3D-MAM: 3D Morphable Appearance Model for Efficient Fine Head Pose Estimation from Still Images

Markus Storer, Martin Urschler and Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology
Inffeldgasse 16/II, 8010 Graz, Austria
{storer, urschler, bischof}@icg.tugraz.at

Abstract

Identity-invariant estimation of head pose from still images is a challenging task due to the high variability of facial appearance. We present a novel 3D head pose estimation approach, which utilizes the flexibility and expressibility of a dense generative 3D facial model in combination with a very fast fitting algorithm. The efficiency of the head pose estimation is obtained by a 2D synthesis of the facial input image. This optimization procedure drives the appearance and pose of the 3D facial model. In contrast to many other approaches we are specifically interested in the more difficult task of head pose estimation from still images, instead of tracking faces in image sequences. We evaluate our approach on two publicly available databases (FacePix and USF HumanID) and compare our method to the 3D morphable model and other state of the art approaches in terms of accuracy and speed.

1. Introduction

Automatically determining the head orientation from images has been a challenging task for the computer science community for decades [19]. The head can be assumed to be modeled as a rigid object, and therefore the pose of the head can be characterized by *pitch*, *roll* and *yaw* angles as illustrated in Figure 1. Head pose is also essential for understanding the eye's gaze, i.e., to determine the viewing direction of a person. In [15] it is shown, that the gaze direction is a combination of head pose and eye's gaze.

Many exciting and important application areas for head pose estimation emerged in the last decade. One major interest is human-computer interaction to determine the gaze as well as interpreting head gesturing, i.e., the meaning of head movements like nodding. This enables very direct means to interact with virtual worlds, especially in the computer gaming industry.

A second area of application is automotive safety. To avoid vehicle collisions, the driver's head is monitored to recognize driver distraction and inattention. A 3D head tracking approach for a driver assistance system is presented in [18].

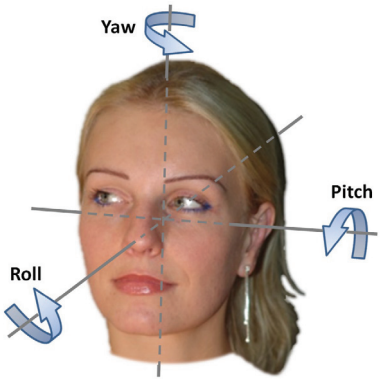


Figure 1. The three degrees of freedom of a human head described by *pitch*, *roll* and *yaw* angles.

Biometrics also utilizes the important task of head pose estimation. One direction of impact is the rectification of non-frontal facial images to the frontal pose to improve the accuracy of face recognition [5]. Recently, extraordinary attempts to person surveillance for far-field distance views were announced.

Even though there is a broad area of applications and a high demand for accurate systems, research on identity-invariant head pose estimation shows fewer evaluated systems and generic solutions compared to face detection and face recognition.

Related Work

Murphy-Chutorian and Trivedi [19] give a recent and extensive survey on head pose estimation in computer vision. They arrange the methods published over the last 14 years in several categories: appearance template based methods, detector arrays, manifold embedding, flexible models, geometric methods, tracking methods and hybrid methods.

2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops
978-1-4244-4441-0/09/$25.00 ©2009 IEEE

We focus on flexible model based head pose estimation to overcome problems of other approaches of being not adaptive to unseen facial images. Our proposed approach is related to the well known Active Appearance Model (AAM) of Cootes et al. [10]. The AAM is a widely used method for model based vision showing excellent results in a variety of applications. As a generative model it describes the statistical variation in shape and texture of a training set representing an object. AAM model fitting is performed in a gradient descent optimization scheme, where the cost function is defined as the L2 norm of the intensity differences. For optimization the Jacobian is often approximated either by a regression to learn the dependency between model parameter updates and intensity differences [10] or alternatively by a canonical correlation analysis [13]. The fitting procedure is very fast, but one major drawback is its non-robustness to viewpoint changes. In the case of facial images, the AAM fitting is not appropriate for adopting to faces which exhibit pose variations. Cootes et al. [9] extend their AAM approach for multi pose fitting by combining a small number of 2D AAM models.

Blanz and Vetter [6], [7] overcome the drawbacks of the 2D AAM by creating a 3D Morphable Model (3DMM). The 3DMM is a statistical model of shape and texture based on data acquired from a laser scanner. The approach shows amazing image synthesis results but the fitting procedure is computationally very expensive. One attempt to fit a 3DMM more efficiently was proposed in [20], but the fitting of one facial image still takes several minutes.

