Synthesis and Recognition of Face Profiles

Frank Wallhoff and Gerhard Rigoll

Technische Universität München, Institute for Human-Machine-Interaction
Arcisstraße 21, 80290 München, Germany
Email: {wallhoff,rigoll}@mmk.ei.tum.de

Abstract

Research on biometrical systems and especially on face recognition systems has become of high interest. Nowadays several approaches exist to recognize frontal views of faces. Under certain constraints the recognition accuracy even for huge databases seems to be acceptable. However, it has shown, that the recognition performance of nearly all state-of-the-art systems dramatically drops down, when faces rotated in-depth or even profile views are presented.

In this work we present our actual experiments and results of neural network based approaches for face profile recognition systems. One of the main challenges is to implement a self learning approach that does not need any direct 3D information. The key idea of our approaches is to derive the correspondences between head profiles and frontal views by examples automatically. With this approach we are able to synthesize profile views of heads by presenting frontal views. The quality of the resulting systems can be measured with Hidden Markov Models (HMM) and an extension to the well-known Eigenfaces approach. The performances are evaluated on the MUGSHOT and the FERET database.

1 Introduction

Face recognition technology (FERET as introduced by the US Army Research Laboratory, ARL) has become increasingly important for several fields of applications, such as controlling who is entering a building (access control) or detection of violent criminals and terrorists in airports or other public places. Although many different approaches have been presented for the frontal face recognition problem [1, 2], it seems not to be solved yet for real world applications on huge databases, as also stated in [3] and [4]. Especially the recognition of faces rotated in depth is not generally solved.

Therefore this paper addresses our actual research on recognizing synthesized faces rotated by 90 degrees in depth with neural and statistical approaches. When we started our work on this challenging task, we noticed, that there has not been done much investigation in this area before. We found several approaches, which concentrated on generating synthesized 3D images of faces using 3D data [5, 6]. Other, newer approaches are mainly based on so called head meshes or wire-frames to model 3D information into planar images [7, 8].

However, all the approaches above make use of expensive gathered or labeled 3D-data. Our goal is to engineer a *self-learning* approach using sufficient planar training material for the synthesis instead. For this task we deploy a neural network to learn the rotation process from training examples. Another property of such an example is that it can synthesize regions of the profile that were occluded in the frontal view by the use of learned knowledge such as the hair behind the ears.

To classify the quality of the profile views, we examined HMM based classification approaches and an extension of the well-known Eigenface approach.

The paper is structured as follows. Two different approaches for face profile synthesis using neural networks are introduced in the successive section, which is followed by the classification techniques using HMMs and an Eigenmugshots. In the next section the obeyed databases are briefly introduced. Hereafter the classification results of the presented systems are discussed. The paper closes with a summary and outlook.

2 Generation of Synthesized Profile Views

As mentioned above one of the main goals in the present work is to find neural structures that are able to derive the relations between the frontal and the profile view of a face. After the training phase of such a network we are then able to generate profiles of unknown persons by presenting the corresponding frontal view. Although it would also be possible to generate frontal views from given profiles, we mainly concentrated on the first task. However, we do not expect major changes in the performances of such systems.

In the following sections the obeyed neural structures for the generation of profiles are introduced in more detail.

2.1 Synthesis Using Neural Networks

The first approach for the rotation process is based directly on the gray level intensities of the obeyed image pairs. The simple underlying assumption for the network structure is the fact that the same point in two orthogonal images from a rotated head is located in the same plane (if they are visible at all). Figure 1 illustrates this relation.

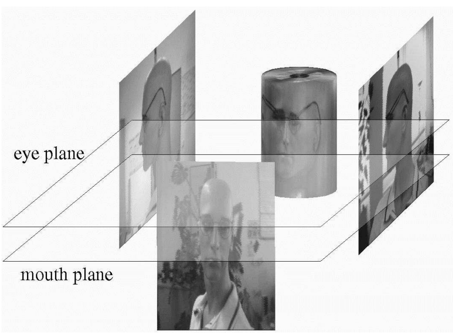

Figure 1: Orthogonal 2D images from 3D object

From this conclusion we derive MLP-networks of the following structure: Rows in the 2 dimensional output layer are derived over a hidden layer from rows in the input layer at the same height (illustrated in figure 2 for a typical image pair with a dimension of 64×64 pixels). Because the right part of a face has no direct relation the left profile, this information is not used. In order to train a more robust MLP against misalignment of image pairs, information of the neighbored rows are also taken into account. The layers itself are line wise fully connected with each other.

