2009 Advanced Video and Signal Based Surveillance

A 3D Face Model for Pose and Illumination Invariant Face Recognition

Pascal Paysan
pascal.paysan@unibas.ch

Reinhard Knothe
reinhard.knothe@unibas.ch

Brian Amberg
brian.amberg@unibas.ch

Sami Romdhani
sami.romdhani@unibas.ch

Thomas Vetter
thomas.vetter@unibas.ch

Abstract

Generative 3D face models are a powerful tool in computer vision. They provide pose and illumination invariance by modeling the space of 3D faces and the imaging process. The power of these models comes at the cost of an expensive and tedious construction process, which has led the community to focus on more easily constructed but less powerful models. With this paper we publish a generative 3D shape and texture model, the Basel Face Model (BFM), and demonstrate its application to several face recognition task. We improve on previous models by offering higher shape and texture accuracy due to a better scanning device and less correspondence artifacts due to an improved registration algorithm.

The same 3D face model can be fit to 2D or 3D images acquired under different situations and with different sensors using an analysis by synthesis method. The resulting model parameters separate pose, lighting, imaging and identity parameters, which facilitates invariant face recognition across sensors and data sets by comparing only the identity parameters. We hope that the availability of this registered face model will spur research in generative models. Together with the model we publish a set of detailed recognition and reconstruction results on standard databases to allow complete algorithm comparisons.

1. Introduction

Automatic face recognition from a single image is still difficult for non-frontal views and complex illumination conditions. To achieve pose and light invariance, 3D information of the object is useful. For this reason, 3D Morphable Models (3DMM) have been introduced a decade ago [7]. They have become a well established technology able to perform various tasks, most important face recognition [8, 18, 11], but also face image analysis [7] (estimating the 3D shape from a single photograph), expression transfer between individuals [6, 17], animation of faces and whole bodies [6, 1], and stimuli generation for psychological experiments [14] to name a few.

A 3DMM consists of a parameterized generative 3D shape, and a parameterized albedo model together with an associated probability density on the model coefficients. A set of shape and albedo coefficients describes a face. Together with projection and illumination parameters a rendering of the face can be generated. Given a face image one can also solve the inverse problem of finding the coefficients which most likely generated the image. Identification and manipulation tasks in coefficient space are trivial, because the generating factors (light, pose, camera, and identity) have been separated. Solving this inverse problem is termed "model fitting", and was introduced for faces in [7] and subsequently refined in [18]. A similar method has also been applied to stereo data [3] and 3D scans [2].

However, the widespread use of 3DMMs has been held back by their difficult construction process, which requires a precise and fast 3D scanner, the scanning of several hundreds of individuals and the computation of dense correspondence between the scans. Numerous face recognition articles acknowledge the fact that 3DMM based face image analysis constitutes the state of the art, but note that the main obstacle resides in the complications of their construction (e.g. [22, 13, 12, 5]). For example, quoting Zhou and Chellappa [23]: "Its only weakness is the requirement of the 3D models". Hence, there is a demand from the face image analysis community for a publicly available 3D Morphable Face Model. The aim of this paper is to fill this gap.

We describe a 3D Morphable Face Model - the *Basel Face Model* (BFM) - that is publicly available (`http://faces.cs.unibas.ch/`). The usage of the

978-0-7695-3718-4/09 $25.00 © 2009 IEEE
DOI 10.1109/AVSS.2009.58

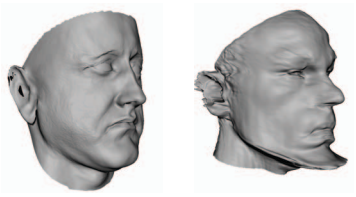
BFM is free for non-commercial purposes. This model not only allows development of 3DMM based image analysis algorithms but will also permit new practices that were impossible before:

First, the 3DMM allows generalization over a variety of different test data sets. Currently, there exist several publicly available face image databases (e.g. CMU-PIE [20], FERET [15], etc.) and databases with unregistered 3D face scans (e.g. UND [9]). Each of the image databases provides gigabytes of face photographs taken at different poses and illumination conditions. These images are either used to train or test new algorithms. Unfortunately, in most cases (e.g. [23, 10]) the same face database is used for both training and testing. Such recognition systems usually have difficulties to generalize from one database to another, because the imaging conditions are too different.

