Constructing Synthetic Faces using Active Appearance Models and Evaluating the Similarity to the Original Image Data

Andre Störmer, Jan Stadermann
Institute of Human-Machine-Communication
Technische Universität München
Arcisstr. 16, 80333 München
{stoermer, stadermann}@mmk.ei.tum.de

Abstract

Active Appearance Models (AAMs) can be used for interpreting face images and image sequences. AAMs combine a statistical shape model and a model of grey-level appearance. They also contain an iterative matching scheme for image interpretation, which only needs an initial estimate of the position and size of the face and results in a set of parameters describing the matched face. In this paper we describe, how to build up the shape model, the grey-level model and how to combine them to an AAM. Then we show, how to construct a search algorithm to derive the model parameters for a given face and investigate how well these parameters can be used to descibe known and unknown face images, measuring the similarity between the model-built faces and the original image data.

1 Introduction

The interpretation of face images is currently of great interest. The possibility to extract some information from a human face could be used for adapting the behaviour of an intelligent human-machine interaction, also the identification of human faces for security issues has gained much attention in the last years.
Since there are many different face interpretation tasks such as identity, gender, age, pose and expression, it would be desirable to have one method being able to handle as many different tasks as possible. If we assume that all the necessary information to solve these tasks is in the image data, and if we can synthesise a good approximation of the original face using the model and some parameters, we can also assume, that the parameters represent all the information necessary. These parameters could therefore be used for classification tasks.

2 Related Work

There are several model-based approaches for the interpretation of face images.
Turk and Pentland [7] use principal component analysis (PCA) to describe face images using a set of

From: SOAVE 2004, VDI

basis functions called eigenfaces. This method is not robust to variations in shape, pose and expression but the eigenface model can be fit easily to a face image using correlation based methods.
Maurer and von der Malsburg [10] use a flexible graph describing local features using garbor jets, which works robust to wide angels of facial orientation.
Cootes, Lanitis and Taylor [6, 5] model local grey-level appearance and shape using Active Shape Models (ASMs) to interprete face images. They use the shape found by the ASM for warping the face texture into a normalised frame. There a model of intensities of the shape-normalised face is used to interprete the image. Later Edwards et al [1, 3] extend this approach using a combined model of shape and grey-level appearance and introduce an iterative matching scheme for finding the parameters on unknown images.

3 Active Appearance Models for face description

In this section we explain how to build an AAM for describing faces, following the approach of Cootes and Taylor [4].

3.1 The Shape Model

For building the shape model we require a set of s labelled face images, where landmarks are set on key positions of the main facial features (outline, nose, eyes, mouth, ...) (see Figure 1).

Figure 1. labelled face image

For each labelled face, the position (x_i, y_i) of the n landmarks is written into a vector $\vec{x}$ with $2n$ elements.

$$\vec{x} = (x_1, y_1, \ldots, x_n, y_n)^T \tag{1}$$

We align all shape-describing vectors $\vec{x}_1 \ldots \vec{x}_s$, translating them into their Center of Gravity and rotating and scaling them, that the mean square error to the mean vector is minimal. This can be done using a simple recursive optimisation scheme [4]. After this is done, we build the statistical shape model applying principal component analysis(PCA). First we calculate the mean $\overline{\vec{x}}$ of the data

$$\overline{\vec{x}} = \frac{1}{s} \sum_{i=1}^{i=s} \vec{x}_i \tag{2}$$

and the covariance of the data:

$$S = \frac{1}{s-1} \sum_{i=1}^{s} (\vec{x}_i - \overline{\vec{x}})(\vec{x}_i - \overline{\vec{x}})^T \tag{3}$$

Then we compute the eigenvectors ϕ_j and corresponding eigenvalues λ_j of S. If Φ_s contains the t eigenvectors $\vec{\phi}_j$ corresponding to the largest eigenvalues, we can approximate any $\vec{x}$.

$$\vec{x} \approx \bar{\vec{x}} + \Phi_s \vec{b}_S \tag{4}$$

where $\Phi_S = (\vec{\phi}_1 | \vec{\phi}_2 | \ldots | \vec{\phi}_{t_S})$ and $\vec{b}_s$ is a t_s-dimensional vector containing shape parameters. For a given $\vec{x}$ the parameters $\vec{b}_S$ can be approximated:

$$\vec{b}_S \approx \Phi_s^T (\vec{x} - \bar{\vec{x}}) \tag{5}$$

The number of eigenvectors t_s to use can be choosen so that the model can represent a certain proportion (e.g. 98%) of the total variance of the training data.
To get an idea, how the shape model works, we show the effect of varying the parameters corresponding to the first three modes (the three eigenvectors with the largest eigenvalues). For better visualisation we connected some of the landmarks with lines. (Figure 2).

