We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



5

Towards Unconstrained Face Recognition Using 3D Face Model

Zahid Riaz[1], M. Saquib Sarfraz[2] and Michael Beetz[1]
[1] *Intelligent Autonomous Systems (IAS), Technical University of Munich, Garching*
[2] *Computer Vision Research Group, COMSATS Institute of Information Technology, Lahore*
[1]*Germany*
[2]*Pakistan*

1. Introduction

Over the last couple of decades, many commercial systems are available to identify human faces. However, face recognition is still an outstanding challenge against different kinds of real world variations especially facial poses, non-uniform lightings and facial expressions. Meanwhile the face recognition technology has extended its role from biometrics and security applications to human robot interaction (HRI). Person identity is one of the key tasks while interacting with intelligent machines/robots, exploiting the non intrusive system security and authentication of the human interacting with the system. This capability further helps machines to learn person dependent traits and interaction behavior to utilize this knowledge for tasks manipulation. In such scenarios acquired face images contain large variations which demands an unconstrained face recognition system.

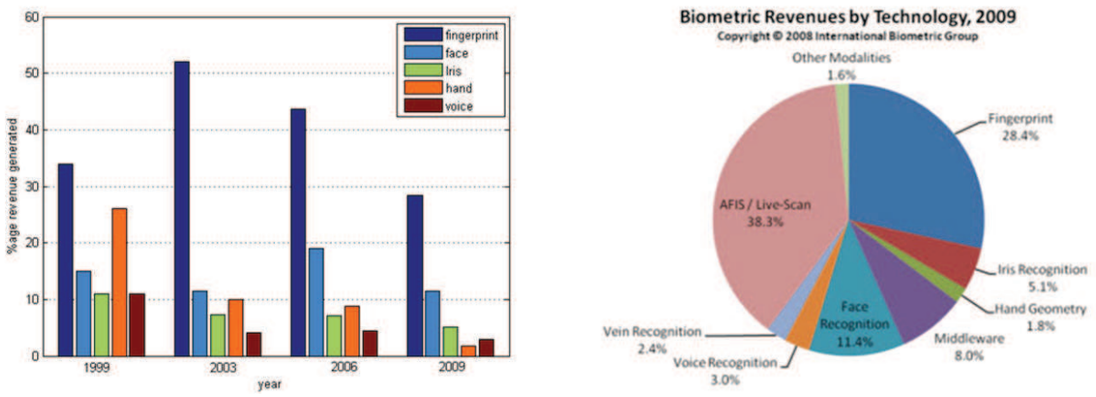


Fig. 1. Biometric analysis of past few years has been shown in figure showing the contribution of revenue generated by various biometrics. Although AFIS are getting popular in current biometric industry but faces are still considered as one of the widely used biometrics.

By nature, face recognition systems are widely used due to high universality, collectability and acceptability. However this attribute has natural challenges like uniqueness, performance and circumvention. Although several commercial biometrics systems use face recognition as a key tool but most of them perform in constrained environment. Real world challenges to face recognition problems involve varying lighting, poses, facial expressions, aging effects, occlusions which also include make up, facial scars or cuts, facial hairs, low resolution images and most recently spoofing. Due to their non intrusiveness faces are the most important biometrics to be employed in real life systems. Figure 1 shows the contribution of the face recognition system in biometrics market. Over the last decade, faces have been consistently used after the usage of fingerprints and AFIS (automated fingerprints identification system)/live scans. Face recognition finds its major applications in document classification, security access control, surveillance, web application and human robot interaction. However fingerprints can easily be spoofed (research is still in progress to deal with this issue using live scans), intrusive, aging effects are more in the sense of damages and finally template security and updation issues.

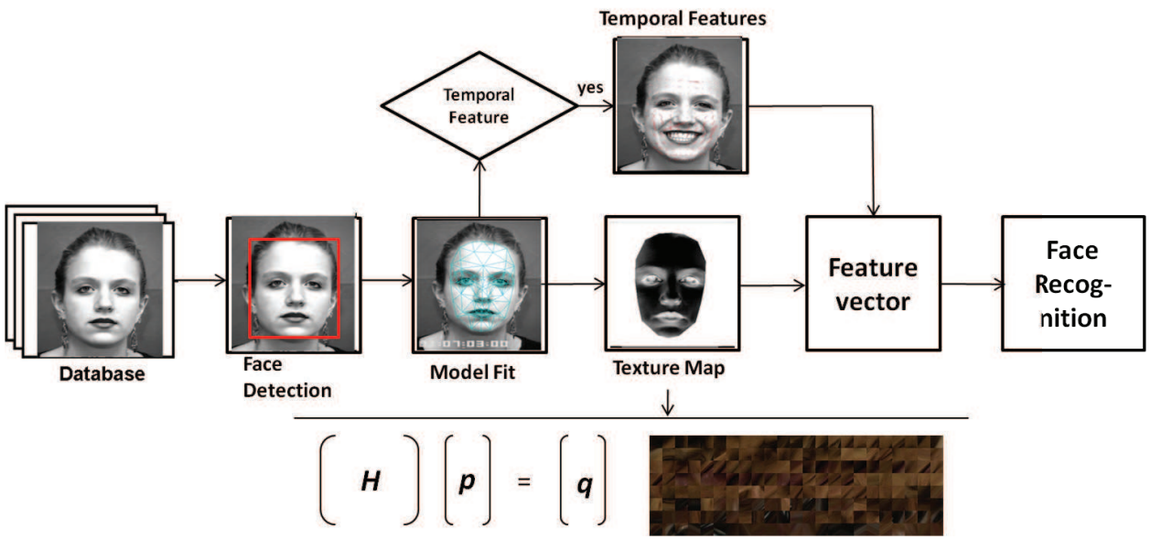


Fig. 2. Detailed process for model based face recognition system. Texture map is generated by using homogeneous 2D point p in texture coordinates and a projection q of a general 3D point in homogeneous coordinates by using transformation $H \times p = q$. Where H is the homography matrix Riaz et al. (2010).

