HUMAN'S FACIAL PARTS EXTRACTION TO RECOGNIZE FACIAL EXPRESSION

Dipankar Das
Department of Information and Communication Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh

ABSTRACT

Real-time facial expression analysis is an important yet challenging task in human computer interaction. This paper proposes a real-time person independent facial expression recognition system using a geometrical feature-based approach. The face geometry is extracted using the modified active shape model. Each part of the face geometry is effectively represented by the Census Transformation (CT) based feature histogram. The facial expression is classified by the SVM classifier with exponential chi-square weighted merging kernel. The proposed method was evaluated on the JAFFE database and in real-world environment. The experimental results show that the approach yields a high recognition rate and is applicable in real-time facial expression analysis.

KEYWORDS

Facial feature points, Facial expression recognition, Census transformation, SVM.

1. INTRODUCTION

Facial expression is one of the most powerful, versatile, and natural non-verbal cues for human beings to communicate their emotion and intention during interaction. Automatic facial expression analysis can be used in various application areas including human-robot interaction, attention level estimation, and data-driven animation. Due to its potential applications, automatic facial expression recognition has gained significant research interests in recent years [1, 2, 3]. However, recognizing facial expression automatically and accurately remains difficult due to the variability and complexity of human faces as well as facial expressions. Automatic facial expression recognition consists of three major steps: face detection, facial feature representation and extraction, and classification into a set of facial expression categories based on the extracted features. Model based approaches for the interpretation of faces [4] are used to extract relevant facial information. The Active Shape Model (ASM) [5] is an effective way to locate facial features, to model both shape and texture, and also to find correlation between them from an observed training set. Thus, in this research we use modified ASM to detect a face from an image and to segment it into different regions based on the detected landmark points on it. In facial feature extraction for expression analysis, there are mainly two types of approaches: geometric feature-based methods and appearance based methods. The geometrical facial features present the shapes and locations of facial components (including mouth, eyes, nose, etc.). In [6], Valstar *et.al.* have demonstrated that geometric feature-based methods give equal or better performance than appearance based methods. However, the performance of the geometric feature-based approach depends on accurate and reliable facial component detection and tracking. In the appearance-based approach, image filters, such as Gabor-wavelet, are applied to either a whole face or specific regions to extract a feature vector. Although Gabor-wavelet features and lower frequency 2D DCT coefficients are widely used in literature [7, 8], they are both time and memory intensive to extract multi-scale and multi-orientation coefficients.

DOI : 10.5121/ijit.2014.3307

In this paper, we propose a person-independent facial expression recognition system using a geometrical feature-based approach. First, it reliably detects facial components by using a modified active shape model. Then the facial regions are represented by the Census Transformation (CT) [9] based feature histograms. The most important properties of CT features are their tolerance against illumination changes and their computational simplicity, with source code publicly available. Moreover, the CT-based feature histogram has several advantages in comparison to other features for facial expression recognition: (i) it has no parameter to tune, (ii) it can be processed extremely fast (>50fps) and applicable for real-time facial expression recognition. Finally, we use the exponential χ^2 weighted merging kernel function with the support vector machine classifier to classify facial expression.

Most existing real-time facial expression recognition systems attempt to recognize facial expression from the data collected in highly controlled environment [10]. In addition, they often need some degree of manual intervention. However, in real-world applications it is important to recognize facial expression in natural environment. In this paper, we present a system that can fully automatically recognize facial expression in real-time from video images taken in natural environment.

2. FACIAL PART SEGMENTATION

We have modified the Active Shape Model (ASM) [5] for our face part detection. The ASM library [11] reliably detects a face region within an image and locates 68 landmark points on the face as illustrated in Fig. 1 (a).

Our modified version of ASM selects 37 points (Fig. 1(b)) among 68 landmark points and locates additional points on the forehead region of the face (Fig. 1(c)). To locate additional points, we consider the forehead as a half-circle above the line l_1, joining between the landmark points p_1 (point `1') and $p_2$ (point `15'). We determine the line l_2, which is perpendicular to l_1 and passes through the point p_3 (point `68'). The center of the half-circle (x_c,y_c) is determined as the intersection point of l_1 and l_2. We determine the additional landmark points (x, y) on the circumference of the circle using the equations:

$$x = x_c + r\cos\theta \tag{1}$$

$$y = y_c + r\sin\theta \tag{2}$$

where r is the radius of the half-circle (Fig. 1 (c)). These points (38 to 44) are used to determine the forehead region on the face. Finally, the combination of the landmark points in Fig. 1(b) and Fig. 1(c) are used to segment the face into 8 parts (R_1 to R_8) as illustrated in Fig. 1(d). It has been shown that most of the facial expression changes occur within these parts. Our feature extraction and subsequent classification are applied to the spatial extent of these face parts.

