International Journal of Advance Research in Science and Engineering
Volume No 06, Special Issue No. (02), September 2017, ICITTESE-17
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Fuzzy Model for Human Face Expression Recognition

Gargishankar Verma[1], K.Ravikant[2],Dharmendra Kaushik[2]

[1,2,3]Computer Science, Columbia Institute of Engineering & Technology, Raipur, (India)

ABSTRACT

This paper present an approach to recognize human face expression and emotions based on some fuzzy pattern rules. Facial features for this specially eye and lips are extracted an approximated into curves which represents the relationship between the motion of features and change of expression. This paper focuses the concepts like face detections, skin color segmentation, face features extractions and approximation and fuzzy rules formation. Conclusion based on fuzzy patterns never been accurate but still our intension is to put more accurate results.

Key words: Face Detection, Skin Color Segmentation, Face Futures, Curve Formation and Approximation, Fuzzy Patterns.

I. INTRODUCTION

Facial expression analysis has been attracted considerable attention in the advancement of human machine interface since it provides natural and efficient way to communicate between humans [2]. Some application area related to face and its expression includes personal identification and access control, video phone and teleconferencing, forensic application, human computer application [5].

Most of the facial expression recognition methods reported to date or focus on expression category like happy, sad, fear, anger etc. For description of detail face facial expression, Face Action Coding System (FACS) was design by **Ekman** [8]**.** In FACS motion of muscles are divided into 44 action units and facial expression are described by their combination. Synthesizing a facial image in model based image coding and in MPEG-4 FAPs has important clues in FACS. Using MPEG-4 FAPs, different 3D face models can be animated. Moreover, MPEG-4 high level expression FAP allows animating various facial expression intensities. However, the inverse problem of extracting MPEG-4 low and high level FAPs from real images is much more problematic due to the fact that the face is a highly deformable object [1].

There have been many advances in face detection; however, the area of expression detection is still in its early stages. There has been a great deal of work done in this area, and even applications of it. For example, Sony cameras have their “Smile Detection” that is supposed to detect when a person in the image is smiling (http://www.gadgetbb.com/2008/02/27/sony-dsc-t300-first-camera-with-smile-detection/). Others who have done work in this field of research include CMU (http://www.pitt.edu/~emotion/research.html) and BMW (“Bimodal Fusion of Emotional Data in an Automotive Environment”, S. Hoch, F. Althoff, G. McGlaun, G. Rigoll). Such research has been focused on detecting when a face *becomes* a particular expression. That is, it video sequences a face and calculates changes in the image from a “neutral” state to determine if the face has become another state – note that there are generally seven categories of expression: *Anger, Disgust, Fear, Happy, Neutral, Sadness,* and *Surprise*.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

The first step was to detect faces within an image, which we hope to classify. To do this we leverage the Viola-Jones face detection algorithm in OpenCV. The OpenCV repository contains a Haar cascade classifier that will find frontal faces. Each of these faces is then saved to a file to be processed by a classifier.

The next step was to create some classifiers for the various expressions we are going to classify. We first began with a simple "Smile" and "No smile" binary classification. To begin to see if it was even remotely possible to classify images, we started with the small class images from CSE 576 project 3. This image "database" contains 17 smiling images and 17 non-smiling images. The next step was to move to a much larger database, provided by CMU, with over 8000 different images with 7 different classifications: Happy, Sad, Anger, Fear, Disgust, Surprise, and Neutral.

The classifiers that we chose to use were K Nearest Neighbor and Support Vector Machine. For each classifier, we chose to use two types of features to train our classifiers. The first feature vector was simply the raw image grayscaled and resized to 25 by 25 (resulting in a feature vector of length 625). This would provide a baseline for all other future classification methods. The second feature vector is an expansion of eigenfaces, using the coefficients of those eigenfaces used to project a face as the feature vector. For this method, we used about 70 images – 10 images from each expression – to create the eigenfaces and saved the top 30.

In order to keep images consistent for classification, we used the Viola-Jones face finder on the training images to crop out the face. These faces would generally have the same bounding box around the face as other faces found from test images. One issue that we ran into by using this method is that Viola Jones finds many false positives so we had to manually delete all the false positives it identified.

After we created all the training images, the next step is to actually train the classifiers. In order to test the classifier, we also wrote a script that for every 10 images, we remove 1 from the classification set and save it for cross validation testing. The classifiers were created using libraries from OpenCV. The KNN library that is provided is fairly straightforward and we simply input each feature vector with a classification number. The SVM library was more complicated and had many configurations that had to be set up before it could be run. For the purposes of our project, we used some basic defaults. This myriad of configurations provides a lot of room to fine-tune the SVM approach to improve the results (for example, we had to change weights for the different classes to account for classes that are less represented).

After setting up the classifiers, the final step was to run tests on some test images – our first test was to use the images we reserved for cross-validation and our second test used images completely unrelated to any images in the training set. We will go into further details about the results we saw in the *Results* section, but we did notice that certain classifications were much more likely than others, such as the neutral face seemed to dominate others. Therefore, we had to change certain weights to avoid misclassifying the non-neutral classes.

