

Automatic Mouth Localization Using Edge Projection

Mohamed Rizon

Department of Biomedical Technology, King Saud University,
P.O. Box 10219, Riyadh 11433, Kingdom of Saudi Arabia

Abstract: Problem statement: This study presented algorithms to detect mouth from color and intensity images. **Approach:** First, this algorithm detected the face region in the image and extracts intensity valleys from the face region. Next, the algorithm extracted iris candidates from the valleys and computed the costs for each pair of iris candidates. Finally, a pair of iris candidates was selected as irises by using the computed costs. Projection based method had been used to detect mouth corresponding to irises location. **Results:** By experiment, the proposed algorithm detected 90% of full mouth region for South East Asian database and 74% for European database. **Conclusion:** The algorithm was considered successful to detect mouth detection under variation of pose, illumination and orientation. For future improvement, more preprocessing steps might be needed to enhance and eliminate the effect of beard, moustache and illumination.

Key words: Face recognition, facial features, face detection, iris detection, mouth detection, edge projection

INTRODUCTION

Over the past decade, face recognition has attracted substantial attention from various disciplines and seen a tremendous growth in researches. Automated face recognition has been applied in two ways: Holistic and feature-based. The holistic approach (Brunelli and Poggio, 1993; Beymer, 1993) treats a face as 2D pattern of intensity variation.

Generally, eyes and mouth are important in understanding the information and feeling conveyed by a person. The feature-based approach (Kawaguchi *et al.*, 2000; Kawaguchi and Rizon, 2003; Brunelli and Poggio, 1993) recognizes a face using the geometrical measurements taken among facial features such as eyes and mouth. Many researches have proposed methods to find the eye (Feng and Yuen, 1998; Zhou and Geng, 2004) and mouth regions (Sobottka and Pitas, 1996; 1998; Li *et al.*, 2004) or to locate the face region in an image. These methods have shown the popularity of using information such as template matching, geometrical and intensity features. Brunelli and Poggio (1993) and Beymer (1993) located eyes using template matching. In this method, an eye template of a person is moved into the input image. Then, a patch of the image that has the best match to the eye template is selected as the eye region. However, template matching and eigenspace method require normalization of the face for variation of size and orientation. A large number of templates are needed for template matching to accommodate varying pose.

For facial feature extraction, one of the very popular methods in geometric-feature based approach is the use of vertical and horizontal projections. The projections can be employed after taking the first derivative of the image (Brunelli and Poggio, 1993) as well as directly on intensity the intensities values (Sobottka and Pitas, 1996). Projection based methods (Baskan *et al.*, 2002; Gourier *et al.*, 2004; Ryu and Oh, 2001; Pantic *et al.*, 2001) have been used particularly to find coarsely the position of the facial features. For eye and mouth localization, valley regions are detected using morphological filtering and component analysis. There are only a few studies give qualitative results about the performance. Most of the sample images used by previous studies (Baskan *et al.*, 2002; Gourier *et al.*, 2004; Ryu and Oh, 2001; Pantic *et al.*, 2001) did not contain thick beard. This is because the intensities level of thick beard might reduce the mouth detection rate of projection based method. A large number of the South East Asians have darker and thicker beard or moustache than Europeans. The combination of methods to detect eyes and mouth proposed in our algorithm has not been applied by previous studies.

In this study, the algorithm first detects the face region in the image and extracts intensity valleys from the region. Then, iris candidates are extracted from the valleys using the feature template of Lin and Wu (1999) and separability filter of Fukui and Yamaguchi (1997). Next, the costs from template matching, separability filter and Hough transform (Kawaguchi *et al.*, 2000) are

computed for iris candidate pairs to determine the irises of both eyes. Finally, horizontal and vertical integral projections have been applied to lower part of the face for mouth detection. The positions of both irises are used to determine the horizontal wide of the mouth region. This study begins by describing each of these phrases. It presents further the experimental results and finally discusses the limitations and future improvement of the proposed method.

MATERIALS AND METHODS

In our experiment, we use two databases to evaluate the performance of our algorithm. One is European database with color images and another is South East Asian database with intensity images. For input image, we assume the image is a head-shoulder image with plain background and head rotation on y-axis is approximately within the interval (-30, 30°).

