# Hybrid Profile Recognition on the MUGSHOT Database

Frank Wallhoff, Dejan Arsić, Björn Schuller, Gerhard Rigoll, *Member IEEE*
Jan Stadermann, Andre Störmer

*Abstract* — **Face recognition is employed in several systems and computer based applications besides security applications. A broad variety of appearance based approaches has been presented and evaluated even on large datasets. However, most of the systems mainly concentrate on images and models that have been taken from the frontal view of the face. Thus they are limited to test views being taken from a similar viewpoint. In the present treatise we introduce several integrated hybrid systems consisting of artificial neural networks and statistical Hidden Markov Models to recognize profile views of unfamiliar persons, although the system has just been trained on frontal material. This problem is called the mugshot recognition task.**

*Keywords* — **Face Recognition, MUGSHOT, hybrid systems, discriminative training objective.**

## I. Introduction

MANY scenarios for flexible and highly reliable face image based identification mechanisms can be considered, such as security applications in airports and railway stations or the capturing of escapees or terrorists or to identify individuals in multimedia documents such as images or video streams.

The numerous in the literature presented systems show different capabilities and strengths in aspects of recognition performance such as those surveyed in [1]. So called appearance based systems have to be trained with

representative examples to constitute robust models for the subsequent recognition. Up to now, most of the presented systems are limited to frontal views or examples with very restricted variations in the head pose.

In the course of the paper we concentrate on the so-called mugshot recognition task, where just two photographs of a candidate are available. One image is taken from the front and the other from the profile. Our first approach to implement a solution for this face recognition task is to keep the frontal face image and apply a transform using neural networks to normalize the pose of the face to the prior learned one. In a following step unknown images can be classified using Hidden Markov Models (HMMs). Since both obeyed approaches need a training step with different optimization strategies, it seems desirable to combine the image based transformation step with the estimation procedure of the face models yielding in a hybrid structure. The emphasis in this work relies on hybrid recognition systems using joint estimation procedures.

With the goal of having a common test-environment for all test systems, a set of 100 images from the NIST Special Database 18 is compiled [2]. The image pairs to be recognized are commonly pre-processed and cropped as depicted in Figure 1. Furthermore a second database containing a disjunctive set of 600 people is collected from which the image transformation is derived, called DB-1. The test set is called DB-2.

The paper is structured as follows: In the next section we briefly summarize the underlying discrete HMM based face recognition framework and it's basic frontal face performance on the FERET- [3] and AR-Database [4].

Hereafter we introduce neural network based approaches to generate artificial images for recognition. In section IV the presented approaches are extended to novel hybrid recognition systems using integrated training paradigms.

The developed systems are evaluated and compared on the MUGSHOT database.

The work closes with our conclusions and an outlook on possible improvements and extensions.

## II. Face Recognition using HMMs

HMMs in conjunction with images, more precisely the

so called pseudo 2 dimensional HMMs (P2DHMMs), have successfully been applied within the face recognition domain. Fundamental work about HMMs can be found in [5]. After image based normalization and pre-processing steps, the most meaningful elements of a pattern are educed in the feature extraction stage. This can be the extraction of pixel intensities or a higher level representation after a DCT-transform, as presented in [6,7]. The resulting features further describe the characteristic information of an individual and are modelled into a double statistical P2DHMM.

Unknown feature sequences or faces can be classified using the following maximum-likelihood (ML) decision

$$M^* = \arg\max_{M \in KnownFaces} P(X \mid M) \qquad (1)$$

where X is the unknown image and M one HMM among all known individuals from a database. The system recognizes the image as belonging to the individual M* whose corresponding model M achieves the highest production probability P(X|M). In the training phase all HMM parameters have to be estimated using the known Baum-Welch estimation algorithm, which iteratively improves the ML-criterion.

To demonstrate the performance of a P2DHMM based system for frontal faces, we measure the recognition scores on two datasets. A more detailed description of the obeyed system using a block based DCT can be found in [8]. The first set is the known *fb versus fa* corpus form the FERET database [3] with 1,195 image pairs showing different facial expressions. The second corpus is the AR-database [4] providing 7 training and test examples for each of the 115 individuals with different facial expressions and lighting conditions at different dates.

TABLE 1: Results for frontal face recognition.

|  | Accuracy [%] | #tests |
|---|---|---|
| AR | 99.15 | 819 |
| FERET | 98.58 | 1195 |

The results demonstrate the high flexibility of the obeyed classification approach. However, applying this approach, which is specialized to frontal faces, to the MUGSHOT recognition task results in a poor recognition score of just 5% correct recognized persons.