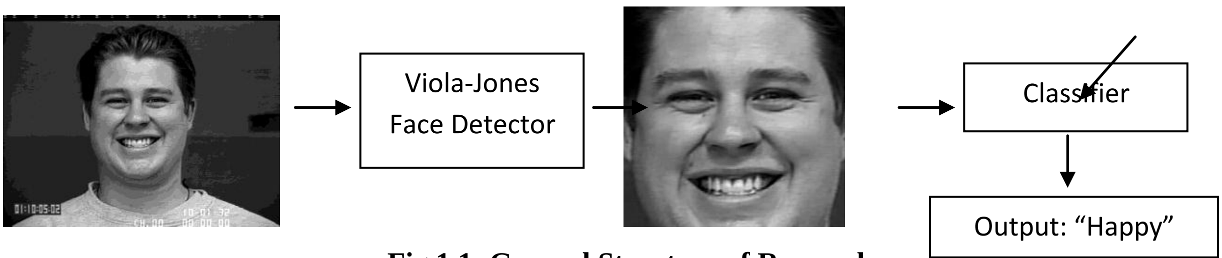


Fig 1.1: General Structure of Research work

We eventually found that with seven classes, it turns out that using raw images with SVM provided the least number of misclassifications while using eigenfaces was less accurate. Since there were still some

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

misclassifications in our best method, we then focused on eliminating misclassifications by improving the "Smile"/"No Smile" classification in the time provided. One of the problems we believe was that neutral faces were overrepresented in the CMU database by an order of magnitude more than any other class. Thus, we reduced the training size of the neutral faces and combined it with angry, fear, disgust, sad, and surprise to form the non-smiling class. This continued to exhibit the same problem as before, classifying nearly everything as neutral. The problem was that certain classes of faces are too similar to happy, particularly in the aspect of an open mouth with teeth showing. This led us to remove anger and fear from the non-smiling class and this provided better results.

Finally, we looked into the use of contours, since this would allow us to really focus on the curve of the mouth. However, we determined that it might not very suitable, since in order to find the mouth, we need to lower the gradient threshold; however, the lowering of the threshold allowed a lot of undesired edges to show up in the image including teeth, creases around the face, shadows, and hair, making classification fairly difficult. At the other end, using a fairly high threshold caused us to miss most of the mouth while we still get a lot of noise from the image (creases in around the mouth and shadows).

Fig 1.2: Segmentation

This is not to say that the method will not work, but it would require careful tuning and possibly more image processing to get good results.

Human communication is a combination of both verbal and nonverbal interactions. Through facial expressions, body gestures and other non-verbal cues, a human can communicate with others. This is especially true in the communications of emotions. In fact, studies have shown that a staggering 93% of affective communication takes place either non-verbally or para-linguistically [Mehrabian, 1971] through facial expressions, gestures, or vocal inflections [Picard, 1998]. As discussed in this paper, when dealing with online shoppers, who are communicating with a computer to potentially make a purchase, detection of emotions can certainly play an important role. The same system has the capability to be used in physical stores as well.

Using computers to analyze human faces has been an area of recent interest in computer science and psychology. Smart rooms with computer systems capable of tracking people and recognizing faces, and which can interpret speech, facial expressions and gestures made by the individuals in the smart room have been proposed before [Pentland, 1996]. According to Pentland, the philosophy behind this technology is that "…computers must be able to see and hear what we do before they can prove truly helpful. What is more, they must be able to recognize who we are and, much as another person or even a dog would, make sense of what we are thinking." [p.68]. Robots that show emotions are being built [Bartlett, et al., 2003] while other similar attempts to include emotion in various computer applications [Sarrafzadeh et al., 2002] are being made. Facial recognition software have even be used to apprehend potential terrorists by using the software to scan a crowd

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

of people going through an airport and looking for facial matches to known terrorists [Kopel and Krause, 2002]. Despite all this, today's computers do not make use of the nonverbal cues common in human communications.

Affective reactions of human subjects in e-commerce environments have been studied [Preira, 2000]. In social psychology researchers have been interested in relating facial expressions with the behavior of a person. So far in marketing there have not been any studies aimed at relating facial expressions with the shopping behavior of customers, and the great potential of such technologies in e-commerce applications have been overlooked. The objective of this paper is exploratory & conceptual and it attempts to relate how the facial expressions of a person might be used to distinguish potential customers from window shoppers when shopping in a physical or online store. Though many people enter a physical or online store, not all end up buying the goods or services from the store. This is especially true about large department stores. The same applies to online shoppers who may browse through shopping sites without actually making purchases. It also describes how, in an e-commerce context, such facial expression recognition could be used to tailor suggestions, or even advertising, to suit the customer's requirements.

People visiting a store can be broadly classified into four categories [Stevens, 1989; Moe, 2003]: The Browsers, who just enjoy wandering through the store and killing their own time and the time of sales personnel. Secondly, we have the Future Customer who may be collecting information about a product or service in order to make a future purchasing decision. Thirdly, there are the Potential Customers, who may have the desire to buy a product but are not handled properly by the sales person and hence do not mature a deal with the store. The fourth category is that of the Buyers who are there specifically to purchase a product straight away. It is not always possible for all the sales personnel to distinguish which of these categories the potential customers fit into and thus design selling strategies tailored to those categories in terms of allocation of time to such customers. There are generally two reasons for sales people failing to make the distinction between different categories of customers: Either they do not possess enough training or skills to identify them, or they do not have enough time at their disposal to attend to those customers on whom it is worth spending the time to mature a deal.

