# Facial Expression Recognition



Shalini Bhatia
University of Canberra

# Facial Expression Recognition

Neeta Sarode
Computer Engineering Dept.
Thadomal Shahani Engineering College,
Bandra (West), Mumbai-400050

Prof. Shalini Bhatia
Computer Engineering Dept.
Thadomal Shahani Engineering College,
Bandra (West), Mumbai-400050

*Abstract*— **Facial expression analysis is rapidly becoming an area of intense interest in computer science and human-computer interaction design communities. The most expressive way humans display emotions is through facial expressions. In this paper a method is implemented using 2D appearance-based local approach for the extraction of intransient facial features and recognition of four facial expressions. The algorithm implements Radial Symmetry Transform and further uses edge projection analysis for feature extraction and creates a dynamic spatio-temporal representation of the face, followed by classification into one of the expression classes. The algorithm achieves an accuracy of 81.0% for facial expression recognition from grayscale image.**

***Keywords- grayscale images; face; facial expression recognition; lip region extraction; human-computer interaction.***

## I. INTRODUCTION

Although the automated recognition of facial expressions has been studied with much interest in the past 10 years [1, 2, 3], it is still a challenging task for a computer program. State-based representation of facial expressions has been investigated by other researchers, such as Tian *et al.* [4]. However, the impact of such representation on recognition robustness has not received much attention. The method introduced in this paper evaluates experimentally, an approach based on a discrete representation of spatially-localized facial dynamics, comprising a small number of states. The approach is anchored on a geometric facial model coupled to the analysis of feature point motion, followed by a state-based representation of the dynamics of expressions. In general there are two approaches to represent the face and consequently the facial features to perform facial expression analysis: the geometric feature-based methods and appearance-based methods. The geometric facial feature-based methods present the shape, texture and/or location information of prominent components such as the mouth, eyes, nose, eyebrow, and chin, which can cover the variation in the appearance of the facial expression. The appearance-based methods, on the other hand, using image filters such as Gabor wavelets, generate the facial feature for either the whole-face or specific regions in a face image. Fiducial points are a set of facial salient points, usually located on the corners of the eyes, corners of the eyebrows, corners and outer mid points of the lips, corners of the nostrils, tip of the nose, and the tip of the chin. Automatically detecting fiducial points can extract the prominent characteristics of facial expressions with the distances between points and the relative sizes of the facial components and form the feature vector. Using fiducial points to model the position of the prominent features one can symbolize the face geometry in a local manner. The number of fiducial points used varies and mainly depends on the desired representation, as it is reported that different positions hold different information regarding the expressions. Additionally, choosing the feature points should represent the most important characteristics on the face and be extracted easily. In other words, the number of feature points should represent enough information and not be too many.

Black and Yacoob [12] use a local parameterized model of image motion obtained from optical flow analysis. They utilize a planar model for rigid facial motion and an affine-plus-curvature model for non rigid motion. Essa and Pentland [9] first locate the nose, eyes and mouth. Then, from two consecutive normalized frames, a 2D spatio-temporal motion energy representation of facial motion is used as a dynamic face model. Cohn *et al.* [5] use feature points that are automatically tracked using hierarchical optical flow method. The feature vectors, used for the recognition, are created by calculating the displacement of the facial points. The displacement of a point is obtained by subtracting its normalized position in the first frame from its current normalized position. Tian *et al.* [4] proposed a feature-based method, which uses geometric and motion facial features and detects transient facial features. The extracted features (mouth, eyes, brows and cheeks) are represented with geometric and motion parameters. The furrows are also detected using a Canny edge detector to measure orientation and quantify their intensity. The parameters of the lower and upper face are then fed into separate neural networks trained to recognize AUs. In most facial expression recognizers, facial feature extractions followed by classification into an expression class. Six basic expression classes, defined by Ekman [7], are often used (Figure 1).

This paper describes a more robust system for facial expression recognition from static images using 2D appearance-based local approach for the extraction of intransient facial features, i.e. features such as eyebrows, lips, or mouth, which are always present in the image, but may be deformed [1] (in contrast, transient features are wrinkles or bulges that disappear at other times). The main advantage of such an approach is low computational requirements.