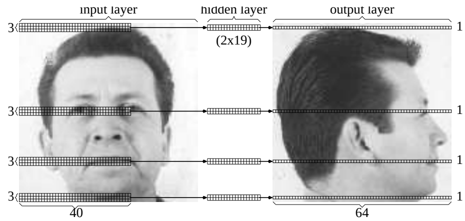

Figure 2: MLP for the generation of a profile view

For the generation of frontal views from given profiles, analog relations can be used. Here the back part of the head does not contain relevant information to the frontal part of the face.

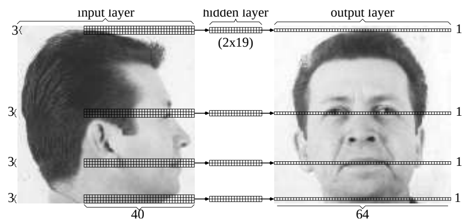

Figure 3: MLP for the generation of a frontal view

For the MLPs activation functions sigmoid functions were chosen and for the training of the parameters and weights we use the well-known RPROP-algorithm. Examples for the training phase of the MLP will be introduced later. However, the functionality of the presented approaches is demonstrated in the following figure.

Figure 4: Examples from real frontal and profile views together with synthesized images

The figure shows that parts of the generated images are noisy, which inspired us to implement another approach for the synthesis.

2.2 Synthesis with Eigenmugshots

In this second approach we decided to use so-called Eigenfaces which where first introduced by Sirovich and Kirby [9] and later extend by Turk and Pentland to the well known Eigenfaces [10]. As mentioned above, we try to make the synthesized images look more naturally using this technique. We call this extension the Eigenmugshot approach.

The main idea behind this is to project all faces into an universal Eigenspace using the principal component analysis. The weights W of the Eigenfaces can be used to measure the distance between two given images directly. The following figure shows the first ten Eigenfaces of each frontal and profile view.

Figure 5: First ten Eigenmugshots of training material

For our purposes this means, that we have to compute the weights of the given examples for frontal views and profile views separately. In the training phase of a single layer MLP the weights of the projected frontal views are learned together with the weights of the corresponding profile views. After this the network is able to estimate the weights of unknown Eigenmugshots. The weights of the Eigenmugshots in the Eigendomain can easily be projected back into the image-domain again. Some examples of typical Eigenmugshots are depicted below. A comparison with figure 4 shows that the visual quality of the second approach is unexpectedly not as good as that of the first one.

Figure 6: Examples of synthesized Eigenmugshots

3 Classification

To judge the quality of the systems presented above and to test the usability for face recognition tasks, we classify the synthesized profiles with the real ones. For this purpose we studied 4 different classifiers, which can be divided into HMM-based and Eigenmugshot-based approaches.

3.1 Face Recognition with HMMs

Hidden Markov Models have shown their excellent performance in several domains of pattern recognition such as speech recognition, handwriting recognition and even face recognition [11, 12].

3.1.1 Classical 1D-HMMs

In most HMM recognition approaches, continuous classical 1-dimensional left to right first order models as described by Rabiner in [13] are used.

An arbitrary HMM λ is completely described by

- the number of internal emitting states q_j given by J,
- a state transition matrix $\overleftrightarrow{A}$
- the (continuous) production probability vector $\vec{b}$, with the elements $b_1 \ldots b_J$.

The production probability b_j in a certain state j for a D-dimensional observation $\vec{x_j}$ is given by a multivariate Gaussian distribution consisting of a mean value vector $\vec{\mu}_j$ and a covariance matrix Σ_j.

