LIP DETECTION BASED-ON NORMALIZED RGB CHROMATICITY DIAGRAM

Aryuanto Soetedjo[1], Koichi Yamada[2], F. Yudi Limpraptono[1]
[1]Department of Electrical Engineering, Faculty of Industrial Technology
ITN Malang, Indonesia
[2] Department of Management Information System Science
Nagaoka University of Technology, Japan
email : aryuanto@gmail.com[1], yamada@kjs.nagaokaut.ac.jp[2], fyudil@yahoo.com[1]

ABSTRACT

This paper presents a new lip detection method based-on normalized RGB chromaticity diagram. The method consists of three stages: face detection, lip region localization and lip detection. The popular Viola-Jones face detection technique is employed in the face detection stage. In the lip detection stage, lip color is extracted using our novel color segmentation method that exploits the distribution of lip color on the RGB chromaticity diagram. From the experiment using 100 face images, the detection rate of 97% is achieved.

Keywords: Face detection, lip detection, color segmentation, chromaticity diagram.

1 INTRODUCTION

Lip detection plays an important role in AVSR (Audio Visual Speech Recognition System), a system which combines the audio and visual information for speech recognition. By incorporating visual information, i.e. lip movement, the performance of speech recognition is improved, especially in the noisy environment [1]. Lip detection is also used in another application such as for facial expression analysis or recognition.

There were many researches on the lip detection/segmentation [2-11]. In [2], a lip detector based on adaptive thresholding for hue-transformed face images was proposed. It first transformed RGB color image to the hue/saturation color space due to the fact that the lip region is fairly uniform for a wide range of lip colors in the hue/saturation color space. The adaptive thresholding using fuzzy membership function was proposed to control the threshold for detecting lip region.

The rule-based lip detection technique based on normalized RGB color space was proposed in [3]. They defined a cresent area on *r-g* plane using a quadratic polynomial discriminant function for detecting lip pixels. In [4], a new color transformation based on RGB color spase was proposed for robust lip segmentation under non-uniform lighting conditions. They defined the chromatic curve to increase the discrimination between lips and skin.

The Adaboost approach with Haar features was employed for lip detecting in [5]. The system first uses cascaded classifier to detect face using Haar-like features, then two mouth classifiers will locate the mouth region in the lower part of face. The Kalman filter is employed for tracking the lip by estimating the center of mouth region in the next frame. In [6], the watershed segmentation was used to detect lip directly without needing preliminary face localisation. The $H^{\sim}$ tracking system was used to track the lip region.

In [7] they used some features based on R/G color space by considering the second order statistics in the adjacent pixels, and employed an unsupervised clustering technique based on the expectation maximisation (EM) algorithm. In [8], an adaptive color space training for lip and non-lip region using Gaussian mixture model (GMM) was employed to extract the exact boundary of lip. The lip segmentation based on *rg*-color histogram was proposed in [9]. They built a gaussian model of the skin pixels distribution to obtain the optimal threshold.

A Spatial Fuzzy C-Means (SFCM) clustering algorithm was proposed in [10] for lip segmentation. The technique considers both the distributions of data in feature space and spatial interactions between neighboring pixels. In [11], they proposed the modified Fuzzy C-means clustering to exploit the spatial interactions through modification of initial membership function.

In this paper we propose a new lip segmentation technique based-on the normalized RGB chromaticity diagram. At first, the face detector followed by a simple geometric rule is employed to localize lip region from image. Then a new developed grayscale image is created based-on the chromaticity color space. Finally the Otsu

ISSN 2085-1944

thresholding technique is adopted to detect the lip region.

The rest of paper is organized as follows. Section 2 describes the lip detection technique. Section 3 presents the experimental results. Conclusion is covered in section 4.

2 LIP DETECTION

2.1 Overview

Figure 1 illustrates our proposed lip detection system. In this research, we deal with frontal face images as input to the system. The algorithm starts with detecting face using a popular face detection technique proposed by Viola-Jones [12]. Since the Viola-Jones's face detection technique works with grayscale image, the gray conversion process is employed to convert RGB color image into grayscale image. After face is detected, then lip region is found by a simple geometric rule.