In [11] an extension to the classical AAM approach is proposed. They build a 3D anthropometric muscle based active appearance model using a generic 3D face shape. They adopt the 3D shape model so that the projected 3D vertices best fit to a facial 2D image. Using several adoptions to different facial images, these obtained shapes can be taken to create a shape eigenspace using PCA. According to the foregoing shape adoption the texture of the 2D facial images is warped back onto the 3D model, i.e., they get several textures to create a texture eigenspace. This generation of training data is a cumbersome work. The model fitting is similar to the original AAM fitting procedure. They apply their approach for head tracking and facial expression recovery [12].

Xiao et al. [23] propose a real time combined 2D+3D AAM to fit 3D shapes to images. They also investigate, that the 2D AAM could generate illegal model instances, which do not have a physical counterpart. They show how to constrain an AAM by incorporating 3D shape information so that the AAM can only generate valid model instances. Their work focuses on tracking applications and experiments show excellent performance, however, in their setup the generative model is always built from the same person that is tracked later. Chen and Wang [8] describe a similar model for human-robot interaction.

In the domain of medical image analysis, 3D Active Appearance Models are used with great success for segmentation [17], [21], [3] and modeling of shape or pathological variations of a population. These approaches are specifically targeted to 3D volumetric data, where the notion of efficiency becomes even more important due to the increased dimensionality.

Due to the two major drawbacks of model based vision, the non-robustness to viewpoint changes and the inefficient fitting procedure, we present a novel 3D Morphable Appearance Model (3D-MAM) for head pose estimation, which utilizes the flexibility and expressibility of a dense generative 3D facial model in combination with a very fast fitting algorithm. The efficiency of the head pose estimation is reached by a 2D synthesis of the facial input image. This optimization procedure drives the appearance and pose of the 3D facial model. In contrast to many other approaches we are specifically interested in the more difficult task of head pose estimation from still images, instead of tracking faces in image sequences. Much effort is undertaken to build a fair evaluation scheme for our approach. That is, the data for building and training of our 3D-MAM was totally independent of the datasets used for the evaluations.

The paper is structured as follows: In Section 2 we introduce and discuss our 3D-MAM approach in terms of model building and model fitting. In Section 3 we present our results by evaluating the head pose estimation accuracy and speed of our approach on two publicly available databases (FacePix and USF HumanID) and compare our method to state of the art approaches. Finally, we discuss our findings and conclude our work in Section 4.

2. 3D Morphable Appearance Model

We build a generative 3D Morphable Appearance Model (3D-MAM) based on registered laser scans of human heads. The advantage of a dense 3D model is its flexibility to express generic faces and the ability to adopt to non-rigid deformations exhibited by faces. Only with a dense model, normal vectors of the surface can be computed and therefore depth can be estimated correctly. This resembles human perception of 3D objects. A human is only able to estimate depth correctly in the presence of shadowed surfaces.

To overcome the slow fitting performance exhibited by many approaches using dense 3D models, we perform our fitting step in 2D (Section 2.2). To sum up, we combine the advantages of dense 3D models and the very efficient fitting speed gained in the 2D domain.

2.1. 3D Morphable Appearance Model

We utilize laser scans of human heads and register them using a nonrigid Iterative Closest Point (ICP) algorithm [1].

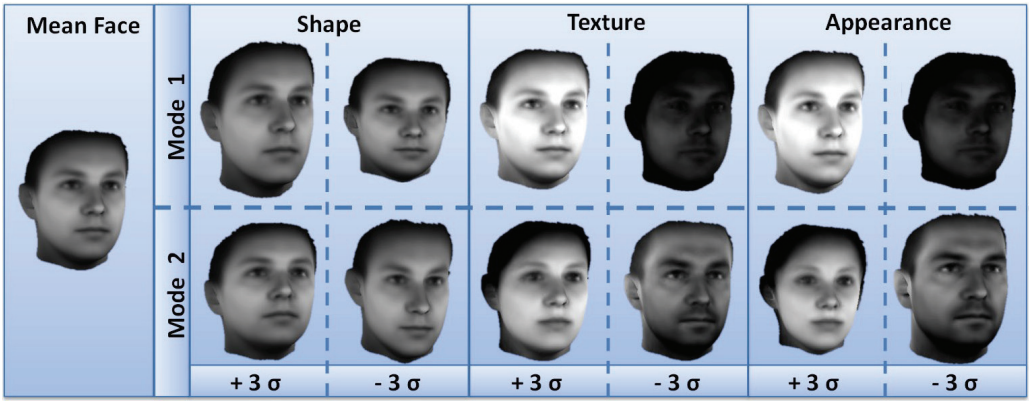


Figure 2. 3D Morphable Appearance Model. Effect of varying the shape, texture and appearance parameters of the first and second mode by ± 3 standard deviations.

