ACTIVE APPEARANCE MODEL FITTING UNDER OCCLUSION USING FAST-ROBUST PCA

Markus Storer, Peter M. Roth, Martin Urschler, Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, 8010 Graz, Austria

Josef A. Birchbauer
Siemens Biometrics Center, Siemens IT Solutions and Services, Strassgangerstrasse 315, 8054 Graz, Austria

Keywords: Active appearance model, Fast robust PCA, AAM fitting under occlusion.

Abstract: The Active Appearance Model (AAM) is a widely used method for model based vision showing excellent results. But one major drawback is that the method is not robust against occlusions. Thus, if parts of the image are occluded the method converges to local minima and the obtained results are unreliable. To overcome this problem we propose a robust AAM fitting strategy. The main idea is to apply a robust PCA model to reconstruct the missing feature information and to use the thus obtained image as input for the standard AAM fitting process. Since existing methods for robust PCA reconstruction are computationally too expensive for real-time processing we developed a more efficient method: fast robust PCA (FR-PCA). In fact, by using our FR-PCA the computational effort is drastically reduced. Moreover, more accurate reconstructions are obtained. In the experiments, we evaluated both, the fast robust PCA model on the publicly available ALOI database and the whole robust AAM fitting chain on facial images. The results clearly show the benefits of our approach in terms of accuracy and speed when processing disturbed data (i.e., images containing occlusions).

1 INTRODUCTION

Generative model-based approaches for feature localization have received a lot of attention over the last decade. Their key advantage is to use a priori knowledge from a training stage for restricting the model while searching for a model instance in an image. Two specific instances of model-based approaches, the Active Appearance Model (AAM) (Cootes et al., 2001) and the closely related 3D Morphable Model (3DMM) (Blanz and Vetter, 1999), have proven to show excellent results in locating image features in applications such as face detection and tracking (Matthews and Baker, 2004), face and facial expression recognition (Blanz and Vetter, 2003), or medical image segmentation (Mitchell et al., 2001; Beichel et al., 2005).

Despite its large success, the AAM model has one main limitation. It is not robust against occlusions. Thus, if important features are missing the AAM fitting algorithm tends to get stuck in local minima. This especially credits for human faces since the large variability in the image data such as certain kinds of glasses, makeups, or beards can not totally be captured in the training stage. Similar difficulties also arise in other areas of model-based approaches (e.g., in the medical domain (Beichel et al., 2005)).

In the recent years some research was dedicated to generative model-based approaches in the presence of occlusions by investigating robust fitting strategies. In the original AAM approach (Cootes et al., 2001) fitting is treated as a least squares optimization problem, which is, of course, very sensitive to outliers due to its quadratic error measure (L_2 norm). To overcome this problem, the work of (Edwards et al., 1999) extended the standard fitting method (a) by learning the usual gray-value differences encountered during training and (b) by ignoring gray-value differences exceeding a threshold derived from these values during fitting. But the main drawback of this method is that the required threshold depends on the training conditions, which makes it improper for real-life situations. In contrast, in (Dornaika and Ahlberg, 2002) a RANSAC procedure is used for the initialization of the AAM fitting in order to get rid of occlusions due to differing poses. However, since the AAM fitting

Storer M., M. Roth P., Urschler M., Bischof H. and A. Birchbauer J. (2009).
ACTIVE APPEARANCE MODEL FITTING UNDER OCCLUSION USING FAST-ROBUST PCA.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 129-136
DOI: 10.5220/0001768701290136

remains unchanged this approach has still problems with appearance outliers.

Another direction of research was dedicated to replacing the least-squares error measure by a robust error measure in the fitting stage (Gross et al., 2006). Later this approach was further refined by comparing several robust error measures (Theobald et al., 2006). The same strategy is also used in (Romdhani and Vetter, 2003) and was adapted to a statistical framework in (Yu et al., 2007). But the latter approach is limited in several ways: (a) a scale parameter is required, which is hard to determine in general, (b) the framework around the inverse compositional algorithm is specifically tailored to tracking, and (c) the face models are built from the tracked person, which limits its applicability for general applications.