We have proved that it is possible to automatically detect facial expressions by developing facial expression analysis software that is capable of detecting six different facial expressions. Based on this proof, we are proposing an intelligent sales assistant that can automatically scan the store and guide sales staff towards potential buyers and suggest suitable sales strategies. Such a tool would be a great asset for large department stores and will enhance sales while decreasing costs. The same is true for online sales activities.

With on-line sales activities, there are currently systems available that will suggest products to customers based on their previous purchases, or even on their browsing behavior. An intelligent facial expression recognition system could provide further depth to such systems by allowing the computer to suggest further products to the customer based on the expressions detected while the customer is browsing through the product catalogue. In many e-commerce situations it is also possible to compare several products to see how they differ in features. In such a scenario, facial expression could also be very useful in detecting which products the customer has an unspoken preference for. The customer is therefore taken to products they have been recognized to be more interested in. The ability of the system to direct the customer to the right products is a desirable feature and results in considerable savings in time as well as reduced frustration. This is a very important issue when one is searching an online shop for the right product.

The facial expression recognition feature of the sales assistant is a general tool for detecting facial expressions. In the context of online sales the facial expression recognition system works well as there is only a single person the system is dealing with and the system has a nearly frontal view of the face. Although meant for use within the sales assistant, this feature of the system has many potential business and other applications such as targeted advertising and marketing both on line and otherwise. Among other applications are assessment of audience mood in television debates and security applications.

II LITERATURE REVIEW

Designer of FACS, Ekman himself as pointed out some of these action units as unnatural type facial movements. Detecting a unit set of action units for specific expression is not guaranteed. One promising approach for recognizing up to facial expressions intensities is to consider whole facial image as single pattern [4]. Kimura and his colleagues have reported a method to construct emotional space using 2D elastic net model and K-L expansions for real images [7]. Their model is user independent and gives some unsuccessful results for unknown persons. Later Ohba proposed facial expression space employing principle component analysis which is person dependant [9].

Although there are various studies on the applications of computers in areas related to our proposed intelligent sales systems in the form of e-commerce and m-commerce [Okazaki, 2005], personalized online product selection, and web-based shopping systems we have, so far, found no conceptual or empirical studies in this field. This paper is conceptual in nature and therefore aimed at exploring and initiating the debate in this field.

It has been well researched that facial expressions do reflect cognitive behavior, and that individuals observe other's facial expressions and then use these to regulate their own behavior in social interactions. We do find some studies relating facial expression to some aspects of marketing but none on facial expression and shopping behavior. Howard and Gengler [2001] found that favorable facial expressions do have a positive bias on consumer product attitudes. Sirakaya and Sonmez [2000] used facial expression to study the gender images in Government tourism brochures. Derbaix [1995] also used facial expression to investigate the effect of television advertisement on attitude towards advertisement and brand attitude. Yuasa et al. [2001] developed a computer network negotiation support tool using facial expression to negotiate an agreement strategy between seller and buyer. They argue that, if players select a happy face, then there is a greater chance they will reach an agreement. Lee [2002] developed computer software programs to implement an e-shopping authentication scheme that used facial recognition instead of user name and password to control access to the web site in a modern consumer e-shopping environment. Consumers would not need to remember their user name and password to get access to the web site. The computer, based upon the facial features of the user, would recognize faces and allow access to the web site to authorized users.

A recent study by Ghijsen et al. [2005] in the context of computer based tutoring systems has shed some light on the types and frequencies of facial expressions exhibited by students during interactions with a computer. These are the same as the results that were found in another recent study of human tutoring sessions [Alexander et al., 2005]; we might tentatively suggest that the expressions displayed by students are not significantly affected by whether the tutor is human or artificial. This would concur with the Media Equation of Reeves and Nass [1996] which argues that interactions between computers and humans are inherently social, and that the same rules for

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

interactions between humans also apply to interactions between humans and computers. However, much more complete studies would be required before firm conclusions could be reached on various aspects of the facial expressiveness of customers.

The authors of this article have developed hardware and software to analyze different facial expressions and then relate these to the behavior of a person in an educational setting [Sarrafzadeh et al., 2003]. The outcome of this research has some useful applications in the field of marketing, especially in sales (online or physical) and the training of sales people. There is also an added advantage that, if sales people can quickly identify the potential customers from window shoppers with the intervention of the system, then they can spend more time on potential customers and convert these potential customers into buyers. It will save the sales people time and the customer will feel well attended to. This will also reduce the expenses on sales forces as the stores will employ less people as they can identify the potential customer and devote sufficient time on them to convert them into buyers instead of spending unreasonable time on window shoppers. Using the same technology, online shoppers can be directed to appropriate products which could result in increased sales.