3. FEATURE EXTRACTION

Feature extraction converts pixel data into a higher-level representation to describe the spatial configuration of the face and its components. We need a feature descriptor to capture the general structural properties of the face. Our Census Transformation based histogram is a holistic representation that captures the structural properties by modeling distribution of local structures.

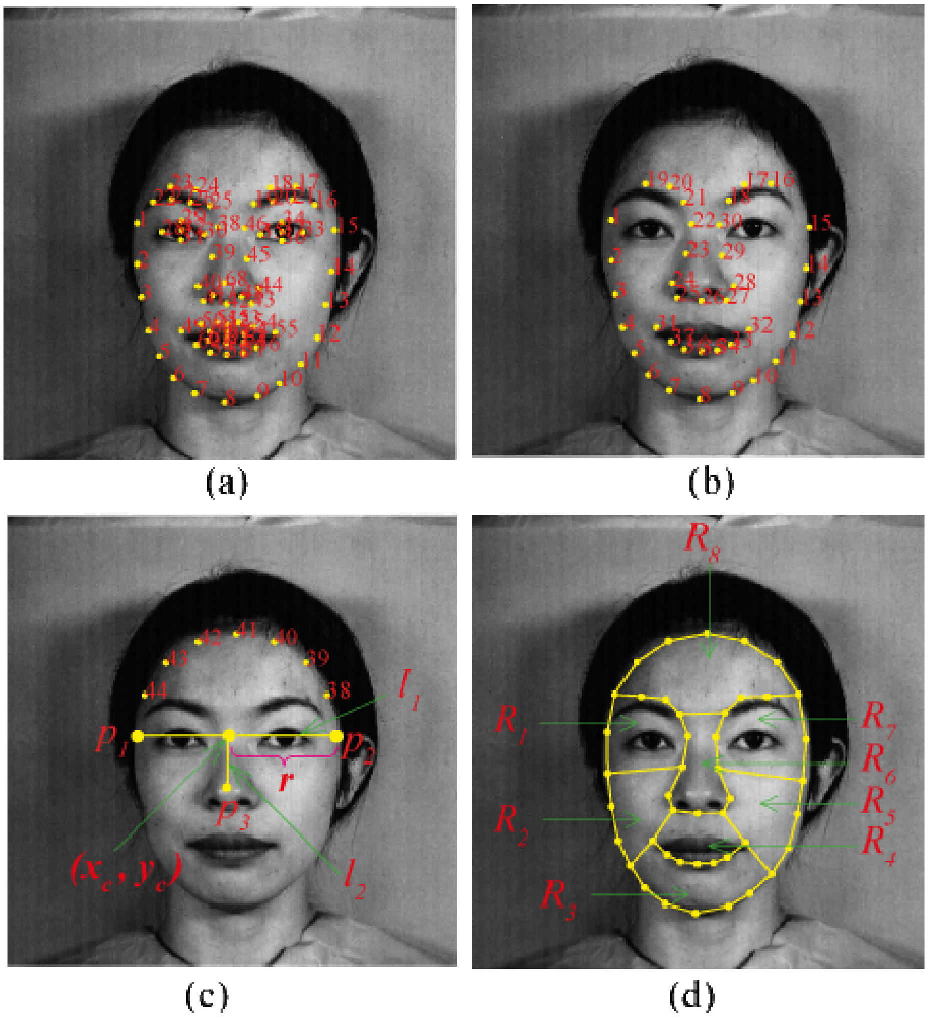
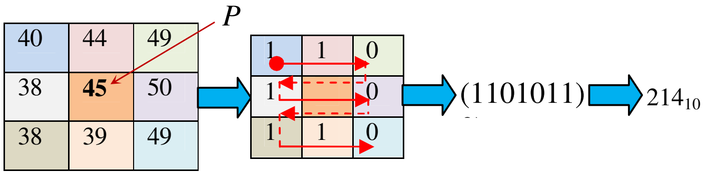


Fig. 1: Facial Region Segmentation: (a) landmark points using ASM lib, (b) selected points, (c) detected additional landmark points, and (d) segmented regions.