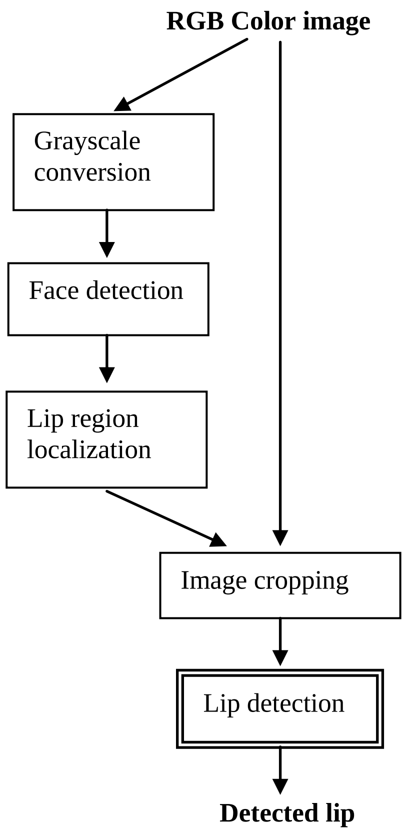

Figure 1. Overview of proposed lip detection.

The next step is to crop the original color image around lip region. This cropped image is then fed to the lip detection module to extract the lip based-on the normalized RGB chromaticity diagram as discussed in the following section.

2.2 Face Detection and Lip Region Localization

The Viola-Jones's face detection technique [12] has three important aspects, i.e. a) image feature called integral image; b) learning algorithm based-on Adaboost; c) combining classifiers using a cascade scheme.

Integral image is an intermediate representation for the image which is expressed as [12] :

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

where *int_i*(*x*,*y*) is integral image at location *x*,*y* and *i*(*x*,*y*) is the original image. From Eq. (1) we 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 Eq. (1) could be computed using the following formulas [12]:

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

$$int\_i(x,y) = int(x-1,y) + row\_sum(x,y) \qquad (3)$$

where *row_sum*(*x*,*y*) is the cumulative row sum.

Using the integral image, the rectangle features as illustrated in Fig. 2 could be computed efficiently. Figs. 2(a) and 2(b) show the two-rectangle feature which is calculated as the difference of the sum of the pixels within two rectangular regions. Fig. 2(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. Fig. 2(d) shows the four-rectangle feature which is calculated as the difference between diagonal pairs of rectangles.

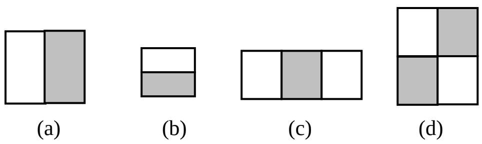

Figure 2. Three kinds of rectangle features.

To learn a classification function that could classify the image given a set of features and positive and negative training images, the AdaBoost classifier is employed [13]. The Adaboost learning algorithm is used to select a small set of features and train the classifier. 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

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

Where x is a 24x24 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. The scheme is to use simpler classifiers which reject the majority of sub-windows before more complex classifiers to achieve low false positive rates.

After face is detected, a lip region is localized by considering the lower part of the face region defined empirically as follows:

Lip_left=*Face_left*+(*Face_width*/4) (5)
Lip_top=*Face_top*+(2*(*Face_height*)/3) (6)
Lip_width=*Face_width*/2 (7)
Lip_height=*Face_height*/3 (8)

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.

The lip region found using Eqs. (5)-(8) is used to crop the original color image to obtain the lip image in which the lip will be detected using the algorithm described in the following.

2.3 Lip Detection

Since lip colors are distinguishable from skin colors, the lip could be extracted using their colors by defining a lip region on the r - g color space [3]. They defined the two discriminant functions for extracting lip pixels as

$$l(r) = -0.776r^2 + 0.5601r + 0.2123, \quad f_{lower}(r) = -0.776r^2 + 0.5601r + 0.1766 \quad (9)$$

where

$r=R/(R+G+B)$ (10)
$g=G/(R+G+B)$ (11)

Then the lip pixels are detected by the following rule

$$L = \begin{cases} 1 & if\ f_{lower}(r) \le g \le l(r)\ and \\ & R \ge 20\ and\ G \ge 20\ and\ B \ge 20 \\ 0 & otherwise \end{cases} \quad (12)$$

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, we propose a new lip detection method as described in the following.

