

Automatic Facial Expression Analysis and Methods

Pranav Kumar¹

¹Department of CSE,
Birla Institute of Technology,
Mesra, Ranchi-835215, India

Kanchan Bala²

²Department of CSE,
Banasthali University, Jaipur
Rajasthan- 304022, India

G. Sahoo³

³Department of IT,
Birla Institute of Technology,
Mesra, Ranchi-835215, India

Abstract - Past few years, facial expression recognition becomes an active research area. Recent advances in image analysis and pattern recognition, it becomes possible to automatic detection and classification of emotion facial signal. This paper surveys the technique present in literature in the past decade for facial expression analysis by a computer in facial image and image sequences.

Key Words— Facial expression recognition; Facial feature extraction; Facial expression classification.

1. INTRODUCTION

Human express their emotion by facial expression, gesture etc. but facial expression play important role in prediction of emotion. The development of artificial intelligence, people pays more attention to facial expression recognition which is important technology of human-machine interaction.

In 1872 Darwin demonstrated facial expression in man and animal are specific inborn emotion. 1971 Ekman and Friesen define six basic emotions: anger, sadness, fear, disgust, surprise and happiness. Then later 1978 the Facial Action Coding System (FACS) unit developed for detection of changes in facial expression. This system divided face into different type of Action unit (AU) to describe the facial movements. Mase and Pentland used 8 directions of optical flow changes to detect the movement of FACS in 1991. Recently there are many methods for automatic recognition of facial expression in image. Facial motion analysis[1], measurement of the facial feature and their spatial arrangement[2], gray level pattern analysis using spatial filter[3], relating face images to physical models of facial skin and musculature[5] and principle component analysis(PCM)[4] in holistic spatial pattern analysis are approaches that have been explored. More general, distinguished the facial expression analysis system by different methods used in face detection, facial feature extraction and expression classification. This paper more emphasize on facial feature extraction and expression classification and goal is to survey work done in automatic facial expression analysis in image and image sequence.

2. FACE DETECTION

Face detection is special case of object-recognition and general case of face localization. Different kinds of images as input such as facial images, arbitrary images in

face detection. Detecting face in image or image sequence has two approaches- holistic approach, face is determined as whole unit and second approach, analytic approach; face is detected by some facial feature first then locates other feature corresponding to detected feature [6].

Face detection in face image, to represent the face by point distribution model (PDM) [7]. Detect the face as a whole by Pantic and Rothkrantz [8] they use algorithm based on HSV color model. Yoneyama [9], Kobayashi and Hara [10] use analytic approach to detect the face in face image. Kimura and Yachida et al. [11] use potential net for face representation.

Face detection in arbitrary image uses different type of algorithm. Spatial-temporal filtering, stereo algorithm is used by Hong to track face in arbitrary image. Principal Component Analysis (PCA) is used by Pentland to locate the face in arbitrary image.

3. FACIAL EXPRESSION DATA EXTRACTION

Generally, three type of face representation in facial expression analysis: Holistic, analytic and hybrid. Template based or feature based method face model applied for facial expression data extraction. Template based method used in holistic based method while feature based method used in analytic face model in input image.

In holistic approach, PDM by Huang, labeled graph by Hong, Random block eigen vectors by Padgett, optical flow by Black and Otsuka, AAM by Timothy are methods used while in analytic approach, Optical flow by Cohn, Frontal view point based model by Zhao, Dual-view point based model by Pantic FCP's model and 13 vertical lines by Hara are method used.

In hybrid approach labeled graph by Wang, Potentail net by Kimura, optical flow (whole face) by Essa, Fiducial point and Gabor wavelet by Zhang, Quadratic grid by Yoneyama, Fiducial grid and Gabor wavelets by Lyons are method used.

3.1 Static Image: Template-based method

Timothy et al [12] use Active Appearance Model (AAM) to represent the face. The model is built by shape variation with model of texture variation. Eigen analysis is applied to build a texture model. Procreates analysis is used for shape model. Finally, the correlation between shape and texture generate Active Appearance Model (AAM). Huang and Huang [7] use point distributed model (PDM) to

represent the face. A modified PDM facial feature extraction parameter analysis of each facial feature (eyes, mouth, eyebrows) is used. A modified PDM is generated from 90 facial feature points manually located on 90 images of Chinese subjects. Proposed method uses combination of PSM and mouth template. Thus proposed method is as close to a feature based as template based model.