Extraction of the face regions from intensity images: This proposed algorithm extracts the face region from an intensity image using a similar method to that shown in (Brunelli and Poggio, 1993). First, we apply Sobel edge detector to the original intensity image $I(x, y)$. Let $E(x,y)$ denote the obtained edge image where $E(x,y) = 1$ if (x,y) is an edge pixel and otherwise $E(x,y) = 0$. Next, for each column x and each row y , $v(x)$ and $H(y)$ are computed by:

$$V(x) = \sum_{y=0}^{N-1} E(x, y), H(y) = \sum_{x=0}^{M-1} E(x, y) \tag{1}$$

The x -positions x_L and x_R of the left and right boundaries of the head are given by smallest and largest values of x such that $V(x) \geq V(x_0)/3$ where x_0 denotes the column x with the largest $V(x)$. The y -position y_{min} of the upper boundary of the head is given by the smallest y such that $H(y) \geq 0.05(x_R - x_L)$. Finally, we give the y -position y_{max} of the lower boundary of the head by $y_{max} = y_{min} + 1.2(x_R - x_L)$.

Extraction of the face regions from color images: The proposed algorithm first creates a skin-color model for the face region detection from color images. Let (R, G, B) denote a color vector in RGB color space. Then, the normalized color vector (r, g) is given by:

$$r = \frac{R}{R + G + B}, g = \frac{G}{R + G + B} \tag{2}$$

Using the rg color space, we create a skin-color model by (Kawaguchi and Rizon, 2003):

1. Select skin-color pixels (x,y) whose color value $v = (r,g)$ satisfy $g(v) \geq \epsilon$
2. Estimate the mean $\mu = (\mu_r, \mu_g)$ and covariance matrix Σ of the color distribution of the pixels in the selected regions
3. Computes the Gaussian distribution model where the probability density function is given by:

$$p(v) = (2\pi)^{-1} |\Sigma|^{-1/2} g(v) g(v) = \exp\left\{-\frac{1}{2}(v - \mu)^T \Sigma^{-1} (v - \mu)\right\} \tag{3}$$

where, $v = (r,g)$ denotes a random color vector of a pixel in an image.

4. Apply a closing and opening of the mathematical morphology to the regions of skin-color pixels
5. Find a connected component of skin-color pixels with the largest area
6. If (x,y) denotes the coordinates of the pixels obtained in Step 5, let X_1 and X_2 be the smallest and largest of x_i and Y_1 be the smallest of y_i . Then, the face region produced by two vertical lines $x = X_1$ and $x = X_2$, two horizontal lines $y = Y_1$ and $y = Y_1 + X_2 - X_1$

Extraction of valleys in the face region and preprocessing: The proposed algorithm applies grayscale closing (Sternberg, 1986) to the face region to extract valleys. $V(x, y) = G(x, y) - I(x, y)$ where $G(x, y)$ and $I(x, y)$ denote its intensity value and value obtained by applying grayscale closing. Then, region consisting of pixels (x, y) such that $V(x, y)$ is greater than or equal to a threshold value are determined to be valleys. Histogram equalization and light spot deletion are performed to enhance the quality of image and reduce illumination effect in this algorithm.

Detection and selection of iris candidates: The proposed algorithm performs the similar method as proposed by (Kawaguchi and Rizon, 2003):

- Computes costs $C(x,y)$ for all pixels in the valleys:

$$C(x,y) = C_1(x,y) + C_2(x,y) \tag{4}$$

Then, the algorithm selects m pixels according to non-increasing order that give the local maxima of $C(x,y)$.

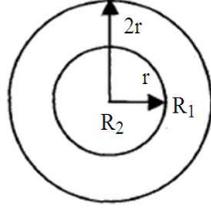


Fig. 1: An eye template to detect blob

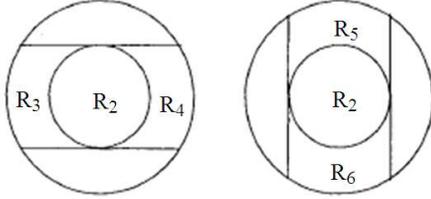


Fig. 2: The templates used to compute $C_2(i)$ and $C_3(i)$

- Places the template of Fig. 1 at each candidate location and measures the separability between two regions R_1 and R_2 given by :

$$\eta = \frac{B}{A} \quad (5)$$

- Applies Canny (1986) edge detector to the face region and measures the fitness of iris candidates to the edge image using Hough transform (Kawaguchi *et al.*, 2000). We give the equation of a circle by:

$$(x-a)^2 + (y-b)^2 = r^2 \quad (6)$$

where, (a,b) is the circle center and r is the radius.

