Model Based Face Recognition across Facial Expressions

Zahid Riaz, Christoph Mayer, Matthias Wimmer, and Bernd Radig

***Abstract*—This paper describes a novel idea of face recognition across facial expression variations using model based approach. The approach follows in 1) modeling an active appearance model (AAM) for the face image, 2) using optical flow based temporal features for facial expression variations estimation, 3) and finally applying binary decision trees as a classifier for facial identification. The novelty lies not only in generation of appearance models which is obtained by fitting active shape model (ASM) to the face image using objective but also using a feature vector which is the combination of shape, texture and temporal parameters that is robust against facial expression variations. Experiments have been performed on Cohn-Kanade facial expression database using 61 subjects of the database with image sequences consisting of more than 4000 images. This achieved successful recognition rate up to 91.17% using decision tree as classifier in the presence of six different facial expressions.**

***Index Terms*— Active Appearance Models, Face Recognition, Image Classification, Decision Trees**

I. INTRODUCTION

Since last three decades of face recognition technology, there exists many commercially available systems to identify human faces, however face recognition is still an outstanding challenge against different kinds of variations like facial expressions, poses, non-uniform light illuminations and occlusions. Meanwhile this technology has extended its role to Human-Computer-Interaction (HCI) and Human-Robot-Interaction (HRI). Person identity is one of the key tasks while interacting with the robots, exploiting the un-attentional system security and authentication of the human interacting with the system. This problem has been addressed in various scenarios by researchers resulting in commercially available face recognition systems [1,2]. However, other higher level applications like facial expression recognition and face tracking still remain outstanding along with person identity. This gives rise to an idea for generating a framework suitable for solving these issues together.

As cameras are widely used and mounted on computer screens, embedded into mobiles and installed into everyday living and working environments they become valuable tools for human system interaction. A particular important aspect of this interaction is detection and recognition of faces and interpretation of facial expressions. These capabilities are deeply rooted in the human visual system and a crucial building block for social interaction. Consequently, these capabilities are an important step towards the acceptance of many technical systems.

Although faces are the most important and natural way for human-human interaction but some outstanding challenges like uniqueness made its market value a bit less than the other biometrics in year 2003. But in 2006 Face recognition technology again raised up to 19% of the biometric market as shown comparably in Figure 1.

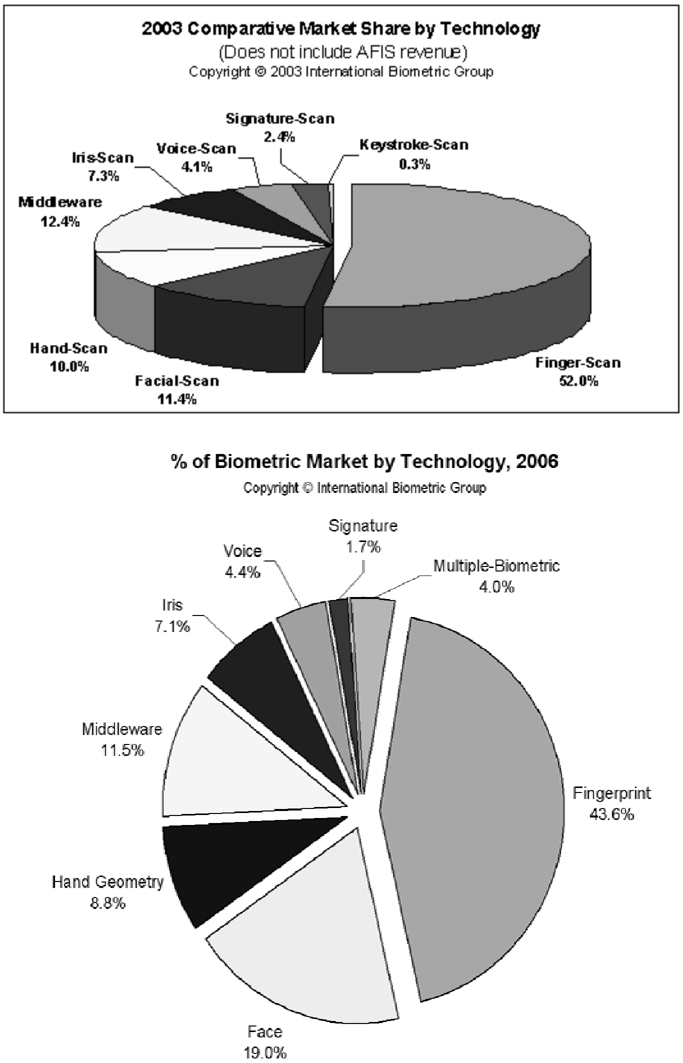

Figure 1 Face Recognition Technology compared to other Biometrics for years 2003 and 2006

Zahid Riaz, Department of Informatics, Technische Universität München, 85748 Garching bei München, Germany, riaz@in.tum.de
Christoph Mayer, Department of Informatics, Technische Universität München, 85748 Garching bei München, Germany, mayerc@in.tum.de
Matthias Wimmer, Department of Informatics, Technische Universität München, 85748 Garching bei München, Germany, matthias.wimmer@in.tum.de
Bernd Radig, PhD, Department of Informatics Technische Universität München, 85748 Garching bei München, Germany, radig@in.tum.de

This publication focuses on one of aspects of natural human-computer interfaces: our goal is to build a real-time system for facial recognition that could robustly run in real-world environments. We develop it using model-based image interpretation techniques, which have proven its great potential to fulfill current and future requests on real-world image understanding. Our approach comprises methods that robustly localize facial features, seamlessly track them through image sequences, and finally infer the facial recognition.

II. Related Work

A formal method of classifying the faces was first proposed by Francis Galton [2] in 1888. Galton proposed collecting facial profiles as curves, finding their norm, and then classifying other profiles by their deviations from the norm. The classification was resulting in a vector of independent measure that could be compared with other vectors in the database. Traditional recognition systems have the abilities to recognize the human using various techniques like feature based recognition, face geometry based recognition, classifier design and model based methods. Principal Components Analysis (PCA) was firstly used by Sirvovich and Kirby [3], which were latterly adopted by M. Turk and A. Pentland introducing the famous idea of eigenfaces [4,5]. This paper focuses on the modeling of human face using a two dimensional approach of shape, texture and temporal information and then utilizing this model for recognition purposes. The type of models using shape and texture parameters are called Active Appearance Models (AAMs), introduced by Cootes et al [6]. In [7] authors have used weighted distance classifier called Mahalanobis distance measure. However, Edwards et al [8] isolated the sources of variation by maximizing the interclass variations using Linear Discriminant Analysis (LDA), the technique which was holistically used for Fisherfaces representation [9] which is similar to the eigenface approach resulting in out performance of previous approach. In [10] authors have utilized shape and temporal features collectively to form a feature vector for facial expressions recognition. Lanitis et al [11] used separate information for shape and gray level texture. These models utilize the shape information based on a point distribution of various landmarks points marked on the face image. In our approach a predefined shape model consisting of 134 points in two dimensional space used by Wimmer et al [12]has been utilized.

III. Our Approach

The purpose of this paper is the hierarchal utilization of shape model fitting, texture mapping and estimating optical flow-based parameters for feature vector extraction. The feature vector consists of the shape, texture and temporal variations, sufficient for considering local variations in shapes. All the subjects in the database are labeled for identification. A fitted face model, on the training images is then used for defining the reference shape in our experiments. This reference shape is calculated by finding the mean shape of the all shapes in the database.

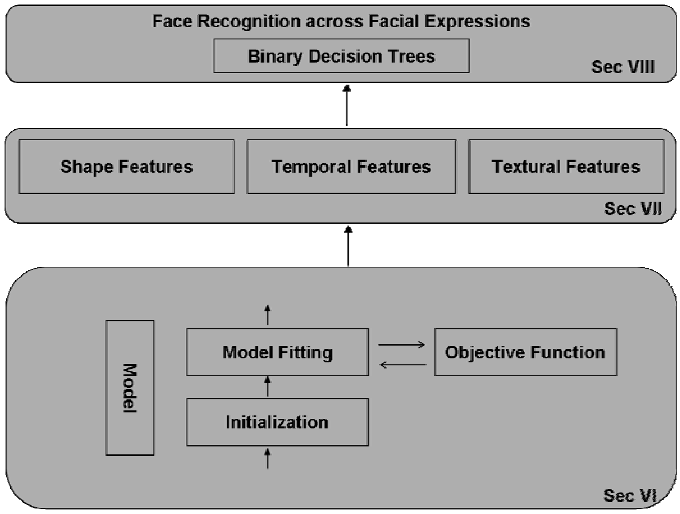

Figure 2 Model-based techniques split the challenge of image interpretation into computationally independent modules. The upper right corners refer to the sections with detailed explanation.

In this paper an explicit 2D face model is used to develop a baseline for feature extraction. A point distribution model (PDM) is used as an active shape model. This model comprises of 134 points that prominently defines the location of local face features like eyes, nose and lips in 2D space. Face is localized in the image. An objective function is learned for fitting this model to the faces. After fitting the model to the example face image, texture information is extracted from the example image on a reference shape which is the mean shape of all the shapes of database. Image texture is extracted using planar subdivisions of the reference and the example shapes. Texture warping between the subdivisions is performed using affine transformation. The resulting image texture is normalized to remove the global lighting effects. This image texture is now normalized both in the sense of shape and varying illuminations effects, making the image robust for shape and illumination. Principal Component Analysis (PCA) is used to obtain the texture and shape parameters of the example image. This approach is similar to extracting Active Appearance Model (AAM) parameters. In addition to AAM parameters, temporal features of the facial changes are also calculated. Local motion of the feature points is observed using optical flow. We use reduced descriptors by trading off between accuracy and runtime performance. A binary decision tree is used as classifier for person identification. A detailed process flow of our approach is shown in Figure 2.

Our approach achieves real-time performance and provides robustness for real-world applicability. This computer vision task comprises of various phases for which it exploits model-based techniques that accurately localize facial features, seamlessly track them through image sequences, and finally infer facial expressions visible. We specifically adapt state-of-the-art techniques to each of these challenging phases.

The remainder of this paper is divided in four sections. Section V deals with the localization of the face in an image. In section VI model based image interpretation is described along with the model fitting process. Sections VII discusses about the model based feature extraction technique comprising shape and appearance along with the temporal features. Section VIII deals with feature classification. Finally the experimentation is performed in section IX of this paper along with conclusions and future directions in section X.

IV. BENCHMARK DATABASE

The Cohn-Kanade-Facial-Expression database (CKFE-DB) contains 488 short image sequences of 97 different persons performing the six universal facial expressions [25]. It provides researchers with a large dataset for experimenting and benchmarking purpose. Each sequence shows a neutral face at the beginning and then develops into the peak expression. Furthermore, a set of action units (AUs) has been manually specified by licensed Facial Expressions Coding System (FACS)-experts for each sequence. Note that this database does not contain natural facial expressions, but volunteers were asked to act. Furthermore, the image sequences are taken in a laboratory environment with predefined illumination conditions, solid background and frontal face views. Algorithms that perform well with these image sequences are not immediately appropriate for real-world scenes.

V. FACE LOCALIZATION

Localizing a face inside an image can be performed using distinct feature of the face that is, skin colour. Although skin colour varies over the regions and ethnic groups in the world, but still is applied successfully by the many researchers [13]. For initialization, we used Viola and Jone's Haarlike features [14]. This defines a face inside a box to locate it in an image as shown in Figure 3. These are the efficient feature to calculate and are able to localize the face in an image under various poses and multi faces per image.

These features are useful for locating other local features like eyes in a face image as shown in Figure 4.

VI. MODEL-BASED IMAGE INTERPRETATION AND FITTING

Model-based techniques consist of four components: the model, the initialization algorithm, the objective function, and the fitting algorithm.

Our approach makes use of a statistics-based deformable model, as introduced by Cootes et al. [6]. The model contains a parameter vector **p** that represents its possible configurations, such as position, orientation, scaling, and deformation. Models are mapped onto the surface of an image via a set of feature points, a contour, a textured region, etc. Referring to [6], deformable models are highly suitable for analyzing human faces with all their individual variations.

Its parameters $p = (tx, ty, s, \theta, b)^T$ comprise the translation, scaling factor, rotation, and a vector of deformation parameters $b = (b_{s,1}, \ldots, b_{s,m})^T$. The latter component describes the configuration of the face, such as the opening of the mouth, roundness of the eyes, raising of the eye brows, see Figure 4.

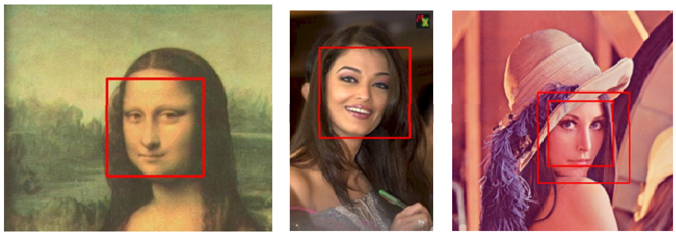

Figure 3 Face Localization

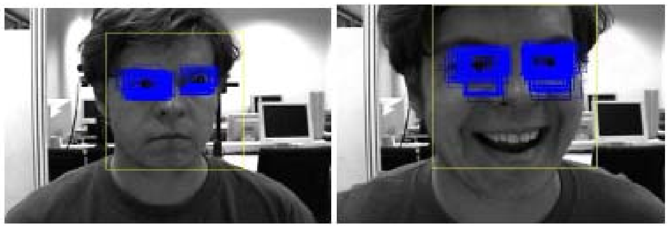

Figure 4 Eyes locator using Viola and Jone's algo

The initialization algorithm (see Figure 4) automatically starts the interpretation process by roughly localizing the object to interpret. It computes an initial estimate of the model parameters that needs to be further refined by the subsequent fitting algorithm. Our system integrates the approach of Viola and Jones, which is able to detect the affine transformation parameters (tx, ty, s, and θ) of frontal faces.

In order to obtain higher accuracy, we apply a second iteration of the Viola and Jones object detector to the previously determined image region of the face (see section V). This extension allows to roughly estimate the deformation parameters b_s as well, because we learn the algorithm to localize facial components, such as eyes and mouth. In the case of the eyes, our positive training examples contain the images of eyes, whereas the negative examples consist of image patches in the vicinity of the eyes, such as the cheek, the nose, or the brows. Note that the resulting eye detector is not able to robustly localize the eyes in a complex image, because it usually contains a lot of information that was not part of the training data. However, it is highly appropriate to determine the location of the eyes within a pure face image or within the face region of a complex image, see Figure 4 for some examples.

The objective function f(I, p) yields a comparable value that specifies how accurately a parameterized model *p* describes an image I. It is also known as the likelihood, similarity, energy, cost, goodness, or quality function. Without losing generality, we consider lower values to denote a better model fit. Traditionally, objective functions are manually specified by first selecting a small number of simple image features, such as edges or corners, and then formulating mathematical calculation rules. Afterwards, the appropriateness is subjectively determined by inspecting the result on example images and example model parameterizations. If the result is not satisfactory the function is tuned or redesigned from scratch. This heuristic approach relies on the designer's intuition about a good measure of fitness. Earlier works [15, 16] show that this methodology is erroneous and tedious. This traditional approach is depicted to the top in Figure 5.

To avoid these drawbacks, we recently proposed an approach that learns the objective function from annotated example images [15]. It splits up the generation of the objective function into several tasks partly automated. This provides several benefits: firstly, automated steps replace the labor-intensive design of the objective function. Secondly, this approach is less error prone, because giving examples of good fit is much easier than explicitly specifying rules that need to cover all examples. Thirdly, this approach does not rely on expert knowledge and therefore it is generally applicable and not domain-dependent. The bottom line is that this approach yields more robust and accurate objective functions, which greatly facilitate the task of the fitting algorithm. For a detailed description of our approach, we refer to [15].

The **fitting algorithm** searches for the model that best describes the face visible in the image. Therefore, it aims at finding the model parameters that minimize the objective function. Fitting algorithms have been the subject of intensive research and evaluation, e.g. Simulated Annealing, Genetic Algorithms, Particle Filtering, RANSAC, CONDENSATION, and CCD, see [17] for a recent overview and categorization. We propose to adapt the objective function rather than the fitting algorithm to the specifics of our application. Therefore, we are able to use any of these standard fitting algorithms, the characteristics of which are well-known, such as termination criteria, runtime, and accuracy. Due to real-time requirements, our experiments in Section VIII have been conducted with a quick hill climbing algorithm. Note that the reasonable specification of the objective function makes this local optimization strategy nearly as accurate as a global optimization strategy, such as Genetic Algorithms. Figure 6 shows the fitting results.

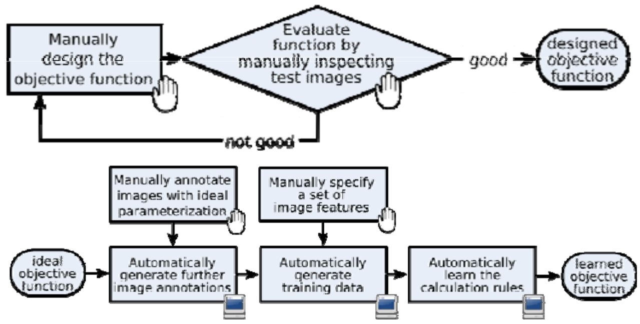

Figure 5 The traditional procedure for designing objective functions (top), and the proposed method for learning objective functions (bottom)

Figure 6 Fitting Shape Model

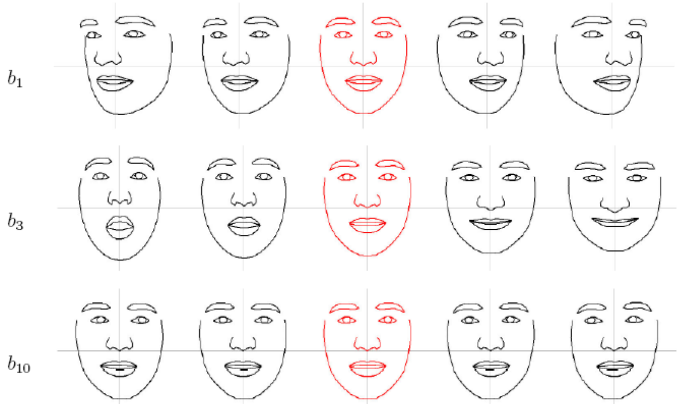

Figure 7 Deformation by change of just one parameter in each row. Topmost row: b1 rotates the head. Middle row: b3 opens the mouth. Lower-most row: b10 moves pupils in parallel

VII. Feature Extraction

A. *Active Shape Models*

Different kind of shape models have been introduced by researchers depending upon the application. Some are landmark based models [18,19,20] defining some fixed points annotated on the images and then defining the boundaries around the objects. However some rely on the contour based approach. Different contours define the shape of the object for outlining it along with covering the feature inside an object [21]. Landmark based models however provide the exact location of the features inside the object.

Fitting of this shape model on the face is performed by training an objective function (see section VI). The shape model used in this paper is shown in Figure 7 with different types of deformations against different parameters.

The model is parameterized using PCA to form the shape feature vector.

$$x \approx x_m + P_s b_s \qquad (1)$$

where the shape x is parameterized by using mean shape x_m and matrix of eigenvectors P_s to obtain the parameter vector b_s [22] (see Appendix-I).

B. *Appearance Model*

For the various images of the same person different types of variations are required to be modeled. For example, shape deformations including both facial expression changes and pose variations along with the texture variations caused by illuminations. For this reason, different normalizations are required to be performed at this stage. At first, shape variation is required to be controlled in order to record the texture. This can be achieved by defining a reference shape for a particular object. In our case, this reference image is mean shape, obtained by taking the mean of all the shapes of all persons in our database. Figure 8 (bottom-left) shows the mean shape of the subject in consideration. Since the points distribution defines a convex hull of points in space so a planar subdivision is defined for the reference shape to map image texture. Delaunay triangulation is used to divide the shape into a set of different facets. Figure 8 shows the delaunay triangulations of the reference shape. The delaunay triangulation is a triangulation which is equivalent to the nerve of the cells in a voronoi diagram, i.e., that triangulation of the convex hull of the points in the diagram in which every circumcircle of a triangle is an empty circle [23]. The detail of warping process is given in Appendix-II.

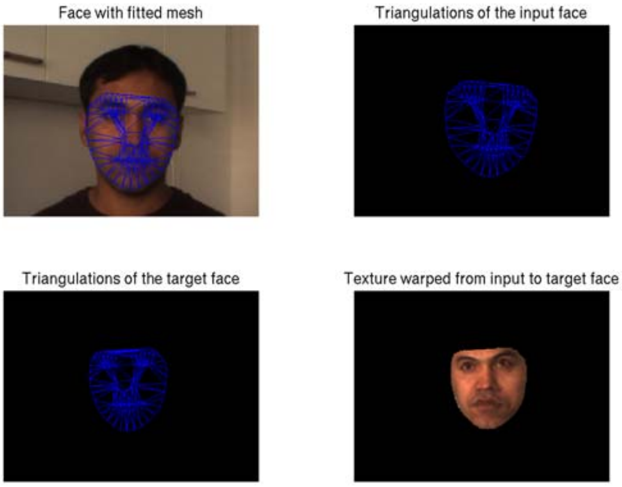

Figure 8 Shape Subdivisions and Texture Warping

Once the texture is extracted it could be parameterized using PCA as,

$$g = g_m + P_g b_g$$

Where the texture g is parameterized by using mean texture g_m and matrix of eigenvectors P_g to obtain the parameter vector b_g [22] (see Appendix-II).

C. Temporal Features

Since facial expressions emerge from muscle activity, the motion of particular feature points within the face gives evidence about the facial expression. These features further help the classifier to learn the motion activity. Real-time capability is important, and therefore, a small number of feature points is considered only. The relative location of these points is connected to the structure of the face model. Note that we do not specify these locations manually, because this assumes a good experience of the designer in analyzing facial expressions. In contrast, we automatically generate G feature points that are uniformly distributed, see Figure 9. We expect these points to move descriptively and predictably in the case of a particular facial expression. We sum up the motion $g_{x,i}$ and $g_{y,i}$ of each point $1 \leq i \leq G$ during a short time period. We set this period to 2 sec to cover slowly expressed emotions as well. The motion of the feature points is normalized by the affine transformation of the entire face (tx, ty, s, and θ) in order to separate the facial motion from the rigid head motion. In order to determine robust descriptors, PCA determines the H most relevant motion patterns (principal components) visible within the set of training sequences. A linear combination of these motion patterns describes each observation approximately correct. This reduces the number of descriptors ($H \leq 2G$) by enforcing robustness towards outliers as well. As a compromise between accuracy and runtime performance, we set the number of feature points to G = 140 and the number of motion patterns b_t to H = 14 containing. Figure 9 visualizes the obtained motion of the feature points for some example facial expressions.

The overall feature vector then becomes as in equation 3.

$$u = (b_{s,1}, \ldots.., b_{s,m}, b_{t,1}, \ldots.., b_{t,H}, b_{g,1}, \ldots.., b_{g,n})$$

Where b_s, b_t and b_g are shape, temporal and textural parameters respectively.

VIII. FEATURE CLASSIFICATION

With the knowledge of feature the vector u, a classifier infers the correct facial identity. We learn a Binary Decision Tree [24], which is a robust and quick classifier. However, any other multi-class classifier that is able to derive the class membership from real valued features can be integrated as well, such as a k-Nearest-Neighbour classifier. We take 66% of the image sequences of the CKFE-DB as the training set and the remainder as test set, the evaluation on which is shown in the next section.

The feature vector obtained consists of three different kinds of characteristics of the face image. There could be two approaches as this level to unify different types of parameters. One approach can define weights for the features so that they could be unified in one vector regardless of their diverse nature. This is achieved similar to traditional AAM parameters extraction techniques. However, the other approach is to use a classifier which can handle this issue by scaling. The latter technique is valid in our approach.

IX. EXPERIMENTS

Experiments have been performed on Cohn-Kanade facial database for human faces. Since this database consists of six standard facial expressions, it is useful to perform person identity experiment against these expressions. For experimental purposes, image sequences of 61 persons have been used which consists of overall 4060 images. A binary decision tree is trained as classifier in 22.99 sec. We used 1381 images for testing the recognition results and successfully recognized 1259 images. The recognition rate achieved was 91.17% in the presence of facial expressions. Figure 10 and 11 show true positive and true negative for the database respectively.

Figure 9 Model-based image interpretation for facial expression recognition: Fitting a deformable face model to images and inferring different facial expressions by taking shape and temporal image features into account.

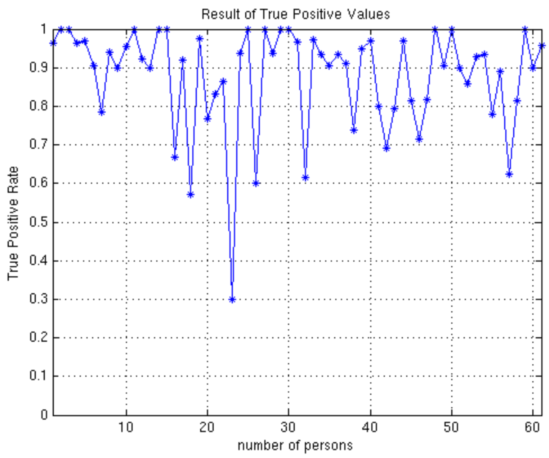

Figure 10 True Positive for 61 persons in experiments

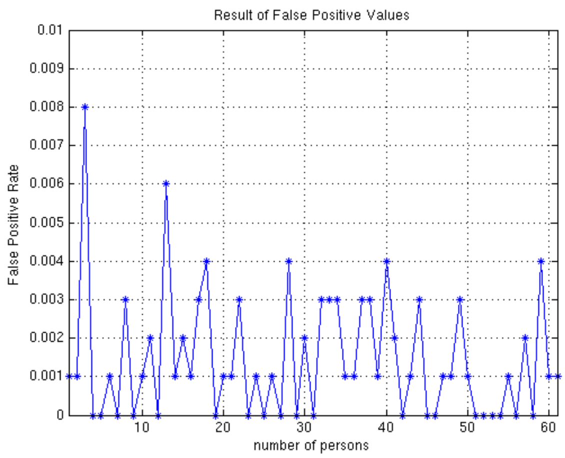

Figure 11 False Positive for 61 persons in experiments

Experiments have been performed on the standard AMD Athlon™ XP 2000+ processor with 1GB of RAM. The system developed is capable to work in real time.

Facial expressions tend to be mild during texture warping to the reference shape. This accounts for the views of the facial expression towards neutral face. This is shown in Figure 8.

X. Conclusion

We introduced an idea to develop a feature vector which consists of three types of facial variations and is robust against the expressional changes in the human faces in real environments. Since the training set consists of the facial expressions information of a person, it can recognition the person even under various expressions. A binary decision tree is efficient to train and classify. However the benchmarked database consists of frontal faces only. This technique is capable of working in real time environment and can be applied to HRI very efficiently. It can keep the person identity information even under the presence of facial expressions which could originate under human machine interaction scenarios. Further extensions of this work could improve the result using some different classifiers for instance, computing intra-class and inter-class variations. However, in real time environment the system can work by further improving it for light illuminations and using 3D information. The approach is initially developed for still images, which is applied to sequence set of images in this paper. We tested the feature extraction using our approach also work efficiently for real time images and videos.

Appendix-I

In case of shape model: 2D points are aligned in a column to create a point distribution model vector as follows:

$$x = (x_1, x_2, ..., x_n, y_1, y_2, ..., y_n)^T$$

Once the points are captured, following approach is utilised

Calculate the mean shape that is, reference shape:

$$\bar{x} = \frac{1}{s}\sum_{i=1}^{s} x_i$$

Covariance matrix is calculated as:

$$S = \frac{1}{s-1}\sum_{i=1}^{s}(x_i - \bar{x})(x_i - \bar{x})^T$$

Eigenvectors and eigenvalues for covariance matrix S are calculated. A decision for the eigenvalues is taken to truncate the eigenvectors as per requirement. If Φ_i is i^{th} eigenvector of the covariance matrix then we can write:

$$x \approx \bar{x} + P_s b_s$$

Where b_s is t-dimensional vector given by

$$b_s = P_s^T (x - \bar{x})$$

and P_s is

$$P_s = (\Phi_1 | \Phi_2 | .... | \Phi_t)$$

For image texture: A vector consisting of gray values is used to obtain the textural parameters. The process is similar.

Appendix-II

Given a set of shape feature points 'x' of the input example image and 'x_{avg}' of the average image, we can find the texture vector g_{im} as follows:

a) Compute the pixel position in the average shape.

b) Find the relative position in example image using affine transformation.

c) Sample the texture values at the points inside convex hull of the average image forming texture vector g_{im}.

The texture vector is normalized to remove global lighting effects. This is performed by applying the linear transformation,

$$g = (g_{im} - \beta 1)/\alpha$$

Where

$$\beta = (g_{im}.1)/n$$

$$\alpha = |g_{im}|^2 / n - \beta$$

We apply PCA approach this time to find texture parameters,

$$g = \bar{g} + P_g b_g$$

Where $\bar{g}$ is mean normalized texture vector, P_g is a set of eigenvectors and b_g is a set texture parameters.

Piecewise affine transform is used to warp the texture of example images on the normalized image. If x_1, x_2 and x_3 are the vertices of a triangle then any point x lying inside the triangle can be written as:

$$x = \alpha x_1 + \beta x_2 + \gamma x_3$$

Where

$$\alpha + \beta + \gamma = 1$$

For a given point (x, y) the values of α, β and γ are given by:

$$\alpha = 1 - (\beta + \gamma)$$

$$\beta = \frac{yx_3 - x_1 y - x_3 y_1 - y_3 x + x_1 y_3 + xy_1}{-x_2 y_3 + x_2 y_1 + x_1 y_3 + x_3 y_2 - x_3 y_1 - x_1 y_2}$$

$$\gamma = \frac{xy_2 - xy_1 - x_1 y_2 - x_2 y + x_2 y_1 + x_1 y}{-x_2 y_3 + x_2 y_1 + x_1 y_3 + x_3 y_2 - x_3 y_1 - x_1 y_2}$$

ACKNOWLEDGMENT

This work is mainly supported by cluster of excellence (CoTeSys) at Technical University of Munich, Germany, for human robot joint activities. Further we are thankful to Higher Education Commission (HEC) of Pakistan.