The registered 3D laser scans form the basis for building a parameterized generative 3D model, which will later be used for head pose estimation.

2.1.1 3D Head Laser-Scans

The 3D facial model is built from 350 facial laser scans. The scans were acquired by a *Cyberware*™ laser scanner, which captures the 3D information (vertices) in cylinder coordinates with radius $r(h, \phi)$, 512 equally-spaced angles ϕ and 512 equally-spaced vertical steps h. Additionally, the RGB-color information $R(h, \phi), G(h, \phi), B(h, \phi)$ is recorded for each vertex.

The obtained 3D database consists of 350 different subjects exhibiting a wide variability in race, gender and age. Most of the subjects show neutral facial expression. The raw facial scans have to be post-processed to remove certain parts of the scans, e.g., hair and shoulders. Often those areas can not be captured very well, because of the fine structure of the hair or due to self occlusions. More specifically, the scans are cut vertically behind the ears and cut horizontally to remove the hair and shoulders. Additionally, laser scanning artifacts, like holes or spikes, are removed manually.

We reduce the amount of vertices from about 100,000 to 10,000 for the purpose of decreasing the computational effort in the model fitting procedure, see Section 2.2. This simplification of the 3D data at regions with little details is performed by a structure preserving surface simplification approach [14].

To build a generative model (Section 2.1.2), the individual laser scans have to be non-rigidly registered. Blanz and Vetter [6] register their data using a modified gradient-based optical flow algorithm. We use the more sophisticated *optimal step nonrigid ICP* method [1] to establish correspondence between a pair of 3D scans. This method's runtime is slower compared to optical flow, but yields more robust registration results, e.g., filling of holes due to missing data.

2.1.2 Model Building

We create a statistical model of shape, texture and appearance similar to [10] with the difference of using 3D laser scanner data, instead of annotated 2D images. The laser scanner data have to be registered (Section 2.1.1) to allow the construction of a generative model utilizing Principal Component Analysis (PCA).

The registered 3D shapes are composed of the 3D positions of the vertices, and the texture consists of the intensity values of the vertices. Taking N training shape and texture tuples with sample mean $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ correspondingly, they are used to build statistical models of shape and texture by using PCA

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{U}_s \mathbf{p}_s \quad (1)$$

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{U}_t \mathbf{p}_t \quad (2)$$

Here $\mathbf{U}_s$ and $\mathbf{U}_t$ are shape- and texture eigenvectors, which describe the modes of variation derived from the training set. By adjusting the parameters $\mathbf{p}_s$ and $\mathbf{p}_t$, new instances of shape $\mathbf{s}$ and texture $\mathbf{t}$ can be generated.

To remove correlations between shape and texture variations, we apply a further PCA to the data. The shape- and texture eigenspaces are coupled through

$$\begin{pmatrix} \mathbf{W}_s \mathbf{p}_s \\ \mathbf{p}_t \end{pmatrix} = \mathbf{U}_c \mathbf{c} \quad (3)$$

to get the statistical model of appearance (combined model), where $\mathbf{W_s}$ is a diagonal scaling matrix to compensate for the different measure units of shape and texture. $\mathbf{W_s}$ is defined as the ratio of the total intensity variation to the total shape variation [10]. This appearance model is controlled

by parameter $\mathbf{c}$ to obtain new instances of facial shape and texture

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{U}_s \mathbf{W}_s^{-1} \mathbf{U}_{c,s} \mathbf{c} \quad (4)$$

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{U}_t \mathbf{U}_{c,t} \mathbf{c} \quad (5)$$

$$\mathbf{U}_c = \begin{pmatrix} \mathbf{U}_{c,s} \\ \mathbf{U}_{c,t} \end{pmatrix} \quad (6)$$

Figure 2 shows the effect of varying the shape, texture and appearance parameters of the first and second mode by ± 3 standard deviations obtained from the training set.