Figure 1: Basic Facial Expressions

To place the work in a larger context of face analysis and recognition, the overall task requires that the part of the image involving the face be detected and segmented. It is assumed that a near-frontal view of the face is available. Tests on a grayscale image database demonstrate a good recognition rate for four facial expressions (happy, surprise, sad and anger against neutral).

The rest of the paper is organized as follows. A description of image segmentation and feature extraction is given in the next section, which is followed by a section dealing with expression classification. Experimental results are then presented and discussed. Finally, conclusions are drawn.

## II. IDENTIFICATION OF FEATURE S AND FEATURE EXTRACTION

### A. Feature Identification

First, the face area was filtered by a gaussian filter to cope with noise and lighting variations. A rotationally symmetric gaussian lowpass filter is applied. We get a gaussian filtered image. After that, the left and right eye regions are roughly segmented and the radial symmetry transform is applied.

The transform is calculated over a set of one or more ranges *N* depending on the scale of the features one is trying to detect. The value of the transform at range $n \in N$ indicates the contribution to radial symmetry of the gradients a distance *n* away from each point. Whilst the transform can be calculated for a continuous set of ranges this is generally unnecessary as a small subset of ranges is normally sufficient to obtain a representative result. At each range *n* an *orientation projection image* $O_n$ and a *magnitude projection image* $M_n$ are formed. These images are generated by examining the gradient g at each point p from which a corresponding *positively-affected pixel* P+ve(P) and *negatively-affected pixel* P−ve(P) are determined, as shown in Figure 2.

The *positively-affected pixel* is defined as the pixel that the gradient vector g(P) is pointing to, a distance *n* away from p, and the *negatively-affected pixel* is the pixel a distance *n* away that the gradient is pointing directly away from.

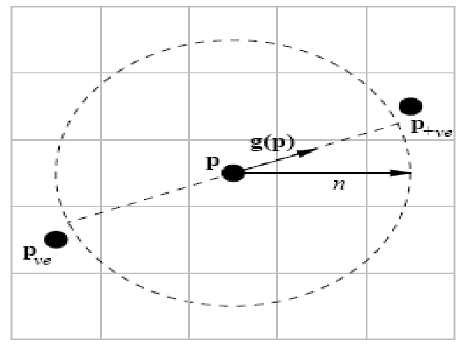


Figure 2: The locations of pixels P+ve(P) and P−ve(P) affected by the gradient element g(P) for a range of n = 2. The dotted circle shows all the pixels which can be affected by the gradient at P for a range n.

The coordinates of the positively-affected pixel are given by

$$P + ve(P) = P + round\left(\frac{g(P)}{\|g(P)\|}n\right) \qquad (1)$$

while those of the negatively-affected pixel are

$$P - ve(P) = P - round\left(\frac{g(P)}{\|g(P)\|}n\right) \qquad (2)$$

where 'round', rounds each vector element to the nearest integer.

The orientation and projection images are initially zero. For each pair of affected pixels the corresponding point P+ve in the orientation projection image $O_n$ and magnitude projection image $M_n$ is incremented by 1 and $\|g(P)\|$ respectively, while the point corresponding to P−ve is decremented by these same quantities in each image.
That is

$$O_n(P + ve(P)) = O_n(P + ve(P)) + 1 \qquad (3\text{ a})$$

$$O_n(P - ve(P)) = O_n(P - ve(P)) - 1 \qquad (3\text{ b})$$

$$M_n(P+ve(P)) = M_n(P+ve(P)) + \|g(P)\| \quad \text{(3 c)}$$

$$M_n(P-ve(P)) = M_n(P-ve(P)) - \|g(P)\| \quad \text{(3 d)}$$

The radial symmetry contribution at a range *n* is defined as the convolution:

$$S_n = F_n * A_n \quad (4)$$

Where,

$$F_n(P) = \left\| \tilde{O}_n(P) \right\|^{(\alpha)} \tilde{M}_n(P),$$

$$\tilde{O}_n(P) = \frac{O_n}{\max_p\{\|O_n(P)\|\}},$$

$$\tilde{M}_n(P) = \frac{M_n}{\max_p\{\|M_n(P)\|\}},$$

$\alpha$ is the radial strictness parameter, and *An* is a two-dimensional Gaussian.