4. HISTOGRAM OF CENSUS TRANSFORMATION

The Census Transformation is a form of non-parametric local transform originally developed for finding correspondence between local patches [9]. The CT maps the local neighborhood surrounding a pixel P, to a bit string representing the set of neighboring pixels in a square window, as illustrated in Fig. 2.



40	44	49
38	**45**	50
38	39	49

1	1	0
1		0
1	1	0

Fig. 2: Census Transformation of a pixel.

If P is greater than or equal to one of its neighbors, `*1*' is set in the corresponding location. Otherwise, `*0*' is set. The eight bits generated from the intensity comparison can be put together in any order. Here we arrange the bits from left to right and from top to bottom. The converted CT value is a base-10 number in the range $[0,\ 2^8 - 1]$. The census transform relies solely upon the set of comparisons, and is therefore robust to illumination changes, gamma variation, etc.

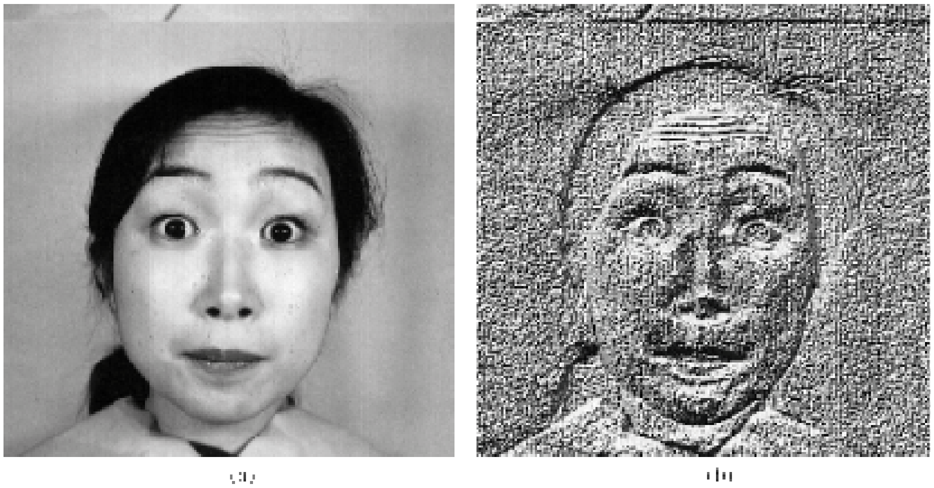


Fig. 3: Census Transformation of an image: (a) original image, and (b) CT image.

The CT retains the global structures of an image besides capturing the local structures. For example, Fig. 3 shows an original image and its CT image. CT values of neighboring pixels are highly correlated and there exist some direct and indirect constrains posed by the center pixels. Such constrains propagate to pixels that are far apart because of its transitive property. The propagated constrains make CT values and their histograms implicitly contain information for describing global structures. As shown in Fig. 2, the CT operation maps any *3*×*3* pixels into one of 2^8 cases each corresponding to a special type of local structure of pixel intensities. Here the CT value acts as an index to these different local structures. We compute histograms of CT values for different regions on the face image. Finally, these histograms are used as the visual descriptors of facial expression.

5. Facial Expression Classification

We have developed a person-independent facial expression recognition system using CT histogram features. It is observed that some local facial regions contain more useful information for expression classification than others. For example, facial features contributing to facial expressions mainly lie in such regions as eyes, mouth, and forehead. Thus, a weight can be assigned for each sub-region based on its importance. We adopt the SVM classifier with exponential χ^2-weighted merging kernel for training and testing purposes. The kernel is defined as:

$$K(X,Y) = e^{[-\sum_{i=1}^{8} \alpha_i \{(X_i - Y_i)^2 / (X_i + Y_i)\}]} \qquad (3)$$

where α is the weight for the i-th facial region R_i. To find the similarities between feature sets X_i

and Y_i in R_i. Our kernel functions use χ^2 distance on the feature histogram, as it is demonstrated to be a good distance measure for histogram comparisons than other kernels. We use the LIBSVM package in multi-class mode for our experiment.

6. EXPERIMENTAL RESULT

We performed experiments to evaluate the facial expression recognition performance of the proposed method by using the JAFFE database [12]. It consists of 213 images of Japanese female facial expressions. Ten expressers posed 3 or 4 times each of the seven basic expressions: angry, disgust, fear, happy, neutral, sadness, and surprise. The image size is 256×256 pixels.