The 3DMM, however, can generate face images at any pose and under any illumination. As mentioned before the face rendering can be used directly in an analysis by synthesis approach [7], to fit the model to images. Or it can be used indirectly to generate training or test images at any imaging condition. Hence, in addition to being a valuable model for use in face analysis it can also be viewed as a *meta-database* which allows the creation of an infinity of accurately labeled synthetic training and testing images.

Registered scans of ten individuals, which are not part of the BFM training set, are provided together with a fixed test set of 270 renderings with pose and light variations. Additionally, synthetic faces can be generated from random model coefficients. This flexibility can also be used to test specific aspects of face image analysis algorithms: For instance, how the departure from the Lambertian assumptions affects the performance of an algorithm (with or without cast shadows and specular lobe, with a sparse set of lights or an environment map, etc.). The pose can be varied continuously such that the extent of the pose generalization of an algorithm can be easily analyzed. For example, test images for stereo algorithms with variable baseline, or for photogrammetric stereo with programmable light directions can also be easily generated. The bottom line is that it is now possible and easy to test the limitations of face image analysis algorithms in terms of pose and illumination generalization.

Secondly, the vast majority of face recognition articles provide results in terms of percentage of rank-1 correct identification or False Acceptance Rate (FAR) / False Rejection Rate (FRR) curves on a standard database. However, these numbers do not fully describe the behavior of an algorithm and leave the reader with open questions such as: Is the algorithm able to



BFM MPI Model

Figure 1. Correspondence artifacts caused by the surface parametrization occur especially for larger values of model coefficients. Rendering the MPI model (right) and the BFM (left) with the same coefficient vector ($c_i \sim \mathcal{N}(0,(2.5)^2)$) shows that the new BFM displays less artifacts.

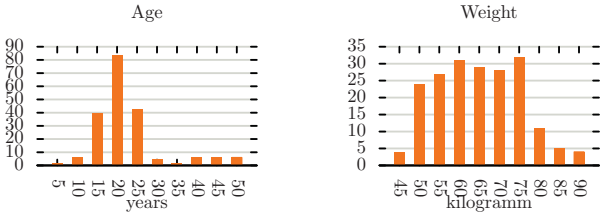


Figure 2. The BFM was trained on 200 individuals (100f/100m). Age (avg 25y) and weight (avg 66kg) are distributed over a large range, but peak at students age.

reproduce the input images and how accurate is the reconstruction? Do the coefficients of different images of the same individual cluster nicely in model space? Can the recognition be improved with a different metric in model space? These and similar questions can only be answered if the model coefficients are made public.

Moreover, numerous face image analysis articles describe algorithms but do not release training data. This hinders reproducible research and fair comparison with other algorithms. To address these two restrictions we provide both the training data set (the BFM) and the model fitting results for several standard image data sets (CMU-PIE, FERET and UND) obtained with the state of the art fitting algorithms [18, 2]. We hope that researchers developing future algorithms based on the BFM will likewise provide the coefficients to enable deeper algorithm comparison and accuracy analysis.

Currently, to the best of our knowledge, there exist only two comparable 3DMMs of faces: the Max-Planck-Institut Tübingen (MPI) MM [7] and the University of South Florida (USF) MM [19]. Compared with these, the BFM is superior in two aspects: Our 3D scanner (ABW-3D) offers a higher resolution and precision in shorter scan time than the Cyberware (TM) scanner (used for the MPI and USF models). This results in a more accurate model. Secondly, a different registration method is used yielding less correspondence artifacts (Fig. 1). Additionally, renderings of fitting results are more realistic with the new model (Sec. 3). We now describe the construction of the

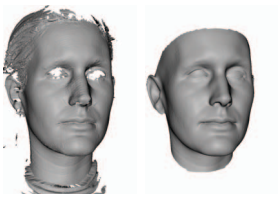


Figure 3. Registration establishes a common parametrization between the original scans (left) and fills in missing data (right).

model followed by baseline experiments against which to compare future computer vision algorithms. We describe results obtained with the BFM in terms of identification experiments and visual quality of the fittings.