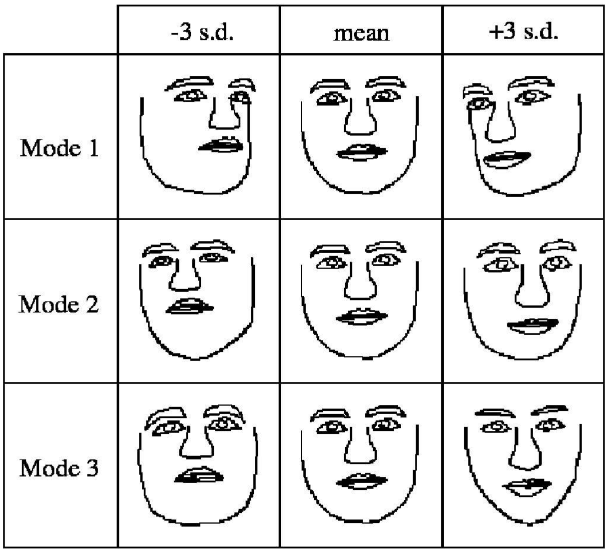

Figure 2. Effect of varying the first three shape parameters in turn between ± 3 *standard derivations*

3.2 Warping

As a next step we want to explain the basic principles of the warping algorithm. It is needed during the model building for warping all the training images to a normalised reference shape (the mean shape). These shape-normalised images are the basis to build the grey-level model. Later the warping algorithm is also used to synthesise faces.

The intention of this step is to warp the texture from a given shape to another one. For this we triangulate (e.g. Delaunay triangulation) the points listed in the shape vectors. After we have done this, we can map the texture data from the triangles in the source image to the corresponding triangles referring to the shape of the target image (Figure 3).

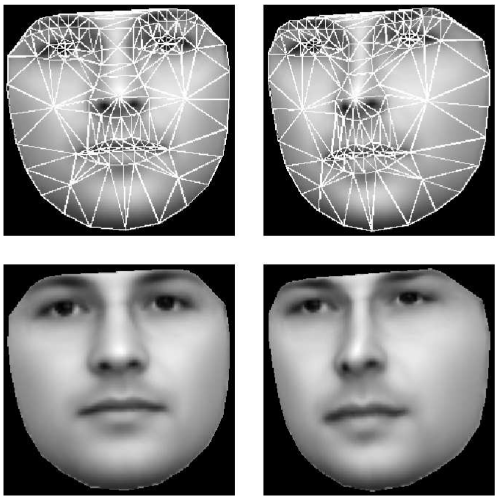

Figure 3. Warping a face from one shape to another by mapping the texture of the triangles

3.3 Grey-Level Model

For building the grey-level model we sample the grey-level information from the shape-normalised images. To minimise global lighting effects, the mean is subtracted from each sampled vector. A scaling factor α is choosen to minimise the mean square error to the mean grey-level vector, normalised, that its variance is 1. This can be done by scaling each grey-level vector $\vec{g}$ with

$$\alpha = \frac{\bar{\vec{g}}\,\bar{\vec{g}}}{\vec{g}\,\bar{\vec{g}}} \tag{6}$$

Scaling all the grey-level vectors changes their mean vector, so calculation of a new mean grey-level vector and aligning the dataset to it has to be repeated until there is no significant chance. After this alignment all grey-level vectors have zero mean and aligned variance.
After that, a PCA is applied on the normalised image data, so that any grey-level vector $\vec{g}$ can be approximated with a set of grey-level parameters $\vec{b}_g$ (see Figure 4).