The full transform is defined as the sum of the symmetry contributions over all the ranges considered,

$$S = \sum_{n \in N} S_n \quad (5)$$

If the gradient is calculated so it points from dark to light then the output image S will have positive values corresponding to bright radially symmetric regions and negative values indicating dark symmetric regions. The symmetry maps are then filtered and we get the location of right and left eye. Once the left eye and right eye are located, the in-plane rotation of the head is first normalized so that the two eyes are positioned at the same horizontal line. The search region of nose and mouth can also be roughly located with reference to the eye locations and the distance between eyes.

Let the distance between eye centers be denoted as DEyes, with its top left corner positioned at the lower bottom boundary of the left eye, the width and height of search region are set as 1.25*DEyes and DEyes, respectively. Figure 3 shows illustration of the search region for nose and mouth location, where the region has been marked with a black rectangle. To efficiently locate nose and mouth in the search region, the following method is implemented.

1. The Sobel edge operator is applied to the search region.

2. Then, horizontal projection of the edge strength is performed. Since strong intensity variations exist at nostril and mouth, peaks occur at the positions of nostril and mouth.

3. Integral projections of the edge map of the search region are been used for identification and extraction of facial features using the following equations.

Let I (x, y) be the input image. Vertical and horizontal projection vectors in the rectangle [x1, x2] x [y1, y2] are defined as:

$$V(x) = \sum_{y=y1}^{y=y2} I(x,y) \quad (6)$$

$$H(y) = \sum_{x=x1}^{x=x2} I(x,y) \quad (7)$$

4. Median filtering followed by Gaussian smoothing smooths the projection vectors so obtained. Higher value of projection vector at a particular point indicates higher probability of occurrence of the feature.

Eyebrow and eye are segmented after being located on the face. Its edges form closed contours, obtained by applying *Laplacian of Gaussian* operator at zero threshold. These contours are filled and the resulting image containing masks of eyebrow and eye is morphologically filtered. Similarly the nose and the mouth are also extracted. For tracking the extreme points of the lips the bounding box is different from that of the eye and eyebrow.

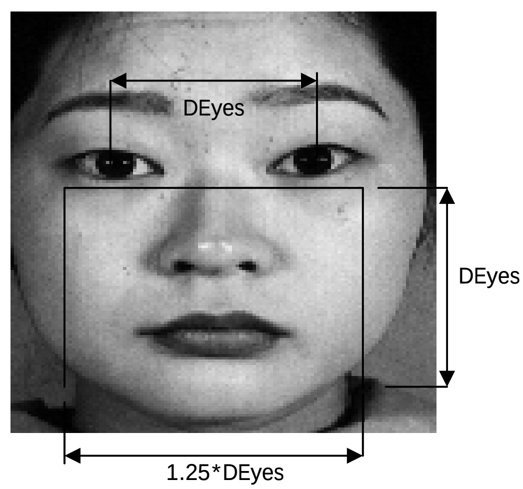


Figure 3: Search regions for nose and mouth location

## III. FEATURE VECTOR AND CLASSIFICATION

A spatio-temporal representation of the face is created, based on geometrical relationships between features using Euclidean distance (Figure 4). Each geometric parameter is used to create a set V of 5 spatio-temporal features for each image:

$$V = \{Vd0 \;\; Vd1 \;\; Vd2 \;\; Vw \;\; Vh\}; \qquad (8)$$

where,
Vdo = distance of eyebrows,
Vd1 = distance between right eyebrow and nose tip,
Vd2 = distance between left eyebrow and nose tip,
Vw = mouth width,
Vh = mouth height.