In the context of medical image analysis a robust AAM fitting approach was presented in (Beichel et al., 2005). In their method, which is based on the standard AAM fitting algorithm, gross disturbances (i.e., outliers) in the input image are avoided by ignoring misleading coefficient updates in the fitting stage. For that purpose, inlier and outlier coefficients are identified by a Mean Shift based analysis of the residual's modes. Then, an optimal sub-set of modes is selected and only those pixels covered by the selected mode combination are used for actual residual calculation. The Robust AAM Matching (RAAM) approach shows excellent results on a number of medical data sets. However, the mode selection is computationally very complex. Thus, this method is impractical for real-time or near real-time applications.

To overcome these drawbacks we introduce a new efficient robust AAM fitting scheme. In contrast to existing methods the robustness (against occluded features) is not directly included in the fitting step but is detached. In fact, we propose to run a robust pre-processing step first to generate undisturbed input data and then to apply a standard AAM fitting. Since the robust step, which is usually computationally intensive, has to be performed only once (and not iteratively in the fitting process), the computational costs can be reduced.

In particular, the main idea is to robustly replace the missing feature information from a reliable model. Thus, our work is somehow motivated by (Nguyen et al., 2008) and (Du and Su, 2005), where beards and eye-glasses, which are typical problems when applying an AAM approach, are removed. In (Du and Su, 2005) a PCA model was built from facial images that do not contain any eye-glasses. Then, in the removal step the original input images are reconstructed and the regions with the largest reconstruction errors are identified. These pixels are iteratively replaced by the reconstruction. But this approach can only be applied if the absolute number of missing pixels is quite small. In contrast, in (Nguyen et al., 2008) two models are computed in parallel, one for bearded faces and one for non-bearded faces. Then, in the removal step for a bearded face the detected beard region is reconstructed from the non-bearded space.

Since both methods are restricted to special types of occlusion or limited by a pre-defined error level, they can not be applied for general tasks. Thus, in our approach we apply a robust PCA model (e.g., (Rao, 1997; Black and Jepson, 1996; Leonardis and Bischof, 2000)) to cope with occlusions in the original input data. For that purpose, in the learning stage a reliable model is estimated from undisturbed data (i.e., without any occlusions), which is then applied to robustly reconstruct unreliable values from the disturbed data. However, a drawback of these methods is their computational complexity (i.e., iterative algorithms, multiple hypothesis, etc.), which hinders practical applicability. Thus, as a second contribution, we developed a more efficient robust PCA method that overcomes this limitation.

Even though the proposed robust AAM fitting is quite general, our main interest is to apply it to facial images. Thus, this application is evaluated in the experiments in detail. However, we also note that it is necessary that the image patch, where the robust PCA is applied has to be roughly aligned with the feature under consideration. In the case of our face localization this can be ensured by using a rough face and facial component detection algorithm inspired by the Viola-Jones algorithm (Viola and Jones, 2004). Moreover, the applied PCA model can handle a wide variability in facial images.

This paper is structured as follows. In Section 2 we introduce and discuss the novel fast robust PCA (FR-PCA) approach. In addition, we performed experiments on the publicly available ALOI database, which show that our approach outperforms existing robust methods in terms of speed and accuracy. Next, in Section 3, we introduce our robust AAM fitting algorithm that is based on the new robust PCA scheme. To demonstrate its benefits, we also present experimental results on facial images. Finally, we discuss our findings and conclude our work in Section 4.

2 FAST ROBUST PCA

If a PCA space $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}]$ is estimated from n samples, an unknown sample $\mathbf{x} = [x_1, \ldots, x_m]$, $m > n$, can usually be reconstructed to a sufficient degree of accuracy by p, $p < n$, eigenvectors:

$$\tilde{\mathbf{x}} = \mathbf{U}_p\mathbf{a} + \bar{\mathbf{x}} = \sum_{j=1}^{p} a_j\mathbf{u}_j + \bar{\mathbf{x}}\,, \qquad (1)$$

where $\bar{\mathbf{x}}$ is the sample mean and $\mathbf{a} = [a_1,\ldots,a_k]$ are the linear coefficients.

But if the sample $\mathbf{x}$ contains outliers (e.g., occluded pixels) Eq. (1) would not yield a reliable reconstruction; a robust method is required (e.g., (Rao, 1997; Black and Jepson, 1996; Leonardis and Bischof, 2000)). But since these methods are computationally very expensive (i.e., they are based on iterative algorithms) they are often not applicable in practice. Thus, in the following we propose a more efficient robust PCA approach.