In this chapter we focus on an unconstrained face recognition system and also study other soft biometric traits of human faces which can be useful for man machine interaction applications. The proposed system can be applied in major biometrics application equally well. The goal of this chapter is two folded. Firstly, it serves as self-contained and compact tutorial for face recognition systems from design to its development. It describes a brief history of face recognition systems, major approaches used and challenges. Secondly, it describes in detail an approach towards the development of a robust face recognition system against varying poses, facial expressions and lighting. This approach uses a 3D wireframe model which is useful for such applications. In order to proceed, a comprehensive overview of the face models currently used in the area of computer vision applications is provided. This covers an overview about deformable models, point distribution models, photorealistic models and finally wireframe

models. In order to provide an automatic generation of a realistic human face model, a 3D wireframe model called Candide-III Ahlberg (2001) is used. This model has benefit over the others models in the sense of its speed, textures realization and well-defined of facial animation units. On the other hand realizing this model is more challenging since it is less detailed as compared to 3D morphable model which consists of a dense point distribution of facial points acquired from laser scans. Candide-III consists of a coarse mesh containing only 184 triangulations and defines facial action coding system (FACS) Ekman & Friesen (1978), which can easily be used for facial animations. Figure 2 show the overview of different modules working in our designed system.

2. System overview

The system mainly comprise of different modules contributing towards the final feature extraction. It starts with a face detection module followed by a face model fitting. Candide-III is fitted to face images using robust objective functions Wimmer et al. (2006). This fitting algorithm is compared with other algorithms and has been adopted here due to their efficiency in real time and robustness. The model is fitted to any face image with arbitrary pose, however pre-requisite for this module is face detection. If face detection module fails to find any face inside the image then system searches again for a face in next image. After model fitting, extracted structural information from this image contains shape and pose information. The fitted model is used for extracting the texture from given image. We use graphic tools to render texture to model surface and warp the texture to standard template. This standard template is a block based texture map shown in Figure 2. Texture variations are majorly caused by illuminations and shadows, which are dealt with textural feature and image filtering. We use principal component analysis (PCA) to parameterize extracted textures. Cognitive science explains that temporal information is an essential constituent for human perception and learning Sinha et al. (2006). In this regard, we choose local descriptors on the face image and track their motion to find temporal features. This motion provides local variations and deformations in the face image caused by facial expressions. Finally we construct a feature vector set by concatenating these three different features for any given image.

$$\boldsymbol{u} = (\mathbf{b_s}, \mathbf{b_g}, \mathbf{b_t}) \tag{1}$$

Where b_s, b_g and b_t are the structural, textural and temporal features. For texture extraction we comparatively perform different methods including discrete cosine transform (DCT), principal components analysis (PCA) and local binary patterns (LBP). From these three types of textural feature, it is observed that PCA outperforms other two feature set and hence used for further experimentation. The results are given in detail in section 10.

3. Related work

Face recognition has always been a focus of attention for the researchers and has been addressed in different ways by using holistic and heuristic features, dense and sparse feature representation, parametric and non-parametric models and content- and context-aware methods Zhao et al. (2003)Zhao & Chellappa (2005). Due to the universality, collectability non-intrusiveness, face recognition is currently used as a baseline algorithm in several biometric systems either as a standalone technology or together with other biometrics, called multibiometric systems. In 3D space, faces are complex objects consisting of a regularized

structure and pigmentation and mostly observed performing various action and conveying a set meaningful information O'Toole (2009). Besides these meanigful set of information, faces convey several challenges which are under consideration by the research community. Human facial recognition system is unconstrained and provides stability against varying poses, facial expressions, changing illuminations, partial occlusions (including facial hair, scars and make ups) and temporal effects (aging factors).
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. In Zhao et al. (2003) the authors give a comprehensive survey of face recognition and some commercially available face recognition software. Subspace projection method like PCA was firstly used by Sirvovich and Kirby Sirovich & Kirby (1987) , which were latterly adopted by M. Turk and A. Pentland introducing the famous idea of eigenfaces Turk & Pentland (1991). This chapter focuses on the modeling of human face using a three dimensional model for shape model fitting, texture and temporal information extraction and then low dimensional parameters for recognition purposes. The model using shape and texture parameters is called Active Appearance Model (AAMs), introduced by Cootes et. al. Cootes et al. (1998)Edwards, Taylor & Cootes (1998). For face recognition using AAM, Edwards et al Edwards, Cootes & Taylor (1998) use weighted distance classifier called Mahalanobis distance. In Edwards et al. (1996) the authors used separate information for shape and gray level texture. They isolate the sources of variation by maximizing the interclass variations using discriminant analysis, similar to Linear Discriminant Analysis (LDA), the technique which was used for Fisherfaces representation Belheumeur et al. (1997). Fisherface approach is similar to the eigenface approach however outperforms in the presence of illuminations. In Wimmer et al. (2009) the authors have utilized shape and temporal features collectively to form a feature vector for facial expressions recognition. These models utilize the shape information based on a point distribution of various landmarks points marked on the face image. Blanz et al. Blanz & Vetter (2003) use state-of-the-art morphable model from laser scaner data for face recognition by synthesizing 3D face. This model is not as efficient as AAM but more realistic. In our approach a wireframe model known as Candide-III Ahlberg (2001) has been utilized. In order to perform face recognition applications many researchers have applied model based approach. Riaz et al Riaz, Mayer, Wimmer, Beetz & Radig (2009) apply similar features for explaining face recognition using 2D model. They use expression invariant technique for face recognition, which is also used in 3D scenarios by Bronstein et al Bronstein et al. (2004) without 3D reconstruction of the faces and using geodesic distance. Park et. al. Park & Jain (2007) 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.
In Riaz, Mayer, Beetz & Radig (2009a) author introduced spatio-temporal feature for expressions invariant face recognition. This chpater is an extended version of the similar approach but with improved texture realization and texture descriptors. Further the feature set used in this chapter are stable against facial poses.