### III. RECOGNITION USING ARTIFICIAL IMAGES

As reported in our previous work [9], it is possible to train a Multi Layer Perceptron (MLP) that is able to transform a given image to a new viewpoint. To perform this difficult task without using 3D information the network parameters, respectively their weights **W**, have to be learned from a set of image pairs during the training phase by minimizing the training error between the input examples **y** and their target values **ŷ**:

$$W^* = \arg\min_{t \in DB-1} | y_t - \hat{y}_t | \qquad (2)$$

Fig. 1 shows the underlying matching problem originated by a change of the azimuth by three planar representations of a three dimensional object.
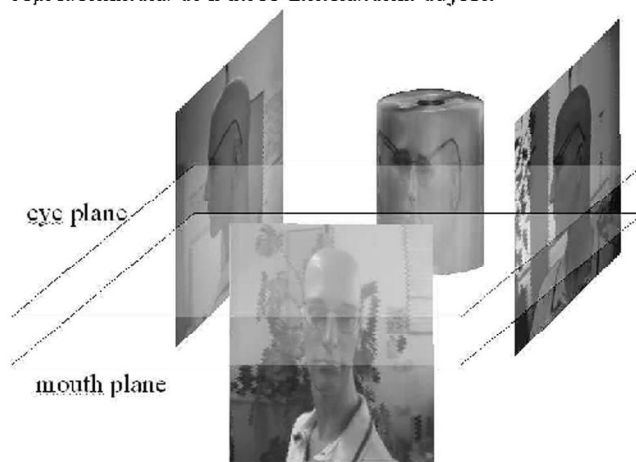


**Fig. 1** In-plane pixel correspondences.

One successful implementation of a MLP architecture to transform views using direct pixel intensities is shown in Fig. 2. Intuitively each line in the output layer has to be linked over a hidden layer to a input region in the same plane respectively height.



Fig. 2 MLP structure for synthesizing profiles.
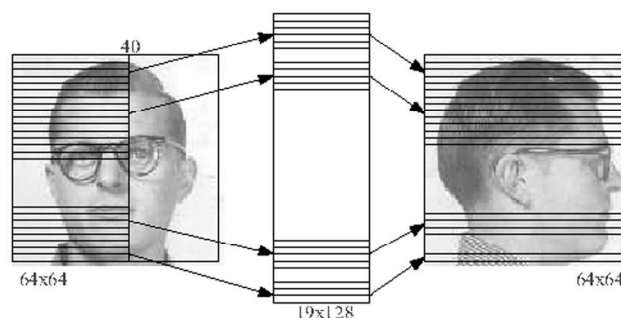
Feeding an unknown image to the MLP's input layer constitutes a corresponding image at the output layer. Examples of frontal and profile views from the MUGSHOT test set are depicted together with artificial views in Fig. 3.



Fig. 3 Examples of image pairs in the database.

As can be seen from the examples above, it is basically possible to achieve the desired views from the new

viewpoint. However, the images are noisy and do not have photo realistic quality. In addition to the pixel intensities we therefore have demonstrated in [10] that it is also possible to do the image transform using Eigenmugshots, i.e. the PCA weights of the images resulting in smoother outputs. Examples of synthesized images are shown in Fig. 4.



Fig. 4 Examples of synthesized Eigenmugshots.

## IV. HYBRID SYSTEMS

Now, being able to transform images to the new view point, images can be classified with the proposed P2DHMM approach. As an extension, it further becomes now possible to estimate the parameters of continuous HMMs directly using challenging hybrid structures. A typical resulting connectionist HMM/MLP system is depicted in Fig. 5 and can be summarized as follows:

1. Estimation of profile HMMs for people in DB-1

2. MLP training with HMM's means as target values

3. Propagation of faces in DB-2 to estimate models
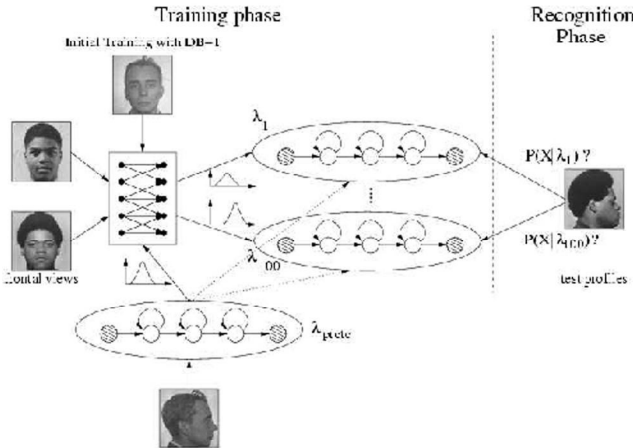
4. Regular HMM classification of unknown profiles