Hong [14] uses Elastic Graph Matching Method in which feature of image is described in the form of a vector called jet. A jet contains 40 components. Each component is filter response of Gabor wavelets of specific 5 frequencies and 8 orientations. Jets are taken at different feature points of the image. These jets form a graph which represents the object in the image. In order to automatically create a graph for a new image, General Face knowledge (GFK) structures are used. GFK are consisting of a small gallery of images whose graphs were created with nodes placed at the exact position of feature point in the image.

Padgett and Cottrell [15] utilize three distinct representation of face, a whole face representation (Eigen face), a more localized representation of face based on the eyes and mouth areas (Eigen feature) and finally eye and mouth areas obtained from 32*32 random image patches (Eigen vector). They use facial emotion database of 97 digitalized images of 12 individuals (6 male, 6 female). Each image was linearly stretched over the 8 bit grayscale range. To achieve scale invariance, each image was scaled. Mouth and eye templates were constructed from images. Most correlated template was used to localized the feature. Area around the each eye was divided into two vertically overlapping 32*32 blocks and mouth area divided into three horizontally overlapping randomly selected from images. In order to obtain eigenvector, PCA was applied on randomly selected pixel block. NN (Neural Network) was used to classify the facial expression.

Zhengyong Zhang et al [16] represent the facial expression by hybrid method. They use 213 images of female facial expression. All are frontal pose. Each image was represented in two ways first use 34 fiducial points selected by manually. Second way was 2D Gabor transforms. A Gabor kernel of 3 spatial frequencies and 6 orientations was used.

3.2 Static Image: Feature based method

Hiroshi et al [17] use 30 x- and y- coordinates of facial characteristic points representing 3 face components (eyes, mouth and eyebrows). They take 172 facial images of 6 basic facial expressions from 30 clients by CCD camera. Among 44 Action Unit (AU) only 39 AUs are directly associated with movements of eyes, eyebrows and mouth. By these three components determine the facial characteristic points (FCP). These FCP are representative of the boundary between these three components and skin. In their later work [10] they use CCD camera in monochrome mode for brightness distribution. Select 13 vertical lines across FCP in area of face organ such as eyes, eyebrows and mouth. Normalize the face image by affine transformation so that distance between left and right irises becomes 20 pixels.

Pantic and Rothkrantz [8] use a point-based model with two 2D facial views (Frontal and Side view). Frontal view facial model is divided into two groups. The first one is formed by 25 features. Second one is formed by five features. Side view is composed of 10 face points. Points are corresponding with the peak and valleys of the curvature of the contour function. After localization of contour of prominent feature, extract the model features in dual view. They apply multiple feature detectors for each prominent facial feature (eyebrows, eyes, nose and mouth).

3.3 Image Sequences: Template-based method

Facial expression analysis from image sequences uses a holistic or hybrid approach to represent face and applies a template-based method for facial data extraction from image sequence.

Black and Yacoub [18] utilize the local parameter model of image motion for facial expression analysis. Parametric flow models are used for estimating motion in ridge scenes. Local region in space and time, such model used for non-ridge facial motion accurately. These parameters are related to motion of facial feature during facial expression. The image motion of facial features (eyes, eyebrows and mouth) are modeled with respect to head motion by different parametric models. For the eyes, affine model is used. For the eyebrows and mouth, an affine model is augmented with additional curvature parameter is used. The eyes, eyebrows and mouth are localized. Then estimate ridge motion of face region between two frames by planar motion model. The motion estimates of facial feature and face are used to predict their location in next frame.

Otsuka and Ohya [19] use Hidden Markov Model (HMM) with continuous output probabilities to extract a temporal pattern of facial motion. They use 180 image sequences of three male subjects. A face image sequence is transformed into a feature vector sequence. A feature vector is obtained by applying wavelet transformed to an image in image sequence. Then the feature vector sequence is converted into symbol sequence by codebook and codebook is obtained by vector quantization (VQ). By HMM parameters select the best match sequence. For effectiveness of the proposed method, facial motion image sequences taken by a camera fixed to the helmet worn by subject. The average frames in images in sequence are 50 to 60 at 30 frames per second. From each frame, right eye and mouth with 256*256 pixels are cut out.

