RECOGNITION OF FACE PROFILES FROM THE MUGSHOT DATABASE USING A HYBRID CONNECTIONIST/HMM APPROACH

Frank Wallhoff, Stefan Müller, Gerhard Rigoll

Department of Computer Science
Faculty of Electrical Engineering
Gerhard-Mercator-University Duisburg, Germany
{wallhoff,stm,rigoll}@fb9-ti.uni-duisburg.de

ABSTRACT

Biometrical systems have been the focus of concentrated research efforts in recent years. These systems can be used to identify a person or to grant a person access to something, e.g. a room. Face recognition technology has reached a level of performance at which frontal-view recognition of faces with slightly different facial expressions, view angles or head poses can be considered nearly solved. In this paper we present a novel hybrid ANN/HMM approach to recognize a person from that person's *profile* view (90°) although the recognition system is trained with only one single *frontal* view of the person. Such a system can be useful for mugshot identification where a victim or witness has seen the criminal from the side only. Our approach uses neural methods in order to synthesize a profile out of the frontal view using no additional knowledge about the 3D shape and structure of a human head. The classification of the generated images is accomplished using a statistical HMM-approach.

1. INTRODUCTION

Face recognition has emerged as an important topic within the field of pattern recognition, with a variety of interesting applications in areas such as security or indexing of image and video databases.

Due to the wide degree of variety in facial expressions, a high performance face recognition system requires the use of sophisticated algorithms capable of handling patterns with large variations within one single class. As research in recent years has shown, the task of recognizing faces with slight variations in head pose and facial expression even for large databases is nearly solved and high recognition scores near 100% were presented with different approaches. For some further details about these approaches see [1], [2], [3] and [4].

As one can see further in these publications, the recognition or classification of larger variations in head pose, such as the relations between the frontal view to the profile view of a person turns out to be a very difficult task. For that reason this paper addresses a problem that can still be considered as a major new challenge in face recognition, namely the problem of recognizing faces that are rotated by 90 degrees and therefore appear as profiles in the image.

Some work on this topic was done in [5] and [6] for example. The success of the approach presented in [5] is dependent on a large quantity of 3D-data collected and processed in a preliminary procedure. In this paper we demonstrate that an ANN structure can generate effective synthetic profile views and thereby eliminates the need to supplement the system with additional, possibly expensive to gather, 3D-data.

In our case, the data for training the face models consists of frontal views as in [1]. Thus, the system has to accomplish a task that can be also considered as a certain challenge for humans, namely to recognize the profile of a person whose face is just known from the front. It turns out that this is indeed a major problem for a standard face recognition system optimized on the recognition of frontal views or views with only moderate tilting [6].

While various databases exist for such simpler tasks, fortunately also for the problem of profile recognition a new database has been constructed and provided by NIST. This database, known as the MUGSHOT database [7], provides the necessary material for setting up a system for profile recognition, and therefore will be described in more detail in the following section.

2. DESCRIPTION OF THE MUGSHOT DATABASE

The MUGSHOT database contains the faces of about 1500 persons, where each person is usually represented by only two photographs: one showing the frontal view of the person's face and the other showing the person's profile. The photographs are provided from archives of the FBI. The figures 1, 2 and 4 show a few examples for persons in that database.

Further examples of images in this database are available at **http://www.nist.gov/srd/nistsd18.htm**. Since there are usually only two samples for each person, it is thus possible, to either train person models from the frontal views and recognize the person by presenting his profile, or vice versa. We have decided to concentrate on the first option, namely to use the frontal views for training. It turned out that a considerable number of images in the MUGSHOT database are of very bad quality, with distortions in the images, numbers printed in the background of various images or severe under-exposure of the photograph. It was therefore decided to manually pre-select and preprocess the images in the following way: Photographs with unusual high distortions, perturbations or under-exposure were discarded, while each image considered for inclusion in the experimental database was manually preprocessed so that all faces appeared in the center of an image with a moderate amount of background and with similar size of the faces. In order to constrain the experiments, we selected 100 persons from the remaining mugshots to form a smaller database we called DB-1. DB-1 contains the frontal views and profiles of the 100 persons we selected for training face models from the information provided by the frontal views and for testing by presenting the profiles of these 100 persons for recognition. The resulting images consist of 64×64 pixels.

Furthermore, beside the 100 persons contained in DB-1, an additional database with upgraded images of another 600 people was constructed, which was named DB-2. This database was not intended for providing additional training data, but instead has been foreseen for training an ANN which produces synthetic profiles.

