin-color region implies a better outstanding feature for face.

Owing to the skin characteristic, we will take the different color space and threshold to decide the skin region. As previously researched, many color models, *YCbCr* or *HSV* or *HIS*, provide the advantage information in skin-color detection, in this paper, we use *YCbCr* color space to detect the skin-color area. The YCbCr function is defined as follow:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & -0.587 & 0.114 \\ -0.168 & -0.331 & 0.5 \\ 0.5 & -0.418 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

where *Y*, *Cb*, and *Cr* denote Luma, blue, and red chrominance components, respectively. *R*, *G*, and *B* present the color pixel values.

Because *Y* is Luma influenced by light factor, we ignore *Y* value and adopt the chrominance components *Cb* and *Cr* to decide the skin range. Using the adaptive threshold assignment to decide the skin region, the related works have been published [7] [8].

Chai and Ngan [7] defined this skin condition as follows.

$$\text{Skin} = \begin{cases} 1, \text{if} \begin{cases} 77 < Cb < 127 \\ 133 < Cr < 173 \end{cases} \\ 0, \text{otherwise} \end{cases} \quad (2)$$

Wu [8] expressed as:

$$\text{Skin} = \begin{cases} 1, \text{if} \begin{cases} 60 \le Y \le 255 \\ 100 \le Cb \le 125 \\ 135 \le Cr \le 170 \end{cases} \\ 0, \text{otherwise} \end{cases} \quad (3)$$

According to (2) and (3), the skin range for *Cb* and *Cr* components can be obtained. Because this range is too strict, it will lose some advantage information. Based on this reason, we will modify this range to fit the skin area more feasibly. Here, the range is defined as follows.

$$\text{Skin} = \begin{cases} 1, \text{if} \begin{cases} 90 \le Cb \le 124 \\ 136 \le Cr \le 180 \end{cases} \\ 0, \text{otherwise} \end{cases} \quad (4)$$

After skin segmentation, we will calculate the ratio of number of white pixels (skin color) and number of total pixels. Next, we use threshold to filter out non-face region and to reduce mistake. The results are presented in Fig. 2.
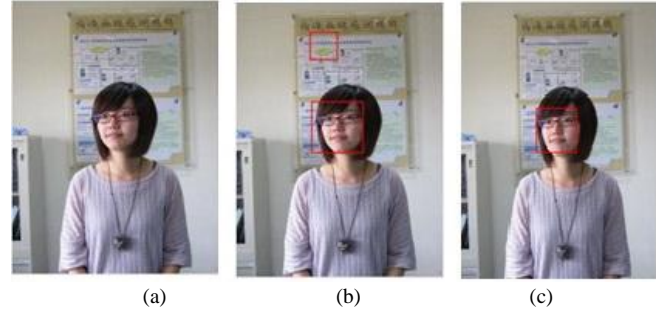

(a)   (b)   (c)

Fig. 2. Result of filtering out non-faces. (a) Original image. (b) Face detection by V-J detector. (c) Face detection by skin ratio.

D. Face Mask

In general, the face posture has two types of frontal face and profile face. For mouth location, it existed higher wrong on biased face image because it includes skin color and non-skin color. In order to solve this situation, we propose a face mask to reduce mistake. The face mask is obtained by morphologic processing.

1) Skin-color segmentation

This procedure is to transform RGB into skin color before product face mask. The skin color model is to normalize RGB because it can discriminate non-skin color clearly. The normalized RGB function is expressed as (5) and many related researches have been presented [9]-[11].

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B} \quad (5)$$

where the *r* and *g* are the normalized pixels.

Here, we present the results of the different skin-color segmentation methods compared with Soriano *et al.*'s method [9], Huang's method [10], Chen's method [11], and our proposed method. Figure 3 shows the compared result. From Fig. 3 (d) shows that Chen's method is better than other author's methods on facial contour and noise, but the median image of Fig.3 (d) has little defects which non-skin region as skin color such as hair. Hence, based on Chen's method, in this paper, we will further improve this method to redefine the skin constraint expressed as

$$\text{Skin} = \begin{cases} 1, & \text{if} \begin{cases} Q_-(r) < g < Q_+(r) \\ (R-G) \ge 25 \\ w > 0.001 \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where

$$\begin{cases} Q_+(r) = A_u r^2 + b_u r + c_u \\ Q_-(r) = A_d r^2 + b_d r + c_d \\ w = (r - 0.33)^2 + (g - 0.33)^2 \end{cases}$$

The upper bound quadratic coefficients are $A_u = -1.3767$, $b_u = 1.0743$, $c_u = 0.1452$, while the lower bound coefficients are $A_d = -0.776$, $b_d = 0.5601$, and $c_d = 0.1766$ [9].