The pose of the appearance model in 3D can be altered by six degrees of freedom (DOF), i.e., three angles of rotation and the three directions of translation. We map the 3D points to the 2D image coordinates by a weak perspective projection. That is, the rendering of the 3D model to the image plane is given by the two rotation angles θ_{pitch} and θ_{yaw} and the two translations u_x and u_y in the image plane. For linearity, the *scaling* and the *roll* angle is represented as $sr_x = (scale \cos\theta_{roll} - 1)$ and $sr_y = scale \sin\theta_{roll}$. The concatenation of those single parameters yields the pose parameter vector $\mathbf{p}_{pose} = (sr_x,\ sr_y,\ \theta_{pitch},\ \theta_{yaw},\ u_x,\ u_y)$.

2.2. Model Fitting

The model is fitted iteratively in an analysis-by-synthesis approach, see Figure 4. A direct optimization of the appearance model parameters $\mathbf{c}$ and pose parameters $\mathbf{p}_{pose}$ is computationally not feasible for real time applications. Hence, we precompute a parameter update matrix [10], which will be used to incrementally update the parameters $\mathbf{p} = (\mathbf{c},\ \mathbf{p}_{pose})$ in the fitting stage. The patch used for synthesizing a new input image is restricted to an area of the head, where most of the vertices are visible for slight pose variations (Figure 3).

Starting from the 3D mean shape, we project the positions of the vertices from the 3D patch to the 2D input image using a weak perspective projection. The texture from the input image underlying the projected points in 2D is then warped to a shape free (mean shape) representation. Simultaneously, the texture of the model patch is rendered also to the same shape free representation. Now, the texture from the input image $\mathbf{t}_s$ and the rendered texture from the model patch $\mathbf{t}_m$ (both in the same shape free representation) can be subtracted to get a residual image

$$\mathbf{r}(\mathbf{p}) = \mathbf{t}_s - \mathbf{t}_m. \quad (7)$$

A first order Taylor expansion of (7) gives

$$\mathbf{r}(\mathbf{p} + \delta\mathbf{p}) = \mathbf{r}(\mathbf{p}) + \frac{\partial \mathbf{r}}{\partial \mathbf{p}} \delta\mathbf{p} \quad (8)$$

In the fitting stage, $\|\mathbf{r}(\mathbf{p} + \delta\mathbf{p})\|^2$ is minimized by computing

$$\delta\mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p}) \quad \text{where} \quad \mathbf{R} = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{p}}^T \frac{\partial \mathbf{r}}{\partial \mathbf{p}} \right)^{-1} \frac{\partial \mathbf{r}}{\partial \mathbf{p}}^T \quad (9)$$

In a direct optimization scheme, the Jacobian matrix $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ has to be recomputed in every iteration step yielding poor runtime performance. Hence, the parameter update matrix $\mathbf{R}$ is assumed to be fixed and can therefore be precomputed by numeric differentiation. The numeric differentiation is accomplished through a perturbation scheme, i.e., each parameter is displaced from a known optimal value. More details can be found in [10].

In the fitting stage, texture residuals are computed in the same way as in the training stage of the parameter update matrix $\mathbf{R}$. This residual in combination with the update matrix gives the parameter update $\delta\mathbf{p}$ for driving the parameters $\mathbf{p}$ of the 3D model. That is, the appearance and pose of the 3D model is iteratively fitted to the 2D input image. The whole fitting procedure is illustrated in Figure 4.

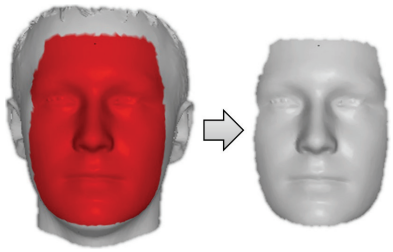


Figure 3. Patch extracted from the whole head used for synthesizing a new input image.

3. Experimental Results

We build a 3D-MAM (Section 2.1.2) and keep 90% of the eigenvalue energy spectrum for the shape, 85% for the texture and 90% of the appearance variation to represent our compact model. We precompute the parameter update matrix (Section 2.2) with a resolution of the fitting patch of 60x80 pixels.

The head pose estimation is evaluated on two different publicly available datasets (USF Human ID 3D face database and FacePix). Those data sets are independent of the data used for model building.