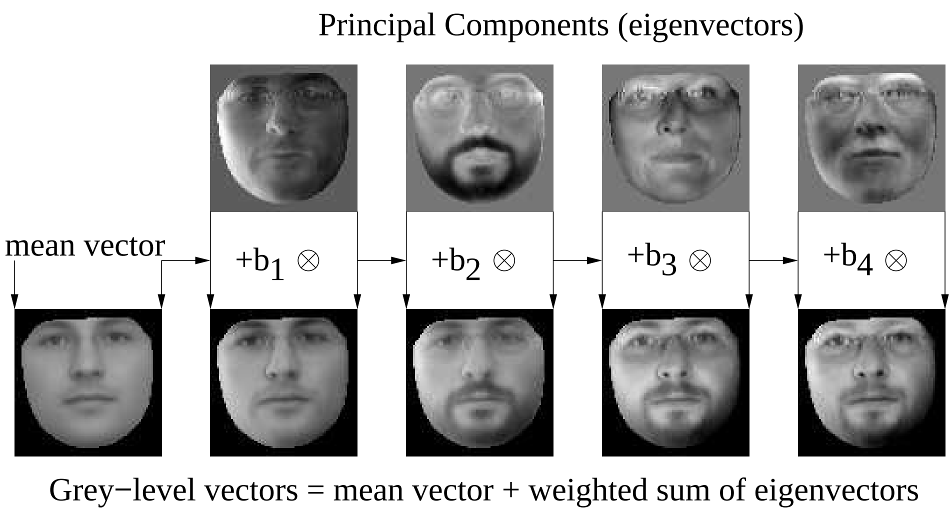

Figure 4. Synthesis of grey-level vectors

$$\vec{g} \approx \overline{\vec{g}} + \Phi_g \vec{b}_g \tag{7}$$

where the matrix Φ_g contains the t eigenvectors $(\vec{\phi}_1|\ldots|\vec{\phi}_{t_g})$ corresponding to the t_g largest eigenvalues. For a given grey-level vector $\vec{g}$, the grey-level parameters $\vec{b}_g$ can be computed using

$$\vec{b}_g \approx \Phi_g^T(\vec{g} - \overline{\vec{g}}) \tag{8}$$

Again the number of eigenvectors t_g to use can be choosen that the model can represent a given proportion (e.g. 98%) of the total variance of the training data.

3.4 Combined Appearance Model

Any example of the training data can be described with the shape and grey-level parameters $\vec{b}_s$ and $\vec{b}_g$. For each we create the combined column vector

$$\vec{b} = \begin{pmatrix} W\vec{b}_f \\ \vec{b}_g \end{pmatrix} = \begin{pmatrix} W\Phi_f^T(\vec{x} - \overline{\vec{x}}) \\ \Phi_g^T(\vec{g} - \overline{\vec{g}}) \end{pmatrix} \tag{9}$$

where the matrix W weights the shape parameters so that the variance of the combined vector $\vec{b}$ is not dominated by one of the vectors $\vec{b}_s$ or $\vec{b}_g$. This can be done by choosing $W = wI$ (I is identity matrix) with

$$w = \frac{\sum_{i=1}^{t_g} \lambda_{g,i}}{\sum_{i=1}^{t_f} \lambda_{s,i}} \tag{10}$$

where λ_g and λ_s are the eigenvalues of the used grey-level and shape eigenvectors. Since shape and grey-level appearance could be correlated the PCA is applied. After this is done, any combined vector $\vec{b}$ can be approximated

$$\vec{b} \approx \overline{\vec{b}} + \Phi\vec{c} \tag{11}$$

where Φ is the matrix containing the t eigenvectors $(\vec{\phi}_1|\ldots|\vec{\phi}_t)$ with the t largest eigenvalues. The t-dimensional vector $\vec{c}$ contains the appearance parameters. For a given combined vector $\vec{b}$ we can approximate the appearace parameters

$$\vec{c} \approx \Phi^T(\vec{b} - \overline{\vec{b}}) \tag{12}$$

Having this set of appearance parameters we can calculate the shape parameters $\vec{b}_s$ and the grey-level parameters $\vec{b}_g$

$$\vec{b} = \begin{pmatrix} \vec{b}_f' \\ \vec{b}_g \end{pmatrix}; \quad \vec{b}_f = W^{-1}\vec{b}_f' \tag{13}$$

Again, the number t of eigenvectors to use, can be chosen that the model can represent a given proportion (e.g. 98%) of the total variance of the training data.
Having such a set of shape and grey-level parameters we can synthesise a face (D), creating the shape free face (A) from the given grey-level parameters $\vec{b}_g$ and warping this from the mean shape $\overline{\vec{x}}$ (B) to the shape $\vec{x}$ given by the shape parameters $\vec{b}_s$ (C)(as shown in Figure 5).