The above components of the feature vector are created by taking the difference between the geometric coefficients i.e. the geometric relationships using Euclidean distance between feature points. The feature vector formed is just the geometrical model of a facial image. The difference of statistical features is generated. The differences are obtained by subtracting statistical features of neutral expression from statistical features of corresponding person for every expression. The change in feature vector V when the face undergoes change from neutral state to some expressional state is given by:

$$\Delta V = \{\Delta Vd0 \; \Delta Vd1 \; \Delta Vd2 \; \Delta Vw \; \Delta Vh\};$$

Such dynamic characteristic of the feature vector provides shape independence. **ΔF** serves as an input for classification. It is observed that each component in the feature vector that is obtained transforms into one of the possible states: 'increase', 'stable', or 'decrease'.

A single state is calculated for each facial parameter, across an image representing a facial expression. Normally it is observed that for 'happy' expression the mouth width increases against 'neutral' image of the same person. In this way every parameter either increases, decrease or remain stable against the 'neutral' image. Depending on the possible states: 'increase', 'stable', or 'decrease' that is the information which is received from fourth stage, the image is classified as: 'happy', 'surprise', 'sad' or 'angry'.

Distance between eyebrows- stable; distance between right eyebrow and nose tip- stable; distance between left eyebrow and nose tip- stable; mouth width - increase; mouth height - increase; is classified as 'happy' face. Expressions that are to be identified can be done with the minimal information about the spatial arrangement of features as shown in the Figure 5.

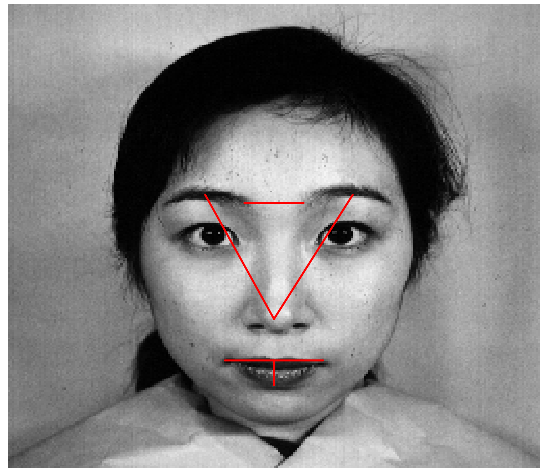

Figure 4. Geometrical parameters of the face, forming the feature vector.

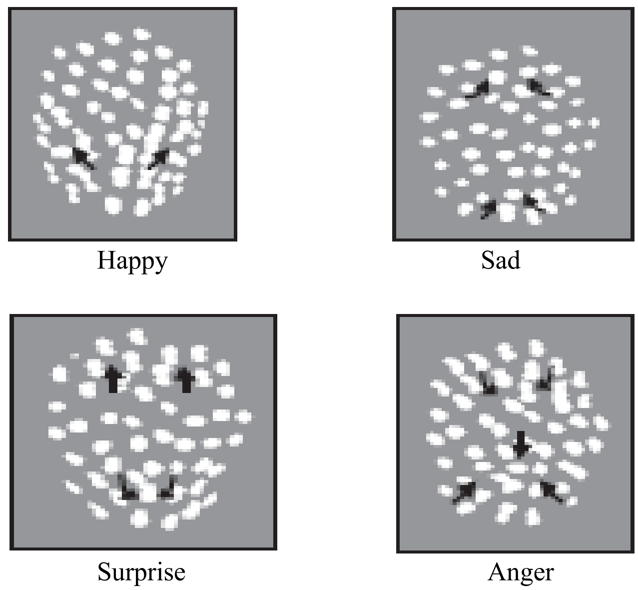


Figure 5. The cues of facial expressions

Figure 6: A) Original image of lips (Happy)
B) Extraction of lips

Figure 7: Some of the images from JAFFE database

## IV. EXPERIMENTAL RESULTS

Experiments are performed on grayscale image databases. Images from Yale facial image database and JAFFE database (Figure 7) are used for experiments.

JAFFE database consists of grayscale images. The database consists of Japanese Female Facial Expressions that have 7 expressions of 10 people including neutral. Each person has 3-4 images of same expression, so the total number of images in the database comes to 213 images.

