Automated System for interpreting Non-verbal Communication in Video Conferencing

Chetana Gavankar
Senior Lecturer,
Department of Information Technology
Cummins College of Engineering for Women
Karve Nagar, Pune - 411052
chetana.gavankar@cumminscollege.in

Dr.Chandragupta Warnekar
Senior Professor,
Department of Information Technology
Cummins College of Engineering for Women
Karve Nagar, Pune – 411052
chandragupta.warnekar@cumminscollege.in

***Abstract*— Gesture is a form of non-verbal, action-based communication made with a part of the body and used instead of and/or in combination with verbal communication. People frequently use gestures for more effective inter-personal communication; out of which nearly 55% come from the facial expressions alone. Facial gestures often reveal when people are trying to conceal emotions such as fear, contempt, disgust, surprise, or even unspoken political tensions. Video conferencing captures such facial signals which can be directly processed by suitable image processing system. Facial gestures are more pronounced via eyes and lip region of human face and hence they form the regions of interest (ROI) while processing the video signal. Some of these concepts are used to develop a system which can identify specific human gestures and use their interpretation towards business decision support.**

Keywords- Non-verbal Communication, Facial Expression, Gesture Recognition, Video Conferencing

I. INTRODUCTION

Gestures form a significant part of business communication. A gesture is a form of non-verbal, action-based communication made with a part of the body and used instead of or in combination with verbal communication. It is thus a form of perceptual information (mostly visual) supporting the interpersonal audio communication. As stated above, nearly 55% of total inter-personal communication comes from facial expressions alone. Dan Hill, an expert in facial coding (President of Sensory Logic and author of a book about facial coding called *Emotionomics: Winning Hearts and Minds*) has suggested certain clues to interpret the facial gestures of the candidates. Accurately interpreting the meanings of nonverbal communications, especially facial expressions, can make effective leaders and business managers, says Paul Ekman, noted psychologist. In business settings, people often don't say what they really think. Facial gestures often reveal the concealed emotions—such as fear, contempt, disgust, surprise, or even unspoken political tensions in board or executive committee meetings. It also better equips them to handle sensitive staffing situations such as performance reviews. Because each emotion has more or less unique signals in the face, facial gestures are more reliable indicators of a person's emotional state than other body language. The state of art digital cameras and camera equipped cell phones are enabling easy capture, store and share of valuable moments or gestures.

Is it possible to describe, classify and quantify every emotional expression? How many different expressions are possible? What is the associated emotional meaning with each? Can a significant Facial Gesture Recognition System be developed to assist Business Communication? The present paper is an attempt to address some of these questions.

II. FACIAL EXPRESSIONS THROUGH VIDEO CONFERENCING

Quick, efficient and accurate communication forms an important part of business today. Emails and audio teleconferences have facilitated this, but the need to meet face to face remains integral to building business relationships. Video conference incorporates face to face communication sans the tiring business travel. According to one report over 87% of world business takes place through video conference.

The body language and facial expressions of participants exposed during such video conferencing; enable better business decision-support via their interpretation. Video conferencing captures the facial signals which can be directly processed by suitable image processing system. Well developed existing procedure could be used to locate the face in the captured video frame. There is now considerable empirical support for the emotional meanings of many facial action patterns (for a review, see Ekman, 1994). The evidence is primarily based on observers' interpretations of facial expressions.

Recent scientific research has shown that facial expressions can be assigned reliably to about six categories, though many other combinations of human emotions are possible and used by philosophers, scientists and others concerned with emotion. The basic emotions are depicted in Table 1.

Gesture		
Emotion	**Description**	**Image**
Anger	Furrows in eyebrows, Swollen Nostrils, Eyes reduced	
Fear	Eyes reduced, mouth open, furrows in eyebrows	
Disgust	Nose raised, eyes reduced	
Surprise	Eyebrows raised, mouth pouch	
Smile	Widened and slightly raised lips	
Sad	Lip corners downwards, furrows in eyebrows	

TABLE 1: Basic types of Facial Expressions

The patterns of regional brain activity coinciding with different facial expressions have also been identified. Facial gestures are more pronounced via eyes and lip region of human face and hence they form the regions of interest (ROI) while processing the video signal. Some of these concepts are used in developing the Nonverbal communication interpretation system.