Our proposed lip detection method is the extension our previous research [14], which utilizes a normalized RGB chromaticity diagram for red color thresholding

Figure 3 shows the chromaticity diagram of normalized RGB, where the chromaticity coordinates are r and g are defined by Eqs. (10) and (11) respectively. From the figure, it is clear that red color could be separated from other colors using a diagonal line as proposed in [14]. The objective is how to select a proper line which is able to separate red color automatically.

Our proposed lip detection as depicted in Fig. 1 is applied to the cropped image which contains lip region. Thus the task is to extract lip color from face color (skin color). Since the colors of lips range from dark red to purple [3], we might extract the lip color using the similar way as our previous work in [14].

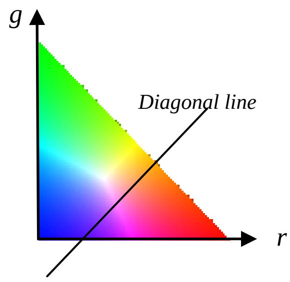

Figure 3. The normalized RGB chromaticity diagram.

The diagonal line in Fig. 3 could be expressed as

$$g = r + C \quad (13)$$

where C is interception on r-coordinate. By choosing the value of C appropriately, we could find a line that separating lip color properly. Then to extract the lip color, the following rule is employed [14]:

If g-r < TR, then assign pixel as lip (14)

TR is a threshold that is the optimal value of C for separating lip color.

To find TR, we develop a new transformed image (grayscale image) called *Igr* using the following equations:

$$Igr = (100 + (100 * g) - (100 * r)) / 200 \quad (15)$$

The value of *Igr* in Eq. (15) could be interpreted as the value of C in Eq. (13). From Fig. 3 and Eq. (13), the value of C is determined by the number of pixels in the image whose chromaticity's values

(g,r) lie along the line. Therefore, if we draw a histogram that counting the number of pixels for each value of C (or g-r) then there will be a peak represents the change of C when the line moves from non-lip color to lip color as depicted in Fig. 4.

In our previous work [14], we find the optimal TR from Fig. 4 using peak/valley analysis. This approach has a drawback, because in some cases the valley or peak could not be determined easily. To overcome this problem, we propose to use the Igr image as depicted in Fig. 5 and employ the Otsu's thresholding method [15] to find the optimal TR.

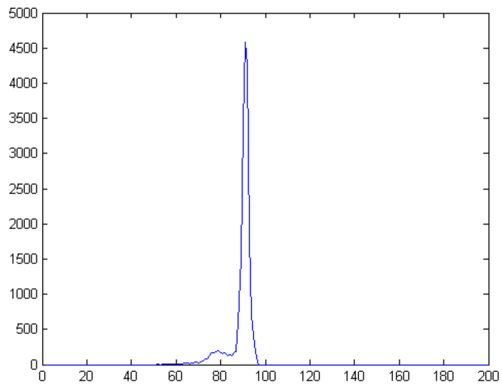

Figure 4. The number of pixels of g-r.

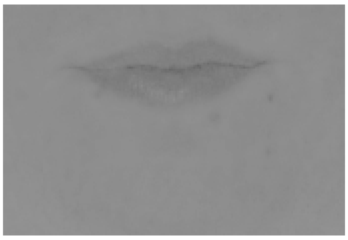

Figure 5. The transformed image (Igr).

3 EXPERIMENTAL RESULT

We implemented our algorithm using MATLAB running on PC. We tested the algorithm using one hundred face images from Caltech Database [16]. The database contains front face color images under different lighting, expressions, and backgrounds. Each face image only contains one face. We also made a comparison with a method proposed in [3] which is described by Eqs. (9) and (12).

Figure 6 shows an example of face image used in the experiments, where the detected face is drawn in red rectangle. Figure 7(a), 7(b), 7(c) show lip region, extracted lip, and bounding box of detected lip images respectively.

Figure 6. Face image with the detected face.

Figure 7. (a) Lip region image; (b) Extracted lip image; (c) Bounding box of the detected lip image.