For classification, 150 images of 10 people having 5 different facial expressions are taken. The images for expressions: 'neutral', 'happy', 'surprise', 'sad', 'anger' is taken. The database is organized in the same sequence given above for each person. This database was used to test the accuracy of the facial expression recognition algorithm. Each grayscale image sequence in the database depicted one of the expression classes (happy, surprise, sad and anger against neutral). The first image in the sequence was a neutral image. Confidence level of each expression was calculated for each of the subsequent images against the neutral image. The calculated vector of confidence levels was added to give total confidence for each of the expressions. It showed an accuracy of about 83% for expressions of happy and surprise and an accuracy of about 78% for expressions of anger and sad (Figure 8).

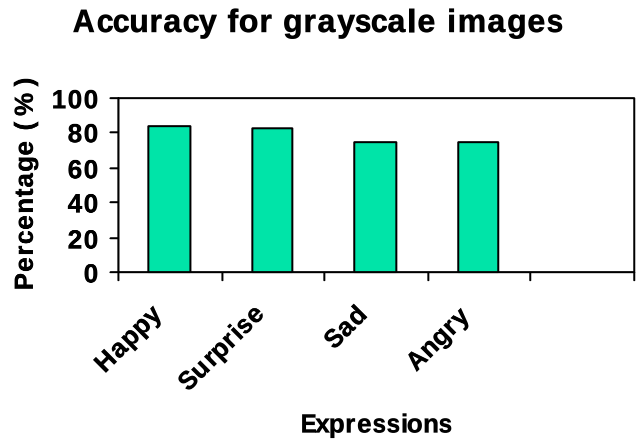


Figure 8: Accuracy of facial expression recognition

## V. CONCLUSION

An efficient, local image-based approach for extraction of intransient facial features and recognition of four facial expressions was presented. In the face, we use the eyebrow and mouth corners as main 'anchor' points. It does not require any manual intervention (like initial manual assignment of feature points). The system, based on a local approach, is able to detect partial occlusions also.

## VI. REFERENCES

[1] G. Donato, M. S. Barlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying Facial Actions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974– 989, October 1999.

[2] M. Pantic and Rothkrantz L. J. M, "Automatic Analysis of Facial Expressions: The State of the Art", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424– 1445, December 2000.

[3] A Samal and A. P. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey", *Pattern Recognition*, 25(1):65–77, January 1992.

[4] Y. Tian, T. Kanade, and J. F. Cohn, " Recognizing Action Units for Facial Expression Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, February 2001.

[5] J. Cohn, A. Zlochower, J. Lien, and T. Kanade, "Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression", *Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 396–401, 1998.

[6] Ashutosh Saxena, Ankit Anand and Amitabha Mukherji, "Robust facial expression recognition using spatially localized geometric model", International Conference on Systemics, Cybernetics and Informatics, February 12–15, 2004.

[7] Paul Ekman, American Psychologist, "Facial Expression and Emotion", April 1993, Vol. 48, No. 4, pp. 384-392.

[8] Liu. C & Wechsler. H (2002), "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", IEEE Trans. Image Processing, Vol. 11.467-476.

[9] I. A. Essa and A. P. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, July 1997.

[10] Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen,& Jian-Cheng Huang.(2005), "A New Facial Expression Recognition Method Based on

Local Gabor Filter Bank and PCA plus LDA", International Journal of Information Technology. 11(11). 86-96.

[11] Guanming Lu, Xiaonan Li, & Haibo Li (2008), "Facial Expression Recognition for Neonatal Pain Assessment", IEEE Int. Conference Neural Ne tworks & Signal Processing, Zhenjiang, China.

[12] M. J. Black and Y. Yacoob, "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion", *International Journal on Computer Vision*, 25(1):23–48, October 1997.

[13] M. Valstar and M. Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis", *Computer Vision and Pattern Recognition Workshop*, June 2006.

[14] Lin-Lin Shen, Zhen Ji, "Modelling Geiometric Features For Face Based Age Classification", Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008.

[15] JAFFE Face Database. Available: [Online] http:// www.kasrl.org/**jaffe**.html