Fig. 5 Hybrid System consisting of a MLP and HMMs

Besides the utilization and processing of direct pixel intensities the well known Eigenface computation (EFC) can be applied to estimate novel views. In contrast to the proposed distance classifiers, such as Euclidean or Mahalanobis, a second MLP for classification is employed. For this purpose a similar structure with a MLP based rotation and classification can be introduced:

1. Train rotation MLP with profile weights W

2. Train classifier MLP with known profiles

3. Merge MLPs to recognize unknown frontals
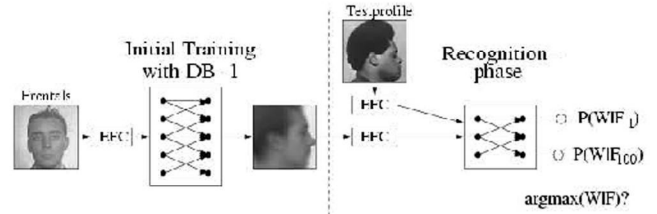
4. Classification



Fig. 6 Hybrid System using Eigenmugshots

The third presented system consists of a strongly combined MLP/RBF compound. As introduced earlier, two different criteria are optimized during the two training steps. Therefore the parameters of the HMM are considered to be the same as those of an equivalent RBF network. This can be easily done by coupling the means and variances of the HMM to the RBF. By fusing the rotation MLP with the classification RBF integrated training paradigms become realizable. In more detail this means that a discriminative gradient descent training algorithm similar to RPROP [11] can be applied such that the parameters **W** are updated according to:

$$W_{\text{Disc.}}^{*} = \arg\max_{i,\kappa \in \text{DB-1}} \frac{P(y_i \mid W)}{P(\hat{y}_{\text{fals}} \mid W)} \qquad (3)$$

After the joint training, the classification can be carried out by HMMs again. The entire structure is depicted in **Fig. 7** and can be described as follows:

1. Estimation of rotation MLP using DB-1
2. HMM training using synthetic views of DB-2
3. Porting to RBF structure and fusion with MLP
4. Discriminative parameter estimation
5. Re-conversion of RBF to HMM and classification

## V. EXPERIMENTS AND RESULTS

Table 2 shows the performance of all proposed recognition architectures, which are measured on the test set DB-2. The hybrid MLP/RBF approach shows a superior performance and can be measured with an accuracy of 60\%. It turns out that due to the low photorealistic quality of the synthesized images, the P2DHMM have a dramatically drop in performance. Furthermore it seems that there is not a sufficiently strong relation between the frontal and the profile weights in their Eigenface domains.

TABLE 2: Comparison of recognition accuracies.

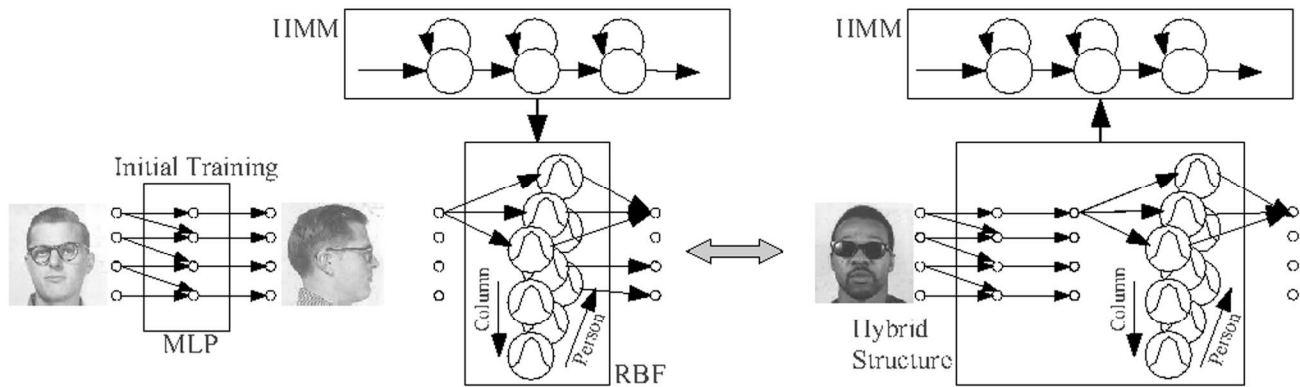|  | Accuracy [%] | Feature |
|---|---|---|
| P2DHMM | 5 | DCT |
| MLP/HMM | 49 | Intensities |
| MLP/MLP | 34 | Eigenfaces |
| MLP/RBF | 60 | Intensities |