From Fig. 3(e), it is clear that our proposed rule of skin color constrain is superior to Chen's method which let the face out line be good.
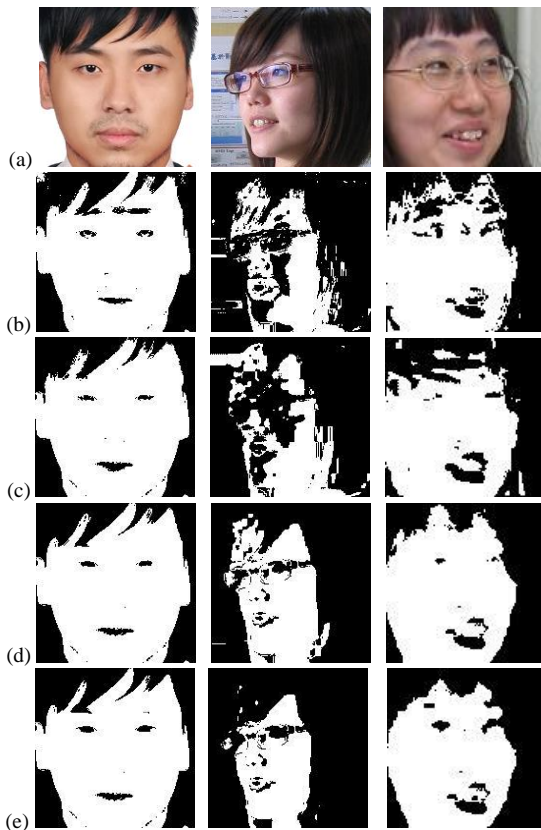

Fig. 3. Skin color detection results. (a) Original images. (b) Results of skin color detection by Soriano *et al's* method. (c) Result of skin color detection by Huang's method. (d) Result of skin color detection by Chen's method. (e) Results of our proposed skin color rule.

2) Design of ratio filter

In order to clearly present the skin-color region in a face image, we design a novel morphological process called ratio filter. The filter aims to emphasize skin pixels and to decrease the non-skin pixel. The ratio filter is to compute number of skin pixels and non-skin pixels on 5×5 mask, and then to compare those of skin pixels.

$$\begin{cases} P_{255} = P_{255} + 1, \ \text{if } X = 255 \\ P_0 = P_0 + 1, \ \text{if } X = 0 \end{cases} \tag{7}$$

where P_{255} is number of skin pixel, P_0 is number of non-skin pixel, and X is a pixel value of the result of skin detection. The current pixel has three conditions on setting as follows.

If P_{255} is greater than P_0, the value of current pixel is as 255.

If P_0 is greater than P_{255}, the value of current pixel is as 0.

If P_0 is equal P_{255}, the value of current pixel is as original pixel.

Fig. 4 presents the result by means of ratio filter. The facial contour is more compact and the non-skin color is reduced.

3) Face mask

The section describes how face mask is produced. The face mask is also called white region. The process consists of two phases. One phase is to find first pixel which value is 255 on horizontal (left to right and right to left) and vertical (top to down and down to top). And then, the pixel is 255 which it between first points (left to right and right to left, top to down and down to top). Finally, the horizontal result is and with vertical that it is called face mask. The diagram of this operation presents in Fig. 5.

Fig. 4. Result. (a) The result without using ratio filter. (b) The result image using ratio filter.
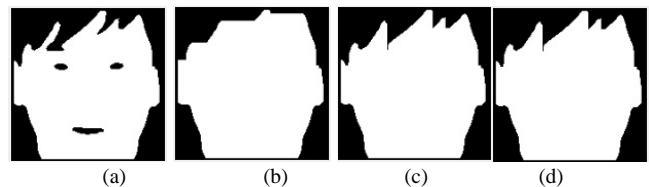

Fig. 5. The face mask results. (a) The binary image by ratio filter. (b) Horizontal result. (c) Vertical result. (d) The result is to compute "And" operation of (b) and (c).

E. Edge Detection using Canny Detector

We adopt Canny detector [12] to obtain the edge information of skin-color region. After processing face mask, we can recover the original face within this mask region, and the takes this original face to make edge detection. Figure 6 presents the results. This edge information can take advantage of the next procedure to locate the mouth position.