2. Model Construction

The construction of a 3DMM requires a training set with a large variety in face shapes and appearance. The training data should be a representative sample of the target population. The training data set for the BFM consists of face scans of 100 female and 100 male persons, most of them Europeans. The age of the persons is between 8 and 62 years with an average of 24.97 years and the weight is between 40 and 123 kilogram with an average of 66.48 kilogram (Fig. 2). Each person was scanned three times with neutral expression, and the most natural looking scan was selected.

2.1. 3D face scanning

Scanning of human faces is a challenging task. To capture natural looking faces, the acquisition time is critical. We use a coded light system with an acquisition time of $\sim 1s$. This leads to more accurate results compared to laser scanners with a acquisition time of around $\sim 15s$. The structured light system was built by *ABW-3D*. It uses a sequence of light patterns which uniquely encode each pixel of the projectors, such that triangulation can be performed even on unstructured regions like the cheeks. To capture the full face the system uses two projectors and three cameras resulting in four depths images. The system captures the facial surface from ear to ear with outstanding precision (Fig. 3, left). The 3D shape of the eyes and hair cannot be captured with our system, due to their reflection properties. The resolution of the geometry measurement is higher than all comparable systems: ABW-3D $\sim$ 200 k, Cyberware $\sim$ 75 k and 3Dmd $\sim$ 20 k.

Simultaneously with each scan, three photos are taken with SLR cameras (sRGB color profile). Three studio flashes with diffuser umbrellas are used to achieve a homogeneous illumination. This ensures a higher color fidelity compared to other systems.

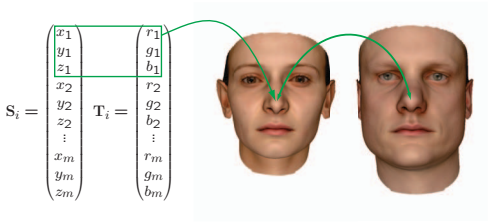


Figure 4. Each entry in the data vectors correspond to the same point on the faces. In this example the first entry corresponds to the tip of the nose

2.2. Registration

To make the raw data usable it needs to be brought in correspondence. This means, that the scans are re-parameterized such that semantically corresponding points (i.e.} the nose tips or eye corners) share the same position in the parametrization domain (Fig. 4). Registration establishes this correspondence for all points of the face, including unstructured regions like the cheek. After bringing the scans into correspondence linear combinations of scans, are again faces. The effect of a bad registration on the model quality can be seen in figure 1. To establish correspondence we use a modified version of the Optimal Step Nonrigid ICP Algorithm [4]. The registration method is applied in the 3D domain on triangulated meshes. It progressively deforms a template towards the measured surface while ensuring a smooth deformation. In addition to establishing correspondence this method fills in missing regions by using a robust distance measure. To improve the model quality we manually added landmarks at the lips, eyebrows and ears.

2.3. Texture Extraction and Inpainting

The face albedo is represented by one color per vertex, which is calculated from the photographs. The information from the three photographs is blended based on the distance from the visibility boundaries and the orientation of the normal relative to the viewing direction. To improve the albedo model we manually removed hair and completed the missing data using diffusion.

2.4. Model

After registration the faces are parameterized as triangular meshes with $m = 53490$ vertices and shared topology. The vertices $(x_j, y_j, z_j)^T \in \mathbb{R}^3$ have an associated color $(r_j, g_j, b_j)^T \in [0, 1]^3$. A face is then repre-

sented by two $3m$ dimensional vectors

$$\mathbf{s} = (x_1, y_1, z_1, \dots x_m, y_m, z_m)^T \quad (1)$$
$$\mathbf{t} = (r_1, g_1, b_1, \dots r_m, g_m, b_m)^T \quad .$$