2.1 Fast Robust Training

The training procedure, which is sub-divided into two major parts, is illustrated in Figure 1. First, a standard PCA subspace is generated from all training images. Second, in addition, a large number of smaller subspaces (sub-subspaces) is estimated from small sets of randomly selected data points (sub-sampling). Since occlusions are often considered to be spatially coherent the sub-sampling is done in a smart way. Hence, in addition to the random sampling over the whole image region, the random sampling is also restricted to image slices (vertical, horizontal, quadrant).

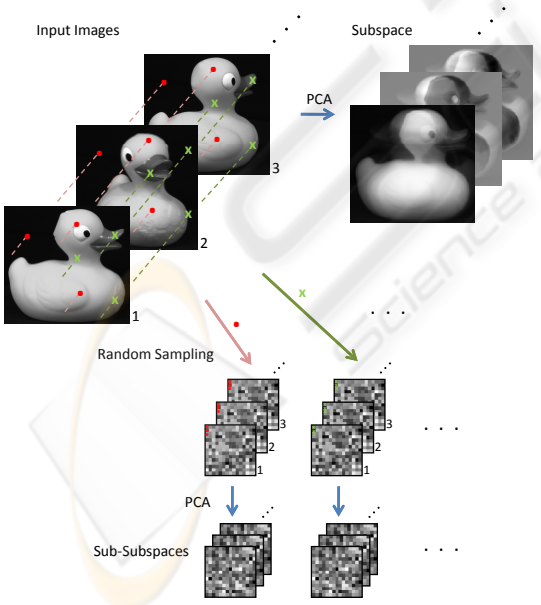


Figure 1: FR-PCA training. Generation of the subspace and the smaller sub-subspaces derived by randomly sub-sampling the input images.

2.2 Fast Robust Reconstruction

Given a new unseen test sample $\mathbf{x}$, the robust reconstruction is performed in two stages. In the first stage (*gross outlier detection*), the outliers are detected based on the sub-subspace reconstruction errors. In the second stage (*refinement*), using the thus estimated inliers a robust reconstruction $\tilde{\mathbf{x}}$ of the whole image is generated.

Assuming that in the training stage N sub-subspaces were estimated as described in Section 2.1, first, in the *gross outlier detection*, N sub-samplings $\mathbf{s}_n$ are generated according to the corresponding sub-subspace. In addition, we define the set of "inliers" $\mathbf{r} = \{\mathbf{s}_1 \cup \ldots \cup \mathbf{s}_N\}$. This set of points is illustrated in Figure 2(a) (green points). Next, for each sub-sampling $\mathbf{s}_n$ a reconstruction $\tilde{\mathbf{s}}_n$ is estimated by Eq. (1), which allows to estimate the (pixel-wise) error-maps

$$\mathbf{e}_n = |\mathbf{s}_n - \tilde{\mathbf{s}}_n|\,, \qquad (2)$$

the mean reconstruction error $\bar{e}$ over all sub-samplings, and the mean reconstruction errors $\bar{e}_n$ for each of the N sub-samplings.

Based on these errors we can detect the outliers by local and global thresholding. For that purpose, the sub-samplings $\mathbf{s}_n$ are ranked by their mean error $\bar{e}_n$. The local thresholds (one for each sub-sampling) are then defined by $\theta_n = \bar{e}_n w_n$, where the weight w_n is estimated from the sub-sampling's rank to remove less outliers from first ranked sub-samplings. The global threshold θ is set to the mean error $\bar{e}$. Then, all points $s_{n,(i,j)}$ for which

$$e_{n,(i,j)} > \theta_n \quad \text{or} \quad e_{n,(i,j)} > \theta \qquad (3)$$

are discarded from the sub-samplings $\mathbf{s}_n$ obtaining $\tilde{\mathbf{s}}_n$. Finally, we re-define the set of "inliers" by

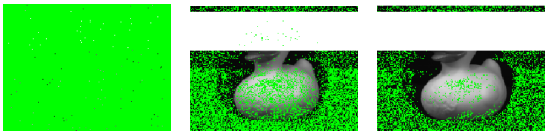
$$\mathbf{r} = \left\{\tilde{\mathbf{s}}_1 \cup \ldots \cup \tilde{\mathbf{s}}_q\right\}\,, \qquad (4)$$

where $\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_q$ indicate the first ranked q sub-samplings such that $|\mathbf{r}| \leq k$ and k is the pre-defined maximum number of points. The thus obtained "inliers" are shown in Figure 2(b).