F. Projection of Horizontal and Vertical Direction from Edge Data

After processing edge detection, we utilize the projection computation to locate mouth position on horizontal and vertical direction. According to the spatial geometric relationship of facial features, the mouth in general is positioned between one third of face region and one fourth of face region; we define this range as follows. The horizontal/vertical region is expressed as

$$\begin{cases} H_{h1} = H \times 0.7 \\ H_{h2} = H \times 0.1, \\ W_h = W \times 0.1 \end{cases} \begin{cases} H_{v1} = H \times 0.4 \\ H_{v2} = H \times 0.1 \\ W_v = W \times 0.1 \end{cases} \qquad (8)$$

where H_{h1}, $H - H_{h2}$, and W_h denote the first position and the last position on image height (H) and the rang for horizontal region on image width (W), respectively. H_{v1}, $H - H_{v2}$, and W_v denote the first position and the last position on H and the rang for vertical region on W, respectively. Figure 7 shows this diagram of horizontal and vertical regions. The results of horizontal projection are shown in Fig. 8.

In order to reduce mistake which doesn't match mouth region, here, we ignore the first edge point computation from left to right and right to left before vertical projection. Figure 9 shows the result of vertical projection.

G. Mouth Localization

Based on the previous processing, we will compute the maximum value of horizontal and vertical projection to decide the mouth position. Two rules are to decide the mouth location describe as follows.

1) Horizontal maximum value

If there are two or more positions generated the equal maximum value, we take the average position for these two positions and media position for more than two positions.

2) Vertical maximum value

Owing to the accuracy of mouth location, it usually deviates the vertical position; hence, we design (9) and (10) to enhance the correct vertical location. After computing (9) and (10), we can obtain a vertical maximum value the same as the horizontal maximum value.

$$\begin{cases} D_1[i] = W[i-1] - W[i] \\ D_2[i] = W[i] - W[i+1] \\ \quad D[i] = W[i] \times 0.5 \end{cases} \qquad (9)$$

$$\begin{cases} W[i] = W[i], \text{if} \begin{cases} D_1[i] >= 0 \\ D_2[i] >= 0 \\ D_1[i] < D[i] \text{ or } D_2[i] < D[i] \end{cases} \\ W[i] = 0, \text{if} \begin{cases} D_1[i] < 0 \text{ or } D_2[i] < 0 \\ D_1[i] >= D[i] \text{ or } D_2[i] >= D[i] \end{cases} \end{cases} \qquad (10)$$

where i is position, $W[i]$ is the projection of current position, $D_1[i]$ is the difference between the forward position and the current position, $D_2[i]$ is the difference between the current position and the next position, $D[i]$ is a threshold value.

III. RESULTS

In this section, we demonstrate our proposed method using Boa database [13] and our database to estimate the performance for mouth localization and face pose. The face images are 355 that contain 94 on Boa database and 261 on our database respectively. All experiments worked in a 2.80 GHz Intel ® Core(TM) is-2300 CPU with 4 GB RAM PC and C# language. And the computational time of mouth detection is about 0.75s with size of 150×150 for a face image.

Based on Chen's rule and our rule about skin-color segmentation as benchmark, Table I shows the accuracy percentage of mouth detection. Because Chen's condition in skin-color condition is lax, it may give a big face mask which includes non-skin color region, it will cause the mistake mouth detection. Hence, the result of mouth detection by our rule is superior to Chen's rule in Bob database. However, the accuracy of mouth detection in our data can achieve 94%. Figures 10-12 show the detection results. From Figs. 10-12, it is clear that our results are more precise than Chen's rule because our condition is strict on face mask. However, there are some results that detect the error, such as Fig. 13.
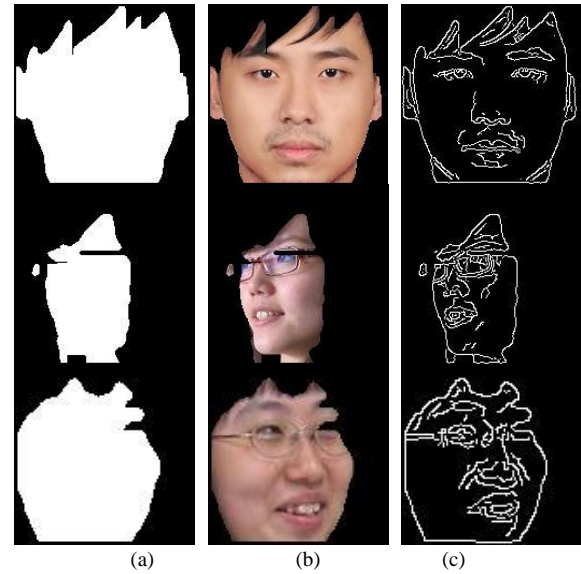