III. FACIAL GESTURE RECOGNITION SYSTEM

The video of facial gesture during corresponding communication often bears important clue towards decision making. This has a special relevance in business communication through video conference. Such Gesture Recognition System is thus based on following facts:

- Facial gestures are more pronounced via eyes and lip region of human face.
- Hence they form the Regions of Interest (ROI) while processing the video signal for facial gesture recognition.
- It is possible to classify and associate emotional meaning with facial gestures via mathematical algorithms
- To create a system which can identify human facial gestures and use their interpretation towards decision support during Business Communication

The basic block diagram of the proposed system is as follows:

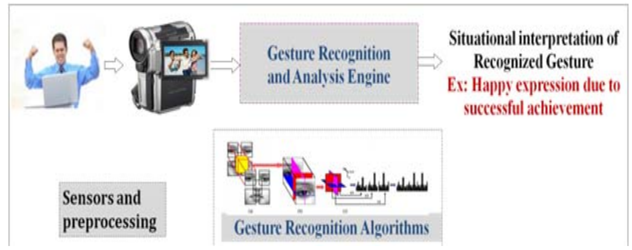

FIGURE 1: Gesture Recognition System

- **Sensors and preprocessing**: An amalgamation of hardware and software to capture gestures. Image Acquisition using IMAQ (MATLAB) video frames capture, frames to image. Preprocessing using IPT (MATLAB) RGB to gray, scaling, smoothing, cropping etc
- **Gesture Recognition and Analysis Engine**: This process sandwiches standard three level image processing pyramid via: Pixel level, Feature level and Global level. Analysis component attempts to map emotional narrative of identified gesture pattern with situational context of video data.
- **Gesture Recognition Algorithm**: Implementation via Feature Extraction (PCA, Gabor, FACS, LBP approaches) and Classification of gestures (SVM, ANN methods) for video data.

IV. FACE GESTURE RECOGNITION PROCESS

1. Image Acquisition using IMAQ (MATLAB) video frames capture, frames to image & locating the face.
2. Preprocessing using IPT (MATLAB) RGB to gray, size reduction, illumination correction, smoothing, crop etc
3. Feature Extraction – holistic approach(template based), Analytic feature based(PCA, Gabor, wavelets, ICA, LDA, ASM, FACS, Eigen face)
4. Classification of gestures(SVM, NN, K nearest neighbor , heuristics, back propagation, bayes classifier etc)
5. Post processing – Interpreting meaning of gestures using context, culture etc

A. FACIAL FEATURE EXTRACTION METHODS

Automatic facial expression recognition involves two vital aspects: facial representation and classifier design. Facial representation is to derive a set of features from original face images to effectively represent the face. The optimal features should minimize within-class variations of expressions while maximize between class variations. If inadequate features are used, even the best classifier could fail to achieve accurate recognition. Psychophysical studies indicate that basic emotions have corresponding universal facial expressions across all cultures. This is reflected by most current facial expression recognition systems that attempt to recognize a set of prototypic emotional expressions including disgust, fear, joy, surprise, sadness and anger.

1) Geometric features

Facial Action Coding System (FACS), which is a human-observer-based system developed to capture subtle changes in facial expressions. With FACS, facial expressions are decomposed into one or more Action Units (AUs). However, the geometric feature-based representation commonly requires accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations. Motions of facial features are measured by simultaneously using an active Infra-Red illumination and Kalman filtering to deal with large variations of head motion.

2) Appearance based

a) Principal Component Analysis [38]

The first step in Principal Component Analysis (PCA) is finding Eigen Values and Eigen Vectors of the frame and deciding on which are significant. Next step is forming a new coordinate system defined by the significant Eigen vectors. PCA gives a high compression rate. Its performance is good even when noise is present. However performance is very bad if scale of image is changed.

b) Independent Component Analysis

Independent Component Analysis (ICA) minimizes both second-order and higher-order dependencies in the input data and attempts to find the basis along which the data (when projected onto them) are - *statistically independent*

ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed nongaussian and mutually independent and they are called the independent components of the observed data. These independent components, also called sources or factors, can be found by ICA.