3. BASELINE SYSTEM FOR FACE PROFILE RECOGNITION

In a first experiment, it was decided to use the most straightforward and simple approach for profile recognition, namely to test the performance of a planar Hidden Markov Model (also referred to as Pseudo-2D HMM, P2DHMM) [1], [8] used for recognition of frontal faces for this task. Thus, for each of the 100 persons of DB-1, a separate P2DHMM was trained from the frontal face image of each person. For recognition, each face profile in DB-1 was presented to the 100 P2DHMMs constructed in the previous step and the probability of generating the profile by each P2DHMM has been calculated. Of course, the P2DHMMs were extremely bad models for the profiles presented to them and one could not expect a high recognition rate. This was confirmed by the first experiments yielding extremely poor recognition rates of 10% and less, depending on the configuration of the system.

From this initial experiment, it was obvious that profile recognition is not possible using only the information provided from the frontal face images. Additional knowledge is necessary about the relationship between the frontal view and the profile of a person. There are basically two possibilities for incorporating this knowledge into our baseline recognition system: One obvious option is the use of 3-dimensional head modeling techniques as proposed in [5]. In this case, the geometrical properties of human heads are exploited, which have to be acquired by 3D-scanning of a variety of human heads. In this case, e.g. the right profile of a person can be obtained by mapping the image of his frontal view onto the 3D-model of a human head. By exploiting only the 3D-data belonging to the right half of his face and computing a 2D-projection of the view that would result from looking at this 3D-data from the right side, the right profile can be obtained.

The second possible option is the utilization of a learning algorithm, capable of learning the head rotation process from the presentation of many examples, which is the approach that we pursue in this paper. The reason for this choice is the fact that this should represent an interesting alternative to the previously mentioned approach and that the collection of 3D-data from a large number of heads is a tedious task, that could be simply replaced by evaluating an appropriate number of images with a suitable learning paradigm instead. Since in principle, not the derivation of 3D-data from the frontal view is the most important topic here, but instead the relationship between the frontal image and the profile, it should be basically possible to derive this relationship from examples for those views, without the necessity of deriving 3-dimensional data. An obvious paradigm for learning such a relation is a neural network. Therefore, the basic idea is the utilization of a neural network that should generate the profile of a person by presenting the frontal view to its input.

Setting up the profile recognition system would then consist of the following steps:

1. Use the 600 profiles and frontal views of DB-2 to train a neural net that learns the generation of profiles in its output if the frontal view is presented at the input.
2. Present the 100 frontal views of DB-1 to the trained neural net in order to generate 100 artificial profiles.
3. Use those artificial profiles in order to train a Hidden Markov Model for each profile which will be the template for each person. This concludes training.
4. Recognition: Present each of the 100 real profiles of DB-1 to the HMMs in (3) and use the Viterbi algorithm to find the most probable class (i.e. the original candidate).

This leads to an interesting combination of neural nets and HMMs and thus results in the use of a hybrid connectionist/HMM approach for profile recognition, which is explained in detail in the following section.

4. HYBRID CONNECTIONIST/HMM APPROACH FOR PROFILE RECOGNITION

4.1. ANN-Architectures

In order to synthesize a profile view from a given input image, we had to find an appropriate ANN-architecture. As input- and output features, normalized gray-values of the images (64×64 pixels) were taken directly. In a first experiment we used a fully connected single layer network with 64^2 input and 64^2 output neurons. Not surprisingly, the results were not satisfactory. The 600 examples of DB-2 were not enough to generalize a net with such a high number of parameters.

With respect, that there is no information about the structure and form of a human head, this task seems to be very difficult. One thing that can be roughly said about a point in the frontal view is that it can be found in the same line of the profile view. Because the used images were cropped manually and were taken at different occasions with slightly different facial expressions, the lines in the neighborhood might be of interest, too. The number of the connected lines besides the active one is given by the value n (see figure 1). So, in order to generate the profile view of a face or head we use 64 fully connected single layer nets, one for each line in the generated profile view.

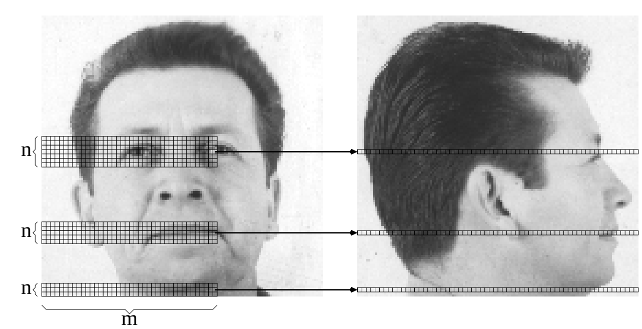

Fig. 1. Subnets for the generation of a profile view

In order to reduce the number of parameters to be estimated, we did not take entire lines from the input image. As one can imagine, the right half of the face is redundant and almost symmetrical to the left half of the face. To make sure that the complete right half-face can be mapped to the output pixels, we used a few more