The *gross outlier detection* procedure allows to remove most outliers (i.e., occluded pixels), thus the obtained set $\mathbf{r}$ contains almost only inliers. To further improve the final result in the *refinement* step, the final robust reconstruction is estimated similar to (Leonardis and Bischof, 2000). In particular, starting from the point set $\mathbf{r} = [r_1,\ldots,r_k]$, $k > p$, obtained from the *gross outlier detection*, an overdetermined system of equations is iteratively solved, where the following least square optimization problem

$$E(\mathbf{r}) = \sum_{i=1}^{k}(x_{r_i} - \sum_{j=1}^{p} a_j\mathbf{u}_{j,r_i})^2 \qquad (5)$$

has to be solved obtaining the coefficients **a**. Hence, the reconstruction $\tilde{\mathbf{x}}$ can be estimated and those points with the largest reconstruction error are discarded from **r** (selected by a reduction factor α). These steps are iterated until a pre-defined number of remaining pixels is reached. Thus, finally, an outlier-free sub-set is obtained, which is illustrated in Figure 2(c), and the robust reconstruction $\tilde{\mathbf{x}}$ can be estimated.



(a) (b) (c)

Figure 2: Data point selection process. (a) Data points sampled by all sub-subspaces. (b) Remaining data points after applying the sub-subspace procedure. (c) Resulting data points after the iterative refinement process for the calculation of the PCA coefficients.

Such a robust reconstruction result obtained by the proposed approach compared to a non-robust method is shown in Figure 3. One can clearly see that the robust method considerably outperforms the standard PCA. Note, that the blur visible in the reconstruction of the FR-PCA is the consequence of taking into account only a limited number of eigenvectors.

(a) (b) (c)

Figure 3: Demonstration of the insensitivity of the robust PCA to occlusions. (a) Occluded image, (b) reconstruction using standard PCA, and (c) reconstruction using the FR-PCA.

In general, the robust estimation of the coefficients is computational very efficient and thus very fast. In the *gross outlier detection* procedure only simple matrix operations (standard PCA) have to be performed, which are very fast; even if hundreds of sub-subspace reconstructions have to be executed. The computationally more expensive part is the *refinement* step, where an overdetermined linear system of equations has to be solved iteratively. Since very few refinement iterations have to be performed due to the preceding *gross outlier detection*, the total runtime is kept small.

2.3 Experimental Results

To show the benefits of the proposed fast robust PCA method (FR-PCA) we compared it to standard PCA (PCA) and the robust PCA approach of (Leonardis and Bischof, 2000) (R-PCA). We have chosen the latter one, because of its proven accuracy and applicability. Our refinement process is similar to theirs.

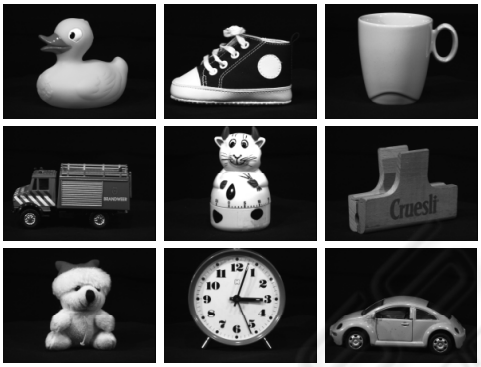


Figure 4: Illustrative examples of objects used in the experiments.

In particular, the experiments were evaluated on the "Amsterdam Library of Object Images (ALOI)" database (Geusebroek et al., 2005). The ALOI database consists of 1000 different objects. Over hundred images of each object are recorded under different viewing angles, illumination angles and illumination colors, yielding a total of 110,250 images. For our experiments we arbitrarily chose 20 categories (018, 032, 043, 074, 090, 093, 125, 127, 138, 151, 156, 174, 200, 299, 354, 368, 376, 809, 911, 926), where an illustrative subset of objects is shown in Figure 4.