The USF Human ID 3D face database [22], [6] consists of 136 individuals, which are recorded by a *Cyberware*™ laser scanner. Each facial model is composed of more than 90,000 vertices and 180,000 triangles. Images of the individuals can be rendered in arbitrary pose. Those rendered images with arbitrary textured background added are used as test images for our approach. First, we evaluated the head pose estimation capability of our approach using the rendered images of the first 50 individuals in the database

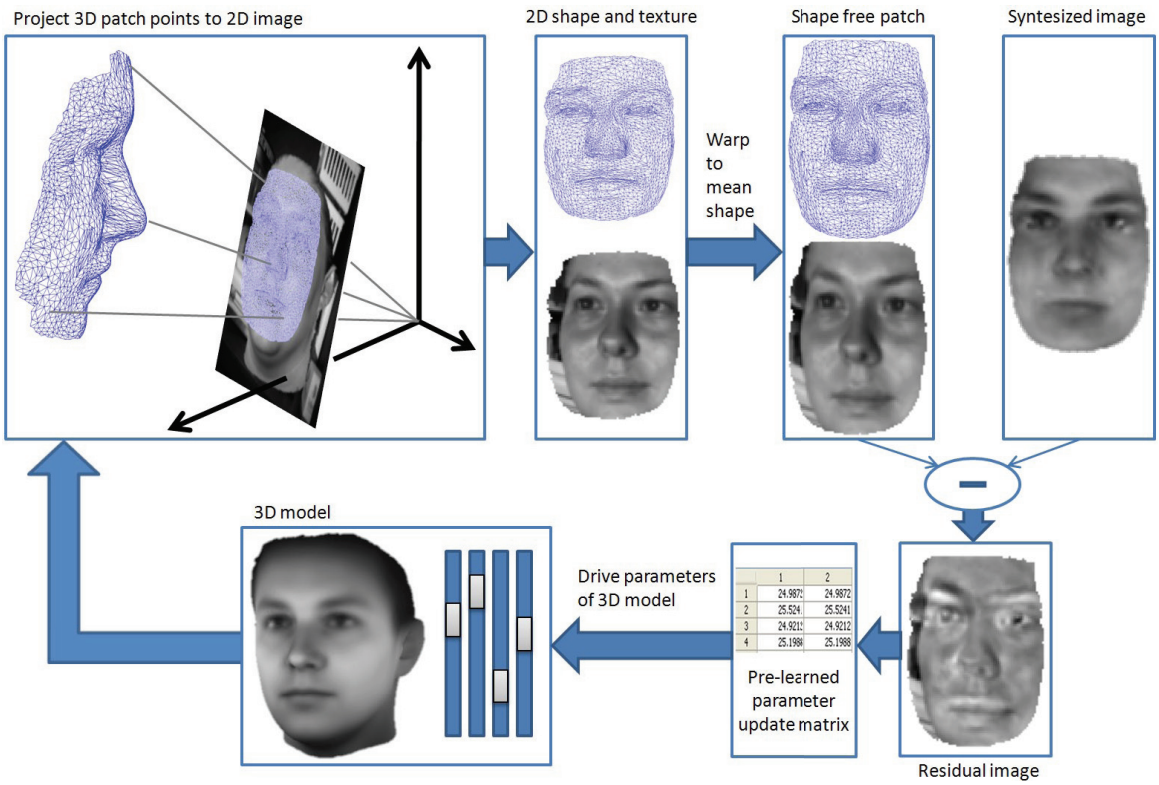
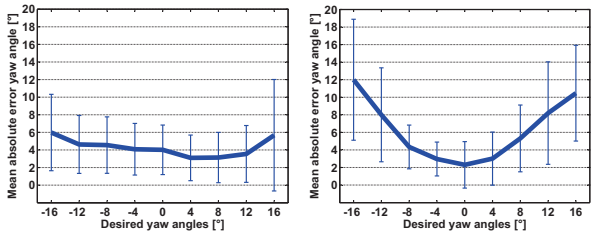


Figure 4. Fitting workflow

by altering only the yaw angles from -16° to +16° in steps of 4° while fixing the roll- and pitch angle of the rendered test images to zero. The 3D-MAM's 2D starting position is roughly initialized manually. In the future, this initialization will be done by an automatic face- and facial feature detection stage. Figure 5a presents the mean and standard deviation of the absolute angular error for the single yaw rotations. Table 1a summarizes the error measures for the whole range of rotations.



(a) (b)

Figure 5. Mean and standard deviation of the absolute yaw angular error for the (a) USF Human ID 3D face database and (b) FacePix database.