The BFM assumes independence between shape and texture, constructing two independent Linear Models as described in [7]. A Gaussian distributed is fit to the data using Principle Component Analysis (PCA), resulting in a parametric face model consisting of

$$\mathcal{M}_s = (\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s, \mathbf{U}_s) \text{ and } \mathcal{M}_t = (\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t, \mathbf{U}_t), \quad (2)$$

where $\boldsymbol{\mu}_{\{s,t\}} \in \mathbb{R}^{3m}$ are the mean, $\boldsymbol{\sigma}_{\{s,t\}} \in \mathbb{R}^{n-1}$ the standard deviations and $\mathbf{U}_{\{s,t\}} = [\mathbf{u}_1, \dots \mathbf{u}_n] \in \mathbb{R}^{3m \times n-1}$ are an orthonormal basis of principle components of shape and texture. New faces are generated from the model as linear combinations of the principal components

$$\mathbf{s}(\alpha) = \boldsymbol{\mu}_s + \mathbf{U}_s \operatorname{diag}(\boldsymbol{\sigma}_s)\alpha \quad (3)$$
$$\mathbf{t}(\beta) = \boldsymbol{\mu}_t + \mathbf{U}_t \operatorname{diag}(\boldsymbol{\sigma}_t)\beta$$

The coefficients are independent and normally distributed with unit variance under the assumption of normally distributed training examples and a correct mean estimation.

The data necessary to synthesize faces (i.e. the model data $\mathcal{M}_s$, $\mathcal{M}_t$ and the triangulation) together with test data and test results are available at our web site. We provide:

- Shape and albedo PCA model $\mathcal{M}_s$, $\mathcal{M}_t$ $(\mathbf{U}, \boldsymbol{\sigma}, \boldsymbol{\mu})$ computed from the 200 face scans.
- Ten additional registered 3D face scans together with 2D renderings of the scans with light and pose variation.
- Vertex indices of MPEG and Farkas points together with 2D projections within the above renderings.
- Model coefficients obtained by the fitting of FERET and CMU-PIE together with 2D renderings of the reconstructed shapes.
- Model coefficients obtained by the fitting of the unregistered 3D shapes of the UND database.
- Mask with four segments (Fig. 5) that is used in the identification experiments (Sec. 3.1).
- Matlab code for own experiments, e.g. generation of random faces.

3. Experiments

With the BFM a standard training set for face recognition algorithms is provided to the public. Together with test sets such as FERET, CMU-PIE and UND, this allows for a fair, data independent comparison of face identification algorithms. We demonstrate that it is not necessary to train a model specifically for each test database by splitting the database into test and training set, instead it is possible to apply the same model to all data sets. With our face identification experiments, we show that the BFM is general enough to be used with different 2D and 3D sensors.

3.1. Face Identification on 2D images

To demonstrate the quality of the presented model we compare it with the MPI model by 2D identification experiments (CMU-PIE/FERET) [18]. To allow a detailed and transparent analysis of the results and to enable other researchers to compare their results with ours, we provide the reconstructed 3D shapes and textures, coefficients and rendered faces for each test image. None of the individuals in the test sets is part of the training data for the Morphable Model. The test sets cover a large ethnic variety.

Test Set 1: FERET Subset The subset of the FERET data set [16], consists of 194 individuals across 9 poses at constant lighting condition except the frontal view taken under a different illumination condition. In the FERET nomenclature these images correspond to the series *ba* through *bk*. We omitted the images *bj* as the subjects present a smile and our model can only represent neutral expressions.

Test Set 2: CMU-PIE Subset The subset of the CMU-PIE data set [20], consists of 68 individuals (28 wearing glasses) at 3 poses (frontal, side and profile) under illumination from 21 different directions and ambient light only. To perform the identification we fit the BFM to the images of the test sets. In the fitting three error terms based on landmarks, the contour and the shading are optimized. To extend the flexibility, four facial regions (eyes, nose, mouth and the rest Fig. 5) are fitted separately and combined later by blending them together. The obtained shape and albedo model parameters for the global fitting $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ and for the facial segments $\boldsymbol{\alpha}_1, \dots \boldsymbol{\alpha}_4$ and $\boldsymbol{\beta}_1, \dots \boldsymbol{\beta}_4$ represent the identity in the model space. These parameters are stacked together into one identity vector

$$\boldsymbol{c} = \left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \dots \boldsymbol{\alpha}_4, \boldsymbol{\beta}_4\right). \quad (4)$$

Similarity of two scans is measured by the angle between their identity vectors.

Table 1 and 2 list the percentages of correct rank 1 identification obtained on the CMU-PIE and the FERET subset, respectively. The overall identification rate with the BFM model is better than the MPI results. For CMU-PIE 91.3% (vs. 89.4%) and for FERET 95.8% (vs. 92.4%) were obtained. As in previous experiments the best results are obtained for frontal views.

