A Unified Features Approach to Human Face Image Analysis and Interpretation

Zahid Riaz, Suat Gedikli, Micheal Beetz and Bernd Radig
Department of Informatics,
Technische Universität München
85748 Garching, Germany
{riaz|gedikli|beetz|radig}@in.tum.de

Abstract

Human face image analysis has been one of the challenging fields over the last few years. Currently, many commercially available systems can interpret face images in an efficient way but are generally limited to only one specific application domain. On the other hand, cameras are becoming a useful tool in human life and are the vital constituent of most of the interactive systems. In this regard, we present a technique to develop a unified set of features extracted from a 3D face model. These features are successfully used for higher level facial image interpretation in different application domains. We extract the feature once and use the same set of features for three different applications: face recognition, facial expressions recognition and gender recognition and obtain a good accuracy level while preserving the generality and efficiency of our feature extraction technique. The proposed technique is easy to implement and real time.

Keywords: Features extraction, Model based face image analysis, Facial behaviors analysis, Human robot interaction (HRI)

1 Introduction

By the recent advancement in the field of robotic technology, future systems would be mainly relying on intelligent machines interacting with the humans in daily life environment. A good example is the assistive robots (1) where robots serve as an attendee or nurse to the elderly people or handicaps. On the other side, cameras are the integral part of the most of the daily life technical systems, for example mobiles, notebooks, ATM machines and access control applications. This role of cameras and quality can further improve future systems by embedding higher level image interpretation. Currently available cameras have face detection capabilities, automatic smile capture and laptops with face recognition systems.

In human robot interaction (HRI) faces play an important role and are the natural way to interact. Human can predict identity, gender, facial expression, behavior, ethnical background and can estimate age almost on very first glance. Whereas the current assistive robots are struggling not only to be structurally humanoid but also to exhibit humanly behavior. For instance, in face recognition applications currently available systems work well under constrained environments however still require improvements. The efficiency and performance is still a trade off in unconstrained systems. Further extracting all possible information from the face images at the same time is quite difficult and yet not fully described on a common platform by the research community. In this paper we focus on the development of such a system which can perform these capabilities at the same time and hence give sufficient time to the robot to continue its job for other interaction activities. We focus on a model based approach to extract multi-features from human face images. These features can be used for face recognition, facial expressions, facial behaviors and gender classification. Further this system is trained to work under different variations like poses and illumination.

In this regard, an intuitive approach is to borrow the concepts of human-human interaction and train the machines for comparable performance in real world situations. Human faces play an important role in daily life interaction. Therefore, learning identity, expressions, gender, age and facial behavior assures better interaction and utilizes improved context information. For example, a system aware of person identity information can better employ user's habits and store current interaction knowledge for improving future interactions. In the recent decade model based image analysis of human faces has become a challenging field due to its capability to deal with the real world scenarios. Further it outperforms the previous techniques which were constrained to user's intervention with the system either to manually interact with system or to be frontal to the camera. Currently available model based techniques are trying to deal with some of the future challenges like developing state-of-the-art algorithms, improving efficiency, fully automated system development and versatility in different applications. In this paper we deal especially with versatility and diversity of the problem. Our aim is to extract features which are fully automatic and versatile enough for different applications. These capabilities of the system suggest to apply it in interactive scenarios like human machine interaction, token less access control, facial analysis for person behavior and person security. Models take benefit of the prior knowledge of the object shape and hence try to match themselves only with the object in an image for which they are designed. Face models impose knowledge about human faces and reduce high dimensional image data to a small number of expressive model parameters. We integrate the three-dimensional Candide-III face model (10) that has been specifically designed for observing facial features variations defined by facial action coding system (FACS) (16). The model parameters together with extracted texture and motion information are utilized to train classifiers that determine person-specific information. Our feature vector for each image consists of structural, textural and temporal variations of the faces in the image sequence. Shape and textural parameters define active appearance models (AAM) in partial 3D space with shape parameters extracted from 3D landmarks and texture from 2D image. Temporal features are extracted using optical flow. These extracted features are more informative than conventional AAM parameters since we consider local motion patterns in the image sequences in the form temporal parameters.

The remainder of this paper is divided in four sections. Section 2 explains the related research work for model based approaches, including 2D and 3D models. Section 3 describes the problem statement in detail followed by the proposed solution in section 4. Section 5 describes feature extraction technique in our case along with the future extensions. These features are used in section 6 for experimentation. In the last section, we describes the analysis of our technique along with the future extensions of this research work.