(a)   (b)   (c)

Fig. 6. The Canny results. (a) The binary image by ratio filter. (b) The recovery result. (c) Canny edge detection.
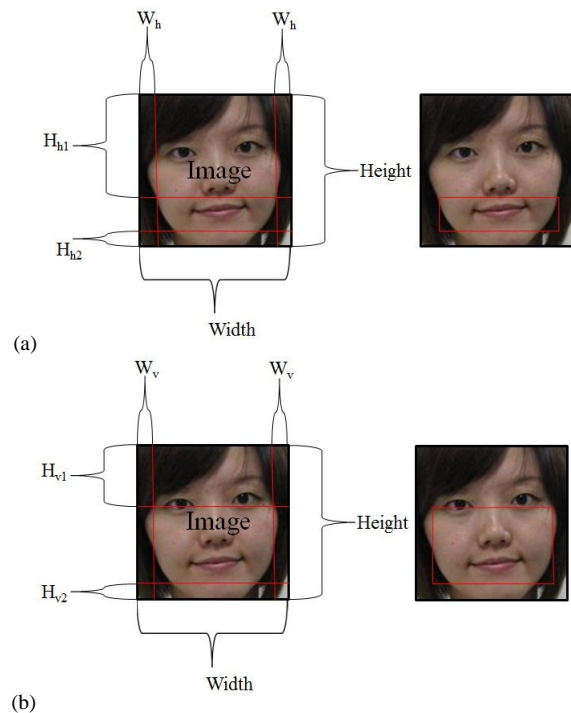

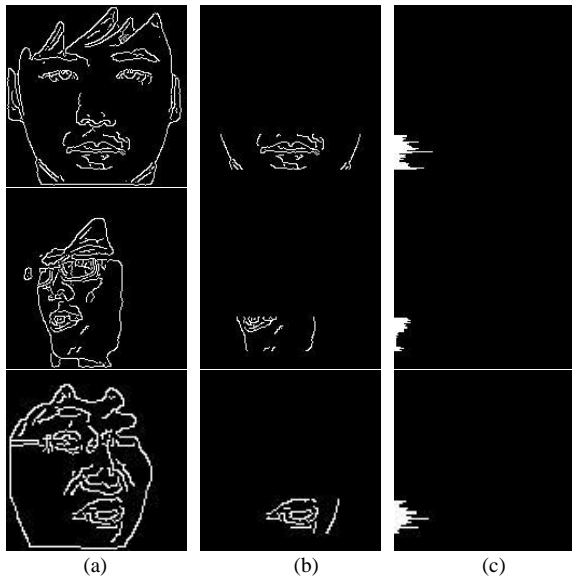Fig. 7. The diagram of mouth region. (a) Horizontal region and (b) vertcal region.

Fig. 8. The horizontal projection. (a) Canny edge detection. (b) The horizontal region of canny image. (c) The horizontal projection.
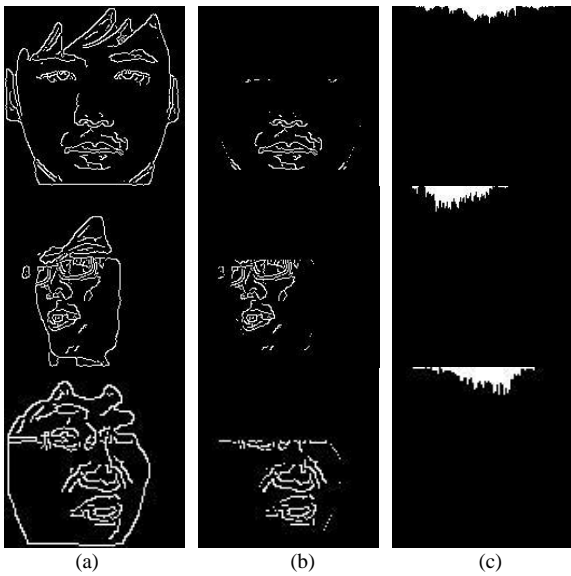