REFERENCES

[1] W. Zhao, R. Chellapa, A. Rosenfeld and P.J. Philips, "Face Recognition: A Literature Survey", *UMD CFAR Technical Report CAR-TR-948*, 2000.

[2] William A. Barrett, "A Survey of Face Recognition Algorithms and Testing Results", *Proceeding of IEEE*, 1998.

[3] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces", *J. Opt. Soc. Am. A, Vol. 4, No. 3*, March 1987, pp 519-524

[4] Turk M.A., Pentland A.P., "Face Recognition using Eigenfaces", *IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp 586-591.

[5] Turk M.A., Pentland A.P., "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience* 3 (1): 1991, pp 71-86.

[6] Edwards, G.J., Taylor, C.J., Cootes, T.F., "Interpreting Face Images using Active Appearance Models", *Proceedings of International Conference on Automatic Face and Gesture Recognition* (1998) 300-305.

[7] G.J. Edwards, T.F. Cootes and C.J. Taylor, "Face Recognition using Active Appearance Models", *in Proc. European Conference on Computer Vision* 1998, vol. 2, pp- 581-695, Springer 1998.

[8] G. J. Edwards, A. Lanitis, C. J. Taylor, and T. Cootes. "Statistical Models of Face Images: Improving Specificity", *In British Machine Vision Conference* 1996, Edinburgh, UK, 1996

[9] P. N. Belheumeur, J.P. Hespanha and D.J. Kreigman, "Eigenfaces vs Fisherfaces: Recognition using Class Specific Linear Projection", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 19, No. 7, July 1997.

[10] M. Wimmer, Z. Riaz, C. Mayer and B. Radig, *Recognizing Facial Expressions Using Model-Based Image Interpretation*, Advances in Human-Computer Interaction, I-Tech Book.

[11] Lantis et al, "A unified Approach to Coding and Interpreting Face Images" *in 5th International Conference on Computer Vision*, Cambridge USA, 1995

[12] Wimmer M, Stulp F, Tschechne S, and Radig B, Learning Robust Objective Functions for Model Fitting in Image Understanding Applications, In Proceedings of the 17th British Machine Vision Conference, pp 1159?1168, BMVA, Edinburgh, UK, 2006.

[13] V. Vladimir et al "A Survey on Pixel Based Skin Color Detection Technique", *in Graphicon* 2003.

[14] Paul Viola and Michael Jones. "Rapid object detection uses a boosted cascade of simple features", *In Computer Vision and Pattern Recognition, volume 1*, pages 511–518, Kauai, Hawaii, 2001.

[15] Matthias Wimmer, Freek Stulp, Sylvia Pietzsch, and Bernd Radig. "Learning local objective functions for robust face model fitting". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 30(8), 2008. to appear.

[16] Wimmer et al. "Learning robust objective functions for model fitting in image understanding applications" *Proceedings of the 17th British Machine Vision Conference (BMVC), volume 3,* pages 1159–1168, Edinburgh, UK, September 2006. BMVA

[17] Robert Hanek. Fitting Parametric Curve Models to Images Using Local Self-adapting Seperation Criteria. PhD thesis, Department of Informatics, Technische Universität München, 2004

[18] Cootes T.F., Edwards G.J., Taylor C.J., Active Appearance Models, Proc. European Conference on Computer Vision 1998 H.Burkhardt & B. Neumann Ed.s. Vol. 2, pp. 484-498, Springer, 1998

[19] Stegmann M.B., Active Appearance Models: Theory Extensions and Cases, Master Thesis, Technical University of Denmark, 2000.

[20] J. Ahlberg, *An Experiment on 3D Face Model Adaptation using the Active Appearance Algorithm*, Image Coding Group, Deptt of Electric Engineering, Linköping University. Gupta, H., R. Chowdhury, A. K., Chellappa, R., Contour-based 3D Face Modeling from a Monocular Video, British Machine Vision Conference 2004.

[21] Stan Z. Li and Anil K. Jain, Handbook of Face recognition, Springer 2005.

[22] Okabe, A.; Boots, B.; and Sugihara, K. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. New York: Wiley, 1992.

[23] Ross Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, California, 1993.

[24] Kanade, T., Cohn, J.F., Tian, Y. (2000). Comprehensive database for facial expression analysis, Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.

[25] Wallace V. Friesen and Paul Ekman. Emotional Facial Action Coding System. Unpublished manuscript, University of California at San Francisco, 1983.