

# Model Based Analysis of Face Images for Facial Feature Extraction

Zahid Riaz, Christoph Mayer, Michael Beetz, and Bernd Radig

Technische Universität München,  
Boltzmannstr. 3, 85748 Garching, Germany  
{riaz,mayerc,beetz,radig}@in.tum.de  
<http://www9.in.tum.de>

**Abstract.** This paper describes a comprehensive approach to extract a common feature set from the image sequences. We use simple features which are easily extracted from a 3D wireframe model and efficiently used for different applications on a benchmark database. Features versatility is experimented on facial expressions recognition, face recognition and gender classification. We experiment different combinations of the features and find reasonable results with a combined features approach which contain structural, textural and temporal variations. The idea follows in fitting a model to human face images and extracting shape and texture information. We parametrize these extracted information from the image sequences using active appearance model (AAM) approach. We further compute temporal parameters using optical flow to consider local feature variations. Finally we combine these parameters to form a feature vector for all the images in our database. These features are then experimented with binary decision tree (BDT) and Bayesian Network (BN) for classification. We evaluated our results on image sequences of Cohn Kanade Facial Expression Database (CKFED). The proposed system produced very promising recognition rates for our applications with same set of features and classifiers. The system is also realtime capable and automatic.

**Key words:** Feature Extraction, Face Image Analysis, Face Recognition, Facial Expressions Recognition, Human Robot Interaction

## 1 Introduction

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 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 under different applications. In this paper we deal with some of these challenges. We focus on feature extraction technique which is fully automatic and versatile

enough for different applications like face recognition, facial expressions recognition and gender classification. These capabilities of the system suggest to apply it in interactive scenarios like human machine interaction, security of personalized utilities like tokenless devices, facial analysis for person behavior and person security.

Model-based image interpretation techniques extract information about facial expression, person identity and gender classification from images of human faces via facial changes. Models take benefit of the prior knowledge of the object shape and hence try to match themselves 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-3 face model [8] that has been specifically designed for observing facial features variations defined by facial action coding system (FACS) [13]. The model parameters together with extracted texture and motion information is utilized to train classifiers that determine person-specific information. A combination of different facial features is used for classifiers to classify six basic facial expressions i.e. anger, fear, surprise, sadness, laugh and disgust, facial identity and gender classification.

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 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 main sections. In section 2, related work to our applications is discussed. In section 3 we discuss our approach in detail. In section 4 higher level features extraction from model based image interpretation is described. This includes description from model fitting to face image to feature vector formation. Section 5 discusses about evaluation of our results on the database. Finally we conclude our results with some future directions.

## 2 Related work

We initiate with a three step approach that has been suggested by Pantic et al. [1] for facial expression recognition. However, the generality of this approach makes it applicable not only to facial expression estimation but also to apply it for person identification and gender classification at the same time. The first step aims at determining the position and shape of the face in the image by fitting a model. Descriptive features are extracted in the second step. In the third step a classifier is applied to the features to determine high level information from the features. Several face models and fitting approaches have been presented in the recent years. Cootes et al. [5] introduced modeling face shapes with Active Contours. Further enhancements included the idea of expanding shape models

with texture information [6]. In contrast, three-dimensional shape models such as the Candide-3 face model considers 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 [7]. 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 [18]. In this work, image features are obtained by approximating the pixel values in a region around a pixel of interest. The learning algorithm used to map image features to objective values is a k-Nearest-Neighbor classifier (kNN) learned from the data. We used similar methodology developed by Wimmer et al. [4] which combines multitude of qualitatively different features [19], determines the most relevant features using machine learning and learns objective functions from annotated images [18]. To extract descriptive features from the image, Michel et al. [14] 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. [15] 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 [13]. This approach is applied by Kotsia et al. [16] to design Support Vector Machines (SVM) for classification. Michel et al. [14] 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 [2] 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 [3]. However they do not discuss face recognition under facial expression. Riaz et al [17] 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 [9] without 3D reconstruction of the faces and using geodesic distance. Park et. al. [10] 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 Our approach

In this section we explain in detail the approach adopted in this paper including model fitting, image warping and parameters extraction for shape, texture and temporal information.

We use a wireframe 3D face model known as *candide-III* [8]. The model is fitted to the face image using objective function approach [4]. After fitting the model to the example face image, we use the projections of the 3D landmarks in 2D for 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. 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.

### 4 Determining High-Level Information

In order to initialize, We apply the algorithm of Viola et al. [20] to roughly detect the face position within the image. Then, model parameters are estimated by applying the approach of Wimmer et al. [4] 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 indicates shape changes and therefore contributes to the recognition of facial expressions.

The shape  $\mathbf{x}$  is parametrized by using mean shape  $\mathbf{x}_m$  and matrix of eigenvectors  $\mathbf{P}_s$  to obtain the parameter vector  $\mathbf{b}_s$  [11].