In our case, face-images are represented by a series of vertical stripes of gray level intensities, grouped in column vectors $X = [\vec{x}_1 \ldots \vec{x}_T]$. This resulting feature sequence X is in a first approximation piecewise stationary.

Figure 7 summarizes such a 1-dimensional left to right HMM with $J = 3$ emitting states as well as the non emitting (hatched) start and end state.

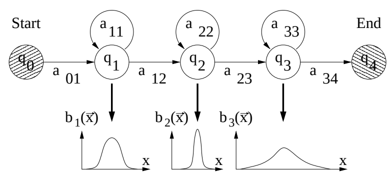

Figure 7: 1-dimensional HMM state transitions and corresponding production probabilities

The unknown HMM parameters $\overleftrightarrow{A}$ and $\vec{b}$ have to be estimated before recognition using the well-known Baum-Welch-Estimation procedure. Unknown images can be classified by a maximum-likelihood decision using the previously trained profile models λ. The HMM with the highest probability score corresponds to the person recognized and to whom the unknown feature sequence or face will be assigned.

3.1.2 Pseudo 2D-HMMs

In addition to the above described modeling technique, we also try to explore the performance of an extension of this approach, which are the so called pseudo 2-dimensional HMMs (P2D-HMMs). These models have been formally introduced in [14] where they have also been applied successfully to problems such as face recognition and handwriting applications. The main difference to 1-dimensional HMMs is based on the state transition sequence, 2-dimensional behavior of the observations can be reconstructed.

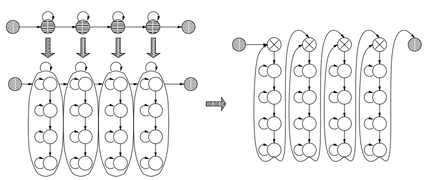

Figure 8: Correspondence of a 1D-HMM, its encapsulated representation and linear representation

3.1.3 Joint-ANN/HMM parameter estimation

In this subsection we present a third HMM based recognition approach. The difference to the systems above is, that we do not need generated synthesized profiles. The HMMs are directly estimated from a given frontal view. For this task we use similar neural structures as described above. But now the target values (outputs) of the MLPs are the mean values $\vec{\mu}_j$ of the production probabilities used by the 1D-HMMs. These are typically the most relevant parameters in the recognition phase.

For this purpose we have to generate appropriate training-material, which is done by the derivation of the feature vector sequences X as described above for each person contained in DB-1 in a first step. Then a common HMM-model λ_p for face profiles is estimated. Hereafter the re-estimation of the mean values $\vec{\mu}_J$ for each individual follows. These computed mean values are further used as the target values of the MLP output as depicted in figure 9. The arrows in the network topology indicate again, that the sub-nets are fully connected. The network architecture is basically the same as in the structure presented before.

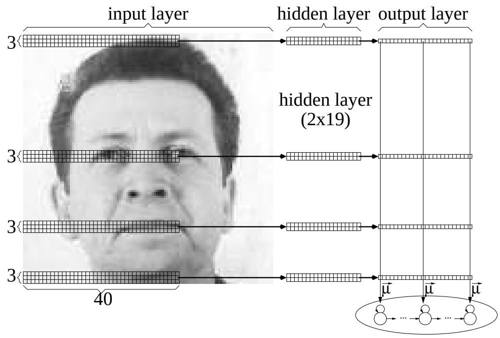

Figure 9: Subnets for the generation of a profile model

In the recognition phase of the system, the models for the unknown individuals are assembled by using the earlier trained common prototype model λ_p and the computed mean values of the MLP without an additional retraining. The recognition procedure is again the same as the one presented for the synthesized profiles. The whole setup is summarized in figure 10.