4. Spatio-temporal Multifeatures (STMF)

We address the problem in which a 3D model can extract a common feature set automatically from face images and performs unconstrained face recognition. This system can not only be used for biometric applications but also for soft biometeric traits document classification,

	Identity	Expressions	Gender	Age	Ethnicity
Structural	*p*	*p*	*p*	*p*	*p*
Textural	*p*	*s*	*p*	*p*	*p*
Temporal	*s*	*p*	*s*	*na*	*na*

Table 1. Contribution of different components to our proposed feature set, p = primary, s = secondary, na = not applicable.

.

access control and medical care. In such applications an automatic and efficient feature extraction technique is necessary to interpret maximum available information from the faces image sequences. Further, the extracted feature set should be robust enough to be directly applied in real world applications. In such scenarios faces are seen from different views under varying facial deformations and poses. This issue is treated by using 3D modeling of the faces. The invariance to facial poses and expressions is discussed in detail in section 10.
For face recognition textural information plays a key role as features Zhao & Chellappa (2005)Li & Jain (2005) whereas facial expressions are mostly person independent and require motion and structural components Fasel & Luettin (2003). Similary facial structure and texture vary significantly between gender classes. On the basis of this knowledge and literature survey we categorize three major features as primary and secondary contributor to three different facial classifications. Table 1 summarizes the significance of these constituents of the feature vector with their primary and secondary contribution towards the feature set formation. Since our feature set consists of all three kinds of information hence it can successfully represent facial indentity, expressions and gender. The results are discussed in detail in section 10. Model parameters are obtained in an optimal way to maximize information within the face region in the presence of different facial pose and expressions. We use a 3D wireframe model however, any other comparable model can be used here.

5. Model fitting and structural features

Our proposed algorithm is initialized by applying a face locator in the given image. We use Viola and Jones face detector Viola & Jones (2004). If a face is found then the system proceeds towards face model fitting. For model fitting, local objective functions are calculated using haar-like features. An objective function is a cost function which is given by the equation 2. A fitting algorithm searches for the optimal parameters which minimizes the value of the objective function. For a given image I, if $E(I, c_i(\mathbf{p}))$ represents the magnitude of the edge at point $c_i(\mathbf{p})$, where $\mathbf{p}$ represents set of parameters describing the model, then objective function is given by:

$$f(I, \mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n} f_i(I, c_i(\mathbf{p})) = \frac{1}{n}\sum_{i=1}^{n}(1 - E(I, c_i(\mathbf{p}))) \tag{2}$$

Where $n = 1, \ldots, 113$ is the number of vertices c_i describing the face model. This approach is less prone to errors because of better quality of annotated images which are provided to the system for training. Further, this approach is less laborious because the objective function design is replaced with automated learning. For details we refer to Wimmer et al. (2008). The geometry of the model is controlled by a set of action units and animation units. Any shape s can be written as a sum of mean shape $\bar{s}$ and a set of action units and shape units.

$$s(\alpha, \sigma) = \bar{s} + \phi_a \alpha + \phi_s \sigma \tag{3}$$

Where ϕ_a is the matrix of action unit vectors and ϕ_s is the matrix of shape vectors. Whereas α denotes action units parameters and σ denotes shape parameters Li & Jain (2005). Model deformation governs under facial action coding systems (FACS) principles Ekman & Friesen (1978). The scaling, rotation and translation of the model is described by

$$s(\alpha, \sigma, \pi) = mRs(\alpha, \sigma) + t \tag{4}$$

Where R and t are rotation and translation matrices respectively, m is the scaling factor and π contains six pose parameters plus a scaling factor. By changing the model parameters, it is possible to generate some global rotations and translations. We extract 85 parameters to control the structural deformation.

6. Textural representation and parameterization

For texture extraction from the face images after model fitting, two different approaches are studied in this chapter. In section 6.1 texture extraction is performed using conventional AAM method, whereas in section 6.2 texture mapping approach is studied. Texture map is formed by storing each triangular patch to a block in memory. This block represents surface texture extracted from 3D surface of the face. Once texture is extracted, it is parametrized by using mean texture $g_m$and matrix of eigenvectors $P_g$to obtain the parameter vector b_g Li & Jain (2005).

$$g = g_m + P_g b_g \tag{5}$$

6.1 Image warping

Once structural information of the image is obtained from model fitting, 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. We use delauny triangulations for the convex hull of the facial landmarks. Texture warping between the triangles is performed using affine transformation. This texture warping is used for CKFE database Kanade et al. (2000) and MMI database Maat et al. (2009). By warping texture to a reference shape, facial expressions are neutralized and hence useful for face recognition.

