Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 46 (2015) 1739 – 1746
Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

An Approach to Detect the Region of Interest of Expressive Face Images

Priya Saha[a,*], Debotosh Bhattacharjee[b], Barin Kumar De[c], Mita Nasipuri[b]

[a]*Department of Computer Science & Engineering,Tripura University (A Central University), Suryamaninagar,Tripura-799022,India*
[b]*Department of Computer Science & Engineering,Jadavpur University, Kolkata-700032,India*
[C]*Department of Physics,Tripura University (A Central University), Suryamaninagar,Tripura-799022,India*

Abstract

On human face, non-rigid facial movements due to facial expressions cause noticeable alterations in their usual shapes, which sometimes create occlusions in facial feature areas making face recognition as a difficult problem. The paper presents an automatic Region of Interest (ROI) detection technique of six universal expressive face images. The proposed technique is a facial geometric based hybrid approach. The localization accuracy was evaluated by rectangular error measure and was tested on Japanese Female Facial Expression (JAFFE) database. The average localization accuracy of all detected facial regions is 94%.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

*Keywords:*Region of Interest (ROI); universal faxial expressions; ideal facial proportions; rectangular error measure; localization accuracy ;

1. Introduction

Now-a-days, Facial Recognition Technology (FRT) has come into view to provide a secure solution in identification and verification of person identity. There are several factors that may produce variations in the facial appearance making face recognition as a challenging problem. Facial expression is one of the influential factors of FRT. The human face endows with a number of signs or observations necessary for conveying social information among individuals. It depicts information about the psychological, physiological and cognitive state of a person.

* Corresponding author. Tel.: +91-9436484151; fax: 0381-2374802.
*E-mail address:*priyasaha.cse@gmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.123

According to Darwin[1], six basic expressions (happiness, sadness, surprise, anger, disgust, fear) are almost universal. Facial expressions cause structural changes in the regions of facial landmarks (eyebrows, eyes, nose, and mouth). In order to detect facial expressions, facial feature localization plays a vital role. Facial landmark localization can be done by first locating facial feature Region of Interest (ROI) to reduce the search space. Well analysis of the located ROIs employs an effective technique to locate facial feature points like eye corners, mouth corners, eye-brow corners. Localization of ROI has been done by several researchers. Popular facial feature ROI localization techniques are based on the learning-based methods; for example, support vector machines[2,3], cascades of boosted classifiers[4] and neural networks. Wavelet image decomposition combined with neural networks has been extensively applied for facial feature ROI localization[5]. Gizatdinova and Surakka[6] introduced edge-based technique of locating ROIs of prominent facial features from up-right facial images. They proposed a new accuracy evaluation method for feature ROI localization.

In this paper, a new expression invariant ROI localization method has been introduced. Initially, gray level intensity based method has been used to extract the shape boundary curvature of eye and mouth region. Here, eye region has been considered as a combination of both eye and eye-brow area. In case of extracting the mouth region, local edge map is also needed to be constructed. The main aim of this paper is to evaluate the efficiency of proposed method on a benchmark facial expression database. Japanese Female Facial Expression (JAFFE) database has been used in this paper. Performance has been evaluated through rectangular error measure strategy[6]. A noticeable ROI localization accuracy has been observed through this proposed method.

The rest of the paper is as follows. Section 2 describes the proposed methodology. Section 3 discusses experiment results and conclusion is drawn in section 4.

2. Methodology

The proposed method has been illustrated in Fig. 1 through the block diagram. The theoretical details have been described below in different stages.

2.1. Pre-processing

In the pre-processing stage, illumination compensation has been performed using logarithm transform to enlarge the values of dark pixels[7].A gray level image $f(x, y)$ can be written as the product of the reflectance $r(x, y)$ and the illumination $e(x, y)$ [8], i.e.

$$f(x, y) = r(x, y).e(x, y) \tag{1}$$

The below equation can be found by taking logarithm transform on (1)

$$\log f(x, y) = \log r(x, y) + \log e(x, y) \tag{2}$$

It can say from (2) that in the logarithm domain, if the incident illumination is $e(x, y)$ and the desired uniform illumination is $e^{'}$ are given ($e^{'}$ is identical for every pixel of an image), we have

$$\begin{aligned} \log f^{'}(x, y) &= \log r(x, y) + \log e^{'} \\ &= \log r(x, y) + \log e(x, y) - \in (x, y) \\ &= \log f(x, y) - \in (x, y) \end{aligned} \tag{3}$$