Table 1: Settings for the FR-PCA (a) and the R-PCA (b) for the experiments.

(a)

FR-PCA	
Number of initial points k	130*p*
Reduction factor α	0.9

(b)

R-PCA	
Number of initial hypotheses H	30
Number of initial points k	48*p*
Reduction factor α	0.85
K_2	0.01
Compatibility threshold	100

In our experimental setup, each object is represented in a separate subspace and a set of 1000 sub-subspaces, where each sub-subspace contains 1% of data points of the whole image. The variance retained for the sub-subspaces is 95% and 98% for the whole subspace, which is also used for the standard PCA and the R-PCA. Unless otherwise noted, all experiments are performed with the parameter settings given in Ta-

Table 2: Comparison of the reconstruction errors of the standard PCA, the R-PCA and the FR-PCA. (a) RMS reconstruction-error per pixel given by mean and standard deviation. (b) RMS reconstruction-error per pixel given by robust statistics: median, upper- and lower quartile. Those results correspond to the box-plots in Figure 5.

(a)

Occlusion	Error per Pixel											
	0%		10%		20%		30%		50%		70%	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
PCA	9.96	5.88	21.30	7.24	34.60	11.41	47.72	14.37	70.91	19.06	91.64	19.78
R-PCA	11.32	6.92	11.39	7.03	11.98	8.02	20.40	19.90	59.73	32.54	87.83	26.07
FR-PCA	10.99	6.42	11.50	6.69	11.59	6.71	11.66	6.88	26.48	23.57	73.20	27.79

(b)

Occlusion	Error per Pixel																	
	0%			10%			20%			30%			50%			70%		
	median	$Q_{.25}$	$Q_{.75}$	median	$Q_{.25}$	$Q_{.75}$	median	$Q_{.25}$	$Q_{.75}$	median	$Q_{.25}$	$Q_{.75}$	median	$Q_{.25}$	$Q_{.75}$	median	$Q_{.25}$	$Q_{.75}$
PCA	9.58	5.77	14.02	21.29	16.56	26.01	34.67	27.71	42.17	47.24	38.22	57.42	70.45	57.03	84.54	89.49	77.55	106.15
R-PCA	10.54	6.39	15.81	10.63	6.50	15.76	10.95	6.60	16.16	13.83	7.96	23.13	62.76	32.47	82.98	87.80	70.64	104.99
FR-PCA	10.46	6.57	15.15	10.97	6.96	15.88	11.01	7.01	16.06	10.98	7.08	16.10	17.25	9.75	36.33	75.04	56.84	92.61

ble 1.

A 5-fold cross-validation is performed for each object category, resulting in 80% training- and 20% test data, corresponding to 21 test images per iteration. The experiments are accomplished for several levels of spatially coherent occlusions. To sum up, 2100 reconstructions are executed for every level of occlusion. Quantitative results for the root-mean-squared (RMS) reconstruction-error per pixel are given in Table 2. In addition, in Figure 5 we show box-plots of the RMS reconstruction-error per pixel for different levels of occlusions.

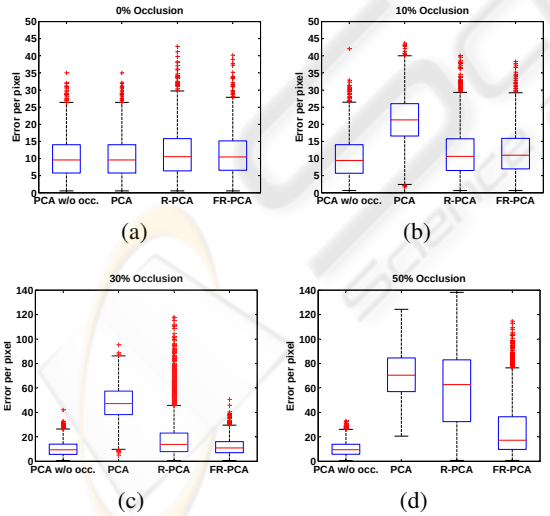


Figure 5: Box-plots for different levels of occlusions for the RMS reconstruction-error per pixel. PCA without occlusion is shown in every plot for the comparison of the robust methods to the best feasible reconstruction result.