6.2 Optimal texture representation

Each triangular patch represents meaningful texture which is stored in a square block of the texture map. A single unit of the texture map represents a trianglular patch. We experiment with three different sizes of the texture blocks and choose an optimal size for our experimentation. These three block sizes include $2^3 \times 2^3$, $2^4 \times 2^4$ and $2^5 \times 2^5$. We calculate energy function from these texture maps of individual persons and observe the energy spectrum of the images in our database for each triangular patch. If N is the total number of images, and p_i be a texel value (which is equal to a single pixel value) in texture map, then we define energy function as:

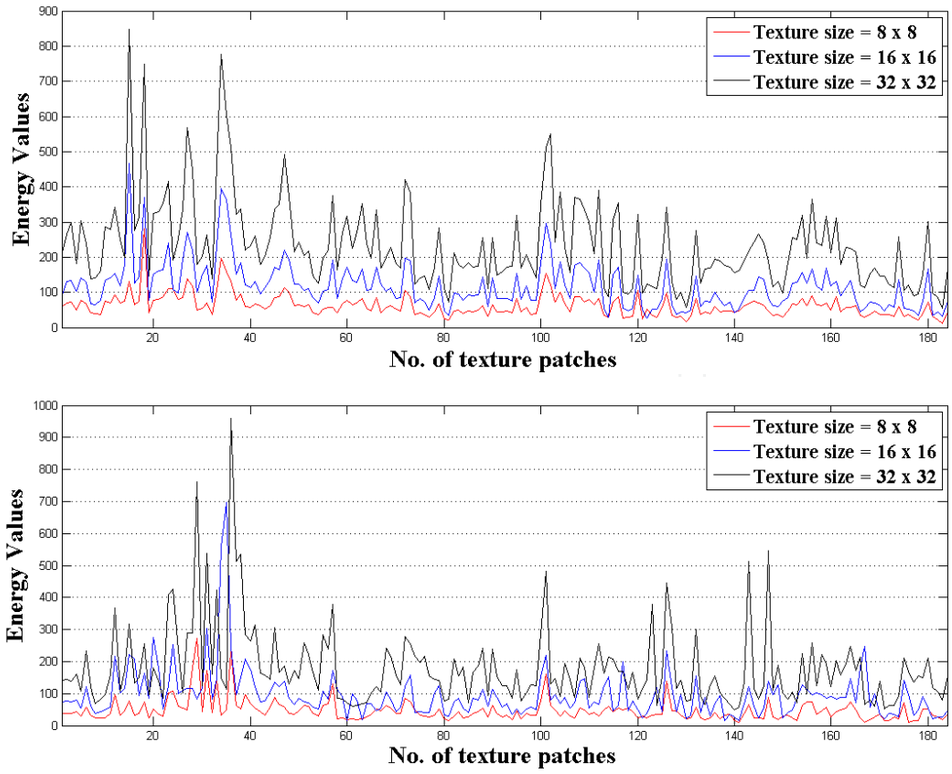


Fig. 3. Energy spectrum of two randomly selected subjects from PIE database. Energy values for each patch is comparatively calculated and observed for three different texture sizes.

$$E_j = \frac{1}{N}\sum_{i=1}^{N}(p_i - \overline{p}_j)^2 \tag{6}$$

Where $\overline{p}_j$ is the mean value of the pixels in j^{th} block, $j = 1 \ldots M$ and $M = 184$ is the number of blocks in a texture map. In addition to Equation 6, we find variance energy by using PCA for each block and observe the energy spectrum. The variation within the given block has similar behavior for two kinds of energy functions except a slight variation in the energy values. Figure 3 shows the energy values for two different subjects randomly chosen from our experiments. It can be seen from Figure 3 that behavior of the textural components is similar between different texture sizes. The size of the raw feature vector extracted directly from texture map increases exponentially with the increase of texture block size. If $d \times d$ is the size of the block, then the length of the raw feature vector is $\frac{d(d+1)}{2}$. This vector length calculation depends upon how texture is stored in the texture map. This can be seen in Figure 4. We store each triangular patch from the face surface to upper triangle of the texture block. The size of raw feature vector extracted for $d = 2^3$, $d = 2^4$ and $d = 2^5$ is 6624, 25024 and 97152 respectively. Any higher value will exponentially increase the raw vector without any improvement in the texture energy. We do not consider higher values due to increase in vector length. The overall recognition rate produced by different texture sizes from eight randomly selected subjects with 2145 images from PIE database is shown in Figure 5. The results are obtained using decision trees and Bayesian networks for classification. The classification

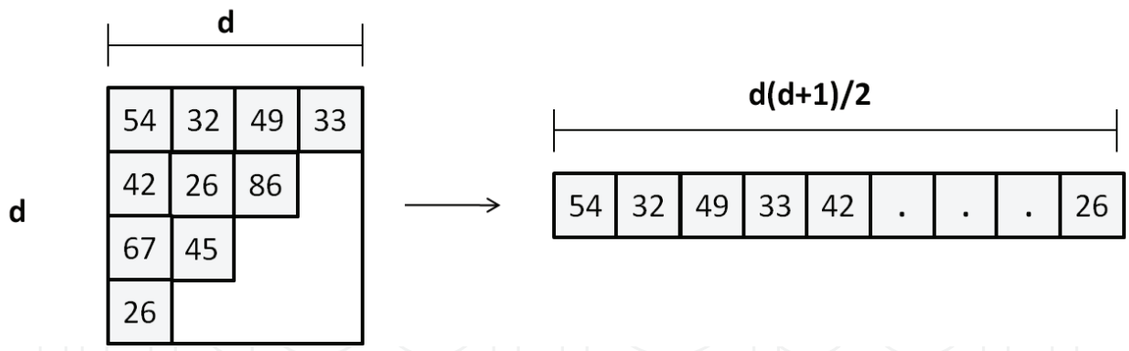


Fig. 4. Texture from each triangular patch is stored as upper triangle of the texture block in texture map. A raw feature vector is obtained by concatenating the pixel values from each block.