Independent component analysis (ICA) is a generalization of PCA which encodes the high-order dependencies in the input in addition to the correlations. Representations for face recognition were developed from the independent components of face images. The ICA representations were superior to PCA for recognizing faces across sessions and changes in expression.

c) Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) finds the vectors in the underlying space that best discriminate among classes. For all samples of all classes the between-class scatter matrix *SB* and the within-class scatter matrix *SW* are defined. The goal is to maximize *SB* while minimizing *SW*, in other words, maximize the ratio det|*SB*|/det|*SW*| . This ratio is maximized when the column vectors of the projection matrix are the eigenvectors of (*SW^-1* × *SB*).

d) Gabor wavelets [25]

With appearance-based methods, image filters, such as Gabor wavelets (GW), are applied to either the whole-face or specific face-regions to extract the appearance changes of the face. Due to their superior performance, the major works on appearance-based methods have focused on using Gabor-wavelet representations However; it is both time and memory intensive to convolve face images with a bank of Gabor filters to extract multi-scale and multi-orientational coefficients, for example, the Gabor-wavelet representation derived from each 48*48 face image has the high dimensionality of O(10^5)

e) Active Shape Model

An Active Appearance Model (AAM) is an integrated statistical model which combines a model of shape variation with a model of the appearance variations in a shape-normalized frame. An AAM contains a statistical model if the shape and gray-level appearance of the object of interest which can generalize to almost any valid example. Matching to an image involves finding model parameters which minimize the difference between the image and a synthesized model example projected into the image.

f) Local Binary Patterns [13]

The most important properties of LBP features are their tolerance against illumination changes and their computational simplicity. LBP features can be derived very fast in a single scan through the raw image and lie in low-dimensional feature space, while still retaining discriminative facial information in a compact representation. Experiments on different image resolutions show that LBP features perform stably and robustly over a useful range of low resolutions of face images. The encouraging performance on compressed video sequences illustrated their promising applications in real-world environments.

The original LBP operator was introduced by Ojala et al. [27], and has proved to be a powerful means of texture description. The operator labels the pixels of an image by thresholding a 3 _ 3 neighborhood of each pixel with the center value and considering the results as a binary number (see Figure 2 for an illustration), and the 256-bin histogram of the LBP labels computed over a region is used as a texture descriptor. The derived binary numbers (called Local Binary Patterns or LBP codes) codify local primitives including different types of curved edges, spots, flat areas, etc

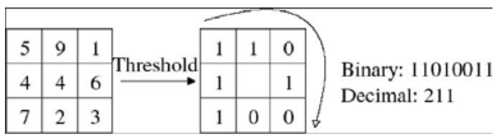

FIGURE 2: LBP code

Operator later was extended to use neighborhood of different sizes. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. See Figure 3 for examples of the extended LBP operator, where the notation (P, R) denotes a neighborhood of P equally spaced sampling points on a circle of radius of R that form a circularly symmetric neighbor set.

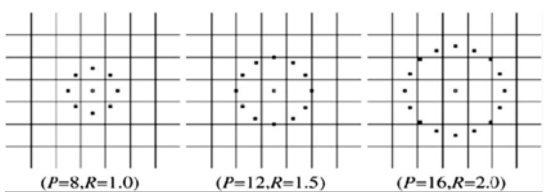

FIGURE 3: Circularly Symmetric Neighbor Set

B. *FACIAL FEATURE CLASSIFICATION METHODS*

1) *Support Vector Machine (SVM)*

Given a set of points belonging to two classes, a Support Vector Machine (SVM) finds the hyperplane that separates the largest possible fraction of points of the same class on the same side, while maximizing the distance from either class to the hyperplane. PCA is first used to extract features of face images and then discrimination functions between each pair of images are learned by SVMs.

2) *Artificial Neural Network*

An artificial neural network (ANN), usually called "neural network" (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks**.** It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

3) *K-nearest neighborhood*

In pattern recognition**,** the *k*-nearest neighbor's algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space. *k*-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. It can also be used for regression. The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors. *k* is a positive integer, typically small. If $k = 1$, then the object is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose *k* to be an odd number as this avoids tied votes. The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its *k* nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. In order to identify neighbors, the objects are represented by position vectors in a multidimensional feature space. It is usual to use the Euclidean distance, though other distance measures, such as the Manhattan distance could in principle be used instead. The *k*-nearest neighbor algorithm is sensitive to the local structure of the data.

4) *Bayes' classifier*

A probabilistic similarity measure is based on Bayesian belief that the image intensity differences are characteristic of typical variations in appearance of an individual. Two classes of facial image variations are defined: *intrapersonal* variations and *extrapersonal* variations. Similarity among faces is measured using Bayesian rule.