Starting from 0% occlusion, all subspace methods exhibit nearly the same RMS reconstruction-error. Increasing the portion of occlusion, the standard PCA shows large errors whereas the robust methods are still comparable to the PCA without occlusion (best feasible case). The FR-PCA presents the best performance of the robust methods over all occlusion levels.

Finally, we evaluated the runtimes[1] for the applied different PCA reconstruction methods, which are summarized in Table 3. It can be seen that compared to R-PCA using FR-PCA speeds up the robust reconstruction by a factor of 18! If more eigenvectors are used or if the size of the images increases, the speed-up factor gets even larger. This drastic speed-up can be explained by the fact that the refinement process is started from a set of data points mainly consisting of inliers. In contrast, in (Leonardis and Bischof, 2000) several point sets (hypotheses) have to be created. The iterative procedure has to run for every set resulting in a poor runtime performance. To decrease the runtime the number of hypotheses or the number of initial points has to be reduced, which decreases reconstruction accuracy significantly. However, the runtime of our approach only depends slightly on the number of starting points, thus having nearly constant execution times. Both algorithms' runtime performance depend on the number

Table 3: Runtime comparison. Compared to R-PCA, FR-PCA speeds-up the computation by a factor of 18.

Occlusion	Mean Runtime [s]					
	0%	10%	20%	30%	50%	70%
PCA	0.006	0.007	0.007	0.007	0.008	0.009
R-PCA	6.333	6.172	5.435	4.945	3.193	2.580
FR-PCA	0.429	0.338	0.329	0.334	0.297	0.307

[1]The runtimes are measured in MATLAB using an Intel Xeon processor running at 3GHz. The resolution of the images is 192x144 pixels.

of eigenvectors used and their length. Increasing one of those values, the gap between the runtimes is even getting larger.

3 ROBUST AAM FITTING

3.1 Active Appearance Model

The Active Appearance Model (AAM) (Cootes et al., 2001) describes the variation in shape and texture of a training set representing an object. By applying PCA to the shape, texture, and the combination of shape and texture, the modes of variation are calculated. By keeping solely a certain percentage of the eigenvalue energy spectrum the model can be represented very compactly and optimally regarding Gaussian noise. The AAM model fitting is performed in a gradient descent optimization scheme. The cost function is defined as the L_2 norm of the intensity differences (between the estimated model and the given test image). To efficiently approximate the Jacobian of the cost function a learned regression model is used that describes the relationship between parameter updates and texture residual images according to (Cootes et al., 2001). A local minimum of the cost function corresponds to a model fitting solution. Since the minimum is local and the parameter space is high dimensional, multi-resolution techniques have to be incorporated and the fitting requires a coarse initialization.

3.2 Robust Fitting

Since the parameter updates for the fitting process are estimated from the texture's residual, the standard AAM is not robust against occlusions. To overcome this limitation, we propose to use our FR-PCA, introduced in Section 2, as a pre-processing step to remove disturbances in the input image and to perform the AAM fitting on the thus obtained reconstruction. Occlusions can not only be of artificial spatially coherent nature, which were taken for the quantitative evaluation of the FR-PCA (Section 2), but also in case of facial images beards or glasses. Those disturbances of facial images influence the quality of the fitting process of AAMs. Thus, for the pre-processing step we trained the FR-PCA using facial images which do not exhibit any disturbances, i.e., no beards and no glasses.

Figure 6, which was taken from the Caltech Faces data set (Caltech, 1999), demonstrates the whole processing chain for robust AAM fitting under occlusion. Figure 6(b) shows the initialization of the AAM on the occluded input image. The rough initialization of the AAM is done using a Viola-Jones face detection approach (Viola and Jones, 2004), several AdaBoost-based classifiers for locating eyes and mouth, and a face candidate validation scheme to robustly locate the rough face position.

Figure 6(c) demonstrates the converged fit of the AAM on the occluded image which failed totally. In contrast, using the FR-PCA as a pre-processing step results in the converged fit exhibited in Figure 6(d). In Figure 6(e), the shape from the fitting process on the reconstructed image is overlayed on the original input image. It can be clearly seen that the AAM can not handle occlusions directly whereas the fit on the reconstructed image is well defined.