We extend the previous experiment by altering the yaw- and pitch angle by [-16° 0° +16°] and [-8° 0° +8°] correspondingly. These nine angle combinations and the 50 individuals per combination yields 450 3D-MAM fitting runs. The mean absolute angular error of the angle combinations is shown in Figure 6(a,b). The error measures are summarized in Table 2a. We compare our pose estimation results with the well known 3D-Morphable Model [6]. We build a 3DMM based on the same laser scanner data as used for our 3D-MAM (Section 2.1.1). To speed up the 3DMM fitting procedure, we use only the first 10 shape- and texture modes, because we want to estimate head pose and do not want to synthesize the test image in every detail. The results

Table 1. Mean, standard deviation, median, upper- and lower quartil of the absolute yaw angular error by only altering the yaw angle for the (a) USF Human ID 3D face database and (b) FacePix database. The results are compared to the 3DMM [6].

(a)

	Absolute Error [°]				
USF Human ID	mean	std	median	$Q_{.25}$	$Q_{.75}$
Our Approach	4.30	3.51	3.35	1.71	6.08

(b)

	Absolute Error [°]				
FacePix	mean	std	median	$Q_{.25}$	$Q_{.75}$
Our Approach	6.29	4.15	4.62	2.30	8.57
3DMM	4.89	3.15	4.48	2.15	7.06

for the 3DMM are shown in Figure 6(c,d) and summarized in Table 2a. The 3DMM exhibits slightly better head pose estimation results at the cost of a much higher runtime per facial fit, see Table 2b.

The second database, CUbiC FacePix(30) database [4], [16], consists of 30 individuals. For each individual, three sets of images are available. The first set contains images taken from the individuals's right to left (only yaw angle is annotated), in one degree increments. The second- and third set is targeted to non-uniform lighting experiments. We are specifically interested in the first set for our pose estimation experiments. We take those images annotated by -16° to +16° in steps of 4° . The mean and standard deviation of the absolute angular error is shown in Figure 5b. Table 1b summarizes the error measures for the whole range of rotations and compares the results to the 3DMM approach. In [2], they also conducted several experiments on the FacePix(30) database using manifold embedding methods. They show better results, ranging from a mean absolute error of 1.44° to 10.41° , but they performed the training and testing on the same database in a cross-validation scheme, which leaves serious doubts about the general applicability of their method on unseen data. Second, they are limited to only estimate the yaw angle of a given test image.

Our model fitting strategy is similar to the AAM approach [10], leading to an excellent runtime performance. The average runtime[1] for our approach is 3.2s at an average number of iterations of 14. We have a comparable implementation of an AAM in C++, which takes about 15ms per facial fit. If we add about 5ms per iteration for a rendering of a facial image using OpenGL, and taking the average number of iterations into account, we would get an estimated average runtime of 85ms per facial fit with an implementation of 3D-MAM in C++. This runtime would enable the usage of our approach for real-time head pose estimation.

Figure 7 shows frames from a 3D-MAM facial fit starting with the mean model. During fitting the patch synthesizes the face and adjusts the 3D model in appearance and pose.

4. Conclusion

Two major shortcomings of existing model based head pose estimation approaches, the non-robustness to viewpoint changes and the inefficient fitting procedure, motivated us to generate a 3D-MAM. It utilizes the flexibility of a dense 3D facial model combined with a very fast fitting algorithm in 2D. In the experiments, we show the applicability of our approach for head pose estimation on two publicly available databases (USF HumanID and FacePix). We compare our results to state of the art head pose estimation algorithms in terms of accuracy and speed.

In the future, we will extend our approach to multi-view fitting to cover a larger range of head pose.

[1]The runtimes are measured in *MATLAB*™ using an Intel Core 2 Duo processor running at 2.4GHz. The resolution of the images is 60x80 pixels.

Table 2. Evaluations for the USF Human ID 3D face database. (a) Mean, standard deviation, median, upper- and lower quartil of the absolute yaw and pitch angular error by altering the yaw- and pitch angle. (b) Average runtime[1] per facial fit. The results are compared to the 3DMM [6].

(a)

	Absolute Error [°]									
	Yaw					Pitch				
USF Human ID	mean	std	median	$Q_{.25}$	$Q_{.75}$	mean	std	median	$Q_{.25}$	$Q_{.75}$
Our Approach	5.78	4.22	4.86	2.57	7.66	5.89	4.68	4.76	2.32	8.23
3DMM	3.90	3.31	2.81	1.55	5.21	5.14	3.66	4.08	2.11	7.55

(b)

	Average Runtime [s]
Our Approach	3.2
3DMM	33.5

Acknowledgements

This work has been funded by the Biometrics Center of Siemens IT Solutions and Services, Siemens Austria. In addition, we want to thank Christoph Gratl for providing the evaluations for the 3DMM.