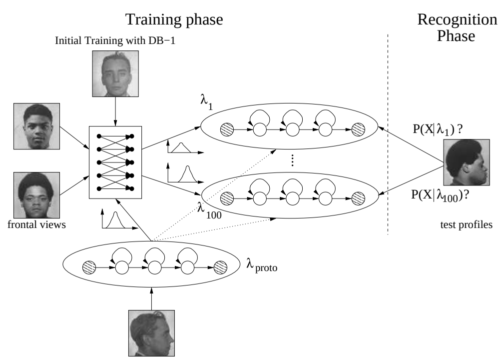

Figure 10: Overview of a recognition system using joint parameter estimation

The advantage of this second structure is the fact, that the computed means need only to be inserted into the previously computed prototype HMMs before the recognition can start. There is no necessity for a complete re-modeling of the synthesized images. This vastly improves computation speed.

3.2 Face Recognition with Eigenmugshots

As mentioned earlier, Eigenmugshots can be classified by using the distance of the weights. For this task a simple Euclidian distance is used. In contrast to the statistical approaches above, this classifier is extremely fast. For the recognition, all distances between the unknown and all known profiles are computed. The recognized person is that with the smallest distance.

4 Description of the obeyed Databases

For the training of the MLPs and testing using HMMs we need well defined sets of image pairs. For this purpose we decided to make use of two publicly available databases called the MUGSHOT and the FERET database to facilitate possible future comparisons. These databases are introduced in more detail in the next two subsections.

4.1 Description of the MUGSHOT database

In the training- and test phase of the systems, we obeyed two disjoint sub-sets of the so-called MUGSHOT-database. The NIST Special Database 18: *Mugshot Identification Data* is available from the National Institute of Standards and Technology (NIST) [15].

The MUGSHOT database contains images of 1573 cases, where most individuals are 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 database contains pairs of mainly male but also female cases at several ages and representatives of various ethnic groups, people with and without glasses or beards and many different hairstyles. The lighting conditions and the background of the photographs also change. The photographs are provided from archives by the FBI. The pictures contained in the database are stored as 8-bit gray scale images with different sizes. Almost all images were scanned at 500 DPI with a Kodak MegaPixel camera. It turns out that a considerable number of images are of bad quality. Images are distorted, contain numbers printed in the background or are severely under-exposure. Such images are excluded from the further processing steps. Figure 11 contains some original examples.

Figure 11: Examples of image pairs from the MUGSHOT database

More information and examples from the MUGSHOT database are available in the World Wide Web at **http://www.nist.gov/srd/nistsd18.htm**.

4.2 Description of the FERET database

The obeyed subset of the FERET database has been assembled between 1993 and 1997 by the US Department of Defense's Counterdrug Technology Development Program through the Defense Advanced Research Agency (DARPA) for performance evaluations of face recognition systems. The final release of the publicly available corpus consists of 14051 eight-bit gray-scale images of human heads with views ranging from frontal to left and right profiles. The images were taken at different locations and times with different lighting conditions and cameras.

Figure 12: Examples of image pairs from the FERET database

A more detailed description of the FERET database is given in [16] and in the World Wide Web at **http://www.nist.gov/humanid/feret**.

4.3 Pre-processing and Subset-Selection

As previously mentioned, our system deploys artificial neural networks to synthesize the rotation process of a given frontal view to a profile view. To estimate the parameters of the MLPs and of the corresponding HMMs, we define several data sets for training and testing purposes.

Therefore the images are semi-automatically pre-selected and pre-processed in the following way: Photographs with unusually high distortions, perturbations or under-exposure are discarded. Each image included in the experimental database is manually labeled, so that all faces appear in the center of an image with a moderate amount of background and with similar size of the faces. All resulting images are re-scaled to a size of 64×64 pixels. Although the images are pre-processed, they still show minor changes concerning the pose and the tilting of about $\pm$ 5 degrees. They are also not normalized regarding contrast or lighting conditions. By this fact the transformation will become more robust and less vulnerable against variations. Preprocessed examples are given in Figure 13.

After the pre-processing the test and training sets are defined. The first set of image pairs is called DB-1 and is exclusively reserved for the *training of the neural structures*. It consists of 600 image pairs taken from the MUGSHOT database.

Then we select another 100 image pairs from individuals in the MUGSHOT database to form a separate sub-set we call DB-2. The third set called DB-3 is taken from the FERET database and consists also of 100 images. Finally we form DB-4, which consists of DB-2 and DB-3 having 200 images. The sets DB-2 to DB-4 are later required as *test corpora* in the evaluation experiments.