From the experiment on 100 face images, our proposed method could detect 97 lips properly, while the method proposed in [3] only detect 65 lips properly. Figure 8 shows some of the experimental results. Two images in the fifth row are the results of the lip detection as proposed in [3], where the lips are not located properly. While two images in the sixth row are the results of our proposed method obtained from the same images used in the fifth row.

4 CONCLUSION

The lip detection method based on RGB chromaticity diagram is presented. The method is used to separate lip color from other colors (face skin color) by a simple effective color segmentation method.

In future we will extend our approach to utilize the lip detection technique for extracting the visual information used in the Automatic Visual Speech recognition System (AVSR).

Figure 8. Results of lip detection.

REFERENCES

[1] Potamianos G, Neti C, Gravier G, Garg A, Senior A W (2003) Recent advances in the automatic recognition of audiovisual speech. Proceedings of the IEEE, Vol. 91, No. 9, pp. 1306-1326.

[2] Kim J Y, Na S Y, Cole R (2006) Lip detection using confidence-based adaptive thresholding. In : Bebis G, et. al. [eds.] Lecture Notes in Computer Science. Vol. 4291. Springer-Verlag Berlin Heidelberg. pp. 731-740.

[3] Chiang C C, Tai W K, Yang M T, Huang Y T, Huang C J (2003) A novel method for detecting lips, eyes and faces in real time. Real-Time Imaging 9. pp. 277-287.

[4] Eveno N, Caplier A, Coulon P Y (2001) A new color transformation for lips segmentation. Proceedings of IEEE 4th Workshop on Multimedia Signal Processing, 3-5 October 2001, Cannes, France, pp. 3-8.

[5] Hoai B C, Hoai V T, Ngoc T H (2010) Lip detection in video using Adaboost and Kalman filtering. Journal on Information and Communications Technologies, Vol. E-1, No. 2 (6). pp. 2-8.

[6] Chin S W, Seng K P, Ang L M, Lim K H (2009) New lips detection and tracking system. Proceedings of the Multi Conference of Engineers and Computer Scientists. March 18-19, 2009, Hongkong.

[7] Lucey S, Sridharan S, Chandran V (2002) Adaptive mouth segmentation using chromatic features. Pattern Recognition Letters. Vol. 23 (11). Pp. 1293-1302.

[8] Ozgur E, Yilmaz B, Karabalkan H, erdogan H, Unel M (2008) Lip segmentation using adaptive color space training. Proceedings of International Conference on Auditory-Visual Speech Processing, Sept. 26-29, 2008, Quensland, Australia. pp. 219-222.

[9] Panning A, Niese R, Al-Hamadi A, Michaelis B (2009) A new adaptive approach for histogram based mouth segmentation. World Academy of Science, Engineering and Technology, Vol. 56. pp. 584-589.

[10] Liew A W C, Leung S H, Lau W H (2003) Segmentation of color lip images by spatial fuzzy clustering. IEEE Transactions on Fuzzy Systems, Vol. 11, No. 4. pp. 542-549.

[11] Sudhavani G, Sathyaprasad K (2009) Segmentation of lip images by modified fuzzy C-means clustering algorithm. International Journal of Computer Science and Network Security, Vol. 9, No. 4. pp. 187-191.

[12] Viola P and Jones M (2001) Rapid object detection using a boosted cascade of simple features. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Kauau, Hawaii, USA.

[13] Freund Y and Schapire R (1997) A decision-theoritic generalization of on-line learning and an application to boosting. Journal of Computer System Science. Vol. 55, No. 1. pp. 119-139.

[14] Soetedjo A and Yamada K (2007) A new approach on red color thresholding for traffic sign recognition system. Journal od Japan Society for Fuzzy Theory and Intelligent Informatics. Vol. 19, No. 5. pp. 457-465.

[15] Otsu N (1979) A thresholding selection method from gray-level histogram. IEEE Transactions on Systems, Man, and Cybernetics 9(1). pp. 62-66.

[16] http://www.vision.caltech.edu/Image_Datasets/faces/faces.tar

[This page is intentionally left blank]

ISSN 2085-1944