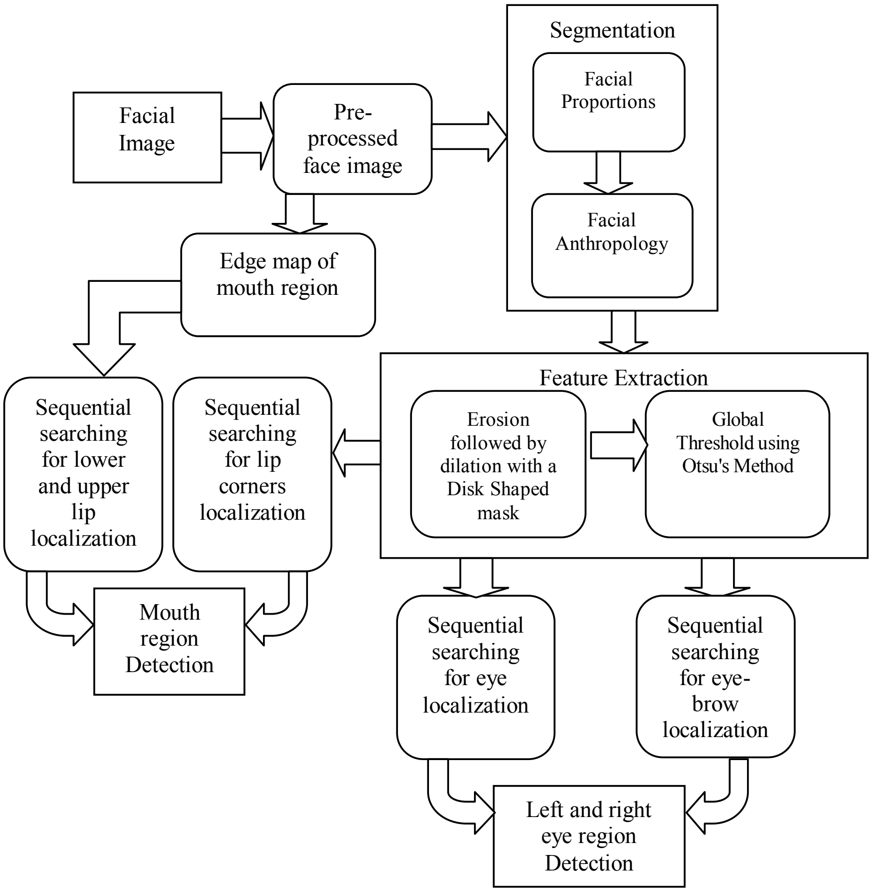

Fig. 1. Block diagram of the proposed method

From (3), it is observed that the compensation term ($\in$) is used to obtain the normalized face image from the original image. Compensation term is defined as the difference between the normalized illumination and the estimated original illumination in the logarithm domain[9]. In this paper, the value of the compensation term varies between 0.10 and 0.11 for different facial expressions. Compensation term is 0.1 for happy, sad and angry face images and 0.11 for surprised, disgusted and fearful face images. Later, a standard mask has been applied over the normalized face image to remove all unnecessary parts of the face image except the central face region. The mask has been formed using centre of the image, and radius has been chosen in such a way that it could include only the face part.

2.2. Segmentation

Segmentation is the process of partitioning an image into regions having a significant effect of easier analysis. For region based segmentation, two ideas are taken: ideal facial proportions and the facial anthropometry.

- Facial Proportions

The face is divided into horizontal three parts based on a particular characteristic of the face. In the case of East Asians people, the middle portion of the face, in general, is greater than the upper part, and that is equal to the lower part, but the upper part is less than the lower part[10]. In Caucasians, the middle third is often less than the upper third, and the middle and upper thirds are less than the lower third[11]. According to this theory, horizontal thirds have been used on input JAFFE face images that are shown in below Fig. 2. It is observed that in most of the JAFFE face images, the middle third is greater than upper third; lower third is less than upper third. Now, anthropometric measurement will be performed based on the facial proportions.

- Anthropometric Measurement

The science behind the analysis of different parts of the human body is known as anthropometry. Facial anthropometric measurement includes different distance measurements of facial landmarks. Farkas[12] described different types of measurements of the human face. Anthropometric assessment begins with the identification of landmark points on the face. Fig. 3 illustrates some of the landmarks on the face used by Farkas[13]. According to facial proportions, we have to measure only three distances on a face based on the anthropological landmarks. Four useful landmarks for our study are (tr, n, sn, gn). Different abbreviations are used to identify the landmarks based on their anatomical terms. Tr for trichion (hairline), n for nasion, sn for subnasale and gn for gnathion (the lowest point on the chin). It has been seen during the distance measurements of these landmarks on JAFFE face images that facial measures slightly differ between subjects. For example, the distance between two landmarks tr-n is varied between 45 and 55. The middle third is in the range between 55 and 65 (i.e. distance between n and sn) and the lower third varies between 45 and 50. Based on these anthropological measurements, we have segmented three facial parts. It is now easier to move towards feature extraction method.