This is a multi-staged project and in the first stage (this paper) the objective is to develop a theoretical and conceptual model to initiate a discussion in this area. The next stage will be to use in-house software developed by the research team that is capable of analyzing facial expressions to relate these different facial expressions to purchasers' behavior while they shop. This will enable the system to differentiate between real and window shoppers. The system can then give advice to sales staff based on this information. One of the strategies that will be followed is to give window shoppers freedom to browse without disturbance from sales staff, while real shoppers will have assistance available as required. We believe that this will result in cost savings and perhaps increased sales for store owners.

In the second stage of this research (which is outside of the scope of this paper) the research plan is to use an in-house camera and the computer software developed by the team to classify the different facial expressions of purchasers while they shop. Before we proceed to use the in-house camera and software to classify the shoppers into four categories of shoppers, we will test our categorizations of shoppers. We intend to use a panel of 20 sales personnel working in durable goods stores as respondents to verify our categorization scheme. Firstly, respondents will be asked if they agree or disagree with our four types of shoppers' categorizations. Secondly, the respondents will be shown a number of customer facial expressions while they shop around and asked to classify them into categories. After empirically testing our categorization scheme, a number of face images will be used to train the neural network based software to enable it to classify store visitors into categories of shoppers.

In addition to being able to interpret facial expressions, the software developed has face recognition abilities. We intend to use this capability to identify past shoppers and relate that to previous purchases made by the same person. The system can then be extended to include shopping behavior and preferences of customers in the processing. Store databases will contain sales data and images of customers.

As stated above, in the second stage of research, a number of face images will be used to train the neural network based software in order to enable it to classify store visitors into the four categories of shoppers discussed earlier. We then intend to empirically test the system on real test subjects.

The third phase of this project is to develop software based on the facial expressions of shoppers that can be used as a tool for training sales staff.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Once these three stages of research are finished, the e-commerce stage will begin in which the system will be further developed to use facial expression to detect product preferences in customers which will then make it beneficial in suggesting specific products to customers based on their expressions when browsing through other products. Such a system would be beneficial for both traditional and on-line shopping experiences.

The same system will then also be tested in new product development scenarios in which several product ideas can be screened by customers, and the preferred option can be chosen based on the customers facial expressions. Customers often have unspoken needs or desires that it may be possible to pick up through changes in their facial expressions.

Though this paper deals primarily with a computerized in-store sales assistant, future development will include e-commerce applications such as those described above. In summary, this paper deals with the first stage of the research which is conceptual in nature.

Significance of facial expressions and Internal State Estimation in Sales and Marketing

In this study, we are using facial expression analysis to estimate the internal state of shoppers and to use that data to identify real shoppers. This is done to save time and to direct sales staff to points where a sales is more likely and to give window shoppers more freedom to browse undisturbed thus increasing the potential for future sales. There is disagreement on the extent to which facial expressions reflect the human internal state [Azar, 2000]. We, however, believe that although facial expressions may not always be the true reflection of emotions, they are one of a number of possible indicators of internal state. Facial expressions may be used as a way of letting others know how one feels about something and can be intentionally deceptive. Hill and Craig [2002] found that they could differentiate between real and faked facial expressions in patients responding to examinations of pain conditions, based on the duration and frequency of the facial expressions.

There is also evidence that the affective state influences processing strategy. Schwarz and Clore [1996] summarize the evidence that during decision-making happy individuals tend to choose top-down strategies, relying primarily on internalized pre-existing knowledge structures, whereas individuals who are sad tend to utilize bottom-up strategies that rely on whatever is being presented at the time of the decision. Since it is relatively easy to detect a happy versus sad affective state through facial expression, the sales staff can then be advised of the strategy that will be the most effective.

Alternative Methods of Detecting Affective State

Although vision based affective state detection is the least intrusive and most applicable way of detecting affect that is possible with the current state of technology there is an alternative. Wearable emotion detection technology [el Kaliouby et al., 2007] can be used for detecting affect. This technology would however require one to attach various devices to the body. Among these devices are data gloves which are very expensive and not feasible for applications similar to the sales assistant application described in this paper. Other devices used for affective state detection purposes include heart rate monitors and pressure sensors. We have developed an intelligent mouse device that is capable of detecting the user’s heart rate while they use the system. It is currently unclear how helpful this device will be in a sales situation but this is an area that is being considered for further study.

Computer-aided Detection of a Shopper’s Intent to Purchase

Given that facial expressions and body gestures are a significant factor in identifying real shoppers, how can we use a computer to detect the affective feedback volunteered by the shopper? Current research focuses on facial

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

expressions as perhaps the most important medium of non-verbal, affective communication. That spontaneous (or unconscious) facial expressions really illustrate an affective state is supported by several recent studies that found a high correlation between facial expressions and self-reported emotions [Rosenberg & Ekman, 1994; Ruch, 1995]. Physiological activity has been found to accompany various spontaneous facial expressions [Ekman & Davidson, 1993] which also bear witness to a link between emotions and facial expressions. We can conclude that both spontaneous facial expressions and their accompanying physiological activity illustrate an underlying affective state.