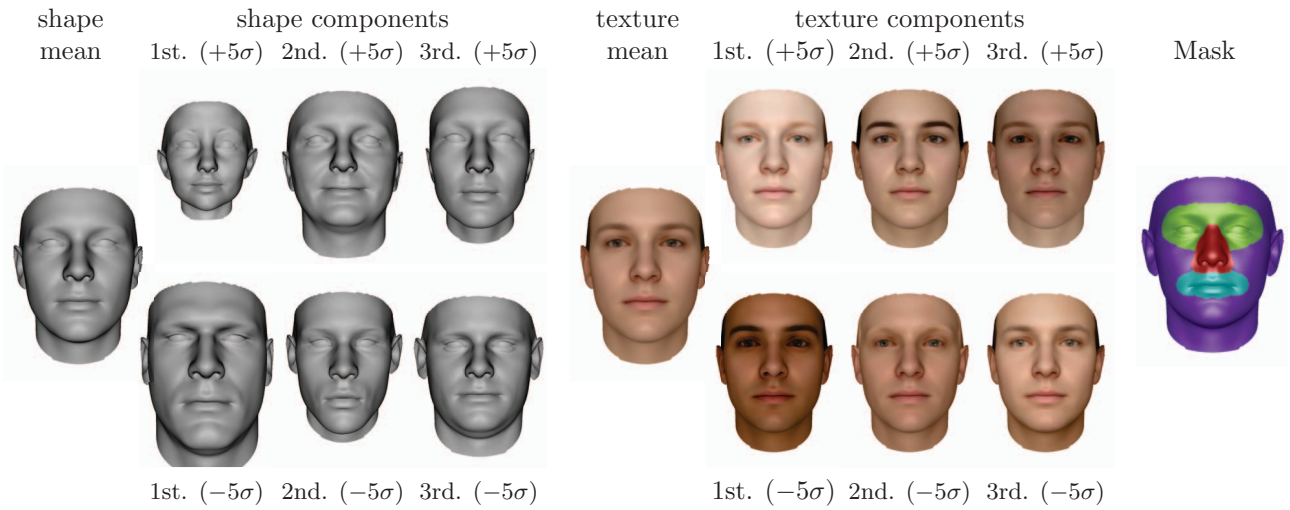


Figure 5. The mean together with the first three principle components of the shape (left) and texture (right) PCA model. Shown is the mean shape resp. texture plus/minus five standard deviations σ. Mask with the four manually chosen segments (eyes, nose, mouth and rest) used in the fitting to extend the flexibility.

Gallery / Probe	front	side	profile	mean
front	98.9 %	96.1 %	75.7 %	90.2 %
side	96.9 %	99.9 %	87.8 %	94.9 %
profile	79.0 %	89.0 %	98.3 %	88.8 %
mean	91.6 %	95.0 %	87.3 %	91.3 %

Table 1. Rank 1 identification results obtained on a CMU-PIE subset. The mean identification rate is 91.3%. With the former MPI model a identification rate of 89.4% was obtained.

Gallery / Probe	Pose Φ	Identification rate
bb	38.9°	97.4 %
bc	27.4°	99.5 %
bd	18.9°	100.0 %
be	11.2°	Gallery
ba	1.1°	99.0 %
bf	-7.1°	99.5 %
bg	-16.3°	97.9 %
bh	-26.5°	94.8 %
bi	-37.9°	83.0 %
bk	0.1°	90.7 %
mean		95.8 %

Table 2. Rank 1 identification results obtained on a FERET subset. The mean identification rate is 95.8%. With the former MPI model a identification rate of 92.4% was obtained.

Compared with the MPI, the visual quality of the BFM fitting results (Fig. 6) is much better since the overfitting in the texture reconstruction has been reduced.

Figure 6. Exemplary fitting result for CMU-PIE with BFM Face Model. Left the original image, middle row the fitting result rendered into the image and right the resulting 3D model.