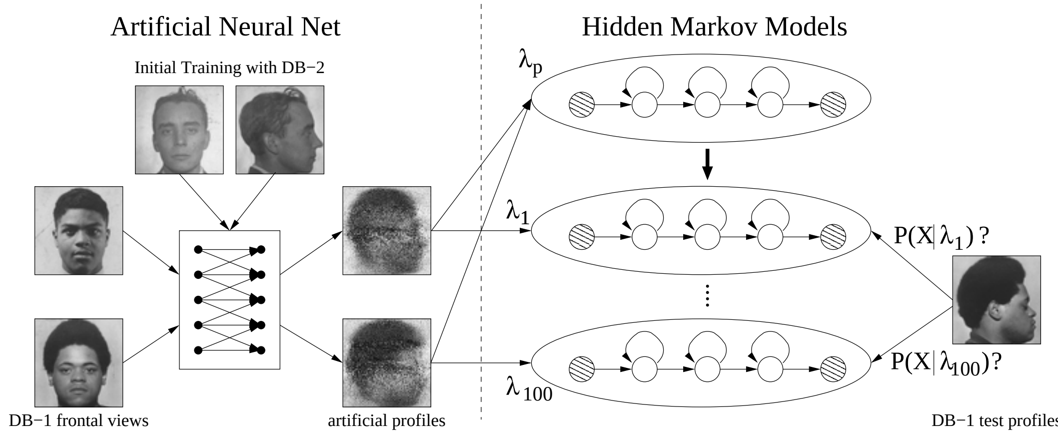

Fig. 2. Overview of the recognition system

pixels from the input image. The number m gives the amount of pixels used. These subnets can be combined to one large net, that generates the profile view of a person. In figure 1 the subnets are shown. We tried out several combinations of the $(n \times m)$ parameters. In the figure $m = 40$ and n varies between 3, 5 and 7. The arrow in figure 1 indicates, that the input neurons are fully connected to the output neurons. For the training procedure the known RPROP-algorithm [9] and for the activation neurons Sigmoid functions were chosen.

4.2. The HMM-Recognition Approach

To recognize the generated images we use classical one-dimensional left to right HMMs, with a vector-dimension of 64, analogous to the approach presented in [10, 11], allowing self-transitions and transitions to the next state only. For recognition the face image is subdivided into 64 vertical stripes, which is then considered to be the observation sequence emitted by the linear Hidden Markov Model. The use of one-dimensional HMMs is preferred here, because of the inter-line connectivity, the resulting images are quite noisy in the vertical direction. This circumstance might be confusing for the vertical dynamic warping-capabilities of the P2DHMM, as was proven by initial results.

Because we have only one training example to estimate the parameters of the resulting HMM, we produce a prototype HMM λ_p with all features from the generated images in a first step. This prototype is intended to be a draft model for an average profile view. After this we retrain the models $\lambda_1 \ldots \lambda_{100}$ for each person in DB-1 using λ_p as the new prototype. For both training-procedures the well-known Baum-Welch-Estimation procedure [12] was used to estimate the appropriate transition and emission probabilities.

After the models are trained, unknown images can be classified by a maximum-likelihood decision. This decision uses the reduced formula of a Bayes-classifier given in 1, where X represents the unknown face, λ one face out of DB-1 and λ^* the best matching model out of DB-1:

$$\lambda^* = \underset{\lambda \in \text{DB-1}}{\operatorname{argmax}} P(X|\lambda) \tag{1}$$

The HMM with the highest probability score indicates the recognized person to whom the unknown face will be assigned. The entire recognition system is summarized in figure 2.

5. RESULTS

In the training phase of the net, the complete net with its weights is stored after every 25th iteration, making it possible to test the overall recognition rate of the system for several states in the training phase of the neural net. After the training of the ANN-part of the entire system, the pictures of the test set (DB-1) are presented to the input neurons of the nets. The resulting images are then modeled and classified with linear 26-state HMMs as described above. In figure 3 the overall recognition rates of such a hybrid system are depicted. In this experiment a net, consisting of 64 subnets, where $m = 40$ and $n = 7$ (see ANN-architecture above) was used. As can be seen in the figure below, our highest score is 56%. In a wide spectrum of iterations the overall recognition rate is mostly over 50%. Several parameter sets were evaluated, which changed the number of HMM-states and the characteristic parameters of the ANN (m, n). It turned out that the variation of these parameters is not too important and just causes smaller changes in the behavior.

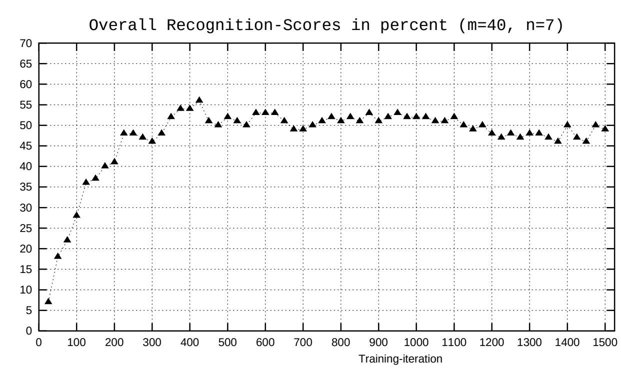

Fig. 3. Overall Recognition Rates

To get an idea of the potential of the HMM-recognition approach we decided to use an Euclidian classifier and compared the results. For the highest recognition score (56%) we repeated the test using the following distance measure between two images P_i and P_j, where $p_{i,j}(x, y)$ are the gray-values of a pixel at (x, y):

$$d(P_i, P_j) = \sum_{x=0}^{64} \sum_{y=0}^{64} (p_i(x, y) - p_j(x, y))^2 \tag{2}$$

In this formula, the square of the gray-value difference of both images is summed up. The winner was the picture with the closest distance to the generated one. For this experiment we achieved a poor recognition rate of just 24%. So the use of the HMMs with its superior modeling capabilities seems to be justified. In figure 4 some examples of the generated profiles (in the middle) are shown together with the frontal view (left) and the real profile (right). These samples show the quality of artificial profiles we are able to generate. Further can be seen in the figure below, that the frontal views do not contain all relevant information of the profile, which is the main reason why this task is so difficult.

Fig. 4. Examples of frontal views, the generated and real profiles

Another criterion of comparing face recognition systems are the cumulative match scores as used in [13]. In the following figure the graph of the cumulated match score of the parameter constellation from above ($m = 40, n = 7$) at the 425th training iteration is shown. As can be seen from this graph, there are already about

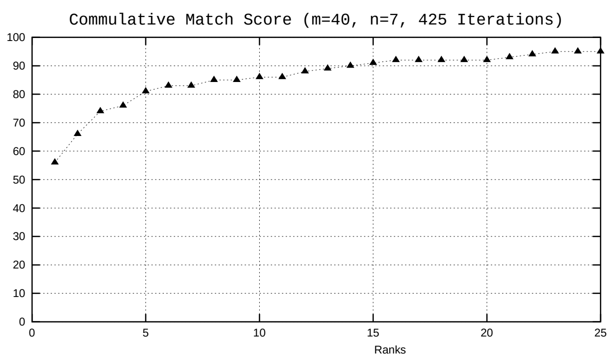

Fig. 5. Cumulative Match Score

75% among the top 3 candidates. This fact is an indicator that the neural network was able to generalize the task it was trained for.

6. CONCLUSIONS

In this paper we presented a novel ANN/HMM connectionist system to recognize profiles of people, using only the frontal views. For this task no additional 3D-information was used. Our system achieves recognition rates up to 56%.

The potential of our approach is impressively demonstrated by the fact that already now, 74% of the correct candidates are among the top 3 ratings. Therefore, we plan further improvements to exploit this recognition potential, which we feel is near 80% by using several networks transforming special face regions, eg. eyes, nose, ears. As mentioned in [6] the recognition rates will not grow near 100% for this even for human beings hard task, because the frontal view does not contain all necessary information of the corresponding profile. Our system can compete with the system which however exploits additional 3D-information presented in [5] and outperforms the approach presented in [6]. In the future, we will use much more sophisticated hybrid training procedures where the HMM and the ANN parameters are trained jointly in order to improve our system. Another goal is to use the FERET-database provided by the Army Research Labs (ARL) for better comparison possibilities.

7. REFERENCES

[1] S. Eickeler, S. Müller, and G. Rigoll, "Improved Face Recognition Using Pseudo 2-D Hidden Markov Models," in *Workshop on Advances in Facial Image Analysis and Recognition Technology (AFIART)*, Freiburg, Germany, June 1998.

[2] S. Eickeler, S. Müller, and G. 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.

[3] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, July 1997.

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

[5] T. Vetter, "Recognizing Faces From A New Viewpoint.," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 1997, pp. 139–144.

[6] T. Maurer and C. van der Malsburg, "Learning Feature Transformations to Recognize Faces Rotated in Depth," in *Proceedings of the International Conference on Artificial Neural Networks*, Paris, France, Oct. 1995.

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

[8] S. Eickeler, S. Müller, and G. Rigoll, "High Quality Face Recognition in JPEG Compressed Images," in *IEEE Int. Conference on Image Processing (ICIP)*, Kobe, Japan, Oct. 1999, pp. 672–676.

[9] M. Riedmiller and H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm," in *Proceedings in IEEE International Conference on Neural Networks*, San Francisco, CA, Apr. 1993.

[10] A. V. Nefian and M. H. Hayes III, "Hidden Markov Models for Face Recognition," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, May 1998, pp. 2721–2724.

[11] F. Samaria and A. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," in *IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida, Dec. 1994.

[12] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.

[13] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *NISTIR 6264*, Oct. 1999.