Computers can interpret facial expressions of emotion with reasonable accuracy due to the recent development of automated facial expression analysis systems. Automated facial expression analysis systems identify the motion of muscles in the face by comparing several images of a given subject, or by using neural networks to learn the appearance of particular muscular contractions [Fasel & Luettin, 2003]. This builds on the classic work of Ekman and Friesen, who developed the Facial Action Coding System for describing the movement of muscles in the face [Ekman & Friesen, 1978]. An affective state can be inferred from analyzing the facial actions that are detected [Pantic & Rothkrantz, 1999]. In the following sections we will describe in-house facial recognition/expression analysis software which we intend to discuss in order to theoretically conceptualize how this work will be used in further stages of research.

The very basis of any recognition system is extracting the best features to describe the physical phenomena. As such, categorization of the visual information revealed by facial expression is a fundamental step before any recognition of facial expressions can be achieved. First a model of the facial muscle motion corresponding to different expressions has to be found. This model has to be generic enough for most people if it is to be useful in any way. The best known such model is given in the study by Ekman and Friesen **Error! Reference source not found.**, known as the Facial Action Coding System (FACS). Ekman has since argued that emotions are linked directly to the facial expressions, and that there are six basic "universal facial expressions" corresponding to happiness, surprise, sadness, fear, anger, and disgust. The FACS codes the facial expressions as a combination of facial movements known as action units (AUs). The AUs have some relation to facial muscular motion and were defined based on anatomical knowledge and by studying videotapes of how the face changes its appearance when displaying the expressions. Ekman defined 46 such action units each corresponds to an independent motion of the face.

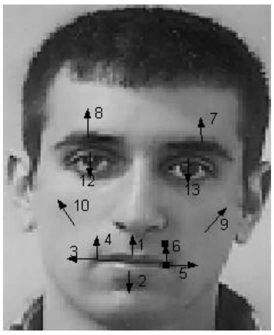


Figure 2.1. AUs extracted by the face

We implemented a face-tracking algorithm. The face tracking algorithm and system are based on the work of Tao and Huang **Error! Reference source not found.** called the Piecewise Bézier Volume Deformation (PBVD) tracker. This system was modified to extract the features for the emotion expression recognition by Chen **Error! Reference source not found.**. The estimated motions are represented in terms of magnitudes of some predefined AUs. These AUs are similar to what Ekman and Friesen **Error! Reference source not found.** proposed, but only 12 AUs are used. Each AU corresponds to a simple deformation on the face, defined in terms of the Bézier volume control parameters. In addition to the 12 AUs, the global head motion is also determined from the motion estimation. Figure 2.1 shows the 12 AUs being measured for emotion expression recognition, where the arrow represents the motion direction of the AU moving away from the neutral position of the face.

Using the measurements of these action units, two types of classifiers were constructed. The first is a frame-based classifier **Error! Reference source not found.**. The second uses the entire facial expression time information.

The frame based classifier makes a decision among the seven classes (happiness, sadness, surprise, anger, fear, disgust, and neutral) for each time frame using a Sparse Network of Winnows (SNoW) classifier **Error! Reference source not found.**. The SNoW classifier transforms the original AUs to higher dimensional feature space, after which the connections between the transformed feature nodes to the output target nodes (the emotion classes in this case) will be sparse. The training uses a multiplicative update rule (Winnow), this in contrast to a neural network that uses an additive update rule. The advantages of using SNoW is that it does not require a large number of training data, and in the sparseness of the connections between the layers, which gives a lower probability of error and speed. For testing, the output target with the highest score is the winning class ("winner-takes-all").

The second classifier is a novel architecture of a multilevel hidden Markov model (HMM) classifier. The multilevel HMM both serves as a classifier of the emotion sequences and does automatic segmentation of the video to the different emotions. The architecture is constructed of a lower level of six emotion-specific HMMs, trained on labeled segmented facial expression sequences, with the observations being the AU measurements of the face tracker. The state sequence of each of the six HMMs is decoded using the Viterbi algorithm, and this state sequence vector (six dimensional) serves as the observation to the high level HMM. The high level HMM consists of seven states, one representing each emotion and a neutral state. The state in which the high level HMM is in at each time can be interpreted as the classification of the time sequence. The high level HMM does both the segmentation and the classification at the same time. Since the observation vector is the state sequence of the lower level HMMs, it also learns the discrimination function between the six HMMs. This is the main difference between this work and the work of Otsuka and Ohya **Error! Reference source not found.**, who used emotion-specific HMMs but did not attempt to use a higher-level architecture to learn the discrimination between the different models.

These algorithms were tested on a database collected by Chen **Error! Reference source not found.**, and the first was also implemented in real time, for person-dependent recognition. The subjects in the database were asked to express different emotions given different stimuli. The database is of 100 subjects of different genders and ethnic backgrounds. It includes sequences of facial expressions only, as well as sequences of emotional speech and video. Testing on this database yielded recognition accuracy of over 90% for both methods, using a person-dependent approach, and a much lower accuracy of around 60-70% for a person independent approach. It was noticed that happiness and surprise are classified very well for both person-dependent and person-independent cases, and the other emotions are greatly confused with each other, especially in the person-

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

independent test. If the number of classes is reduced by combining the classes, disgust, anger, fear, and sadness to one "negative" class, the accuracy becomes much higher for both the person-dependent tests (about 97%) and the person-independent tests (about 90%). Figure 2 shows four examples of the real time implementation of the first method. The label shows the recognized emotion of the user.