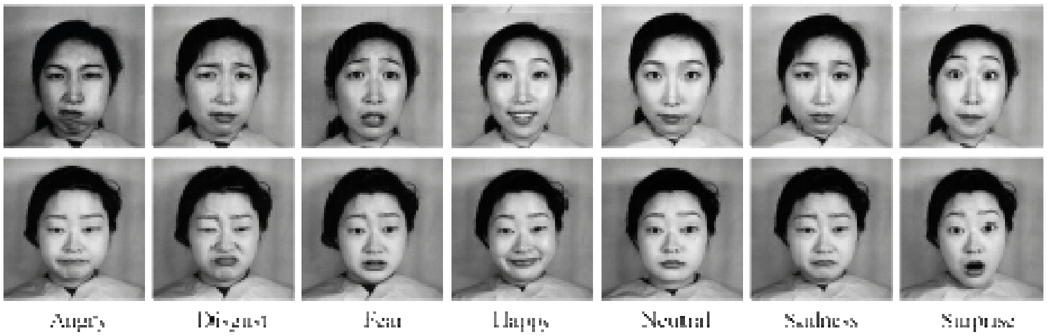


Fig. 4: Sample face expression images from JAFFE database.

7. PERFORMANCE WITHOUT EXTRACTING FACIAL REGIONS

In this experiment, we manually detected the face region and used the CT based histogram to extract features from the whole face area. The SVM classifier was trained using both *rbf* and *exponential* χ^2 kernels. Table 1 shows the SVM classifier performance using both *rbf* and *exponential* χ^2 kernels.

Table 1. Accuracy without extracting facial regions: *rbf* vs χ^2.

Kernels	angry	disgust	fear	Happy	neutral	sadness	surprise	***Average***
Rbf	70.0	72.7	54.6	72.7	50.0	70.0	50.0	**63.0**
χ^2	70.0	72.7	81.8	63.6	60.0	70.0	60.0	**68.5**

The average classification performance with *rbf* kernel is 63.0%, however the performance increases by 5.5 points for the χ^2 kernel.

8. PERFORMANCE WITH EXTRACTING FACIAL REGIONS

In this experiment, the system used images from the database and automatically detected faces and landmark points in the images as shown in Fig. 1. Based on the detected landmark points, the face was segmented into 8 facial regions. Then the CT based histograms from different regions were used to train and test the SVM classifier with *rbf* and exponential χ^2 merging kernel with equal weight. shows the SVM classifier performance using both *rbf* and *exponential* χ^2 merging kernels with equal weight ($\alpha_i; i = 1, \dots, 8$). shows the SVM classifier performance using both *rbf* and *exponential* χ^2 merging kernels with equal weight ($\alpha_i; i = 1, \dots, 8$). The average classification performance with *rbf* and χ^2 merging kernel are 72.1% and 77.3%, respectively.

Table 2. Accuracy with extracting facial regions: *rbf* vs χ^2 with equal weight.

Kernels	angry	disgust	fear	Happy	neutral	sadness	surprise	***Average***
Rbf	74.0	73.7	67.6	81.7	68.8	74.9	63.8	**72.1**
χ^2	79.8	78.7	85.8	79.6	68.7	78.8	69.7	**77.3**

However, using the and χ^2 weighted merging kernel with the weight α_i for the facial region R_i ($i = 1, \dots, 8$) as shown in Table 3, we obtained the best performance. In this case, our classification performance is indicated in the Table 4.

Table 3. Weight for different facial regions.

α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
0.20	0.10	0.13	0.18	0.08	0.04	0.17	0.10

Table 4. Performance of our method with weighted merging kernels.

Kernels	angry	disgust	fear	Happy	neutral	sadness	surprise	***Average***
χ^2	95.5	86.7	93.8	85.2	77.7	87.8	77.5	**77.3**

9. PERFORMANCE COMPARISION WITH OTHER METHODS

The final average recognition rate of our method, using χ^2 weighted merging kernel, is comparable with some of the previous approaches as shown in Table 5. With LBP feature, Shan *et al.* [1] reported the performance of 81% and Liao *et al.* [13] obtained 85.6% classification performance on the JAFFE database. Both previous approaches required some level of manual segmentation of face images and were not applicable for real-time expression recognition.

Table 5. Comparison with other methods.

Approaches	Classification accuracy
Liao *et al.* [13]	85.6
Shan *et al.* [1]	81.0
Authors	86.3

However, we obtain 86.3% accuracy and our approach does not require any manual intervention. The better result is due to the automatic extraction of facial regions and the use of optimized weighted value for each region.