References

[1] B. Amberg, S. Romdhani, and T. Vetter. Optimal step nonrigid ICP algorithms for surface registration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[2] V. Balasubramanian, Y. Jieping, and S. Panchanathan. Biased manifold embedding: A framework for person-independent head pose estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, June 2007.

[3] R. Beichel, H. Bischof, F. Leberl, and M. Sonka. Robust active appearance models and their application to medical image analysis. *IEEE Transactions on Medical Imaging*, 24(9):1151–1169, Sept. 2005.

[4] J. Black, M. Gargesha, K. Kahol, P. Kuchi, and S. Panchanathan. A framework for performance evaluation of face recognition algorithms. In *Internet Multimedia Systems II*, July 2002.

[5] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter. Face recognition based on frontal views generated from non-frontal images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 454– 461, June 2005.

[6] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*, pages 187–194, 1999.

[7] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, September 2003.

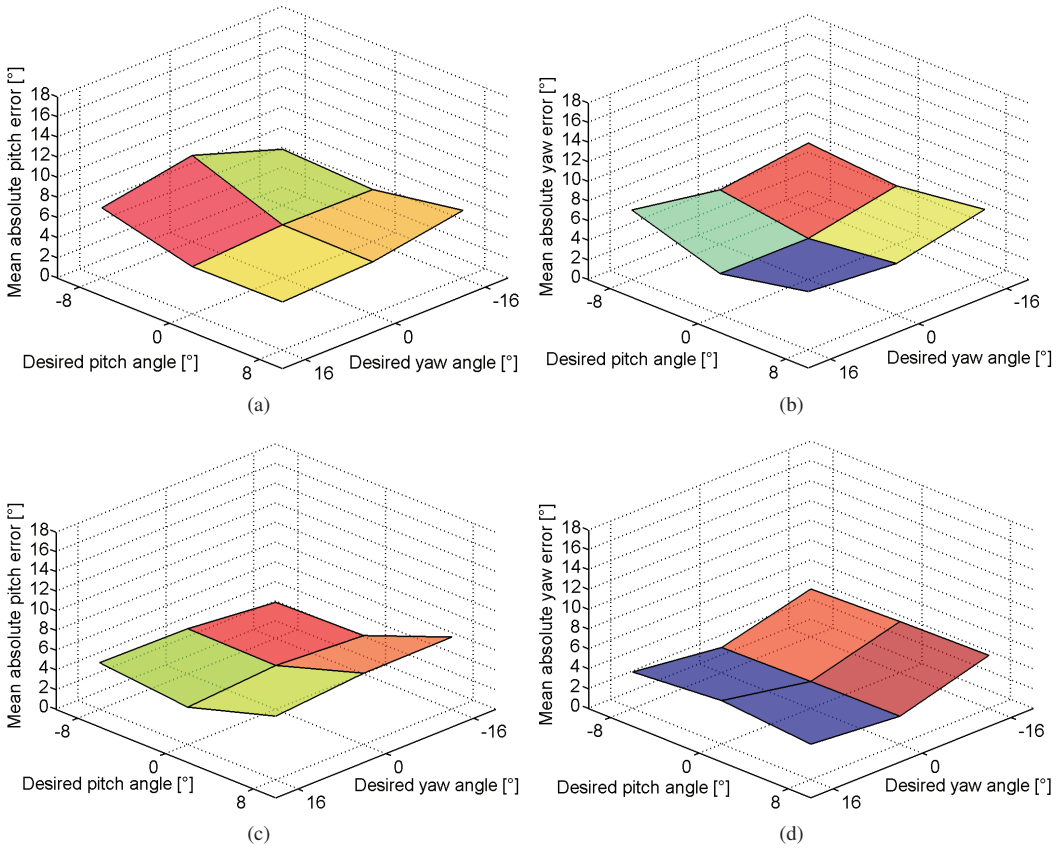


Figure 6. Mean absolute error for pitch- and yaw angle on the USF Human ID 3D face database by altering pitch- and yaw angles for the generation of the test images. (a,b) Our approach (c,d) 3DMM