Emotion Recognition from Audio

Emotions are expressed using the voice as well as through the facial expressions. In the database that was collected, the subjects were asked to read sentences while displaying and voicing the emotion. For example, a sentence displaying anger is: *Computer, this is not what I asked for, don't you ever listen?* The audio is processed on a phrase level to extract prosodic features. The features are statistics of the pitch contour, its derivative, statistics of the RMS energy envelope and its derivatives. A measure of the syllabic rate is also extracted (it expresses the 'rate' of speaking). The features are computed for a whole phrase, since it is unlikely that the emotion changes in the speech very fast. An optimal Naïve Bayes classifier is then used. The overall accuracy using this classifier was around 75% for a person-dependent test, and around 58% for a person-independent test, which shows that there is useful information in the audio for recognizing emotions (pure chance is 1/7=14.29%).

Figure 2.2: Frames from the real time facial expression recognizer

Emotion Recognition from Combined Audio and Video.

There are some inherent differences between the features of the audio and video. Facial expressions can change at a much faster rate than the vocal emotions, which are expressed in longer sequences of a phrase or sentence. To account for these time differences, a classifier is designed for each of the channels, and not a combined classifier that has to wait for the audio to be processed. The combination of the two classifiers is handled using a system that can work in three modes; audio only, video only and combined audio and video. The mode is set using two detectors. An audio detector to recognize that the user is speaking, a video detector determines if the user is being tracked, and if the tracking of the mouth region is reliable enough since when the user talks, the mouth movement is very fast, and then the mouth region is not reliable for expression recognition. In the video and audio mode, only the top region of the face is used for the expression recognition because of the fast mouth movement.

Synthetic Talking Face (iFace) Face Modeling

We have developed a system called iFACE, which provides functionalities for face modeling, editing, and animation. A realistic 3D head model is one of the key factors of natural human computer interaction. In recent years, researchers have been trying to combine computer vision and computer graphics techniques to build

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

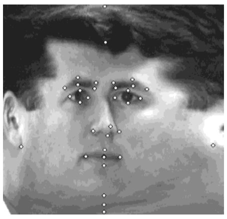
realistic head models **Error! Reference source not found.Error! Reference source not found.**. In our system, we try to make the head modeling processing more systematic. The whole process is nearly automatic with only a few manual adjustments necessary. The generic face model we use is a geometrical model (figure 3(a)) that consists of all the facial accessories such as eyes, teeth, tongue, *etc*. To customize the face model for a particular person, we first obtain both the texture data and range data of that person by scanning his/her head using Cyberware cyberscanner. An example of the cyberscanner data is shown in figure 3(b). 35 feature points (figure 3(c)) are manually selected on the scanned data. Those feature points have their correspondences on the generic geometry face model. We then fit the generic model to the person by deforming it based on the range data and the selected feature points. Manual adjustments are required where the scanned data are missing. Figure 3(d) show an example of a customized head model.

Wired frame Shaded

(a) Generic face model of iFACE

Texture data Range data

(b) Cyberscanner data



(c) 35 Feature points are selected for model fitting

Frontal view. Side view.

(d) The customized head model with texture mapping

Fig 2.3 Face Model

III .PROPOSED METHODOLOGY

This research work consists of following phases:

3.1. Face detection based on skin color

3.2. Face extraction and enhancement

3.3. Face features extraction

3.4. Curve formation using Bezier curve.

3.5. Fuzzy Patterns

3.6. Experiment Results

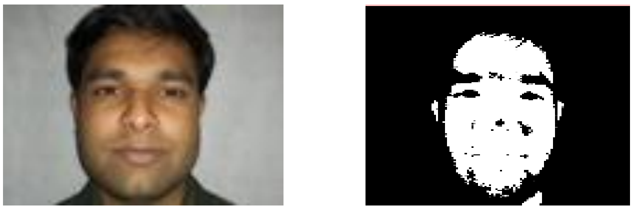
3.1 Face Detection Based on Skin Color:

International Journal of Advance Research in Science and Engineering
Volume No 06, Special Issue No. (02), September 2017, ICITTESE-17
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Skin color plays a vital role in differentiating human and non-human faces. From the study it is observe that skin color pixels have a decimal value in the range of 120 to 140. In this project, we used a trial and error method to locate skin color and non skin color pixels. But many of the times, system fails to detect whether an image contains human face or not (i.e. for those images where there is a skin color background).an image is segmented into skin color and non-skin color pixels

The skin pixels values are set to 1(i.e. #FFFF) and non skin pixels are set to 0(i.e. 0000). The pixels are collected and set

The resultant image becomes as



Original Image Skin and non-skin pixels

Fig. 3.1.1 (Phase I)

3.2 Face Extraction and Enhancement

Literature review point out that, FACS system technique is based on face features extractions like eye, nose, mouth, etc. In this project, we minimize the number of features (i.e. only eyes and mouth) but given the more weight age for fuzzy rules formations from these extracted features. Face extractions consist of following steps

- Let W and H are the width and height of skin and non-pixel image as shown in fig 3.1.1
- Read the pixel at position (0, H/2) which is a middle of i.e. left side of image.
- Travers a distance D_1 = W/6 in horizontal direction to get the start boundary pixel of skin region.
- Travers a distance D_2= H/6 from a pixel position (W/6, H/2) in upward directions. Same may do in downward direction and locate the points X_1, X_2.
- Travers a distance D_3=W/3 from the point X_1 and locate the point X_3. Same do from the point x_2 and locate the point X_4.
- Crop the square image as shown.