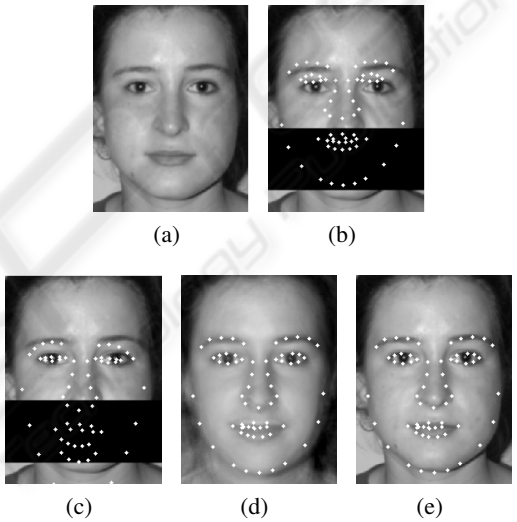


Figure 6: Handling of occlusions for AAM fitting. (a) Test image. (b) Initialization of the AAM on the occluded image. (c) Direct AAM fit on occluded image. (d) AAM fit on reconstructed image. (e) Shape from (d) overlayed on the test image. Image taken from Caltech Faces data set (Caltech, 1999).

3.3 Experimental Results

We trained a hierarchical AAM for facial images on three resolution levels (60x80, 120x160, 240x320). Our training set consists of 427 manually annotated face images taken from the Caltech face database (Caltech, 1999) and our own collection. Taking also the mirrored versions of those images doubles the amount of training data. For model building we keep 90% of the eigenvalue energy spectrum for the lower two levels and 95% for the highest level to represent our compact model.

As described in Section 3.2, we use the FR-PCA as a pre-processing step and perform the AAM fitting on the reconstructed images. Hence, we trained the

Table 4: Point-to-Point error. Comparing the direct fit of the AAM on the test image to the AAM fit utilizing the FR-PCA pre-processing (point errors are measured on 240x320 facial images).

Occlusion	Point-Point Error									
	0%		10%		20%		30%		40%	
	mean	std	mean	std	mean	std	mean	std	mean	std
AAM	4.05	5.77	12.06	11.25	15.19	12.78	18.76	14.89	18.86	13.94
AAM + FR-PCA	5.47	4.97	5.93	5.41	6.06	5.27	9.31	8.75	11.33	9.25

FR-PCA (Section 2.1) using facial images which do not exhibit any disturbances, i.e., no beards and no glasses. The variance retained for the whole subspace and for the sub-subspaces is 95%.

A 5-fold cross validation is performed using the manually annotated images, resulting in 80% training- and 20% test data per iteration. For each level of occlusion, 210 AAM fits are executed. Table 4 shows the point-to-point error (Euclidean distance of converged points to the annotated points) comparing the direct AAM fit on the occluded image to the AAM fit utilizing the FR-PCA pre-processing. Starting from 0% occlusion, the error for the AAM + FR-PCA is slightly larger than the direct fit, because of the unavoidable reconstruction-blur resulting from the FR-PCA reconstruction. When increasing the size of the occlusion, the big advantage of the FR-PCA pre-processing can be seen.

Up to now, to have a steerable environment, we used artificial spatially coherent occlusions. To show the advantage of FR-PCA pre-processing also on natural occlusions such as tinted glasses, occlusions caused by wearing a scarf or by disturbances like beards, Figure 7 depicts some AAM fits on images taken from the AR face database (Martinez and Benavente, 1998). In addition, Figure 8 shows an illustrative result on our own database. The FR-PCA pre-processing step takes around 0.69s per image (150x200) measured in MATLAB using an Intel Xeon processor running at 3GHz.

Figure 7: Examples of AAM fits on natural occlusions like tinted glasses or wearing a scarf. (First column) Test images with AAM initialization. (Second column) Direct AAM fit on the test images. (Third column) AAM fit utilizing the FR-PCA pre-processing. Images are taken from the AR face database (Martinez and Benavente, 1998).