Wang et al [20] propose a technique of quantitative analysis that estimates degree of facial expression in frame of image sequence. Degree of facial expression change can be measured by displacement of facial feature points. In order to achieve this goal, first use B-spline curve to construct expression change model for feature in response of expression change. The use labelled graph matching to track these facial features from an image sequence. Select the end point of facial feature (eyes, eyebrows, and mouth) as feature points (FFP). Wang uses 19 FFP; out of these 7 are work as local topology and remaining 12 FFPs for facial expression. To track these FFPs in the rest of frames, use a system which consist of two layers (memory layer and input layer). FFPs tracked between two consecutive frames by

labeled graph matching. Current frame is treated as input layer where antecedent frame is treated as memory layer.

Klmura and Yachida [21] use a method that is based on elastic net model. Consider the facial image as whole pattern of a potential field. Variation of facial expression is represented as motion vector using elastic net. Then the motion vector is mapped to low dimensional eigen space by applying K-L expansion.

3.4 Image Sequence: Feature based method

Automatic facial expression analysis from image sequence uses an analytic approach to represent the face and for representation of expression extraction information input image uses a feature based method.

Cohn et al [22] propose a model of facial landmark points around the facial feature. These landmark points are marked by manually with the help of mouse of computer in first frame. Rest of frames is examined by hierarchical optical flow method in image sequence. Displacement of landmark points is calculated by subtracting their normal position in first frame from its current frame (all frame of input sequence are normalized manually). Thus displacement vectors are calculated between initial and peak frame. These displacement vectors are used to represent facial information for facial expression recognition. Face should be without glasses/hair and first frame should be an expressionless face. Landmark points are marked on first frame manually.

4. FACIAL EXPRESSION CLASSIFICATIONS

Classification method are classified the facial expression convey by the face. Mechanism of classification can be either rule-based or a neural network based or template based classification method. When a template-based method is applied, facial expression is compared to the templates define for each facial expression category.

A Neural Network (NN) can be performed an expression categorization into multiple classes. Template based method cannot achieve such a performance. In neural network method, a facial expression is classified according to network learn during training phase.

Rule based classification approach classified the facial expression into the basic emotion categories. Emotion categories based on the previously encoded facial action.

4.1 Static Image: Template based method

Edwards et al [23] uses shape and gray parameter and drive appearance Model parameters as classifier. They extract model parameter to identify the individual in such that identification is invariant to confounding factor (lighting, pose and facial expression). To achieve this goal, the Mahalanobis distance measure on training set. This classifier assumes that variation in within class for each individual is very similar. Separate the inter-class variability from the intra-class variability by Linear Discriminate Analysis (LDA). AAM uses 400 images of six basic facial expressions of 25 subjects.

Hong et al [14] use a technique of fit a labeled graph to an input facial image. Then use elastic graph matching for best match in personalized gallery. Best match personalized gallery is used to make judgment on the

category of the observed expression. The personalized galleries of 9 people are used. Each gallery contains 28 images of facial expression (4 images per expression). Purposed method has been tested on images of 25 subjects.

Hung and Hung [7] use first and last picture of the image sequence for recognition process. For the mouth region all pictures in image sequence have to be analyzed due to mouth contour changes a lot between first and last pictures. In recognition process, action parameters (APs) are used to identify facial expressions. APs are generated by difference between the model feature parameters found in an expressionless face and those found in the examined facial expression of same person. They use a set of n training expression. Corresponding sets of n APs in emotion space. Use minimum distance classifier to cluster the APs into six clusters. These clusters are representing six expressions. Each expression has $n/6$ sets of APs. Since principle component distribution of each expression is overlapped with at least two other expressions, tree best matches are selected. The highest score of three best matches determines the final classification of examined facial expression.

4.2 Static Image: Neural network based method

For the automatic analysis of facial expression in static image, Neural Network (NN) work as classifier.

Zhang et al [16] use method of neural network of two-layer perceptron. First layer of perceptron performs a nonlinear reduction of dimensionality of the feature space. They also use desired number of hidden units to represent a facial expression for good recognition rate. At least two hidden units are necessary to code the facial expression. Five to seven hidden units are enough to represent a facial expression.