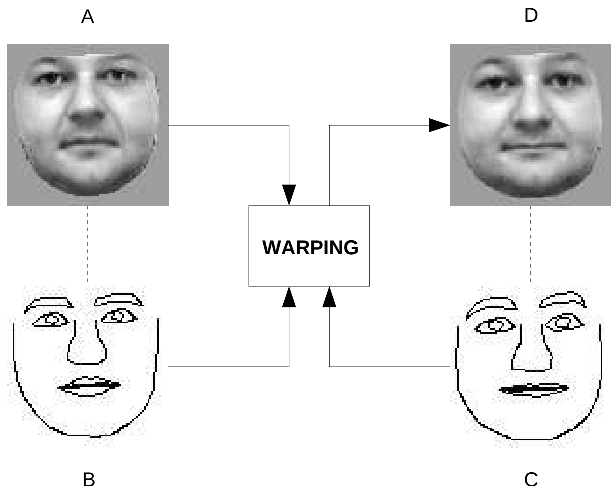

Figure 5. Synthesis of faces using the shape model, grey-level model and warping algorithm

3.5 Active Appearance Model Search (AAM-Search)

To find and interpretate a face in an image, we want to minimise the energy of the difference image between original input data and synthesised face by changing the appearance parameters $\vec{c}$ and the transformation parameters $\vec{t}$ for scaling, rotation and translation. Therefore we calculate the difference image in a normalised reference frame (see Figure 6). We warp the original image data from the estimated shape (A) (synthesisable using the shape parameters and transformation parameters) to the mean shape and align its grey-levels to get a grey-level vector $\vec{g}_{orig}$ (B). Then we subtract the grey-level vector $\vec{g}_{synth}$ (C), which can be synthesised using the estimated grey-level parameters. The difference is the residual (D).

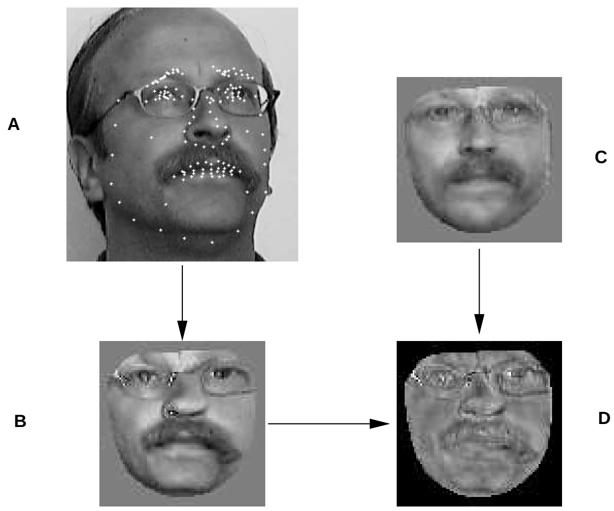

Figure 6. Computing the residual

The residual $\vec{r}$ is

$$\vec{r}(\vec{p}) = \vec{g}_{orig} - \vec{g}_{synth} \tag{14}$$

where $\vec{p} = (\vec{c}, \vec{t})^T$. We apply a first order Taylor expansion

$$\vec{r}(\vec{p}+\delta\vec{p}) = \vec{r}(\vec{p}) + \frac{\partial\vec{r}}{\partial\vec{p}}\delta\vec{p} \tag{15}$$

where $\frac{d\vec{r}_i}{d\vec{p}_j}$ is the ij-th element of the matrix $\frac{\partial\vec{r}}{\partial\vec{p}}$ of partial derivatives.