- Given an iris candidate $B_i = (x_i, y_i, r_i)$ measures the fitness of iris candidates to the intensity image by placing two templates in Fig. 2 and compute the separabilities $\eta_{23}(i)$, $\eta_{24}(i)$, $\eta_{25}(i)$ and $\eta_{26}(i)$ using Eq. 4, where $\eta_{kl}(i)$ denotes the separability between regions R_k and R_l
- Calculate cost for each iris candidate as shown in (Kawaguchi and Rizon, 2003):

$$C(i) = C_1(i) + C_2(i) + C_3(i) + C_4(i) \quad (7)$$

- Computes a cost for each pair of iris candidates, B_i and B_j is given by:

$$F(i,j) = t\{C(i) + C(j)\} + (1-t) / R(i,j) \quad (8)$$



Fig. 3: The templates used for computation of $R(i, j)$

where, $C(i)$ and $C(j)$ are costs computed by Eq. 6. $R(I_i, I_j)$ is the normalized cross-correlation value computed by using eye template in Fig. 3 which is produced by manually cut off the eye region from a face image. t is the weight to adjust two terms of the cost.

Mouth detection: First, we assume this proposed algorithm crops the region containing mouth automatically from the Canny edge detection images by defining:

$$Y_1 = \left(\frac{4}{5} \times H\right) + y_{\min}, Y_2 = y_{\max} - 10 \quad (9)$$

Where:

Y_1 and Y_2 = The upper and lower boundaries for the crop region

H = The height of the face region

y_{\min} and y_{\max} = The upper and lower boundaries for the face region

Then, we apply vertical integral projections method in Eq. 1 on the cropped Canny edge region. This proposed algorithm will search the maxima (the largest sum values in row) in the crop region starting from $Y_2 - Y_1$. Finally, the mouth region will be determined by:

$$Y_{\min} = y_m - 25, Y_{\max} = y_m + 25 \quad (10)$$

Where:

Y_{\min} and Y_{\max} = The upper and lower boundaries of the mouth region

y_m = The y-axis with maxima value

The irises locations obtained in previous steps can be used to determine left and right boundaries of mouth region from horizontal integral projection.

RESULTS AND DISCUSSION

We made the experiments to evaluate the performance of the proposed algorithm. The European database and South East Asian database are used in the experiment. There are 63 color images in European

database with 21 subjects and size of each image is 768×576. South East Asian database contains 60 intensity images with 8 subjects and size of each image is 480×360. The images in both databases are varying in pose, head orientation, illumination and gender. Images with moustache and beard are included as well.

Since the successful detection rate for irises are more than 90% for both databases as shown in (Kawaguchi and Rizon, 2003), we put our concentration on mouth detection. From Table 1, this proposed algorithm scores 74% successful mouth detection. Illumination effect, different gender and the existence of moustache or beard still contribute to the failure detection although there are successful mouth detections from these images in this experiment. However, 15% of half mouths detected by this algorithm from European database can be essential in classification stage using Artificial Intelligence such as Neural Network. This algorithm detected 90% of full mouth region for South East Asian database even with beard or moustache. This may due to the better and more stable illumination environment in intensity images. The processing time for irises and mouth detection for European database is merely 0.2 sec and 0.1 sec for South East Asian database for each image. We can see that intensity image has an advantage over color image in terms of processing time. Figure 4 shows examples of the images for which the proposed algorithm could correctly detect the mouth.

Table 1: The result for successful mouth detection rate for European and South East Asian databases

Database	Success rate of mouth detection (%)		
	Full mouth	Half mouth	No mouth
European	74	15	11
Asian	90	7	3

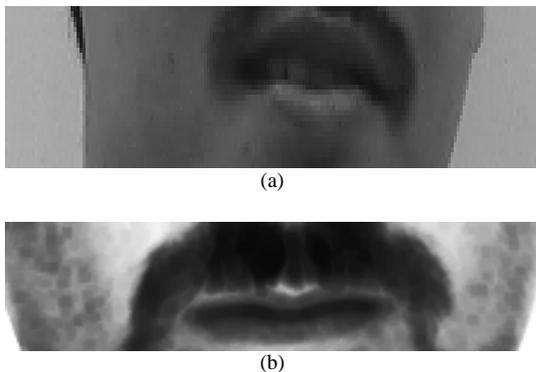


Fig. 4: Example of images (a) East Asian (b) European for which the proposed algorithm could correctly detect the mouth

CONCLUSION

Most of the facial features algorithms previously reported used template matching, eigenspace method or Hough transform. However, template matching and eigenspace method require the normalization of the image face in its size and orientation when large variations occur. These algorithms can detect facial features from faces whose patterns are similar to training samples. In order to be more robust, large amount of facial features' models are required especially when dealing with the moustache, beard and different gender or races. Hough transform algorithms need to estimate the searching windows for the facial features in the face region for example; irises are considerably small as compared to the face size. Thus, these algorithms require complete face region detection.