2.3. Feature Extraction

The high curvature points can be extracted based on intensity variations[14]. The curvature shape of the eye-brow, eye and mouth, which are mostly associated with different facial expressions, appears as a structure of disk. So, to extract these facial features, a disk shaped binary mask has been created. The segmented facial parts have been eroded at first and then dilated using this mask. Mathematically, the process of erosion ε followed by dilation δ of a digital image f can be defined as morphological opening[15] and can be written as

$$\gamma\left(f\right) = \delta(\varepsilon(f)) \tag{4}$$

Then, a global threshold using Otsu's method[16] has been chosen on the resultant facial parts so that binarized facial parts can be obtained.

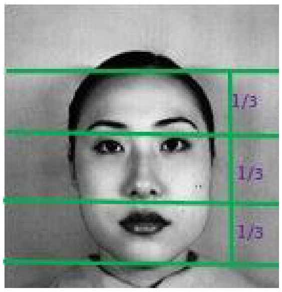

Fig. 2. Horizontal facial thirds

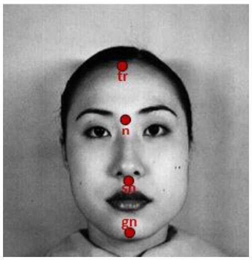

Fig. 3. Anthropometric landmarks on JAFFE face image

2.4. Detection of ROI

- Detection of eye and eye-brow region

According to the measured anthropometric landmark distances, the upper part of the face contains only eye-brows and the middle part contains eyes and nose. The following steps have been performed in the process of eye and eye-brow region detection.

After performing the morphological operations and binarization to the middle part of the face image, sequential search has been performed to find the eye corners. The Fig. 4 shows the eye candidates after performing morphological operations and binarization to the middle part of the face image. As we are only concentrating on eye region, we have discarded the nose area.

After finding the eye corners and eye-centers, it is essential to find the distance between eye and eye-brow. According to the anthropometric measurement, the upper facial part contains eye-brow. Fig. 5 represents the eye-brow after binarization.

Using sequential search from the lower part of the Fig. 5, we have measured the center of the eye-brow, so that the distance can be measured from the eye-center and eye-brow center.

Based on the eye-corners and distance between eye and eye-brow centers, one rectangle has been plotted on the whole eye and eye-brow region.

- Detection of Mouth region

The following steps have been taken to detect the mouth region.

The morphological and binarization operation on the lower facial part of the face image produce the images like shown below in Fig. 6.

Fig. 4. Eye candidates after morphological operations and binarization

Fig. 5. Eye-brow candidates after morphological operations and binarization

Fig. 6. Resultant lower facial part after the morphological and binarization operation

To detect the mouth region, we have to localize four facial feature points of the mouth i.e. two lip corners, upper lip and lower lip. But, it is observed during the experiment that the mouth information excluding lip corners has not been achieved through the previously mentioned operations only. Lip corners can be localized using sequential searching procedure. So, there is a need to take another approach that can detect the upper and lower lip points.

The locations of upper and lower lip parts often change due to different facial expressions. These two parts suffers from patches and shading effect. Due to this reason, we have constructed local edge map of the mouth region to preserve the structural information. In this paper, the local edge map is constructed using canny edge detector in horizontal direction[17]. If the weak edges of an image are linked to the strong edges, canny edge detector includes these weak edges in the output image[18]. So, here we have used canny edge detector. The Fig. 7 is the resultant edge mapped mouth region.

From the middle of the first and last row of the edge mapped mouth region, iterative sequential searching has been performed so that the upper and lower lip point can be determined. Based on the four lip points, a rectangular area has been plotted over the mouth region.