$$\mathbf{x} = \mathbf{x}_m + \mathbf{P}_s \mathbf{b}_s \quad (1)$$

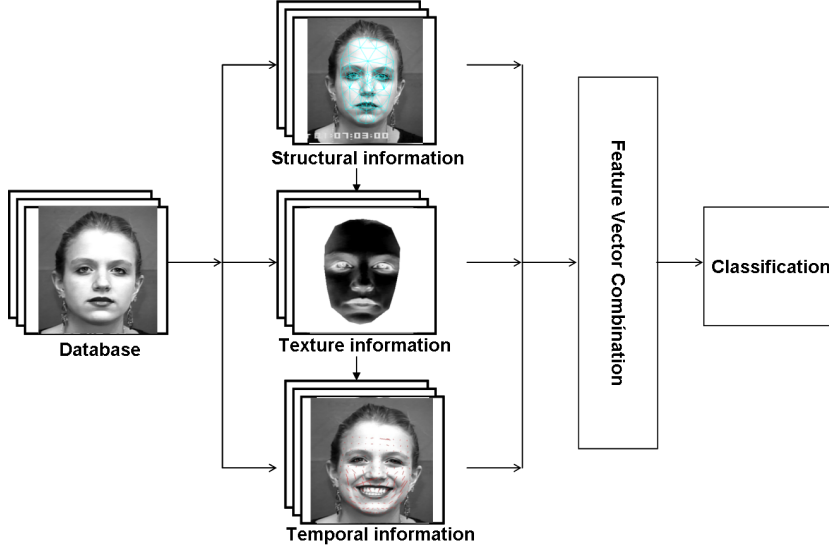


Fig. 1. Our Approach: Sequential flow for feature extraction

The extracted texture is parametrized using PCA by using mean texture  $g_m$  and matrix of eigenvectors  $P_g$  to obtain the parameter vector  $b_g$  [11]. Figure 2 shows shape model fitting and texture extracted from face image.

$$g = g_m + P_g b_g \quad (2)$$

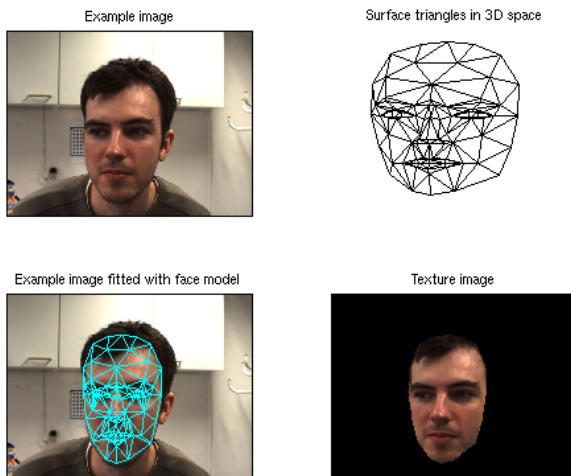
Further, temporal features of the facial changes are also calculated that take movement over time into consideration. Local motion of feature points is observed using optical flow. We do not specify the location of these feature points manually but distribute equally in the whole face region. The number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

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:

$$u = (b_{s1}, \dots, b_{sm}, b_{g1}, \dots, b_{gn}, b_{t1}, \dots, b_{tp},) \quad (3)$$

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

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 binary decision tree (BDT) and bayesian network (BN) for different classifications. The face feature vector consists of the



**Fig. 2.** Texture information is represented by an appearance model. Model parameters of the fitted model are extracted to represent single image information.



**Fig. 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.

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.

## 5 Experimental Evaluations

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 [12]. 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) [13]

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

**Table 1.** Comparison of Extracted Features

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.

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 [21] with 10-fold cross validation. Table 1 shows different recognition rates achieved during experimentations. In all three cases BDT outperforms BN.

## 6 Conclusions

We introduced a technique to develop a set of feature vectors which consist of three types of facial information. The features set is applied to three different applications: face recognition, facial expressions recognition and gender classification, which produced the reasonable results in all three cases for CKFED. We consider different classifiers for checking the versatility of our extracted features. We use two different classifiers with same specifications which evidence simplicity of our approach however, the results can be further optimized by trying other classifiers. The database consists of frontal views with uniform illuminations. Since the algorithm is working in real time, hence it is suitable to apply it in real time environment keeping in consideration the limitation of database. Further extensions of this work is to enhance the feature sets to include information about pose and lighting variations.

## References

1. 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.
2. 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.
3. P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs Fisherfaces: Recognition using Class Specific Linear Projection *IEEE Transaction on Pattern Analysis and Machine Intelligence* Vol 19, No. 7, July 1997.

4. 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.
5. 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.
6. 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.
7. 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.
8. J. Ahlberg. An Experiment on 3D Face Model Adaptation using the Active Appearance Algorithm. *Image Coding Group, Deptt of Electric Engineering* Linköping University.
9. 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
10. Unsang Park and Anil K. Jain 3D Model-Based Face Recognition in Video *2nd International Conference on Biometrics* Seoul, Korea, 2007
11. Stan Z. Li and A. K. Jain. Handbook of Face recognition *Springer* 2005
12. 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 (FG00)*, Grenoble, France, 46-53.
13. P. Ekman and W. Friesen. *The Facial Action Coding System: A Technique for The Measurement of Facial Movement*. Consulting Psychologists Press, San Francisco, 1978.
14. 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.
15. J. Cohn, A. Zlochow, J.-J. J. Lien, and T. Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proceedings of the 3<sup>rd</sup> IEEE International Conference on Automatic Face and Gesture Recognition*, pages 396 – 401, April 1998.
16. I. Kotsia and I. Pita. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transaction On Image Processing*, 16(1), 2007.
17. Riaz Z. et al. A Model Based Approach for Expression Invariant Face Recognition In *3<sup>rd</sup> International Conference on Biometrics*, Italy, June 2009
18. 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.
19. S. Romdhani. *Face Image Analysis using a Multiple Feature Fitting Strategy*. PhD thesis, University of Basel, Computer Science Department, Basel, CH, January 2005.
20. P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
21. Ian H. Witten and Eibe Frank *Data Mining: Practical machine learning tools and techniques* *Morgan Kaufmann*, 2nd Edition, San Francisco, 2005.