2 Related work

The proposed system consists of different modules which work adjacently to each others. We describe related work regarding to each module independently and explain if there exists any dependency among these modules. The overall system consists of face detection and localization for model fitting and structural features extraction, textural features extraction, pose compensation, illuminations compensations, temporal features extraction and finally synthesizing the image for final features extraction. The feature set is then applied different types of applications. Our approach consists of different parallel and sequentially working processes.

Human face image analysis for higher level information extraction is generally performed using appearance based, templates based, model based or hybrid methods based approaches (13). Appearance based analysis comprise of holistic approaches which exploit subspace projection. The extracted features are analyzed at a global level and local facial features are not considered in this regard. These methods mainly involve principal component analysis (PCA), linear discriminant analysis (LDA) and independent component analysis (ICA). The performance of these methods in general, degrades as more variations are considered in the face images like misalignment of local features, pose and lighting variations. In order to overcome these discrepancies, methods of image alignment have been devised by the researchers (13). Several face models and fitting approaches have been presented in the recent years. Cootes et al. (8) introduced modeling face

shapes with Active Contours. Further enhancements included the idea of expanding shape models with texture information (9). In contrast, three-dimensional shape models such as the Candide-3 face model consider the real-world face structure rather than the appearance in the image. Blanz et al. propose a face model that considers both, the three-dimensional structure as well as its texture (2). However, model parameters that describe the current image content need to be determined in order to extract high-level information, a process known as model fitting. In order to fit a model to an image. Van Ginneken et al. learned local objective functions from annotated training images (3). In this work, image features are obtained by approximating the pixel values in a region around a pixel of interest. The learning algorithm which uses to map images features to objective values is a k-Nearest-Neighbor classifier (kNN) learned from the data. We use similar methodology developed by Wimmer et al. (7) which combines multitude of qualitatively different features (22), determines the most relevant features using machine learning and learns objective functions from annotated images (3). To extract descriptive features from the image, Michel et al. (17) extracted the location of 22 feature points within the face and determine their motion between an image that shows the neutral state of the face and an image that represents a facial expression. The very similar approach of Cohn et al. (18) uses hierarchical optical flow in order to determine the motion of 30 feature points. A set of training data formed from the extracted features is utilized to learn on a classifier. For facial expressions, some approaches infer the expressions from rules stated by Ekman and Friesen (16). This approach is applied by Kotsia et al. (19) to design Support Vector Machines (SVM) for classification. Michel et al. (17) train a Support Vector Machine (SVM) that determines the visible facial expression within the video sequences of the Cohn-Kanade Facial Expression Database by comparing the first frame with the neutral expression to the last frame with the peak expression. In order to perform face recognition applications many researchers have applied model based approaches. Edwards et al (5) use weighted distance classifier called Mahalanobis distance measure for AAM parameters. However, they isolate the sources of variation by maximizing the inter class variations using Linear Discriminant Analysis (LDA), a holistic approach which was used for Fisherfaces representation (6). Riaz et al (20) apply similar features for explaining face recognition using Bayesian networks. However results are limited to face recognition application only. They used expression invariant technique for face recognition, which is also used in 3D scenarios by Bronstein et al (11) without 3D reconstruction of the faces and using geodesic distance. Park et. al. (12) apply 3D model for face recognition on videos from CMU Face in Action (FIA) database. They reconstruct a 3D model acquiring views from 2D model fitting to the images.

3 Problem Statement

In the recent decade, model based approaches have attained a huge attention of the research community owing to their compactness and detailed description over the other techniques. Models described a large size image in small set of parameters. These few descriptors are called model parameters. Further they narrow the search domain in an image and precisely look for the object of interest for which they are designed. This reduces false alarms in finding an object. The models used in face image analysis are active shape models (ASM), active appearance models (AAM), deformable models, wireframe models, and 3D morphable models (21). We address the problem in which a robotic system is able to extract a common feature set automatically from face images and capable to classify gender, person identity and facial behavior. In such applications an automatic and efficient feature extraction technique is necessary to be developed which can interpret every possible face information. Currently available systems lack this property. A major reason is that researchers focus on isolating the sources of variations while focusing on a particular application. For example, in face recognition application, many researchers normalize face in order to remove facial expressions variations to improve face recognition results. So the extracted features do not contain facial expressions information. We address an idea to develop a unified feature set which is used for different applications like face recognition, facial expressions and behavior and gender classification.