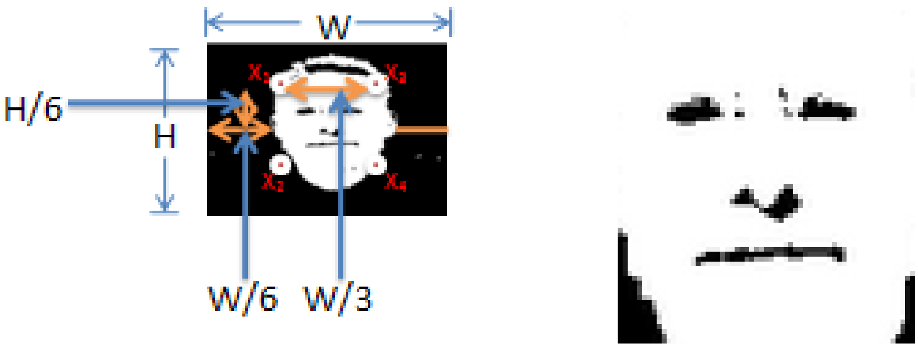


Fig 3.2.2 Face recognition

After face extraction white region pixels (i.e. skin pixels) are filled with skin color. A resultant image with skin color and after enhancement becomes as

Fig 3.2.3 image with skin color and after dimension enhancement

3.3 Face Features Extraction

Human face is made up of eyes; nose, mouth and chine etc. there are differences in shape, size, and structure of these organs. So the faces are differs in thousands way. One of the common methods for face expression recognition is to extract the shape of eyes and mouth and then distinguish the faces by the distance and scale of these organs. The face feature extractions consist of following steps

- Let W and H are width and height of an image shown in Fig 3.2.3
- Mark pixel P_i (W/2, H/2) as centre of image.
- Travers a distance H/8 from the pixel P_i towards upward and mark a point $K_{1.}$
- Travers a distance W/3 from the point K_1 towards leftward and mark a point $K_{2.}$
- Travers a distance H/10 towards downward from the point K_2 and mark a point $K_{3.}$
- Travers a distance W/4 from the point K_3 towards right and mark the point $K_{4.}$
- Travers a distance H/10 from the point K_4 toward up and mark the point $K_{5.}$
- Same steps are repeated for extracting the right eye and mark the point N_2, N_3, N_4, and N5.
- Travers a distance H/8 from the point P_i towards downward and mark the point M_1.
- Travers a distance W/6 towards left and right from the point M_1 and marks the point M_2 and $M_{3.}$
- Start with the point M_2 traverse a distance H/10 towards downward and mark the point $M_{4.}$
- Travers a distance W/6 from the point M_4 towards right and mark the point $M_{5.}$ Same may do from point M_5 and mark the point $M_{6.}$
- Travers the distance H/10 from M_6 towards up that meets to the point $M_{3.}$
- See the below image.

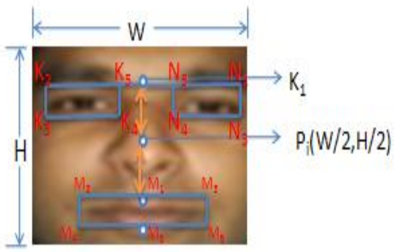


Fig 3.3.1 Feature Extraction

- Dist $|P_i - K_1| = H/8$

- Dist $| K_1 – K_2 |$ = Dist $| M_1 – M_2 |$ = Dist $| M_1 – M_3 |$ =

Dist $| M_4 – M_5 |$ = Dist $| M_5 – M_6 |$ = W/3

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

- Dist | $K_2 - K_3$ | = Dist | $K_4 - K_5$ | = Dist | $N_2 - N_3$ | =

Dist | $N_4 - N_5$ | = Dist | $M_2 - M_4$ | = Dist | $M_1 - M_5$ | =

Dist | $M_3 - M_6$ | = H/10

- Dist | $K_3 - K_4$ | = Dist | $K_5 - K_2$ | = Dist | $N_3 - N_4$ | =

Dist | $N_5 - N_2$ | = W/4

3.4 Curve formation using Bezier curve

Eyes and mouth as shown in fig 3.3.1 are located and extracted. Bezier curve formed from this eyes and mouth as per the equation

$Q(t) = \sum_{i=0}^{n} P_i B_{i,n}(t)$……………….eq. 3.4.1

Where each term in the sum is the product of blending function $B_{i,n}(t)$ and the control point P_i. The $B_{i,n}(t)$ is called as Bernstein polynomials and are defined by

$B_{i,n}(t) = C_i^n t^i (1-t)^{n-i}$…………...eq. 3.4.2

Where C_i^n is the binomial co-efficient given by:

$C_i^n = \frac{n!}{i!(n-i)}$…………………….....eq. 3.4.3

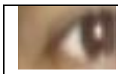
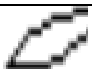
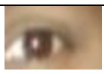
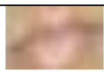
					
Left Eye	Left Eye Bezier Curve	Right Eye	Right Eye Bezier Curve	Mouth	Mouth Bezier Curve

Fig 3.4.1 Bezier Curve

Once the Bezier curve formed features points are located as shown in below image.