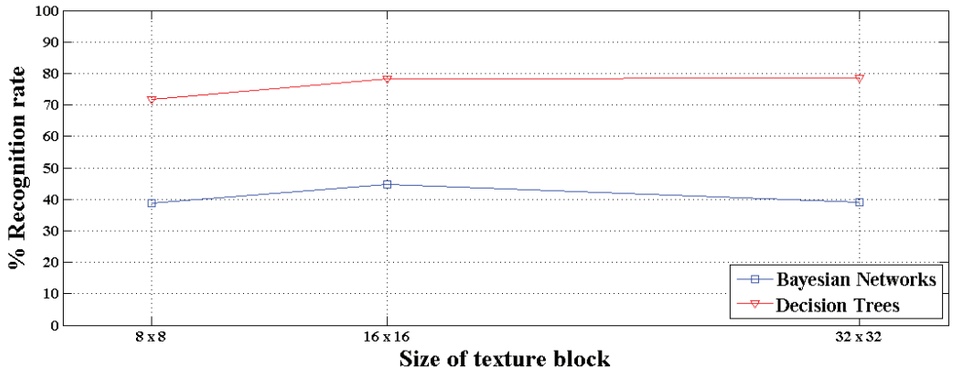


Fig. 5. Comparison over eight random subjects from the database with three different sizes of texture blocks. Recognition rate slightly improved as texture size is increased however causes a high increase on the length of raw feature vector. We compromise on texture block of size 16×16.

procedure is given in detail in next section 10. By trading off between the performance and size of the feature vectors, we choose texture block size to 16×16 during our experiments.

6.3 Affine vs. perspective transformation

We consider perspective transformation because affine warping of the rendered triangle is not invariant to 3D rigid transformations Riaz et al. (2010). In general, texture warping is performed using affine transformation from a given image to a reference shape Cootes et al. (1998)Riaz, Mayer, Wimmer, Beetz & Radig (2009)Riaz, Mayer, Beetz & Radig (2009b)Wimmer et al. (2009). This preserves affinity after the transformation. However, for faces with different views triangular patches on the edges are not well defined and these triangles are tilted such that they contain very less information about the texture as compared to those triangles which are frontal. In order to equally weight all triangles and we use a homogeneous transformation.

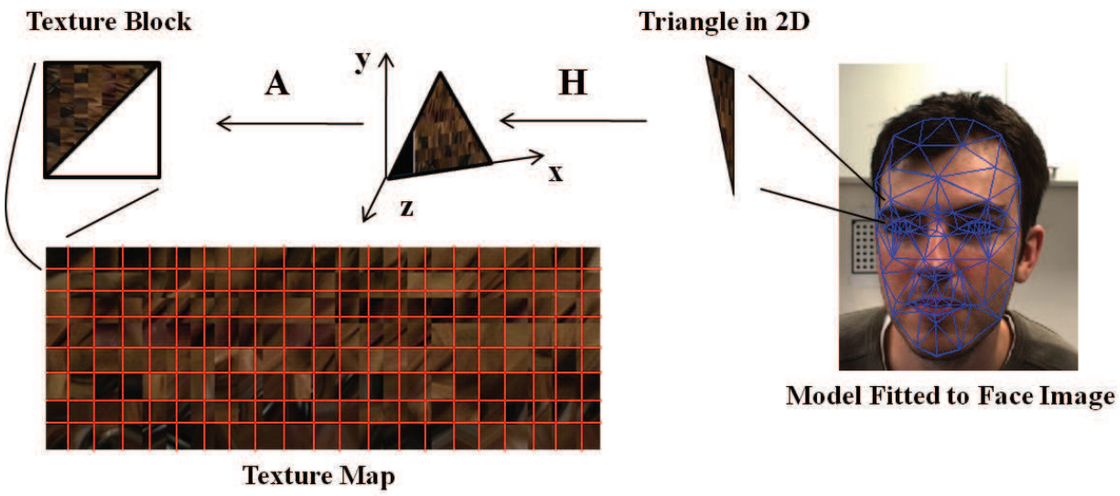


Fig. 6. Detailed texture extraction approach. Each triangular patch from the face surface is stored as a block in a texture map. This texture map is further used for feature extraction.

The homogeneous transformation **M** is the given by

$$M = AK\begin{bmatrix} R & -Rt \end{bmatrix} = AH \tag{7}$$

Where **K** is the camera matrix. An undistorted texture map from the face region is calculated in two steps. Firstly, we find the homography **H**, by obtaining the rotation and translation of the triangle, by supposing that the initial triangle lies on the texture plane, the first vertex lies on the origin and the first edge lies on the x-axis. Secondly, affine transformation **A** is calculated, so that the mapped triangle on the texture plane fits the upper triangle of the rectangular texture block. The lower triangular area is not considered in this regard.
Where **R** and **t** are the unknown to be calculated. In order to fit any arbitrary triangle to this upper triangle, we use an affine transformation **A**. This process is shown in Figure 6. For detail, refer to Riaz et al. (2010).

7. Temporal features

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. Since the motion of the feature points are relative so we choose 140 points in total to observe the optical flow. We again use PCA over the motion vectors to reduce the descriptors.
If **t** is the velocity vector,

$$t = t_m + P_t b_t \tag{8}$$

Where temporal parameters $\mathbf{b_t}$ are computed using matrix of eigenvectors $\mathbf{P_t}$ and mean velocity vectors $\mathbf{t_m}$.

Feature Type	LBP Features	DCT Features	PCA Features
Recognition Rate	78.56%	80.60%	83.06%

Table 2. Comparison of three different feature types for face recognition. The overall recognition rate shows the number of images correctly classified. PCA outperforms the other two feature types and is used for further experimentations.

8. Feature fusion

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:

$$u = (b_{s,1}, \ldots, b_{s,m}, b_{g,1}, \ldots, b_{g,n}, b_{t,1}, \ldots, b_{t,p}) \tag{9}$$

Where b_s, b_g and b_t are shape, textural and temporal parameters respectively with m, n and p being the number of parameters retained from subspace in each case. Equation 9 is called multi-feature. We extract 85 structural features, 74 textural features and 12 temporal features textural parameters to form a combined feature vector for each image. These features are then used for decision tree (DT) and bayesian network (BN) for different classifications. The face feature vector consists of the shape, texture and temporal variations, which sufficiently defines global and local variations of the face. All the subjects in the database are labeled for classification. Since features arise from different sources, it is not quite obvious to fuse them together to get a feature set. This can cause the dominance of the features with higher values and ones with low values are ignored. We use simple scaling of the features in $[0, 1]$. However, any suitable method for feature fusion can be applied here.