4 CONCLUSIONS

The contribution of this paper is twofold. First, we presented a robust method for AAM fitting. In contrast to existing approaches the robustness is not included in the fitting step but is detached in a pre-processing step. The main idea is to robustly reconstruct unreliable data points (i.e., occlusions) in the pre-processing step and to use the thus obtained undisturbed images as input for a standard AAM fitting. To speed up this robust pre-processing step, as the second contribution, we developed a novel fast robust PCA method. The main idea is to estimate a large number of small PCA sub-subspaces from a sub-set of points in parallel. By discarding those sub-subspaces with the largest errors the number of outliers in the input data can be reduced, which drastically decreases the computational effort for the robust reconstruction. In the experiments, we showed that our new fast robust PCA approach outperforms existing methods in terms of speed and accuracy. In addition, the whole process chain (robust pre-processing and AAM fitting) was demonstrated in the field of face normal-

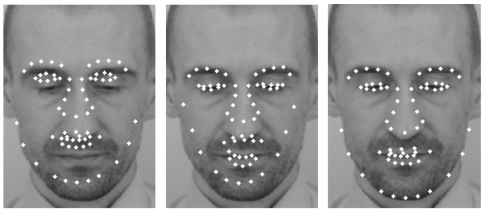


Figure 8: Examples of AAM fits on natural occlusions like beards.

ization in the presence of artificial and natural occlusion noise. The results show that our robust approach can handle such situations considerably better than a non-robust approach. Moreover, due to the very efficient robust pre-processing the proposed robust AAM fitting method is applicable in practice for real-time applications. An immediate idea for future work is the investigation of how to incorporate the FR-PCA approach directly into the AAM fitting procedure.

ACKNOWLEDGEMENTS

This work has been funded by the Biometrics Center of Siemens IT Solutions and Services, Siemens Austria. In addition, this work was supported by the FFG project AUTOVISTA (813395) under the FIT-IT programme, and the Austrian Joint Research Project Cognitive Vision under projects S9103-N04 and S9104-N04.

REFERENCES

Beichel, R., Bischof, H., Leberl, F., and Sonka, M. (2005). Robust active appearance models and their application to medical image analysis. *IEEE Trans. Med. Imag.*, 24(9):1151– 1169.

Black, M. J. and Jepson, A. D. (1996). Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. In *Proc. ECCV*, pages 329–342.

Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3d-faces. In *Proc. SIGGRAPH*.

Blanz, V. and Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *IEEE Trans. PAMI*, 25(9):1063–1074.

Caltech (1999). Caltech face database. www.vision.caltech.edu/html-files/archive.html.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Trans. PAMI*, 23(6):681–685.

Dornaika, F. and Ahlberg, J. (2002). Face model adaptation using robust matching and active appearance models. In *Proc. IEEE Workshop on Applications of Computer Vision*.

Du, C. and Su, G. (2005). Eyeglasses removal from facial images. *Pattern Recognition Letters*, 26(14):2215–2220.

Edwards, G. J., Cootes, T. F., and Taylor, C. J. (1999). Advances in active appearance models. In *Proc. ICCV*, pages 137–142.

Geusebroek, J. M., Burghouts, G. J., and Smeulders, A. W. M. (2005). The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1):103–112.

Gross, R., Matthews, I., and Baker, S. (2006). Active appearance models with occlusion. *Image and Vision Computing*, 24:593–604.

Leonardis, A. and Bischof, H. (2000). Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118.

Martinez, A. and Benavente, R. (1998). The AR face database. Technical Report 24, CVC.

Matthews, I. and Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164.

Mitchell, S. C., Bosch, J. G., Lelieveldt, B. P. F., van der Geest, R. J., Reiber, J. H. C., and Sonka, M. (2001). 3-D active appearance models: Segmentation of cardiac MR and ultrasound images. *IEEE Trans. Med. Imag.*, 21:1167–1178.

Nguyen, M. H., Lalonde, J.-F., Efros, A. A., and de la Torre, F. (2008). Image-based shaving. *Computer Graphics Forum Journal (Eurographics 2008)*, 27(2):627–635.

Rao, R. (1997). Dynamic appearance-based recognition. In *Proc. CVPR*, pages 540–546.

Romdhani, S. and Vetter, T. (2003). Efficient, robust and accurate fitting of a 3D morphable model. In *Proc. ICCV*, volume 2.

Theobald, B.-J., Matthews, I., and Baker, S. (2006). Evaluating error functions for robust active appearance models. In *Proc. FGR*, pages 149–154.

Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154.

Yu, X., Tian, J., and Liu, J. (2007). Active appearance models fitting with occlusion. In *EMMCVPR*, volume 4679 of *LNCS*, pages 137–144.