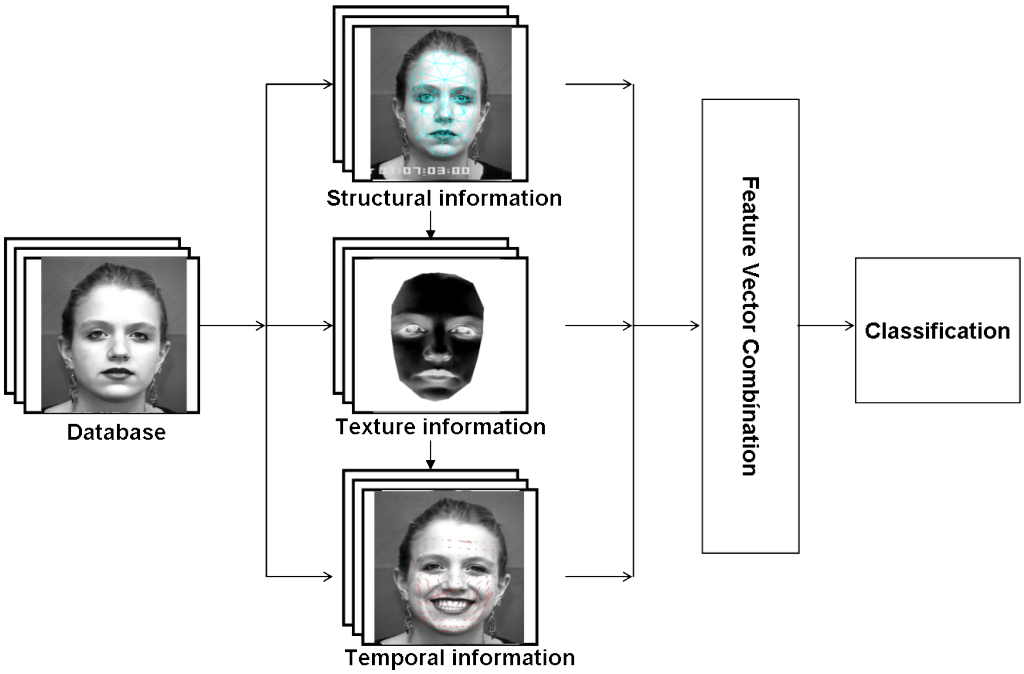

Figure 1: Our Approach: Sequential flow for feature extraction

4 Our approach: proposed solution

We propose a model based image analysis solution to this problem. Model parameters are obtained in an optimal way to maximize information from face region under various factors like facial pose, expressions, illuminations and self occlusions. In this section we explain our approach in different modules including shape model fitting, image warping, pose, illumination and expressions normalizations and finally synthesizing the image to extract model parameters.

We use a wireframe 3D face model known as Candide-III (10). The model is fitted to the face image by learning objective functions. We find correspondences between 2D and 3D points for model fitting and texture mapping. Texture information is mapped from the example image to a reference shape which is the mean shape of all the shapes available in database. However the choice of mean shape is arbitrary. Image texture is extracted using planar subdivisions of the reference and the example shapes. We use delaunay triangulations of the distribution of our model points. Texture warping between the triangulations is performed using affine transformation. Principal Component Analysis (PCA) is used to obtain the texture and shape parameters of the example image. This approach is similar to extracting 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 run time performance. These features are then used for classification. Our approach achieves real-time performance and provides robustness against facial expressions in real-world scenarios. Currently the system finds the pose information implicitly in structural parameters whereas illuminations are dealt in appearance parameters. This computer vision task comprises of various phases shown in Figure 1 for which it exploits model-based techniques that accurately localize facial features, seamlessly track them through image sequences, and finally infer facial features. We specifically adapt state-of-the-art techniques to each of these challenging phases.

5 Determining High-Level Information

In order to initialize, we apply the algorithm of Viola et al. (23) to roughly detect the face position within the image. Then, model parameters are estimated by applying the approach of Wimmer

et al. (7) because it is able to robustly determine model parameters in real-time.