Hara and Kobayashi [10] utilize the Neural Network (NN) for the recognition of facial expression. In input layer of NN number of units is 234 corresponding to the number of facial information. At hidden layer, number of units is 50 and output layer has 6 units corresponding to 6 facial expressions. They use back propagation for NN training. Network has been trained on 90 images of six basic facial expressions of 15 subjects.

Padgett and Cottrell [4] purpose feed-forward neural network containing a single hidden layer with 10 nodes. Hidden layer employ a non-linear activation function (sigmoid). The output layer of neural network contains seven units, each correspond to one emotion. There are 12 subject but they trained the network on the images of 11 subject. Test the images of 12th subject. By changing the testing and training set, total 132 networks to evaluate the entire database.

4.3 Static Images: Rule based method

Automatic facial expression analysis from static image applies a rule based method to expression classification.

Pantic and Rothkrantz [8] use automatic facial action unit (AUs) from input facial dual view. Facial expressions are classified in emotion category by comparing the AUs-coded description of the expressions of that a particular emotion category. AUs coded description of the

six basic emotional expressions represent in fact production rules which are used for automatic classification of the basic emotional expressions. Classification of an input facial dual-view into facial expression categories is performed by comparing AU-coded description of shown facial expression to AU-coded description of six basic emotional expressions. Overall performance of automatic facial expression classification has been tested on 265 dual-view images. Dual views have been recorded under constant illumination and subject had not wear glasses, a beard and moustache.

4.4 Image Sequence: Template based method

Automatic facial expression analysis from facial image sequence applies a template-based method for expression classification.

Cohn et al [22] state that first digitalized frame; key feature points are manually marked around the facial landmarks. These feature points are automatically track by optical flow method in the image sequence. Displacements of feature points are calculated by subtraction of its normalized position in the first frame from its normalized position in current frame in the image sequence. These displacements are known as predictors. Facial Action recognized by discrimination function which employs on eyes, eyebrows and mouth in facial region. There are two discrimination function for three actions of eyebrow, two discrimination functions for three actions of eye and five discrimination functions for nine facial actions of mouth and nose. For classification purpose separate group variance – covariance matrix are used. Database is consisting of 504 image sequences. 872 Action units of 100 subjects are used. Data are randomly divided into training and test sets of image sequences.

Kimura and Yachida [21] use technique of potential net of two dimension mesh to each frame in examined facial image sequence. In potential net most exterior nodes are fixed while other nodes are forced to move by image and elastic force. When potential net put on potential field then its force deformed net sample. Thus the deformed net samples whole potential field and represent the specific pattern. The pattern of the deformed net is compared to the pattern of expressionless face. Variation observed in net nodes is used further processing. Emotion space is built by applying PCA on six image sequence of three expressions (anger, happiness and surprised) of single person. Input image is projected in emotional space for quantified emotional classification purpose. Purposed method is not workable for unknown subjects.

Wang et al [20] use facial feature point, for represent the face. There are 19 points to represent the face. Among 7 FFPs are not used in expression recognition. These 7 FFPs are used as local topology for rigid motion of head. So, 7 FFPs known as the facial reference points (FRPs). Remaining 12 FFPs are known as facial expression feature points (FEFPs). These 12 FEFPs are used for facial expression recognition. By using B-spline curve, construct expression model. Thus each curve describes the expression the relation between the expression change and displacement of the corresponding FEFPs. Degree of

expression change is determined by displacement of the FEFPs in image sequence. Eight subjects with three expression, and collect the 29 image sequences at 30 frames per second with resolution of 640*480.

4.5 Image Sequence: Rule based method

Black and Yacoub [18] purpose local parameterized model of image motion for representation of rigid head motions. Motion parameters are used to drive the midlevel predicates which describe the motion of the facial features. Midlevel predicates are rules. Left part of rule is a comparison of a motion parameter. Right part of rule is the derived predicate. In purposed method, the facial expression emotion classification is temporal consistency of the midlevel representation predicates. A model is develop for six basic emotion expressions and represented by a set of rules for detecting the beginning and ending of the expression. Purposed method is tested on to image sequences containing 145 expressions of 40 subjects.

5. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi. During this work we have collaborated with many colleagues for whom we have great regard, and we wish to extend our warmest thanks to all those who have helped us with our work.

6. REFERENCES

1. I. Essa and A. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions" IEEE Trans Pattern Analysis and Machine Intelligence vol .19, no. 7, pp 757-763, July 1997.
2. A Lanitis, C Taylor and T Cootes, "Automatic Interpretation and Coding of Face Images using Flexible Models" IEEE Trans. Pattern Analysis and Machine intelligence, Vol. 19, no. 7, pp. 743-756 July 1997.
3. Z. zhang, "Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiments with a Multi-Layer Perceptron" Pattern Recognition and Artificial Intelligence, in press.
4. C. Padgett and G. W. Cottrell, "Representing Face Images for Emotion Classification" Proc. Conf. Advances in Neural Information Processing System, pp. 894-900, 1996.
5. H. Li P. Roivainen, and R. Forchheimer, "3D Motion Estimation in Model-Based Facial Image Coding" IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, no. 6, pp 545-555,1993.
6. Maja Pantic and Leon J.M Rothkrantz, "Automatic Analysis of facial expression: The State of the Art. IEEE Trans. on Pattern Analysis and Machine Intelligence vol. 22, no. 12, December 2000.
7. C. L. Hung and Y. M. Hung, "Facial Expression Recognition Using Model Based Feature Extraction and Action Parameters Classification", Journal of Visual Communication and Image Representation, Vol. 8, No. 3, pp. 278-290 September 1997.
8. M. Pantic and L. J. M Rothkrantz, "Expert System for Automatic Analysis of Facial Expression", Image and Vision Computing, J, vol. 18, no. 11 pp 881-905, 2000
9. M. Yoneyama, Y. Iwano, A. Ohtake and K. Shirai, "Facial Recognition Using Discrete Hopfield Neural Networks", Proc, Int' Conf. Information Processing vol 3, pp. 117-120, 1997.
10. H. Kobayashi and F Hara, "Facial Interaction between Animated 3D Face Robot and Human Beings", Proc. Int' Conf. System, Man, Cybernetics, pp. 3,732-3,737,1997.
11. S.Kimura and M.Yachida, "Facial expression Recognition and Its Degree Estimation", Proc. Computer Vision and Pattern Recognition, pp. 295-300, 1997.

12. T. F. Cootes, Gareth J. Edwards, and C. J. Taylor, "Active Appearance Models", IEEE Tran. on Pattern Analysis and Machine Intelligence, Vol. 23, No. 6, June 2001.
13. H. Hong, H. Neven, and C. von der Malaburg, "Online Facial Expression Recognition Based on Personalized Galleries", Proc. Int' Conf. Automatic Face and Gesture Recognition, pp.354-359, 1998.
14. C. Padgett and G. Cottrell, "Representing Face Images For Emotion Classification" Proc. Conf. Advances in Neural Information Processing Systems, pp.894-900,1996.
15. Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison Between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multilayer Perceptron ", Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 454-459,1998.
16. H. Kobayashi and F. Hara, "Recognition of six Basic Facial Expressions and their Strength by Neural Network", proc. Int. Workshop Robot and Human Comm., pp 381-386,1992.
17. M. J Black and Y. Yacoob, "Recognizing Facial Expression in Image Sequence Using Local Parameterized Models of Image Motion", Int. Journal of Computer Vision, Vol. 1, pp. 23-48, 1997.
18. T. Otsuka and J. Ohya, "Recognition of Facial Expressions Using HMM with Continuous Output Probabilities", Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 442-447, 1998.
19. M. Wang, Y. Iwai, and M. Yachida, "Expression Recognition from Time-Sequential Facial Images by Use of Expression Change Model", Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 324-329, 1998.
20. S. Kimura and M. Yachida, "Facial Expression Recognition and Its Degree Estimation", Proc. Computer Vision and Pattern Recognition, pp. 295-300, 1997.
21. J. F. Cohn, A. J. Zlochow, J. J. Lien, and T Kanade, "Feature Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression", Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 396-401, 1998.
22. G. J. Edwards, T.F. Cootes and C.J. Taylor, "Facial Recognition Using Active Appearance Models" Proc European Conf. Computer Vision, Vol. 2, pp. 581-695,1998.

IJERT