3. Experimental Results and Discussions

The performance evaluation of facial feature localization methods proposed in the literature has been given either visual inspection of the detection result or error measure. The localization error can be measured using the distance between manually annotated and automatically detected feature points. The distance is calculated in terms of Euclidean pixel distance. According to Jesorky et al.[19], if the distance between manually annotated and automatically detected feature point location is less than 1/4th of the annotated inter eye distance, then the detection result will be considered as correct. This criterion will be appropriate for feature point localization. But, for detection of feature ROI, another criterion is to be needed. Gizatdinova and Surakka[6] proposed a new ROI detection rate evaluation measure, which has been adapted in this paper for performance evaluation. According to their proposed evaluation measure, four points have to be selected to define the eye and eye-brow region, mouth region. These points define the right, left, top and bottom of these corresponding features ROI.

$$\max(d(p_{tl}, \overline{p}_{tl}), d(p_{br}, \overline{p}_{br})) \leq R \quad (5)$$

where p_{tl} and p_{br} define the coordinates of the top-left and the bottom-right boundaries of the annotated feature location $\overline{p}_{tl}$ and $\overline{p}_{br}$ define the coordinates of the automatically located feature positions. $d(p_{tl}, \overline{p}_{tl})$ and $d(p_{hr}, \overline{p}_{hr})$ are Euclidian pixel distances. If $\overline{p}_{tl}$ or $\overline{p}_{hr}$ is found inside the annotated feature bounded rectangular box, then it confirms the automatically detected feature positions within the bounded box. For a given R (R = 0, 1, 2 . . . is a real number of pixels), correct localization must satisfy the criterion from Eq. (5). The ratio between the total number of correctly located features (for a given R) and the number of testing images is the average localization rate.

The entire experiment is carried out on JAFFE face database[20]. The database contains 30 neutral and 176 expressive face images of 10 Japanese females. The performance of the proposed method has been shown in tabular form on JAFFE expressive face images. The value of R has been taken as ¼ of the inter eye distance. The average ROI localization accuracies have been listed in table 1.Some ROI detected face images are shown in Fig. 8.

Fig. 7. Edge mapped mouth region

Table 1.Performance evaluation.

Dataset	Left eye-region	Right eye-region	Mouth region
JAFFE	97%	96%	89%

This proposed method has some limitations like it cannot cope up the facial wrinkles, facial marks, shadows for which localization has been influenced. Some examples of the misclassifications have been shown in Fig. 9. Most of the researches in facial feature localization have been performed based on point error measure. But, we have measured performance based on rectangular error measure. Gizatdinova and Surakka[6] also have made their comparative study based on point error measure. They have illustrated rectangular error measure through graphical plots. From the graphs, it is observed that the average ROI detection of eye and mouth on JAFFE expressive face images is near about 90-92%. So, by comparing results, we can say that our proposed ROI detection method generates noticeable results on JAFFE expressive face images.

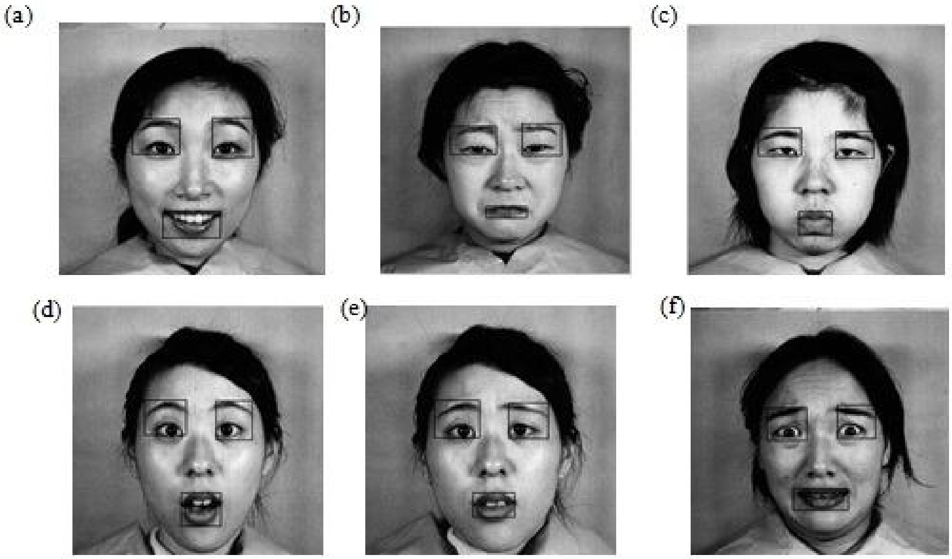

Fig. 8. ROI detection on JAFFE face images (a) happy; (b) sad; (c) angry; (d) surprise; (e) disgust; (f) fear