9. Comparative texture descriptors

From above texture representations, features are extracted using three different approaches, a) PCA, b) discrete cosine transform (DCT) and c) local binary pattern (LBP). Each texture map consists of 184 texture blocks where each texture block corresponds to texture in a triangular surface. The size of each block is 16×16 pixels. This size is chosen by trading off between accuracy and efficiency. DCT coefficients are extracted in a zig-zag pattern from top-left corner of each block. We extract five coefficients per block and obtain a feature set of length $5 \times 184 = 920$. The advantage of using DCT over conventional approach is two fold, 1) it reduces the dimensions to a great extent, 2) DCT coefficients contain low frequency information which are robust to distortions and noise. For LBP descriptor, we consider those pixels for coding which lie inside face area. An LBP histogram of 255 gray levels and color histogram are calculated and used as texture features. The results of three different features types on all subjects of PIE database session from November 2000 to December 2000 Terence et al. (2002) is shown in Table 2. A J48 decision tree from Weka Witten & Frank (2005) is used as classifier. The detail about classifier specification is given in section 10. It can be seen that PCA outperforms the other two features types. For further experimentation, we use PCA for feature extraction.

9.1 Local energy based descriptors

From section 6.2, we have texture extracted from the face images and stored in a texture map in triangular patches of same sizes. Each patch represents a specific area of the face. Instead

of using conventional subspace learning which do not preserve localization information, we find heuristic features representative of each patch. If training data consists of M images, then we can calculate the variance energy of each patch using:

$$e_j = \frac{1}{M}\sum_{i=1}^{N}(p_i - \overline{p_{ji}})^2 \tag{10}$$

Where $i = 1, \ldots, N$ and $N = 136$, which is the number of pixel in triangle from texture map, $\overline{p_j}$ is the average value of jth patch, with $j = 1, 2, \ldots, 184$. Finally a feature vector E is formed by finding energy descriptor for each triangular patch and is written as:

$$E = \{e_1, e_2, \ldots, e_j\} \tag{11}$$

There are three major benefits of using patch based representation.

- The extracted feature vector although is compact and sufficient to perform well in view invariant face recognition.
- It avoids subspace learning and new faces can be added easily in the database.
- Since such representation presevers localization of the facial features, it outperforms the conventional AAM and holistic approaches.

10. Experimental evaluation

In order to validate the extracted feature, we have used different subjects from three different databases called, CMU-PIE database Terence et al. (2002), MMI database Maat et al. (2009) and Cohn-Kanade facial expressions database (CKFED) Kanade et al. (2000). These databases consist of face images with different variations, like varying poses, facial expressions, gender information and talking faces. MMI and CKFED contain image sequences with temporal information. CKFED consists of 97 subjects range in age from 18 to 30 years. Sixty-five percent are female, 15 percent are African-American and three percent Asian or Latino. The MMI facial expression database holds over 2000 videos and over 500 images of about 50 subjects displaying various facial expressions on command. In case of CKFE and MMI databases we compute spatio-temporal feature from the image sequences. CMU-PIE database is collected between October and December 2000 consisting of 41,368 images of 68 people. Each person is captured with 13 different poses, 43 different illumination conditions, and with 4 different expressions. We take spatial feature and test them against frontal and half-profile poses. The texture extracted in this case is stored as a texture map after removing perspective distortion. During all experiments, we use two-third of the feature set for building the classifier model with 10-fold cross validation to avoid overfitting. The remaining feature set is used for testing purpose. We use all subjects from MMI and CKFED and partially use PIE database to perform face recognition and person dependent facial expressions and gender classification.
Since STMF set arises from different sources, so decision tree (DT) is applied for classification. However, other classifiers can also be applied here depending upon the application. We choose J48 decision tree algorithm for experimentation which uses tree pruning called subtree raising and recursively classifies until the last leave is pure. We use same configuration for all classifiers trained during the experiments. The parameters used in decision tree are: confidence factor C = 0.25, with two minimum number of instances per leaf and C4.5 approach for reduced error-pruning Witten & Frank (2005). For further validation, we use random

PIE - Session (Oct-Nov 2000)			
Vector Size	181	257	**184**
Random Forest	64.45%	70.12%	**89.12%**
Decision Tree	61.29%	69.65%	**75.13%**
PIE - Session Nov-Dec 2000			
Vector Size	337	295	**184**
Random Forest	90.15%	90.60%	**98.50%**
Decision Tree	88.42%	89.53%	**94.89%**
Feature Type	**Shape free AAM**	**3D Surface Texture**	**Energy feature**

Table 3. Comparison of three different approaches used for face recognition. The overall recognition rate shows the number of images correctly classified. PCA representation of the surface texture from a face outperforms conventional AAM approach, however energy based descriptors are not only compact but perform even better than two other approaches.

	BN	**DT**
CKFED	90.66%	98.50%
MMI	90.32%	99.29%

Table 4. Face recognition across facial expressions on two different databases using Bayesian networks (BN) and decision tree (DT).

forests for classification with default Weka Witten & Frank (2005) parameters and 10-fold cross validation. The results coincide with those from decision trees. During all experimentation, we use same training and testing approach. For subspace learning, one-third of the database is used while the remaining part is projected to this space. We retain 97% of the eigenvalues during the subspace learning.