For a given estimate of parameters $\vec{p}$ with corresponding residual $\vec{r}(\vec{p})$ we want to choose $\delta\vec{p}$ that $|\vec{r}(\vec{p}+\delta\vec{p})|^2$ will be minimised. By equating (15) to zero we obtain:

$$-\vec{r}(\vec{p}) = \frac{\partial\vec{r}}{\partial\vec{p}}\delta\vec{p} \tag{16}$$

thus we get

$$\delta\vec{p} = -R\vec{r}(\vec{p}) \qquad \text{where} \qquad R = \left(\frac{\partial\vec{r}}{\partial\vec{p}}^T \frac{\partial\vec{r}}{\partial\vec{p}}\right)^{-1} \frac{\partial\vec{r}}{\partial\vec{p}}^T \tag{17}$$

To avoid the expensive calculation of R at every step, we assume, that it can be considered approximately fixed. Thus we can estimate it, measuring residuals by synthesising faces from random filled parameter vectors (eg in a range of ± 3 standard derivations of each parameter), changing them by known values (eg ± 0.5 standard derivations for the appearance parameters and $\pm 10\%$ for scaling, translation and rotation) and calculate the average differential quotient for each parameter p_j.

$$\frac{dr_i}{dp_j} \approx \frac{1}{k}\sum_k \frac{1}{\delta p_{jk}}(\vec{r}_i(\vec{p}+\delta\vec{p}_{jk}) - \vec{r}_i(\vec{p})) \tag{18}$$

Having all partial derivations we can calculate R using (17).

3.5.1 Iterative Search Algorithm

Given a method for predicting the correction of the parameters to minimise the energy of the residual, a simple iterative algorithm is constructed. One step of the iterative procedure is as follows:

- Evaluate the current residual $\vec{r}_0 = \vec{g}_{orig} - \vec{g}_{synth}$
- Evaluate the current Error $E_0 = |\vec{r}_0|^2$
- Compute the predicted parameter change, $\delta\vec{p} = R\delta\vec{g}_0$
- Set $k = 1$
- Let $\vec{p}_1 = \vec{p}_0 - k\delta\vec{p}$
- Sample $\vec{g}_{orig}$ from the estimated position , synthesise the estimated grey-level vector and compute the new residual $\vec{r}_1$ (see Figure 6)
- If $|\vec{r}_1|^2 < E_0$ accept the new estimate $\vec{p}_1$
- Otherwise try at $k = 1.5, k = 0.5, k = 0.25$ etc.

This procedure is repeated until no improvement is made, convergence is declared.

4 Results

4.1 Measuring the similarity

For a wide range of interpretation tasks, we wish to achieve a high similarity between the original face image and the face synthesised using the estimated parameters. To measure the similarity we compare the grey-level vector $\vec{g}_{orig}$, sampled from the original image at the coordinates of the shape given by the approximated shape parameters, and the synthesised grey-level vector $\vec{g}_{synth}$, using the similarity measurement q.

$$q = \frac{\vec{g}_{orig}}{|\vec{g}_{orig}|} \frac{\vec{g}_{synth}}{|\vec{g}_{synth}|} \tag{19}$$

This measurement is similar to a normalised cross correlation in the shape normalised space. The grey-level vectors used for the calculation of this measurement have zero mean and aligned variance and are available as an intermediate result of the search algorithm.
The length of the vectors is normalised to avoid evaluation of grey-level scaling, which could be removed easily. The measurement gives us values between $(0..1)$, the nearer it is to 1 the higher is the similarity.
To get an idea about the values of this measurement, we show some examples of original face images and corresponding synthesised faces using the resulting parameters of the search algorithm (Figure 7).