In this proposed algorithm, iris detection rate is higher than 90% as shown in (Kawaguchi and Rizon, 2003) even in incomplete face region. This is due to the higher intensities of the irises compared to other features in face region. Our algorithm is considered successful in South East Asian database with 90% successful rate for mouth detection under variation of pose, illumination and orientation. This may be due to the better illumination environment offered by intensity images compared to color images. In European database, the illumination environment and existence of other features such as nose, beard and moustache have contributed to the failure detections. In addition, the difference of facial texture for Europeans and Asians might also affect the mouth detection rate. Note that, the half mouth detected in both databases 15% for European database and 7% for South East Asian database can be useful as training samples in classification stage. This is because these regions contain essential geometric information which is useful for face recognition purpose.

For future improvement, more preprocessing steps are needed to enhance and eliminate the effect of beard, moustache and illumination. Geometrical distances according to the detected irises locations can be added to increase the successful rate of mouth detection. Artificial Intelligence techniques such as Neural Network can be used as classifier by using half mouth regions as training samples to increase the performance of face recognition system.

REFERENCES

Baskan, S., M.M. Bulut and V. Atalay, 2002. Projection based method for segmentation of human face and its evaluation. *Patt. Recog. Lett.*, 23: 1623-1629. DOI: 10.1016/S0167- 8655(02)00037-5

- Beymer, D.J., 1993. Face recognition under varying pose. Memorandum Report. <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA290205>
- Brunelli, R. and T. Poggio, 1993. Face recognition: Features versus templates. *IEEE Trans. Patt. Anal. Mach. Intell.*, 15: 1042-1052. DOI:10.1109/34.254061
- Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. Patt. Anal. Mach. Intell.*, 8: 679-698. DOI: 10.1109/TPAMI.1986.4767851
- Feng, G.C. and P.C. Yuen, 1998. Variance projection function and its application to eye detection for human face recognition. *Patt. Recog. Lett.*, 19: 899-906.
- Fukui, K. and O. Yamaguchi, 1997. Facial feature point extraction method based on combination of shape extraction and pattern matching. *Trans. IEICE Jap., J80-D-II*, 8: 2170-2177.
- Gourier, N., D. Hall and J.L. Crowley, 2004. Facial features detection robust to pose, illumination and identity. *Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, (SMC'04)*, IEEE Xplore Press, USA., pp: 617-622.
- Kawaguchi, T. and M. Rizon, 2003. Iris detection using intensity and edge information. *Patt. Recog.*, 36: 549-562. DOI: 10.1016/S0031-3203(02)00066-3
- Kawaguchi, T., D. Hidaka and M. Rizon, 2000. Detection of eyes from human faces by Hough transform and separability filter. *Proceedings of the IEEE 3rd International Conference Image on Processing*, Sept. 10-13, IEEE Xplore Press, Vancouver, BC., Canada, pp: 49-52. DOI: 10.1002/ecjb.20178
- Li, C.M., Y.S. Li, Q.D. Zhuang and Z.Z. Xiao, 2004. The face localization and regional features extraction. *Proceedings of the 3rd International Conference on Machine Learning and Cybernetics*, Aug. 26-29, IEEE Xplore Press, USA., pp: 3835-3840. DOI: 10.1109/ICMLC.2004.1380504
- Lin, C.H. and J.L. Wu, 1999. Automatic facial feature extraction by genetic algorithms. *IEEE Trans. Image Process.*, 8: 834-845.
- Pantic, M., M. Tomc and L.J.M. Rothkrantz, 2001. A hybrid approach to mouth features detection. *Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, (SMC'01)*, IEEE Xplore Press, USA., pp: 1188-1193.
- Ryu, Y.S. and S.Y. Oh, 2001. Automatic extraction of eye and mouth fields from a face image using eigenfeatures and multilayer perceptrons. *Patt. Recog.*, 34: 2459-2466. DOI: 10.1023/A:1016160814604
- Sobottka, K. and I. Pitas. 1996. Face localization and facial feature extraction based on shape and color information. *Proceeding of the International Conference on Image Processing*, Sept. 16-19, IEEE Xplore Press, Lausanne, pp: 483-486. DOI: 10.1109/ICIP.1996.560536
- Sobottka, K. and I. Pitas, 1998. A novel method for automatic face segmentation, facial feature extraction and tracking, signal processing. *Image Commun.*, 12: 263-281.
- Sternberg, S.R., 1986. Grayscale morphology. *Comput. Vis. Graph. Image Process.*, 35: 333-355.
- Zhou, Z.H. and X. Geng, 2004. Projection functions for eye detection. *Patt. Recog.*, 37: 1049-1056.