Figure 13: Examples of pre-processed images

5 Results and Discussion

In the first test-row the performance of a system using synthesized images and a 1D-HMM classifier is evaluated. During the training of the MLP with DB-1, all net parameters are stored after every 5th iteration. This enables us to evaluate the recognition performance for the test-sets at different progresses during the training phase.

As can be seen, our highest score is 60% using synthesized profile views for DB-2, which consists of 100 images. The performance for DB-3 is always worse than the performance of DB-2 and has a maximum peak of 40%. For DB-4 we can measure a performance that lies between DB-2 and DB-3 and has a peak of 42.5%. The lower recognition rates for the test-set DB-3 can be explained by the fact that the images are from another database (FERET) and may differ more from the training examples. The results of DB-4 furthermore depend on the higher confusion in this set (200 examples). As a 3 best list of the recognition scores of this system shows, already now there are about 80% correct among the top 3 candidates. The results are comparable to a system that uses traditionally generated profile views as presented in [5].

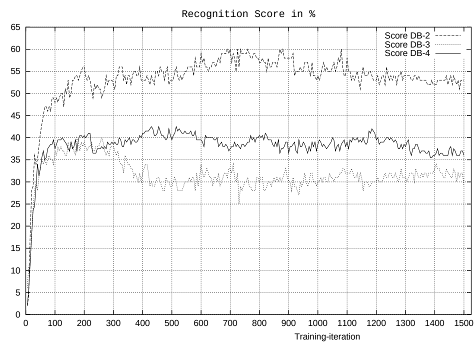

Figure 14: Recognition scores in % for 1D-HMMs

Beside the modeling with 1D-HMMs, we also evaluated the performance using P2D-HMMs for the same images. However, because we just measured approximately one fourth of the recognition rates (correctness) of the 1D-HMMs, this approach was not further prosecuted. The reason for the lower performance is probably relied on the fact that the generated images are too noisy. This confuses the warping capabilities of the P2DHMMs. The noisy pictures violate the assumption of an at least piecewise stationary signal.

The second test-row uses the joint parameter estimation idea. Figure 15 contains the recognition rates of this joint system. As can be seen, the highest score is 49% using the directly estimated HMMs. This lower recognition rates may be explained by the fact that the variances of the common model are not sufficient for a proper profile model. Furthermore, the different training criteria (error minimization of the MLP and maximum likelihood for the HMM) could be a problem.

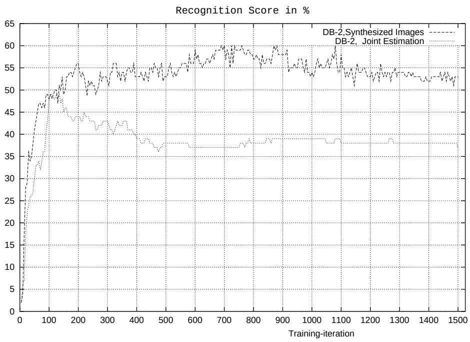

Figure 15: Recognition scores in % of both systems for DB-2

In contrast to the observations of the first system regarding the location of the training minimum error of the MLP and the maximum recognition score for DB-2, both extremes are around iteration 100 for this second approach. This indicates, that the MLP generalizes the HMM's mean values in an optimal way. In table 1 the 1D-HMM based results are summarized for a better overview.

data-set	no. of pairs	from database	highest score	highest score
DB-1	600	MUGSHOT	-	-
DB-2	100	MUGSHOT	60%	49%
DB-3	100	FERET	40%	-
DB-4	200	DB-{2+3}	42.5%	-

Table 1: Overview of the results

Analog to the experiments above, we trained neural structures for the generation of the Eigenmugshots. Here we again stored the outputs of the MLPs at every fifth training iteration. After this we classified the retransformed images using Euclidian distances as well as 1D-HMMs.