3.2. Face Identification on 3D scans

For the 3D identification experiments, we fit the BFM to shape data without using the texture. The fitting algorithm [2] is a variant of the nonrigid ICP work in [4]. We initialize the fitting by locating the tip of the nose with the method of [21]. As test set we use the UND database [9] that consists of 953 unregistered 3D scans, with one to eight scans per subject. As for the 2D experiments, we measure the similarity between two faces as the angle between their coefficients in Mahalanobis space. The recognition performance for different distance thresholds is shown in Fig. 7.

4. Conclusion

We presented a publicly available 3D Morphable Model of faces, together with basic experiments. The model addresses the lack of universal training data for

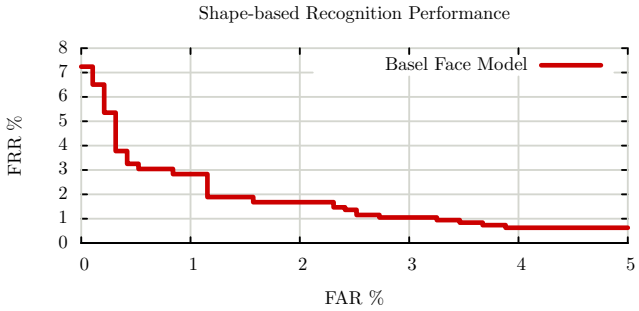


Figure 7. Identification results obtained on the UND database of unregistered 3D shapes. Varying the distance threshold leads to varying false acceptance rates (FAR) and false rejection rates (FRR).

face recognition. Although many test data sets exist, there are no standard training data sets. The reason is that such a training set must be general enough to represent all appearance of faces under any pose and illumination condition. Since we believe that a standard training set is necessary for a fair comparison, we make the model publicly available. Due to its 3D structure it can be used indirectly to generate images with any kind of pose and light variation or directly for 2D and 3D face recognition. It is planned to extend the data collection further and provide it on the web site. We also plan to provide results of experiments and renderings with more complex illumination models. Using these standardized training and test sets makes it possible for researchers to focus on the comparison of algorithms independent of the data. We trained our previously published face recognition algorithm and provide detailed results (parameters for the model). Other researchers are invited to use the same standardized test set and present the results on our web site (`http://faces.cs.unibas.ch/`).

4.1. Acknowledgment

This work was funded in part by the Swiss National Science Foundation (200021-103814, NCCR CO-ME 5005-66380) and Microsoft Research.

References

[1] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *SIGGRAPH '03*.

[2] B. Amberg, R. Knothe, and T. Vetter. Expression invariant 3D face recognition with a morphable model. In *FG'08*, 2008.

[3] B. Amberg, S. Romdhani, A. Fitzgibbon, A. Blake, and T. Vetter. Accurate surface extraction using model based stereo. In *ICCV '07*, 2007.

[4] B. Amberg, S. Romdhani, and T. Vetter. Optimal step nonrigid ICP algorithms for surface registration. In *CVPR '07*.

[5] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. *CVPR '05*, 1, 2005.

[6] V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. In *EuroGraphics*, 2003.

[7] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *SIGGRAPH '99*.

[8] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *PAMI*, 25(9), 2003.

[9] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multimodal 2D+3D face biometrics. *PAMI*, 27(4), 2005.

[10] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *PAMI*, 26(4):449–465, 2004.

[11] B. Heisele, T. Serre, and T. Poggio. A component-based framework for face detection and identification. *IJCV*, 74(2), 2007.

[12] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang. Automatic 3D reconstruction for face recognition. *fg*, 0, 2004.

[13] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. *CVPR*, 01, 2003.

[14] D. A. Leopold, A. J. O'Toole, T. Vetter, and V. Blanz. Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4(1):89–94, 2001.

[15] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *PAMI*, 22(10), 2000.

[16] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *PAMI*, 22, 2000.

[17] S. Romdhani. *Face Image Analysis Using a Multiple Features Fitting Strategy*. PhD thesis, 2005.

[18] S. Romdhani and T. Vetter. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *CVPR '05*.

[19] S. Sarkar. USF HumanID 3D face dataset, 2005.

[20] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *PAMI*, 25(12), 2003.

[21] F. B. ter Haar and R. C. Veltkamp. A 3D Face Matching Framework. In *Shape Modeling Int. '08*.

[22] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4), 2003.

[23] S. Zhou and R. Chellappa. Illuminating light field: image-based face recognition across illuminations and poses. *FG*, 2004.