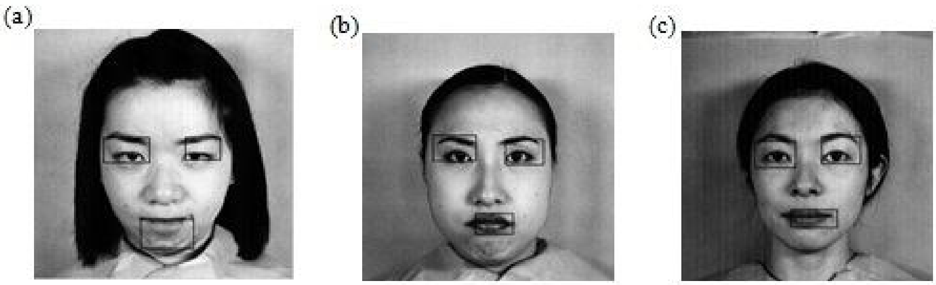

Fig. 9. Sample misclassified face images

4. Conclusion and Future Work

The paper presents an automatic ROI localization method that effectively localizes ROI on expressive face images. Illumination effect has been compensated to some extent using logarithm transform, but it cannot handle the conditions like shading effect. These problems will be considered in future work and proposed method will be evaluated on large scale expressive face datasets.

Acknowledgements

The work presented here is being conducted in the Biometrics Laboratory of Tripura University, under the research project supported by the Grant No. 12(2)/2011-ESD, dated 29/03/2011, from DeitY, MCIT, Government of India. The first author is thankful to the Chief Investigator of Biometrics Laboratory, Dr. Mrinal kanti Bhowmik for giving his kind support. The first author is also grateful to Department of Science and Technology (DST), Government of India for providing her Junior Research Fellowship-Professional (JRF-Professional) under DST INSPIRE fellowship program (No. IF131067).

References

1. Darwin C. *The expression of emotion in man and animals*. New York: Oxford University Press; 1872.
2. Campadelli P, Lanzarotti R, Lipori G, Salvi E. Face and Facial Feature Localization. Proc Int Conf Image Analysis and Processing, Italy. 2005;1002–9.
3. Heisele B, Serre T, Poggio T. A component-based framework for face detection and identification. *Int J Comput Vision* 2007;**74**:167–81.
4. Wilson P, Fernandez J. Facial feature detection using Haar classifiers. *J Comput. Sci Coll* 2006;**21**:127–33.
5. Feris R, Gemmell J, Toyama K, Krüger V. Hierarchical Wavelet Networks for Facial Feature Localization. Proc Int Conf Automatic Face and Gesture Recognition, Washington DC, 2002;118–123.
6. Gizatdinova Y, Surakka V. Automatic edge-based localization of facial features from images with complex facial expressions. *Patt Recog Letters* 2010;**31**:2436–2446
7. Adini Y, Moses Y, Ullman S. Face recognition: the problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell* 1997;**19**:721–32.
8. Horn BKP. *Robot Vision*. Cambridge: MIT Press; 1986.
9. Chen W, Joo ErM, Wu S. Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain. *IEEE Trans Sys Man Cyber* 2006;**36**:458 - 466.
10. Sim RST, Smith JD, Chan ASY. Comparison of the aesthetic facial proportions of Southern Chinese and white women. *Arch Facial Plast Surg* 2000;**2**:113–20.
11. Farkas LG, Hreczko TA, Kolar JC. Munro IR. Vertical and horizontal proportions of the face in young adult North American Caucasians: revision of neoclassical canons. *Plast Reconstr Surg* 1985;**75**:328–38.
12. Farkas L. *Anthropometric Facial Proportions in Medicine*. Thomas Books; 1987.
13. Farkas L. *Anthropometry of the Head and Face*. Raven Press; 1994.
14. Tuytelaars T, Mikolajczyk K. Local Invariant Feature Detectors: A Survey. *Computer Graphics and Vision* 2007;**3**:177-280.
15. Serra J. *Image Analysis and mathematical morphology*. London: Academic Press; 1982.
16. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Sys Man Cyber* 1979;**9**:62–6.
17. Canny J. A Computational Approach to Edge Detection. *IEEE Trans Patt Anal and Mach Intell* 1986;**8**:679–98.
18. Phimoltares S, Lursinsap C, Chamnongthai K. Face detection and facial feature localization without considering the appearance of image context. *Image and Vis Compu* 2007;**25**:741–53.
19 Jesorky O, Kirchberg KJ, Frischholz RW. Robust face detection using hausdorff distance. Proc. of the 3rd Int. Conf. Audio and Video –Based Person Authentication, Sweden, 2001;251-254.
20. Lyons MJ, Akamatsu S, Kamachi M, Gyoban J. Coding facial expressions with Gabor wavelets. Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition. 1998.