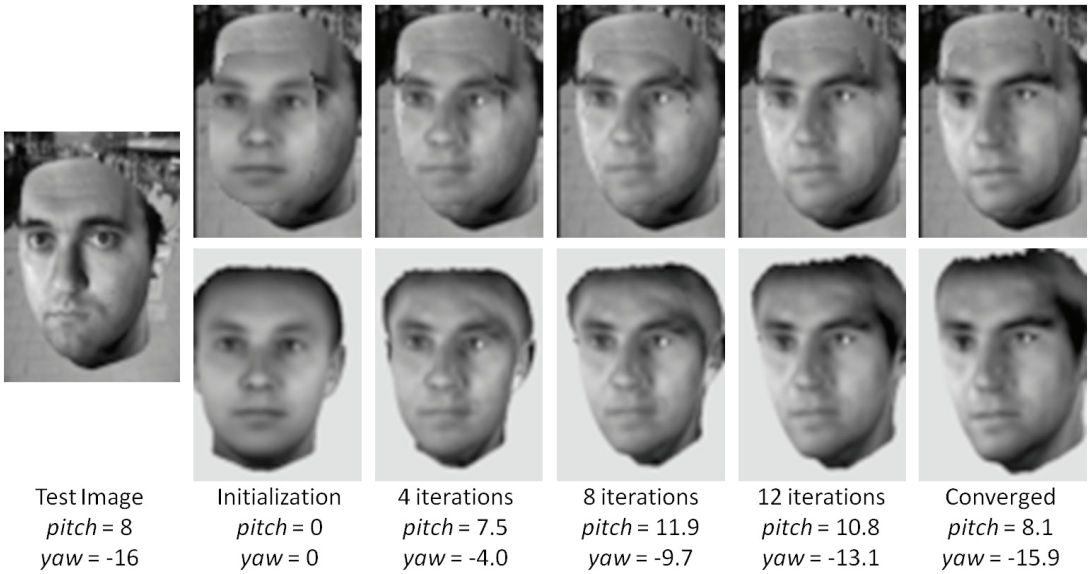


Figure 7. Analysis-by-Synthesis fitting example. The lower row shows the adjustment of the 3D model. In the upper row, the corresponding fitting patch is illustrated. This model fitting converges after 17 iterations.

[8] C.-W. Chen and C.-C. Wang. 3D active appearance model for aligning faces in 2D images. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3133–3139, September 2008.

[9] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models. In *Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 227–232, 2000.

[10] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[11] M. D. Cordea and E. M. Petriu. A 3-D anthropometric-muscle-based active appearance model. *IEEE Trans. on Instrumentation and Measurement*, 55(1):91–98, February 2006.

[12] M. D. Cordea, E. M. Petriu, and D. C. Petriu. Three-dimensional head tracking and facial expression recovery using an anthropometric muscle-based active appearance model. *IEEE Trans. on Instrumentation and Measurement*, 57(8):1578–1588, August 2008.

[13] R. Donner, M. Reiter, G. Langs, P. Peloschek, and H. Bischof. Fast active appearance model search using canonical correlation analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1690–1694, 2006.

[14] M. Garland and P. S. Heckbert. Surface simplification using quadric error metrics. In *Proc. 24th Conference on Computer Graphics and Interactive Techniques*, pages 209–216, 1997.

[15] S. R. H. Langton, H. Honeyman, and E. Tessler. The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception and Psychophysics*, 66(5):752–771, 2004.

[16] G. Little, S. Krishna, J. Black, and S. Panchanathan. A methodology for evaluating robustness of face recognition algorithms with respect to changes in pose and illumination angle. In *ICASSP 2005*, March 2005.

[17] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. van der Geest, J. H. C. Reiber, and M. Sonka. 3-D active appearance models: Segmentation of cardiac MR and ultrasound images. *IEEE Trans. on Medical Imaging*, 21(9):1167–1178, September 2002.

[18] E. Murphy-Chutorian and M. M. Trivedi. HyHOPE: hybrid head orientation and position estimation for vision-based driver head tracking. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 512–517, June 2008.

[19] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(4):607–626, April 2009.

[20] S. Romdhani and T. Vetter. Efficient, robust and accurate fitting of a 3D morphable model. In *Proc. 9th International Conference on Computer Vision*, volume 1, pages 59–66, October 2003.

[21] M. B. Stegmann and D. Pedersen. Bi-temporal 3D active appearance models with applications to unsupervised ejection fraction estimation. In *Proc. International Symposium on Medical Imaging*, volume 5747, pages 336–350, 2005.

[22] USF DARPA Human-ID 3D Face Database. Courtesy of Prof. Sudeep Sarkar, University of South Florida.

[23] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2D+3D active appearance models. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 535–542, June 2004.