The results from these experiments show, that the simple Euclidian distance approach outperforms the HMMs based classifier. Recognition scores of 50% for the Euclidian and poor 24% correct for the HMMs were measured. We conclude, that slight variations of the weights have a critical impact on the retransformed image. The quality of the images seems to be essential on the success of the HMM based approach. The still lower performance can be explained by the fact that the obeyed NN-topologies are not as mature as those of the first systems yet.

As mentioned by Maurer in [17] the recognition rates for the task to recognize unknown faces rotated by 90, will not rise close to 100%. However, the achieved score of 60% for 100 *unfamiliar* individuals is not too far away from the recognition performance of human beings for this special task. A comparison with other psychophysical studies in a related work [18] has shown that even human beings will not come even close to the perfect recognition score. Some preliminary tests in our laboratory confirmed such a human recognition rate of 70%-80% with several test under comparable test-conditions. Considering this fact, we strongly believe that the rates we currently obtain already represent a respectable result and that we can obtain further improvements in the future.

6 Conclusions And Outlook

In the presented work we introduced different approaches to recognize profiles of people, just using their frontal views. For this task no additional 3D-information was used. We reported recognition scores of up to 60% for a test set of 100 people in the MUGSHOT database, up to 40% of a test-set of 100 individuals from the FERET database and up to 42.5% correct for a test-set consisting of 200 individuals using synthesized profile views.

Furthermore, a stronger MLP/HMM combination was introduced using jointly estimated parameters. Although the computation speed was drastically increased, we obtained lower recognition scores of just up to 49% were obtained.

An Eigenmugshot approach was introduced, which resulted in first recognition scores of up to 50%.

The performance of the profile recognition systems we presented competes with the systems previously presented in the literature. To obtain similar or even better recognition rates than with our first approach, alternative and improved network structures have to be examined. In the future we will use connectionist approaches, which will also estimate the variances of the models.

7 Acknowledgement

Parts of this work were funded by a German DFG research project with the title ”Face Profile Recognition using Hybrid Pattern Recognition Techniques”.

References

[1] 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.

[2] D. Beymer and T. Poggio, “Face Recognition From One Example View,” in *MIT A.I. Memo No. 1536 C.B.C.L. Paper No. 121*, Sept. 1995.

[3] S. Cass and M. J. Riezenmann, “Improving Security, Preserving Privacy,” *IEEE Spectrum: Critical Challenges 2002*, pp. 44–49, Jan. 2002.

[4] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E Tabassi, and J.M. Bone, “FRVT 2002: Evaluation Report,” *http://www.frvt.org/FRVT2002/documents.htm*, March 2003.

[5] T. Vetter, “Synthesis of Novel Views from a Single Face Image.,” *International Journal of Computer Vision*, pp. 103–116, 1998.

[6] F. Lavagetto, R. Pockaj, and M. Costa, “Smooth Surface Interpolation and Texture Adaptation for MPEG-4 Compliant Calibration of 3D Head Models,” *Image and Vision Computing Journal, Special Issue on Facial Image Analysis*, vol. 18, no. 4, pp. 345–354, Mar. 2000.

[7] R. Lengagne, P. Fua, and O. Monga, “3D Face Modeling from Stereo and Differential Constraints,” *Image and Vision Computing Journal, Special Issue on Facial Image Analysis*, vol. 18, no. 4, pp. 337–343, Mar. 2000.

[8] C.M. Cheng and S-H. Lai, “An Integrated Approach to 3D Model Reconstruction from Video,” in *IEEE Proceedings of the second Int. Workshop on Recognition , Analysis and Tracking of Faces and Gestures in Real Time Systems in Conjunction with the Int. Conference on Computer Vision 2001*, Vancouver, Canada, July 2001, pp. 16–22.

[9] L. Sirovich and M. Kirby, “Low-dimensional Procedure for the Characterization of Human Faces,” *Journal of the Optical Society of America*, vol. 4, March 1987.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

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

[17] 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, Oct. 1995.

[18] D. Valentin, H. Abdi, and B. Edelman, “What represents a Face : A Computational Approach for the Integration of Physiological and Psychological Face Data,” in *The effects of distinctiveness in recognising and classifying faces.*, 1986, pp. 525–536.