**Fig. 7** Hybrid System consisting of a MLP, RBF/HMM

In order to have a reference and a better understanding of this difficult recognition task, we carried out a perception test with people having comparable test conditions as the computer. Thus a considerably high number of wrong classifications has been observed and an accuracy between 70% - 80% was measured.

Therefore the achieved score of 60% for 100 *unfamiliar* individuals can already be regarded as an acceptable performance.

In [12] Maurer et al. reported a similar observation, and that recognition rates for this task will not rise against the perfect performance.

## I. CONCLUSIONS AND OUTLOOK

In the presented work we introduced different three hybrid approaches to solve the mugshot recognition task without using any 3D information. Besides an Eigenface approach, a stronger MLP/RBF/HMM combination was introduced using jointly estimated parameters. Applying the proposed systems, we achieved a recognition score of up to 60% for a test set of 100 unfamiliar people in the MUGSHOT database, due to the superiority of the optimized training criterion.

To obtain similar or even better recognition rates than with our third approach, alternative and improved network structures have to be examined. In the future we will weight face parts in the final recognition stage, since not all predicted areas are of the same importance and quality.

Due to the limitation of profile to frontal view recognition, it is furthermore planned to derive artificial surface maps for a complete view independent processing as depicted in Fig. 8



**Fig. 8** Surface map for view independent recognition

REFERENCES

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Computing Surveys (CSUR), vol. 35, no. 35, pp. 339 458, December 2003.

[2] C. I.Watson, "Mugshot Identification Data - Fronts and Profiles," Reference Data of NIST Special Database 18, Dec. 1994.

[3] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The feret evaluation methodology for face-recognition algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090 1034, Oct. 2000.

[4] A. Martinez and R. Benavente, "The ar face database," Tech. Rep. 24, Purdue University, 1998.

[5] Lawrence R. Rabiner, "A tutorial on HMM and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257 286, Feb. 1989.

[6] F.S. Samaria, Face Recognition Using Hidden Markov Models, Ph.D. thesis, University of Cambridge, 1994.

[7] Stefan Eickeler, Stefan Müller, and Gerhard Rigoll, "Recognition of JPEG Compressed Face Images Based on Statistical Methods," Image and Vision Computing Journal, Special Issue on Facial Image Analysis, vol. 18, no. 4, pp. 279 287, Mar. 2000.

[8] Frank Wallhoff, Stefan Eickeler, and Gerhard Rigoll, "A Comparison of Discrete and Continuous Output Modeling Techniques for a Pseudo-2D Hidden Markov Model Face Recognition System," in IEEE Int. Conference on Image Processing (ICIP), Thessaloniki, Greece, Oct. 2001.

[9] Frank Wallhoff, Stefan Müller, and Gerhard Rigoll, "Hybrid Face Recognition Systems for Profile Views Using The MUGSHOT Database," in Workshop RATFG in conjunction with the IEEE Int. Conference on Computer Vision (ICCV), Vancouver, Canada, July 2001.

[10] F. Wallhoff and G. Rigoll, "Synthesis and recognition of face profiles," in Tagungsband 8th Intern. Fall Workshop Vision, Modelling, and Visualization (VMV) 2003, München, B. Girod et al. T. Ertl, Ed., nov, 2003, pp. 545 55, Aka Verlag, Berlin, IOS Press, Amsterdam.

[11] Frank Wallhoff, Daniel Willett, and Gerhard Rigoll, "Frame Discriminative and Confidence-Driven Adaptation for LVCSR," in IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, June 2000, pp. 1835 1838.

[12] T. Maurer and C. von der Malsburg, "Learning feature transformations to recognize faces rotated in depth," in Proceedings of the International Conference on Artificial Neural Networks, Paris, October 1995.