10.1 Expression invariant face recognition

MMI and CKFE databases contain six basic facial expressions in the form of image sequences. Although neutral expression is present as a seventh expression but we exclude it during experiments and solve the problem as six class problem. All images are frontal and hence we use spatio-temporal features with texture warped to reference shape. Most of the face recognition information is available in textural components and hence we obtain stability of our feature in face recognition results in the presence of facial expressions. Texture warping neutralizes the effect of facial expressions. The recognition results using decision tree and Bayesian networks are shown in Table 4. However, structural and temporal part of the same feature set contain sufficient facial expression information (refer to Table 1). In this way, a single STMF is representative of facial expression, face recognition and gender information.

10.2 Pose invariant face recognition

In section 6.2, we explained the detailed process for texture extraction. Since the model is defined over a coarse mesh of vertices, so it is useful to consider texture map as an image with undistorted texture patches. In the presence of different facial poses, triangles at the face edges are tilted such that texture information is extremely distorted. In order to solve this problem, we project each triangle on a block of 16×16 pixel size. The block size is chosen by trading off between efficiency and accuracy. In this procedure, each triangle is weighted

Approach	% Accuracy
model based Mayer et al. (2009)	87.1%
TAN Cohen et al. (2003)	83.3%
LBP + SVM Shan et al. (2009)	92.6%
IEBM Asthana et al. (2009)	92.9%
Fixed Jacobian Asthana et al. (2009)	89.6%
STMF + DT	**93.2%**

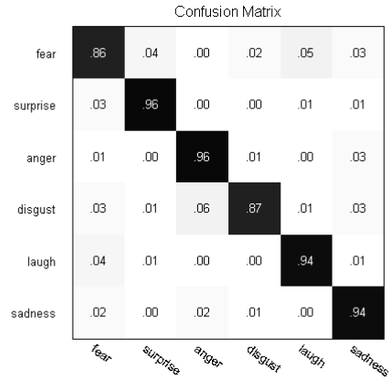


Table 5. Facial expressions recognition in comparison to different approaches Asthana et al. (2009) and Confusion matrix from our results

Approach	Classification rate
Pixels + SVM + Fusion	88.5%
LBP + SVM + Fusion	92.1%
VLBP + SVM	84.5%
EVLBP + AdaBoost	84.6%
our approach+ BDT	**94.8%**

Table 6. Comparison of gender classification in comparison to different approaches in Hadid & Pietikaeinen (2009)

equally toward the feature calculation and face edges are not destroyed but rather provide the detailed texture information that might be lost during conventional image warping approach. The recognition results with different approaches are given in Table 3.

10.3 Facial expressions and gender classification

Facial expressions recognition is performed on CKFED with six universal facial expressions: anger, disgust, fear, laugh, sadness and suprise. Each video sequence starts from a neutral face and reaches up to the peak of the particular expression. We exclude neutral expression during the experiments because it is included in all image sequences and cause more confusion. However, a neutral expression can also be considered as a seventh expression during classification. It can be automatically segmented using velocity vectors magnitudes. STMF features with their three structural, textural and temporal constituents are used for experiments. Finally, the results are compared in Table 5 with the state-of-the-art approaches which uses comparable system along with the confusion matrix from our experiments.
We further estimated age using FGNet database *FG-NET AGING DATABASE* (n.d.). This database contains 1002 images of 62 subjects with images of different ages ranging from 0 to 69 years. We divide the whole dataset in seven classes. Since the database consists of static images hence we experiment only with shape and textural component of the feature set. A classification rate of 49.70% is achieved with texture whereas the classification rate improved to 57.29% using support vector machine based classification. The mean absolute error (MAE) was 0.769.

11. Conclusions and future work

This chapter explained the STMFfor unconstrained face recognition. The spatial part of this feature set consists of structural (section 5) and textural (section 6) information. Two different types of texture extraction approaches are discussed in detail in section 6.1 and section 6.2. Further a comparative study of three different textural features has been studied which shows that PCA outperforms LBP and DCT. Since PCA is global representation of a face, hence it does not contain local information. In section 9.1 a local representation for each triangular patch is calculated. Since this local representation is added the extracted features, hence it further improves the results and outperforms holitic PCA. Since feature set given in equation 9 is consistent with Table 1 hence it can be used for facial expression recognition, gender classification and age estimation. The results are shown in section 10.1. This chapter provides a comprehensive overview and a compact description of 3D face modeling, face recognition, classifying soft-biometric traits including facial expressions, gender and age. However such systems require more memory and relatively slower as compared to conventional image based approaches. Future goal of this work is to enhance its efficiency to apply it in real time for interactive systems. Further more diverse conditions with large variabilities are required to be tested.

12. References

Ahlberg, J. (2001). An experiment on 3d face model adaptation using the active appearance algorithm, *Image Coding Group, Deptt of Electric Engineering, Linköping University* .

Asthana, A., Saragih, J., Wagner, M. & Goecke, R. (2009). Evaluating aam fitting methods for facial expression recognition, *Affetive Computing and Intelligent Interaction* 1(1): 598–605.

Belheumeur, P., Hespanha, J. & Kreigman, D. (1997). Eigenfaces vs fisherfaces: Recognition using class specific linear projection, *IEEE Transaction on Pattern Analysis and Machine Intelligence* 19(7).

Blanz, V. & Vetter, T. (2003). Face recognition based on fitting a 3d morphable model, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(9): 1063–1074.

Bronstein, A., Bronstein, M., Kimmel, R. & Spira, A. (2004). 3d face recognition without facial surface reconstruction, *Proceedings of European Conference of Computer Vision*.

Cohen, I., Sebe, N., Garg, A., Lawrence, C. & Huanga, T. (2003). Facial expression recognition from video sequences: temporal and static modeling, *Computer Vision and Image Understanding*, Elsevier Inc., pp. 160–187.

Cootes, T., Edwards, G. & Taylor, C. (1998). Active appearance models, *Proceedings of European Conference on Computer Vision* 2: 484–498.