10. REAL-TIME PERFORMANCE ANALYSIS

We checked whether or not the system could work in real time by using video data. Figure-5 shows a snapshot in the experiment with detected regions on the face. The system can recognize facial expression at average 14.86 *fps* (face tracking module run at 21 *fps*, feature extraction (CT) module at 53 *fps*, and recognition (SVM) module at 1176 *fps*) on a 640 ×480 video image.

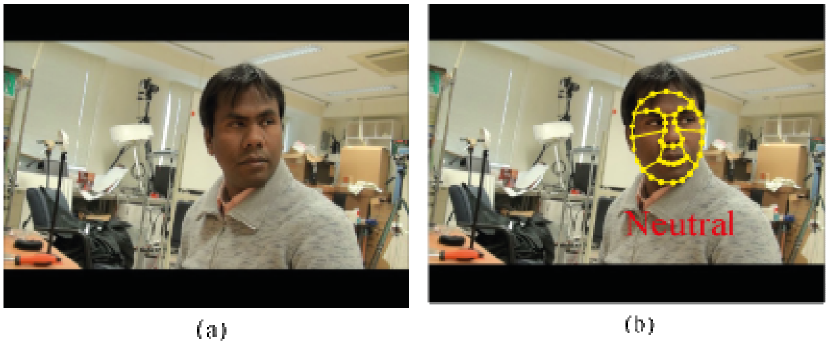


Fig. 5: Real-time facial parts extraction and facial expression recognition.

11. CONCLUSION

In this paper, we present a facial expression recognition system based on the Census Transformation and the geometrical feature-based approach. Extracting different parts of a face and constructing an effective facial representation are vital steps for successful facial expression recognition. The modified ASM successfully extracts different parts of a face. The CT based histogram provides an effective representation of facial expression. Our system performs better than the existing ones for the JAFFE dataset and works in real time on video data. In future, we will compare our system with other approaches on video data.

REFERENCES

[1] C. Shan, S. Gong and P. W. McOwan (2009), "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vision Comput.,* vol. 27, no. 6, pp. 803-816.

[2] G. R. S. Murthy and R.S.Jadon (2009), "Effectiveness of eigenspaces for facial expressions recognition," *Int. Journal of Comp. Theory and Eng.,* vol. 1, no. 5, pp. 638-642.

[3] K.-T. Song and Y.-W. Chen (2011), "A design for integrated face and facial expression recognition," in *Annual Conf. on IEEE Industrial Electronics Society*, Melbourne, Australia.

[4] M. H. Mahoor, M. Abdel-Mottaleb and A.-N. Ansari (2006), "Improved active shape model for facial feature extraction in color images," *J. Multimedia,* vol. 1, no. 4, pp. 21-28.

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham (1995), "Active shape models-their training and application," *Comp. Vis. and Image Understand,* vol. 61, no. 1, pp. 38-59.

[6] M. Valstar and M. Pantic (2006), "Fully automatic facial action unit detection and temporal analysis," in *CVPRW*, New York, NY, USA.

[7] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel and J. R. Movellan (2005), "Recognizing facial expression: Machine learning and application to spontaneous behavior," in *CVPR*.

[8] L. Ma, Y. Xiao, K. Khorasani and R. Ward (2004), "A new facial expression recognition technique using 2d dct and k-means algorithm," in *ICIP*.

[9] R. Zabih and J. Woodfill (1994), "Non-parametric local transforms for computing visual correspondence," in *ECCV*.

[10] K. Anderson and P. W. McOwan (2006), "A real-time automated system for the recognition of human facial," *IEEE Trans. on Systems, Man, and Cybernetics, PartB,* vol. 36, no. 1, pp. 96-105.

[11] Y.Wei (2009), "Research on facial expression recognition and synthesis," in *Master's thesis*.

[12] S. Liao, W. Fan, A. C. S. Chung and D.-Y. Yeung (2006), "Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features," in *ICIP*.

[13] M. J. Lyons, J. Budynek and S. Akamatsu (1999), "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 21, no. 12, pp. 1357-1362.

Authors

Dipankar Das received his B.Sc. and M.Sc. degree in Computer Science and Technology from the University of Rajshahi, Rajshahi, Bangladesh in 1996 and 1997, respectively. He also received his PhD degree in Computer Vision from Saitama University Japan in 2010. He was a Postdoctoral fellow in Robot Vision from October 2011 to March 2014 at the same university. He is currently working as an associate professor of the Department of Information and Communication Engineering, University of Rajshahi. His research interests include Object Recognition and Human Computer Interaction.