5) *Dynamic Baysian Network (DBN classifier)*

Dynamic Bayesian Networks (DBNs) recently were exploited for sequence-based expression recognition. Kaliouby and Robinson proposed a system for inferring complex mental states from videos of facial expressions and head gestures, where a multi-level DBN classifier was used to model complex mental states as a number of interacting facial and head displays. Zhang and Ji explored the use of multisensory information fusion technique with DBNs for modeling and understanding the temporal behaviors of facial expressions in image sequences.

6) Hidden Markov Model [22]

Hidden Markov Models (HMMs) have been widely used to model the temporal behaviors of facial expressions from image sequence. Cohen et al. proposed a multi-level HMM classifier, which allows not only performing expression classification on a video segment, busting also to automatically segment a long video sequence to the different expressions segments without resorting to heuristic methods of segmentation. But HMMs cannot deal with dependencies in observation. Hidden Markov Models (HMM) are a set of statistical models used to characterize the statistical properties of a signal. HMM consists of two interrelated processes: (1) an underlying, unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution and (2) a set of probability density functions associated with each state.

7) Optical flow

Optical flow analysis has been used to model muscles activities or estimate the displacements of feature points. However, flow estimates are easily disturbed by the nonrigid motion and varying lighting, and are sensitive to the inaccuracy of image registration and motion discontinuities

8) Probabilistic method

Chang et al. proposed a probabilistic video-based facial expression recognition method based on manifolds. Lee and Elgammal [21] recently introduced a framework to learn decomposable generative models for dynamic appearance of facial expressions where facial motion is constrained to one dimensional closed manifold. The learned model can generate different dynamic facial appearances for different people and for different expressions, so enabling simultaneous recognition of faces and facial expressions.

V. RESEARCH CHALLENGES

As seen from above discussion Gesture Recognition is challenging research area and constraint based attempts are being reported in recent literature. Some of the research problems include**:**

1. Temporal dispersion of individual actions
2. Variation in level of action intensity
3. Head /Pose variations
4. Frames selection from video
5. Discriminating Feature selection
6. Ambient illumination Issue (Gabor or Infrared camera)
7. Time complexity
8. Space complexity due to huge video-data files
9. Occlusions
10. Low resolution compressed video in real world applications.
11. Subjective deformations in natural shapes of facial features like lips and eyes

This research is an amalgamation of psychology, sensory logic, cybernetics, image processing, pattern recognition and decision making subjects. It goes to form an important decision support system in business communication. Present paper is an attempt in this direction.

REFERENCES

[1] Bociu, I.; Pitas, I., "A new sparse image representation algorithm applied to facial expression recognition," Proceedings of the 14th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing, 2004.

[2] Manikandan, S.; Kumar, R.; Jawahar, C.V., "Tensorial factorization methods for manipulation of face videos," IET International Conference on Visual Information Engineering, 2006

[3] Irene Kotsia, Ioan Buciu, Ioannis Pitas, "An analysis of facial expression recognition under partial facial image occlusion," International Journal on Image and Vision Computing 26 (2008) 1052–1067

[4] Elham Bagherian, Rahmita.Wirza.Rahmat, Nur Izura Udzir, "Extract of Facial Feature Point", International Journal of Computer Science and Network Security, VOL.9 No.1, January 2009

[5] Ashish Kapoor, Gang Hua, Amir Akbarzadeh and Simon Baker, "Which Faces to Tag: Adding Prior Constraints into Active Learning," Proc. of 12th IEEE International Conference on Computer vision, 2009

[6] Yousra ben jemaa, Sana khanfir, "Automatic local Gabor features extraction for face recognition", International Journal of Computer Science and Information Security, Vol. 3, No. 1, 2009

[7] Wei-Kai Liao; Cohen, I., "Belief Propagation Driven Method for Facial Gestures Recognition in Presence of Occlusions", Proc. of IEEE conference on Computer Vision and Pattern Recognition workshop, 2006 Page(s):158 - 158.

[8] Zheng Yang Chin, Kai Keng Ang, Cuntai Guan, "Multiclass Voluntary Facial Expression Classification based on Filter Bank Common Spatial Pattern", Proc. of 30th Annual International IEEE EMBS Conference, 2008

[9] Dornaika, F.; Davoine, F.; "Simultaneous facial action tracking and expression recognition using a particle filter", Tenth IEEE International Conference on Computer Vision, 2005. Page(s):1733 - 1738 Vol. 2

[10] Hadid, A.; Pietikainen, M., "Combining motion and appearance for gender classification from video sequences", Proc. of 19th IEEE International Conference on Pattern Recognition, 2008

[11] Guoying Zhao and Matti Pietik¨ainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions", IEEE Transactions on pattern analysis and machine intelligence, 2007

[12] Ahonen, T.; Hadid, A.; Pietikainen, M., "Face Description with Local Binary Patterns: Application to Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 28, Issue 12, Dec. 2006 Page(s):2037 - 2041

[13] Caifeng Shan A, Shaogang Gong B, Peter W. McOwan. "Facial expression recognition based on Local Binary Patterns: A comprehensive study", Image and Vision Computing (2009) 803–816, Elsevier publications

[14] Timo Ojala, Matti PietikaÈ inen, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, july 2002

[15] Ioannou, S.; Raouzaiou, A.; Karpouzis, K.; Kollias, S., "Adaptation of facial feature extraction and rule generation in emotion-analysis systems". Proc. of IEEE International Joint Conference on Neural Networks, 2004 Volume 1.

[16] Pantic, M.; Rothkrantz, L.J.M., "Facial gesture recognition in face image sequences: a study on facial gestures typical for speech articulation", Proc. of IEEE International Conference on Systems, Man and Cybernetics, 2002 Volume 6.

[17] Aggarwal, G. Chowdhury, A.K.R. Chellappa, R., "A system identification approach for video-based face recognition", Proceedings of the 17th International Conference on Pattern Recognition, 2004.Volume:4, on page(s):175-178.

[18] Zelinsky, A. Heinzmann, J., "Real-time visual recognition of facial gestures for human-computer interaction", Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996.

[19] Brian C. Lovell, Daniel Heckenberg, "Low-Cost Real-Time Gesture Recognition", Proc. of 5th Asian Conference on Computer Vision, 23–25 January 2002, Australia

[22] Jen-Tzung Chien, Chih-Pin Liao "Maximum Confidence Hidden Markov Modeling for Face Recognition," IEEE Transactions on pattern analysis and machine intelligence, Vol. 30, No. 4, pp. 606-616 April 2008

[23] P. Ravindra De Silva, Minetada Osano, Ashu Marasinghe "Towards recognizing emotion with affective dimensions through body gestures", Proc. of 7th International Conference on Automatic Face and Gesture Recognition, 2006

[24] B.Signer, U.Kurmann, M.Norrie, "iGesture: A General Gesture Recognition Framework," Proc. of Ninth International Conference on Document Analysis and Recognition, 2007.

[25] Haiyuan Wu, Y.Yoshida, T.Shioyama, "Optimal Gabor filters for high speed face identification," Proc. of the 16th International Conference on Pattern Recognition, 2002

[26] L.Huang, A.Shimizu, H.Kobatake, "Face detection from cluttered images using Gabor filter features," Proc. of the SICE Annual Conference, 2003

[27] M. Hanif, U.Ali, "Optimized Visual and Thermal Image Fusion for Efficient Face Recognition," Proc. of 9th International Conference on Information Fusion, 2006

[28] S.Mitra, T.Acharya "Gesture Recognition: A Survey", IEEE Transactions on Systems, Man, and Cybernetics, vol. 37, No. 3, pp. 311- 324, May 2007.

[29] Juan P. Wachs, Helman Stern and Yael Edan, "Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 35, No. 6, pp. 932-944, Nov. 2005

[30] J.Kamarainen, K.Kyrki, V. H.Kalviainen "Invariance properties of Gabor filter-based features — Overview and Applications", IEEE Transactions on Image Processing, Vol. 15, No. 5, pp. 1088-1099, May 2006.

[31] Ali, U. Bilal, M. "Video based Parallel Face recognition using Gabor filter on homogeneous distributed systems", Proc. of IEEE International Conference on Engineering of Intelligent Systems, 2006.

[32] Gonzalez, Woods, Eddins, "Digital Image Processing using MATLAB", Pearson Education Publishers

[33] A.K.Jain, "Fundamentals of Digital Image Processing", Pearson Education Publishers

[34] Kehtarnavaz, Mark Gamadia, "Real time Image and Video Processing – from Research to Reality", Pearson Education Publishers

[35] Rudra Pratap, "Getting started with MATLAB 7.0", Oxford University Press

[36] James H. McClellan, M. A. Yoder, Mark Yoder, "DSP First: A Multimedia Approach", Prentice Hall Publishers

[37] Schaum's Outlines, "Probability, Random Variables and Random Processing", Tata McGraw Hill Publishers

[38] M.A.Joshi, "Digital Image Processing -An Algorithmic Approach", Prentice Hall Publishers