To extract descriptive features, the model parameters are exploited. The model configuration represents information about various facial features, such as lips, eye brows or eyes and therefore contributes to the extracted features. These structural features include both, information about the person's face structure that helps to determine person-specific information such as gender or identity. Furthermore, changes in these features indicate shape changes and therefore contribute to the recognition of facial expressions.

The shape $\boldsymbol{x}$ is parameterized by using mean shape $\boldsymbol{x_m}$ and matrix of eigenvectors $\boldsymbol{P_s}$ to obtain the parameter vector $\boldsymbol{b_s}$ (14).

$$\boldsymbol{x} = \boldsymbol{x_m} + \boldsymbol{P_s b_s} \tag{1}$$

Once we have shape information of the image, we extract texture from the face region by mapping it to a reference shape. A reference shape is extracted by finding the mean shape over the dataset. Image texture is extracted using planar subdivisions of the reference and the example shapes. Texture warping between the subdivisions is performed using affine transformation. This extracted texture is parameterized in the same manner as shape information using PCA.

The extracted texture is parameterized after illumination normalization using mean texture $\boldsymbol{g_m}$and matrix of eigenvectors $\boldsymbol{P_g}$to obtain the parameter vector $\boldsymbol{b_g}$ (14). Figure 2 shows shape model fitting and texture extracted from face image.

$$\boldsymbol{g} = \boldsymbol{g_m} + \boldsymbol{P_g b_g} \tag{2}$$

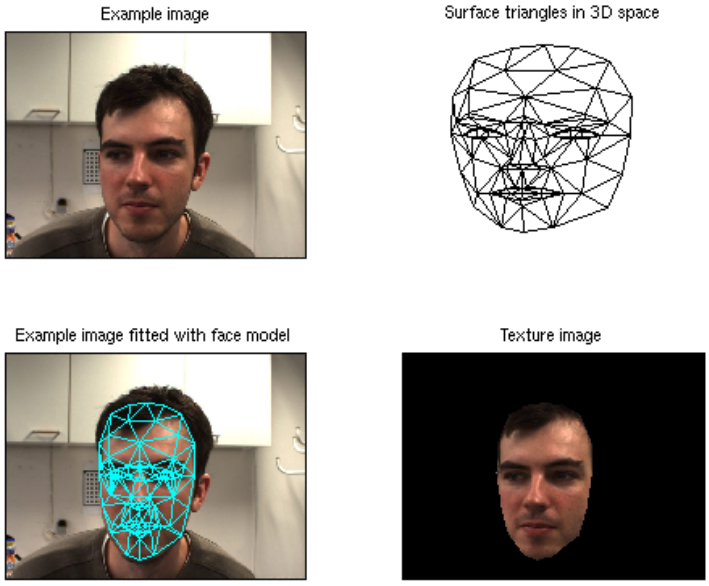

Figure 2: Texture information is represented by an appearance model. Parameters of the fitted model are extracted to represent single image information.

Further, temporal features of the facial changes are also calculated that take movement over time into consideration. Local motion of feature points is observed using optical flow. We do not specify the location of these feature points manually but distribute equally in the whole face region. The number of feature points is chosen in a way that the system is still capable of performing in real time and therefore inherits a tradeoff between accuracy and runtime performance. Figure 3 shows motion patterns for some of the images from database.

	BDT	BN
Face Recognition	98.49%	90.66%
Facial Expressions Recognition	85.70%	80.57%
Gender Classification	99.08%	89.70%

Table 1: Unified features tested with two classifiers Bayesian Networks (BN) and Binary Decision Tree (BDT)

We combine all extracted features into a single feature vector. Single image information is considered by the structural and textural features whereas image sequence information is considered by the temporal features. The overall feature vector becomes:

$$\boldsymbol{u} = (\boldsymbol{b_{s1}}, ..., \boldsymbol{b_{sm}}, \boldsymbol{b_{g1}}, ..., \boldsymbol{b_{gn}}, \boldsymbol{b_{t1}}, ..., \boldsymbol{b_{tp}},) \tag{3}$$

Where $\boldsymbol{b_s}$, $\boldsymbol{b_g}$ and $\boldsymbol{b_t}$ are shape, textural and temporal parameters respectively.

Figure 3: Motion patterns within the image are extracted and the temporal features are calculated from them. These features are descriptive for a sequence of images rather than single images.