Fig. 9. The vertical projection. (a) Canny edge detection. (b) Vertical region of Canny image. (c) Vertical projection.
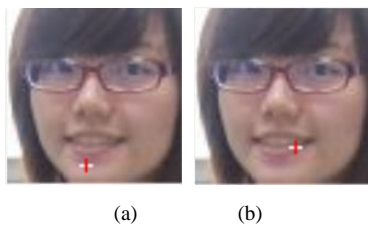

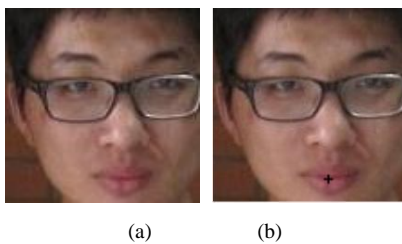Fig. 10. Results of the mouth location. (a) Chen's condition. (b) Our condition.

Fig. 11. Results of the mouth location for frontal face. (a) Original image. (b) Our proposed condition.
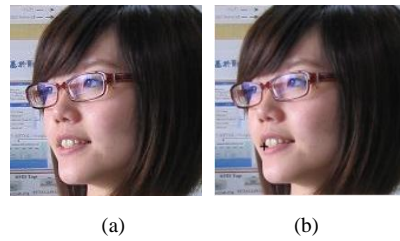

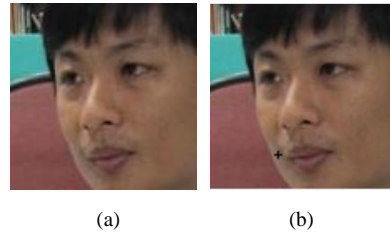Fig. 12. Results of the mouth location for profile face. (a) Original image. (b) Our proposed condition.

Fig. 13. Results of the mouth location. (a) Original image. (b) Wrong result.

TABLE I: THE MOUTH DETECTION

Database	Samples	Accuracy	
		Chen's	The proposed
Bob	94	90%	92%
Our data	261	94%	94%
Total	355	93%	93%

IV. CONCLUSION

In this paper, we have presented a fast mouth detection method based on face localization, skin color segmentation, edge detector, and projection process. The purpose of this approach is to detect mouth position and to raise the correct accuracy based on our proposed rule for skin-color segmentation. Moreover, the results demonstrate that our approach can exactly obtain higher accuracy than Chen's skin-color segmentation condition in the application of mouth detection. Progress in current algorithm about skin-color detection, mouth feature estimation, and face posture estimation will be the major study in the future.

REFERENCES

[1] J. Zhang, Y. Liu, and S. Ha, "A novel approach of face detection based on skin color segmentation and PCA," in *Proc. 9th Int. Conf. Young Computer Scientists*, pp. 1006-1011, 2008.

[2] G. Yang, H. Sun, and H. Li, "Face detection based on adaptive skin model and hough transform," in *Proc. of IEEE Conf. of Int. Conf. on Electronics, Comm. and Control (ICECC)*, pp. 258-260, 2011.

[3] P. Viola and M. Jones, "Robust real-time face detection," in *Proc. of IEEE Conf. of Computer Vision*, vol. 2, pp. 747, 2001.

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. Of the IEEE Computer Vision and Pattern Recognition*, vol. 1, pp. 511-518, Dec. 2001.

[5] R. L. Hsu, M. A. Mottaleb, and A. K. Jain, "Face detection in color Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-706, 2002.

[6] J. He and H. Zhang, "A real time lip detection method in lipreading," *Control Conf. on Chinese*, pp. 516-520, 2007.

[7] D. Chai and K. N. Ngan, " Face segmentation using skin-color map in videophone applications," *IEEE Trans. Circuits and System for Video Technology,* vol. 9, no. 4, pp. 551-564, Jun 1999.

[8] M. W. Wu, "Automatic facial expressions analysis system," Master Thesis, National Cheng King University, 2003.

[9] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Using the skin locus to cope whit changing illumination conditions in color-based face tracking," in *Proc. IEEE Nordic Signal Proc. Symp., Kolmarden, Sweden,* pp. 383-386, 2000.

[10] T. S. Hunag, "A smart digital surveillance system with face tracking and recognition capability," Master Thesis, Chung Yuan Christian University, 2004.

[11] C. T. Chen, "Multiple face recognition based on skin-color regional segmentation and principal component analysis," Master Thesis, National Taiwan Ocean University, 2006.

[12] J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intell.*, vol. PAMI-8, no. 6, pp. 679-698, 1986.

[13] Boa Database. [Online] available: http://www.datatang.com/datares/go.aspx?dataid=604374