Edwards, G., Cootes, T. & Taylor, C. (1998). Face recognition using active appearance models, *European Conference on Computer Vision*, Springer, pp. 581–695.

Edwards, G. J., Taylor, C. J. & Cootes, T. F. (1998). Interpreting face images using active appearance models, *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, FG '98, IEEE Computer Society, Washington, DC, USA, pp. 300–.
URL: *http://portal.acm.org/citation.cfm?id=520809.796067*

Edwards, G., Lanitis, C., Taylor, C. & Cootes, T. (1996). Statistical models of face images: Improving specificity., *British Machine Vision Conference*.

Ekman, P. & Friesen, W. (1978). The facial action coding system: A technique for the measurement of facial movement, *Consulting Psychologists Press* .

Fasel, B. & Luettin, J. (2003). Automatic facial expression analysis: A survey, *PATTERN RECOGNITION* 36(1): 259–275.

FG-NET AGING DATABASE (n.d.). http://www.fgnet.rsunit.com/.

Hadid, A. & Pietikaeinen, M. (2009). Manifold learning for gender classification from face sequences, *Advances in Biometrics*, Springer Verlag., pp. 82–91.

Kanade, T., Cohn, J. & Tian, Y. (2000). Comprehensive database for facial expression analysis, *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition* pp. 46–53.

Li, S. Z. & Jain, A. K. (2005). *Handbook of Face Recognition*, Springer.

Maat, L., Sondak, R., Valstar, M., Pantic, M. & Gaia, P. (2009). Man machine interaction (mmi) database.

Mayer, C., Wimmer, M., Eggers, M. & Radig, B. (2009). Facial expressions recognition with 3d deformable models, *2009 Second International Conferences on Advances in Computer-Human Interactions*, IEEE Computer Society, pp. 26–31.

O'Toole, A. J. (2009). Cognitive and computational approaches to face recognition, *The University of Texas at Dallas* .

Park, U. & Jain, A. K. (2007). 3d model-based face recognition in video, *2nd International Conference on Biometrics*.

Riaz, Z., Gedikli, S., Beetz, M. & Radig, B. (2010). 3d face modeling for multi-feature extraction for intelligent systems, *Computer Vision for Multimedia Applications: Methods and Solutions* 1: 73–89.

Riaz, Z., Mayer, C., Beetz, M. & Radig, B. (2009a). Face recognition using wireframe model across facial expressions, *Proceedings of the 2009 joint COST 2101 and 2102 international conference on Biometric ID management and multimodal communication*, BioID MultiComm'09, pp. 122–129.

Riaz, Z., Mayer, C., Beetz, M. & Radig, B. (2009b). Model based analysis of face images for facial feature extraction, *Computer Analysis of Images and Pattern* pp. 99–106.

Riaz, Z., Mayer, C., Wimmer, M., Beetz, M. & Radig, B. (2009). A model based approach for expression invariant face recognition, *International Conference on Biometrics*, Springer.

Shan, C., Gong, S. & McOwan, P. (2009). Facial expression recognition based on local binary patterns: A comprehensive study, *Computer Vision and Image Understanding*, Elsevier Inc., pp. 803–816.

Sinha, P., Balas, B., Ostrovsky, Y. & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about, *Proceedings of IEEE* 94(II): 1948–1962.

Sirovich, L. & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces, *J. Opt. Soc.* 4(3): 519–524.

Terence, S., Baker, S. & Bsat, M. (2002). The cmu pose, illumination, and expression (pie) database, *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Computer Society, Washington, DC, USA, p. 53.

Turk, M. & Pentland, A. (1991). Face recognition using eigenfaces, *IEEE Computer Society Conference on In Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91* pp. 586–591.

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

Wimmer, M., Riaz, Z., Mayer, C. & Radig, B. (2009). Recognizing facial expressions using model-based image interpretation, *Advances in Human-Computer Interaction*, I-Tech.

Wimmer, M., Stulp, F., Pietzsch, S. & Radig, B. (2008). Learning local objective functions for robust face model fitting, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 30(8): 1357–1370.

Wimmer, M., Stulp, F., Tschechne, S. & Radig, B. (2006). Learning robust objective functions for model fitting in image understanding applications, *Proceedings of the 17th British Machine Vision Conference* pp. 1159–1168.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, San Francisco.

Zhao, W. & Chellappa, R. (2005). *Face Processing: Advanced Modeling and Methods*, Elsevier.

Zhao, W., Chellappa, R., Phillips, P. J. & Rosenfeld, A. (2003). Face recognition: A literature survey.

IntechOpen



Biometric Systems, Design and Applications

Edited by Mr Zahid Riaz

ISBN 978-953-307-542-6
Hard cover, 262 pages
Publisher InTech
Published online 21, October, 2011
Published in print edition October, 2011

Biometric authentication has been widely used for access control and security systems over the past few years. The purpose of this book is to provide the readers with life cycle of different biometric authentication systems from their design and development to qualification and final application. The major systems discussed in this book include fingerprint identification, face recognition, iris segmentation and classification, signature verification and other miscellaneous systems which describe management policies of biometrics, reliability measures, pressure based typing and signature verification, bio-chemical systems and behavioral characteristics. In summary, this book provides the students and the researchers with different approaches to develop biometric authentication systems and at the same time includes state-of-the-art approaches in their design and development. The approaches have been thoroughly tested on standard databases and in real world applications.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Zahid Riaz, M. Saquib Sarfraz and Michael Beetz (2011). Towards Unconstrained Face Recognition Using 3D Face Model, Biometric Systems, Design and Applications, Mr Zahid Riaz (Ed.), ISBN: 978-953-307-542-6, InTech, Available from: http://www.intechopen.com/books/biometric-systems-design-and-applications/towards-unconstrained-face-recognition-using-3d-face-model

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen