International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.2, Issue.4, July-Aug 2012 pp-1449-1453 ISSN: 2249-6645

Facial Expression Recognition for Security

Mrs. Ayesha Butalia, Dr. Maya Ingle, Dr. Parag Kulkarni

Abstract: Facial expressions play an important role in interpersonal relations as well as for security purposes. The malicious intensions of a thief can be recognized with the help of his gestures, facial expressions being its major part. This is because humans demonstrate and convey a lot of evident information visually rather than verbally. Although humans recognize facial expressions virtually without effort or delay, reliable expression recognition by machine remains a challenge.

A picture portrays much more than its equivalent textual description. Along this theory, we assert that although verbal and gestural methods convey valuable information, facial expressions are unparalleled in this regard.

In sustenance to this idea, a facial expression is considered to consist of deformations of facial components and their spatial relations, along with changes in the pigmentation of the same. This paper envisages the detection of faces, localization of features thus leading to emotion recognition in images.

Key Terms: Facial Gestures, Action Units, Neural Networks, Fiducial Points, Feature Contours.

INTRODUCTION

Facial expression recognition is a basic process performed by every human every day. Each one of us analyses the expressions of the individuals we interact with, to understand best their response to us. The malicious intensions of a thief or a person tobe interviewed can be recognized with the help of his gestures, facial expressions being its major part. In this paper we have tried to highlight the facial expressions for security reasons.

In the next step to Human-Computer interaction, we endeavor to empower the computer with this ability — to be able to discern the emotions depicted on a person's visage. This seemingly effortless task for us needs to be broken down into several parts for a computer to perform. For this purpose, we consider a facial expression to represent fundamentally, a deformation of the original features of the face.

On a day-to-day basis, humans commonly recognize emotions by characteristic features displayed as part of a facial expression. For instance, happiness is undeniably associated with a smile, or an upward movement of the corners of the lips. This could be accompanied by upward movement of the cheeks and wrinkles directed outward from the outer corners of the eyes. Similarly, other emotions are characterized by other deformations typical to the particular expression.

More often than not, emotions are depicted by subtle changes in some facial elements rather than their obvious contortion to represent its typical expression as is defined. In order to detect these slight variations induced, it is important to track fine-grained changes in the facial features.

The general trend of comprehending observable components of facial gestures utilizes the FACS, which is also a commonly used psychological approach. This system, as described by Ekman [12], interprets facial information in terms of *Action Units*, which isolate localized changes in features such as eyes, lips, eyebrows and cheeks.

The actual process is akin to a divide-and-conquer approach, a step-by-step isolation of facial features, and then recombination of the interpretations of the same in order to finally arrive at a conclusion about the emotion depicted.

RELATED WORK

Visually conveyed information is probably the most important communication mechanism used for centuries, and even today. As mentioned by Mehrabian [8], upto 55% of the communicative message is understood through facial expressions. This understanding has sparked an enormous amount of speculation in the field of facial gestural analysis over the past couple of decades. Many different techniques and approaches have been proposed and implemented in order to simplify the way computers comprehend and interact with their users.

The need for faster and more intuitive Human-Computer Interfaces is ever increasing with many new innovations coming to the forefront. [1]

Azcarate et al. [11] used the concept of *Motion Units* (MUs) as input to a set of classifiers in their solution to the facial emotion recognition problem. Their concept of MUs is similar to "Action Units" as described by Ekman. [12]. Chibelushi and Bourel [3] propose the use of GMM (Gaussian Mixture Model) for pre-processing and HMM (Hidden Markov Model) with Neural Networks for AU identification.

Lajevardi and Hussain [15] suggest the idea of dynamically selecting a suitable subset of Gabor filters from the available 20 (called Adaptive Filter Selection), depending on the kind of noise present. Gabor Filters have also been used by Patil et. Al [16]. Lucey et. al [17] have devised a method to detect expressions invariant of registration using Active Appearance Models (AAM). Along with multi-class SVMs, they have used this method to identify expressions that are more generalized and independent of image processing constraints such as pose and illumination.

Noteworthy is the work of Theo Gevers et. al [21] in this field. Their facial expression recognition approach enhances the AAMs mentioned above, as well as MUs. Tree-augmented Bayesian Networks (TAN), Native Bayes (NB) Classifiers and Stochastic Structure Search (SSS) algorithms are used to effectively classify motion detected in the facial structure dynamically.

Similar to the approach adopted in our system, P. Li et, al have utilized *fiducial points* to measure feature deviations and OpenCV detectors for face and feature detection.

Moreover, their geometric face model orientation is akin to our approach of polar transformations for processing faces rotated within the same plane (i.e. for tilted heads).

SYSTEM ARCHITECTURE

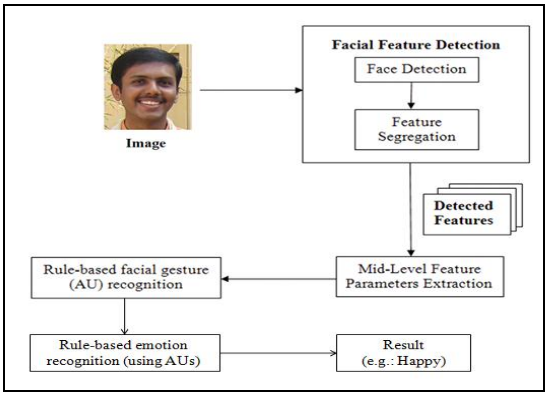

Fig. 1: Block Diagram of the Facial Recognition System

The task of automatic facial expression recognition from face image sequences is divided into the following sub-problem areas: detecting prominent facial features such as eyes and mouth, representing subtle changes in facial expression as a set of suitable midlevel feature parameters, and interpreting this data in terms of facial gestures.

As described by Chibelushi and Bourel [3], facial expression recognition shares a generic structure similar to that of facial recognition. While facial recognition requires that the face be independent of deformations in order to identify the individual correctly, facial expression recognition measures deformations in the facial features to classify them. Although the face and feature detection stages are shared by these techniques, their eventual aim is different.

Face detection is widely applied through the HSV segmentation technique. This step narrows the region of interest of the image down to the facial region¸ eliminating unnecessary information for faster processing.

Analyzing this region (facial) helps in locating the predominant seven regions of interest (ROIs), viz. two eyebrows, two eyes, nose, mouth and chin. Each of these regions is then filtered to obtain the desired facial features. Following up this step, to spatially sample the contour of a certain permanent facial feature, one or more facial-feature detectors are applied to the pertinent ROI. For example, the contours of the eyes are localized in the ROIs of the eyes by using a single detector representing an adapted version of a hierarchical-perception feature location method [13]. We have performed feature extraction by a method that extracts features based on Haar-like features, and classifies them using a tree-like decision structure. This is similar to the process described by Borsboom et. al [7].

The contours of the facial features, generated by the facial feature detection method, are utilized for further analysis of shown facial expressions. Similar to the approach taken by Pantic and Rothkrantz [2], we carry out feature points' extraction under the assumption that the face images are non-occluded and in frontal view. We extract 22 fiducial points, which constitute the action units, once grouped by the ROI they belong to. The last stage employs the FACS, which is still the most widely used method to classify facial deformations.

The output of the above stage is a set of detected action units. Each emotion is the equivalent of a unique set of action units, represented as rules in first-order logic. These rules are utilized to interpret the most probable emotion depicted by the subject.

The rules can be described in first-order predicate logic or propositional logic in the form:

Gesture (surprise):- **eyebrows** (raised) ^ **eyes** (wide-open) ^ **mouth** (open).

For instance, the following diagram shows all the fiducial points marked on the human face. And for a particular emotion, a combination of points are affected and monitored. Like for surprise, the following action units are considered:

- Opening of mouth indicated by pt 19 and pt 18
- Widening of eyes shown by pt 13, pt14, pt 9 and pt 10.
- Raised eyebrows using pt 2 and pt 5

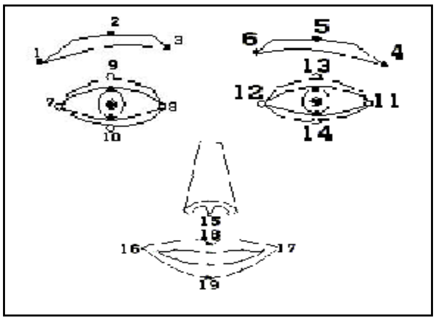

Fig. 2: Facial Points used for the distances definition [13]

Both face and feature localization are challenging because of the face/features manifold owing to the high inter-personal variability (e.g. gender and race), the intrapersonal changes (e.g. pose, expression, presence/absence of glasses, beard, mustaches), and the acquisition conditions (e.g. illumination and image resolution).

Artificial Neural Networks and Fuzzy Sets

We observed that other applications used an ANN [13] with one hidden layer [1] to implement recognition of gestures for a single facial feature (lips) alone. Hence it is but obvious that one such ANN will be necessary for each independent feature or *Action Unit* that will be interpreted for emotion analysis. We intend to utilize the concept of Neural Networks to assist the system to learn (in a semi-supervised manner) how to customize its recognition of expressions for different specialized cases or environments.

Fuzzy logic[22] (and fuzzy sets) is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than accurate. Pertaining to facial expressions, fuzzy logic comes into picture when a certain degree of deviation is to be made permissible from the expected set of values for a particular facial expression.

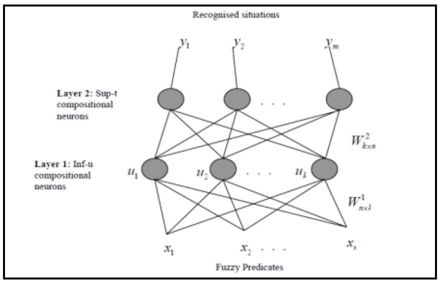

Fig. 3: The Two-Layer Neurofuzzy Architecture [13]

This is necessary because each person's facial muscles contort in slightly different ways while showing similar emotions. Thus the deviated features may have slightly different coordinates, yet still the algorithm should work irrespective of these minor differences. Fuzzy logic is used to increase the success rate of the process of recognition of facial expressions across different people.

A neuro-fuzzy architecture incorporates both the techniques, merging known values and rules with variable conditions, as shown in Fig. 3. Chatterjee and Shi [18] have also utilized the neuro-fuzzy concept to classify expressions, as well as to assign weights to facial features depending on their importance to the emotion detection process.

POTENTIAL ISSUES

Geographical Versatility

One of the fundamental problems faced by any software that accepts images of the user as an input is that it needs to account for the geographical variations bound to occur in it. Though people coming from one geographical area share certain physiological similarities, they tend to be distinct from those who belong to another region. However, these physical differences will create a lot of inconsistencies in the input obtained.

To tackle this issue, the software must employ a technique that helps increase the scale of deviations occurring, so that divergence to a certain degree can be rounded off and the corresponding feature recognized. Support Vector Machines [10] help the software do this.

Fundamentally, a set of SVMs will be used to specify and incorporate rules for each facial feature and related possible gestures. The more the rules for a particular feature, the greater will be the degree of accuracy for even the smallest of micro-expressions and for the greater resolution images.

Originally designed for binary classification, there are currently two types of approaches [10] for multi-class SVMs. The first approach is to construct and combine several binary classifiers (also called "one-per-class" method) while the other is to consider all data in one optimization formulation named one-one way. Also, SVMs can be used to form slightly customized rules for drastically different geographical populations. Once the nativity of the subject is identified, the appropriate multi-class SVM can be used for further analysis and processing.

Thus, this method uses SVMs to take care of the differences and inconsistencies due to the environmental variations.

Personalization and Calibration

The process of emotion detection can be further enhanced to train the system using semi-supervised learning. Since past data about a particular subject's emotive expression can be stored and be made available to the system later, it may serve to improve the machines 'understanding' and be used as a base for inferring more about a person's emotional state from a more experienced perspective.

One approach to supervised learning is on inception of the system, for AU and gesture mapping. Here, the user will give feedback to the system with respect to his/her images flashed on the screen in the form of gesture name. This calibration process may be allowed approximately a minute or two to complete. In this way the user is effectively tutoring the system to recognize basic gestures so that it can build up on this knowledge in a semi-supervised manner later on. This feedback will be stored in a database along with related entries for the parameters extracted from the region of interest (ROI) of the image displayed. The same feedback is also fed into the neuro-fuzzy network structure which performs the translation of the input parameters to obtain the eventual result. It is used to shift the baseline for judging the input parameters and classifying them as part of the processing involved in the system. Now, for further input received, the neural network behaves in tandem with its new, personalized rule set.

The user can also request for training examples to be gathered at discrete time intervals and provide a label for each. This can be combined with the displacements output by the feature extraction phase and added as a new example to the training set. Another way to increase the personalized information is to maintain an active history of the user, similar to a session. This data is used further on to implement a context-based understanding of the user, and predict the likelihoods of emotions that may be felt in the near future.

Calibration and Personalization would further increase the recognition performance and allow a tradeoff between generality and precision to be made. By applying these methods, the system in focus will be able to recognize the user's facial gestures with very high accuracy; this approach is highly suitable for developing a personalized, user-specific application.

Dealing with Missing Values

The uncontrolled conditions of the real world and the numerous methods that can be employed to obtain the information for processing can cause extensive occlusions and create a vast scope for missing values in the input. Back in the history of linear PCA, the problem of missing values has long been considered as an essential issue and investigated in depth. [24][25] Thus in this paper the problem of missing values is tackled by taking into consideration the facial symmetry of the subject.

Though the human faces aren't exactly symmetrical for most expressions, we create perfectly symmetrical faces using the left or right (depending on which one is clearer) side of the face. Let's call it the reference image. The creation of the Reference image is shown below using the left and right sides of the face respectively.

Fig 4: Left: original face, Centre: Symmetrical face created using the right side, Right: Symmetrical face created using the left side.

Now the original image is first scanned for the occluded or blank areas in the face, that constitute the set of missing values. After all the locations of blockages have been identified, the Reference image is super-imposed on this scanned image. During the superimposition the facial symmetry is therefore utilized to complete the input values wherever necessary, and the missing values taken care of.

Overcoming Time Constraints

The main challenge in analyzing the face, for the expression being portrayed, is the amount of processing time required. This becomes a bottleneck in real time scenarios. One way of overcoming this time constraints is dividing the image into various regions of interest and analyzing each in parallel independently. For instance, the face can be divided into regions of the mouth; chin, eyes, eyebrows etc. and the parallel analysis of these can be combined to get the final result. This approach will guarantee a gain in the throughput of the system.

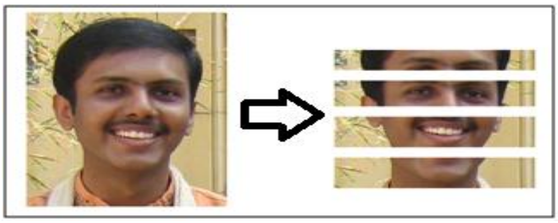

Fig. 5: Delegation of Image Components to Parallel Processors

RESULTS

In this section we present the resulting effectiveness of our systems facial expression recognition capability as well as subsequent emotion classification. The Confusion Matrix below shows that for a dataset of images depicting a particular emotion, how many were identified correctly, and how many were confused with other emotions. The dataset comprised images taken of an expert user, having knowledge of what a typical emotion is expected to show as per universal norms described by Ekman [12]. Our dataset comprised 60 images of Indian faces, with 15 each depicting Joy, Sorrow, Surprise and Neutral emotive states.

Total Accuracy: 71.7%

Table 1: Person-dependent Confusion Matrix for training and test data supplied by an expert user

Expression	Joy	Sorrow	Neutral	Surprise	**Overall**
Joy	**11**	1	1	2	73.3%
Sorrow	2	**10**	2	1	66.7%
Neural	3	2	**10**	0	66.7%
Surprise	2	0	1	**12**	80.0%

As is evident from Table 1, our system shows greater accuracy for Joy and Surprise with 73.3% and 80.0% accuracy rates respectively. The average accuracy of our system is 71.7%.

These results can be further improved by personalizing the system as discussed in the previous section. A personalized system is tuned to its user's specific facial actions, and hence, after a calibration period, can be expected to give an average accuracy of up to 90%.

FUTURE SCOPE AND APPLICATIONS

As human facial expression recognition is a very elementary process, it is useful to evaluate the mood or emotional state of a subject under observation. As such, tremendous potential lies untapped in this domain. The basic idea of a machine being able to comprehend the human emotive state can be put to use in innumerable scenarios, a few of which we have mentioned here.

- The ability to detect and track a user's state of mind has the potential to allow a computing system to offer relevant information when a user needs help – not just when the user requests help, for instance, the change in the Room Ambience by judging the mood of the person entering it.
- Help people in emotion-related research to improve the processing of emotion data.
- Applications in surveillance and security. For instance, computer models obtained up to 71% correct classification of innocent or guilty participants based on the macro features extracted from the video camera footage.
- In this regard, lie detection amongst criminal suspects during interrogation is also a useful aspect in which this system can form a base. It is proven that facial cues more often than not can give away a lie to the trained eye.
- Patient Monitoring in hospitals to judge the effectiveness of prescribed drugs is one application to the Health Sector. In addition to this, diagnosis of diseases that alter facial features and psychoanalysis of patient mental state are further possibilities.
- Clever Marketing is feasible using emotional knowledge of a patron and can be done to suit what a patron might be in need of based on his/her state of mind at any instant.
- Detecting symptoms such as drowsiness, fatigue, or even inebriation can be done using this system. Thus, by helping in the process of Driver Monitoring, this system can play an integral role in reducing road mishaps to a great extent.

Also noteworthy of mention is the advantage of reading emotions through facial cues, as our system does, over reading them through study of audio data of human voices. Not only does the probability of noise affecting and distorting the input reduce for our system, but also there are fewer ways to disguise or misread from visual information as opposed to audio data, as is also stated by Busso et. al [23] . Hence facial expressions would form a vital part of a multimodal system for emotion analysis as well.

CONCLUSION

As is quite evident after plenty of research and deliberation, gaining insight on what a person may be feeling is very valuable for many reasons. The future scope of this field is visualized to be practically limitless, with more futuristic applications visible on the horizon especially in the field of Security.

Our facial expression recognition system, utilizing neuro-fuzzy architecture is 71.7% accurate, which is approximately the level of accuracy expected from a support vector machine approach as well. Every system has its limitations. Although this particular implementation of facial expression recognition may perform less than entirely accurate as per the end users' expectations, it is envisioned to contribute significantly to the field, upon which similar work can be furthered and enhanced. Our aim in forming such a system is to form a standard protocol that may be used as a component in many of the applications that will no doubt require an emotion-based HCI.

Empowering computers in this way has the potential of changing the very way a machine "thinks". It gives them the ability to understand humans as 'feelers' rather than 'thinkers'. This in mind, this system can even be implemented in the context of Artificial Intelligence. As part of the relentless efforts of many to create intelligent machines, facial expressions and emotions have and always will play a vital role.

REFERENCES

[1] Piotr Dalka, Andrzej Czyzewski, "Human-Computer Interface Based on Visual Lip Movement and Gesture Recognition", International Journal of Computer Science and Applications, ©Technomathematics Research Foundation Vol. 7 No. 3, pp. 124 - 139, 2010

[2] M. Pantic, L.J.M. Rothkrantz, "Facial Gesture Recognition in face image sequences: A study on facial gestures typical for speech articulation" Delft University of Technology.

[3] Claude C. Chibelushi, Fabrice Bourel, "Facial Expression Recognition: A Brief Overview", 1-5 ©2002.

[4] Aysegul Gunduz, Hamid Krim "Facial Feature Extraction Using Topological Methods" © IEEE, 2003

[5] Kyungnam Kim, "Face Recognition Using Principal Component Analysis"

[6] Abdallah S. Abdallah, A. Lynn Abbott, and Mohamad Abou El-Nasr, "A New Face Detection Technique using 2D DCT and Self Organizing Feature Map", World Academy of Science, Engineering and Technology 27 2007

[7] Sander Borsboom, Sanne Korzec, Nicholas Piel, "Improvements on the Human Computer Interaction Software: Emotion", Methodology, 2007, 1-8.

[8] A. Mehrabian, Communication without Words, Psychology Today 2 (4) (1968) 53-56.

[9] M. Pantic, L.J.M. Rothkrantz, "Expert System for Automatic for Automatic Analysis of Facial Expressions" Image and Vision Computing 18 (2000) 881-905.

[10] Li Chaoyang, Liu Fang, Xie Yinxiang, "Face Recognition using Self-Organizing Feature Maps and Support Vector Machines", Proceedings of the Fifth ICCIMA, 2003.

[11] Aitor Azcarate, Felix Hageloh, Koen van de Sande, Roberto Valenti, "Automatic Facial Emotion Recognition" Univerity of Amsterdam, 1-10, June 2005

[12] P. Ekman "Strong evidence for universals in Facial Expressions" Psychol. Bull., 115(2): 268-287, 1994.

[13] A. Raouzaiou et. Al, "A Hybrid Intelligence System for Facial Expression Recognition" EUNITE 2002 482-490.

[14] Y. Liu et. al, "Facial Asymmetry Quantification for Expression Invariant Human Identification"

[15] Seyed Mehdi Lajevardi, Zahir M. Hussain "A Novel Gabor Filter Selection Based on Spectral Difference and Minimum Error Rate for Facial Expression Recognition" 2010 Digital Image Computing: Techniques and Applications, 137-140.

[16] Ketki K. Patil et. Al, "Facial Expression Recognition and Head Tracking in Video Using Gabor Filtering", Third International Conference on Emerging Trends in Engineering and Technology, 152-157.

[17] Patrick Lucey, Simon Lucey, Jeffrey Cohn, "Registration Invariant Representations for Expression Detection" 2010 Digital Image Computing: Techniques and Applications, 255-261.

[18] Suvam Chatterjee, Hao Shi, "A Novel Neuro-Fuzzy Approach to Human Emotion Determination" 2010 Digital Image Computing: Techniques and Applications, 282-287.

[19] P. Li et. al, "Automatic Recognition of Smiling and Neutral Facial Expressions" 2010 Digital Image Computing: Techniques and Applications, 581-586.

[20] Irfan A. Essa, "Coding, Analysis, Interpretation and Recognition of Facial Expressions" IEEE Transactions on Pattern Recognition and Machine Intelligence, July 1997, 757-763.

[21] Roberto Valenti, Nicu Sebe, Theo Gevers, "Facial Expression Recognition: A Fully Integrated Approach" ICIAPW '07 Proceedings of the 14th International Conference of Image Analysis and Processing, 2007, 125-130.

[22] L. A. Zadeh, "Fuzzy Sets" Information and Control 8, 1965, 338-353.

[23] Carlos Busso et. al, "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information", ACM 2004.

[24] N.F Troje et al, "How is bilateral symmetry of human faces used for recognition of novel views", Vision Research 1998, 79-89

[25] R Thornhill et al, "Facial Attractiveness", Trans. In Cognitive Sciences, December 1999.