6 EXPERIMENTS

For experimentation purposes, we benchmark our results on Cohn Kanade Facial Expression Database (CKFED). The database contains 488 short image sequences of 97 different persons performing six universal facial expressions (15). Furthermore, a set of action units (AUs) has been manually specified by licensed Facial Expressions Coding System (FACS) (16) experts for each sequence.

In order to experiment feature versatility we use two different classifiers with same feature set on three different applications: face recognition, facial expressions recognition and gender classification. The results are evaluated using classifiers from weka (24) with 10-fold cross validation. Table 1 shows different recognition rates achieved during experimentations using Bayesian Networks (BN) and Binary Decision Tree (BDT). In all three cases BDT outperforms BN. This can be analyzed in the figures. Figure 4 shows true positive and false positive rates for all the subjects in the database for face recognition. Figure 5 shows receiver operating characteristic (ROC) curves for six different facial expressions. Since laugh and fear are mostly confused facial expressions, it

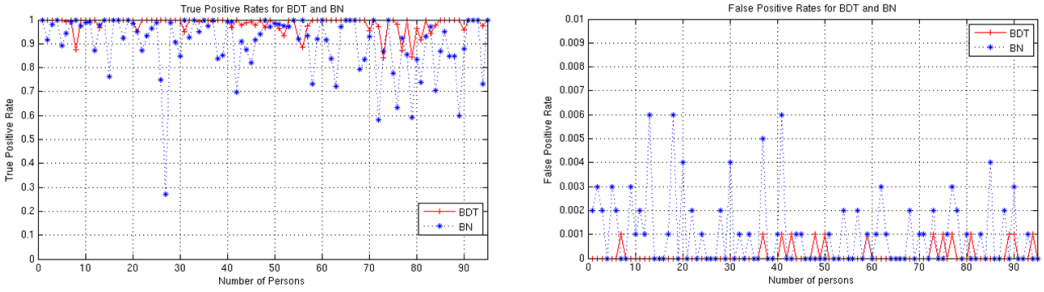

Figure 4: True positive and false psitive rates for face recognition

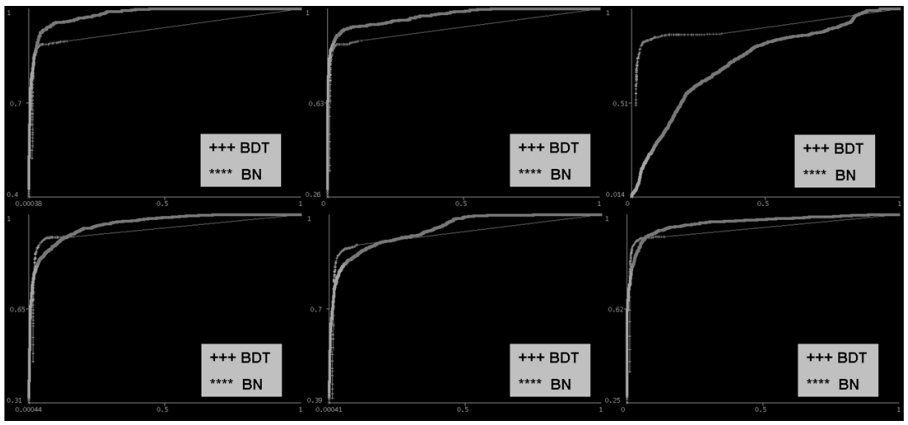

Figure 5: ROC curves for facial expressions Top (left to right): anger, disgust, fear, Bottom (left to right): laugh, sadness, surprise

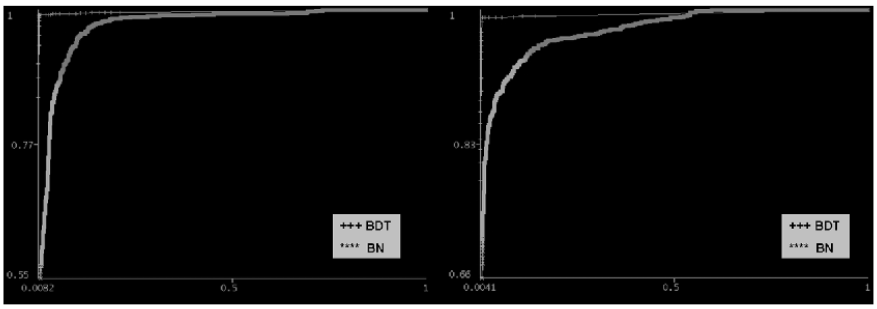

Figure 6: ROC curves for gender classification (left) female, (right) male

can be analyzed from the curves that there exists some confusion between these two expressions for BN classifier. Figure 6 shows gender classification results.