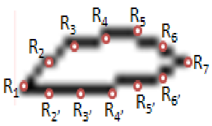
		
Left Eye	Right Eye	Mouth

Fig 3.4.2 Feature Point Location

3.5 Fuzzy Patterns

It is found that expression recognition from the still image never gives a correct output. A one expression id may also false into more than one expression domain. This project forms some fuzzy patterns for expressions. See the set theory diagram below

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

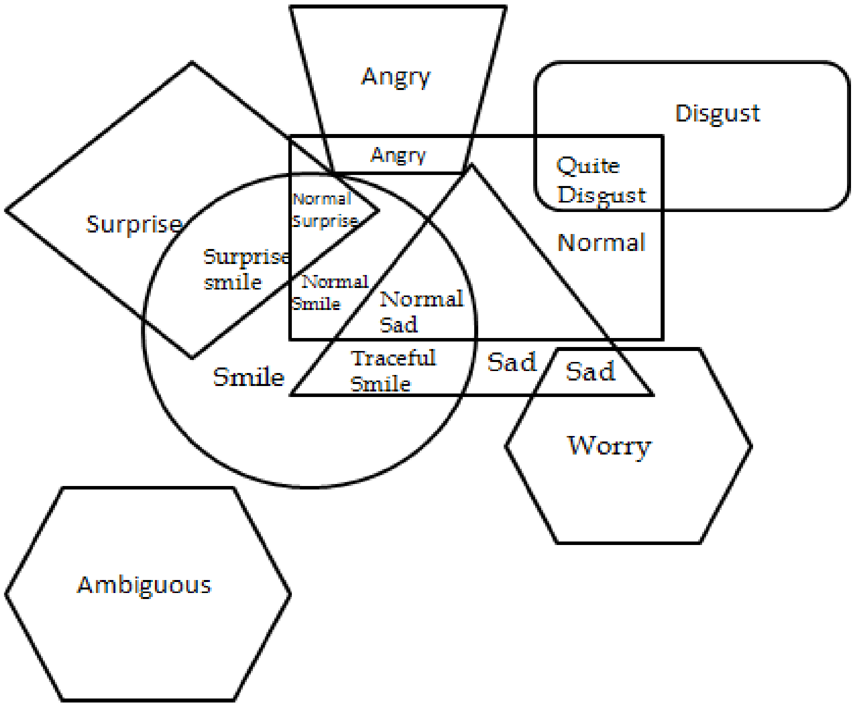


Fig 3.5.1 Fuzzy Expression Patterns

IV. CONCLUSION

This paper proposes a new approach for recognizing the category of facial expression an estimating the degree of continuous facial expression change from time sequential images. This approach is based on personal independent average facial expression model.

V. FUTURE WORK

It is found that, current system fails to recognize an expression of images containing skin color background and multiple faces in single image. A strong system is required to diagnose an expression of such image.

REFERENCES

[1] Y. Yacoob and L.S. Davis, "*Recognizing human facial expressions from long image sequences using optical flow*", IEEE Trans. Pattern Analysis & Machine Intelligence, Vol. 18, No 6, pp. 636-642, 1996.

[2] P. Ekman and W. Friesen, "*Facial Action Coding System*", Consulting Psychologists Press, 1977.

[3] K. Aizawa and T. S. Huang, "*Model-based image coding: Advanced video coding techniques for very low bit-rate applications*", Proc. IEEE, Vol. 83, No. 2, pp. 259-271, 1995.

[4] S. Kimura and M. Yachida, "*Facial expression recognition and its degree estimation*", Proc. Computer Vision and Pattern Recognition, pp.295-300, 1997.

[5] K. Ohba, G. Clary, T. Tsukada, T. Kotoku, and K. Tanie, "*Facial expression communication with FES*", Proc. International Conference on Pattern Recognition, pp. 1376-1378, 1998.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346
[6] M.A. Bhuiyan and H. Hama, "*Identification of Actors Drawn in Ukiyoe Pictures*", Pattern Recognition, Vol. 35, No. 1, pp. 93-102, 2002.

[7] M. B. Hmid and Y.B. Jemaa, *Fuzzy Classification, Image Segmentation and Shape Analysis for Human Face Detection*. Proc. Of ICSP, vol. 4, 2006.

[8] M. Wang, Y. Iwai, M. Yachida, "*Expression Recognition from Time-Sequential Facial Images by use of Expression Change Model*", Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 324 – 329, 1998.

[9] M. I. Khan and M. A. Bhuiyan, "*Facial Expression recognition for Human-Machine Interface*", ICCIT, 2006.