7 Conclusions

In this paper we introduced an idea to solve the face image analysis problem in human robot joint interaction scenarios using a common set of features. The approach is intuitive to human-human interaction where humans can extract these information on a very first look. Another advantage of this approach is to find a robust features set only once during the joint interaction and leaving much of the time for other joint activities. We addressed a feature extraction technique for a face recognition in the presence of facial expressions, recognizing facial expressions and gender. We deal majorly with facial expressions variations on a benchmark database and pose and lighting conditions are considered implicitly in the shape and texture parameters respectively. However, future work includes extension of these features under the presence of other variations. We aim a robust system to apply in real time human robot interaction.

References

[1] Beetz M. et. al. The Assistive Kitchen — A Demonstration Scenario for Cognitive Technical Systems In *Proceedings of the 4th COE Workshop on Human Adaptive Mechatronics (HAM)*,

2007

[2] Blanz V., Vetter T. Face Recognition Based on Fitting a 3D Morphable Model. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol.25 no. 9, pp 1063 - 1074, 2003.

[3] B. Ginneken, A. Frangi, J. Staal, B. Haar, and R. Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924–933, 2002.

[4] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

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

[6] 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.

[7] 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* pp1159–1168, BMVA, Edinburgh, UK, 2006.

[8] Tim F. Cootes and Chris J. Taylor. Active shape models – smart snakes. In *Proceedings of the 3rd British Machine Vision Conference* pages 266 - 275. Springer Verlag, 1992.

[9] Cootes T. F., Edwards G. J., Taylor C. J. Active Appearance Models. In *Proceedings of European Conference on Computer Vision* Vol. 2, pp. 484-498, Springer, 1998.

[10] J. Ahlberg. An Experiment on 3D Face Model Adaptation using the Active Appearance Algorithm. *Image Coding Group, Deptt of Electric Engineering* Linköping University.

[11] Bronstein, A. Bronstein, M. Kimmel, R. Spira, A. 3D face recognition without facial surface reconstruction, In Proceedings of European Conference of Computer Vision Prague, Czech Republic, May 11-14, 2004

[12] Unsang Park and Anil K. Jain 3D Model-Based Face Recognition in Video *2nd International Conference on Biometrics* Seoul, Korea, 2007

[13] W. Zhao, R. Chellapa, A. Rosenfeld and P.J. Philips, Face Recognition: A Literature Survey In *ACM Computing Surveys* Vol. 35, No. 4, December 2003, pp. 399-458.

[14] Stan Z. Li and A. K. Jain. Handbook of Face recognition *Springer* 2005

[15] Kanade, T., Cohn, J. F., Tian, Y. (2000). Comprehensive database for facial expression analysis In *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FGR00)*, Grenoble, France, 46-53.

[16] P. Ekman and W. Friesen. *The Facial Action Coding System: A Technique for The Measurement of Facial Movement*. Consulting Psychologists Press, San Francisco, 1978.

[17] P. Michel and R. E. Kaliouby. Real time facial expression recognition in video using support vector machines. In *Fifth International Conference on Multimodal Interfaces*, pages 258–264, Vancouver, 2003.

[18] J. Cohn, A. Zlochower, J.-J. J. Lien, and T. Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 396 – 401, April 1998.

[19] I. Kotsia and I. Pitaa. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transaction On Image Processing*, 16(1), 2007.

[20] Riaz Z. et al. A Model Based Approach for Expression Invariant Face Recognition In *3rd International Conference on Biometrics*, Italy, June 2009

[21] A. Scheenstra, A. Ruifrok and R. C. Veltkamp, A Survey of 3D Face Recognition Methods In *AVBPA 2005, LNCS 3546* pp. 891-899, 2005

[22] S. Romdhani. *Face Image Analysis using a Multiple Feature Fitting Strategy.* PhD thesis, University of Basel, Computer Science Department, Basel, CH, January 2005.

[23] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[24] Ian H. Witten and Eibe Frank Data Mining: Practical machine learning tools and techniques *Morgan Kaufmann*, 2nd Edition, San Francisco, 2005.