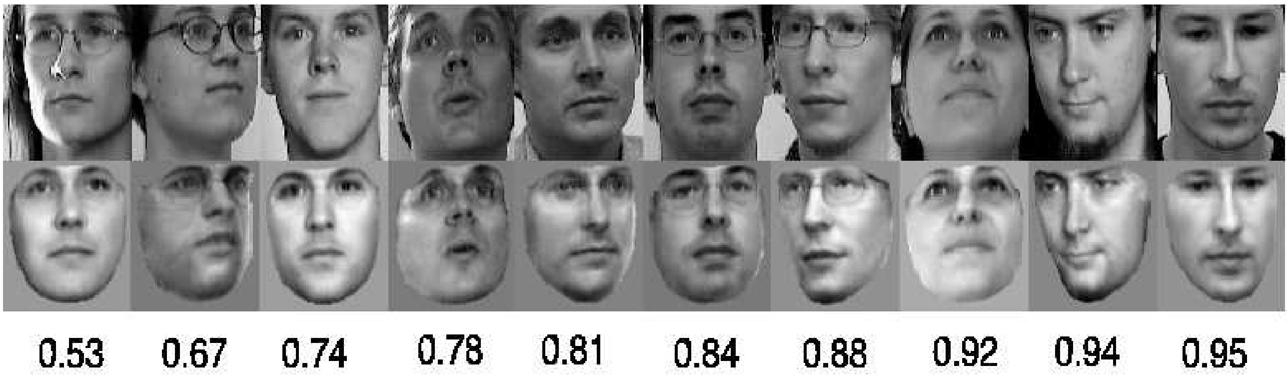

Figure 7. Examples: Similarity measurement for different pairs of original and synthetic images

4.2 Results of the AAM-Search

For the following investigations we built an AAM using 137 face images showing 52 different people (NIFace database, Ilmenau Technical University, Department of Neuroinformatics, [9]), each hand-labelled with 138 landmarks. The shape was described using 40 shape-eigenvectors. The grey-level vectors were sampled from the training images and warped to the mean shape at a resolution of 86×84 pixels and were described using 100 grey-level eigenvectors. The PCA on the combined vectors and the choice of the number of eigenvectors resulted in 70 appearance parameters describing the faces. We computed the matrix R, needed for the AAM-Search, approximating the matrix of partial derivations using differential quotients (computing residuals with random model-built faces and random parameter changes, averaging over 20 residuals per parameter).
Before starting the AAM-Search on the images we used for testing, an initial estimate of the position and size of the face (translation and scaling parameters) was delivered by a face detector (Viola and

Jones, [11]). All other parameters were initially set to zero, which correspondends to the mean value of the parameter sets of the training data. For faces, which have not been detected by the face detector, we set the similarity to 0 for the calculation of the average. To evaluate the influence of the face detector we also measured the similarity on the search results for using the known position and size as initial estimate.
For testing, we used a set of 104 face images showing the same persons as in the training data. In another test we wanted to evaluate, if the model is able to describe faces of persons, which are completely unknown. For this we used 120 images showing 20 different unknown persons.

In table 1 we show an overview of the measured similarities between the original images and the synthesised faces using the parameters derived by the AAM-Search.

Similarity	> 0.90	> 0.85	> 0.80	> 0.75	> 0.70	Average
Training data (known positions)	78.1%	90.5%	97.3%	97.3%	97.3%	0.92
Training data (face detector)	56.2%	69.3%	79.5%	87.6%	95.6%	0.87
Testing data	47.8%	82.5%	81.5%	91.3%	91.3%	0.88
Completely unknown images	12.1%	33.3%	57.5%	78.7%	96.9%	0.81

Table 1. Overview on similarity measures between original image and synthesised face using the results of the AAM-Search

If we look on the results on the training data, we can see, that starting the search algorithm with an initial estimate at the known position and size of the face are somewhat better than that with using a face detector. This leads to the conclusion, that the search algorithm is somewhat sensitive to the initial estimate. The prediction of the tranformation parameters (translation, size, rotation) needs to be enhanced, to compensate this.
We can also see, that the results on the training data and on the testing data are in the same range of quality. The results on the testing data are even slightly better in average. We explain this fact with the quality of the initial estimate delivered by the face detector. But, as we thought, the number of very high similarities (> 0.90) is larger for the training data.
The quality of the results on the unknown face images is not as good as those on the testing data. We expected this, but even for this completely unknown faces the search algorithm was able to derive parameters to synthesise similar faces.
If we look on the synthesised faces and on the similarity to the original samples, we see, that the major part of the found parameters lead to similarities > 0.70. From our viewpoint, this similarity is an indication that they contain some information which could be used in classification task with a low number of classes (e.g. gender). But on the other hand, we see that only a small part of the described faces leads to very high similarities (> 0.90). We think, that for solving some classification task (e.g. identification) such high similarity measures are needed, to ensure that the necessary information is in the synthesised face and thus in the parameters.

5 Conclusion

In this paper we reviewed the Active-Appearance Model approach and evaluated its performance on known and unknown image data. Hand-labelled images are only needed to build up the model, the proposed search algorithm just needs an initial estimate of the position and the size of the face in the

image to compute the Appearance-Parameters.
Considering the fact, that the used implementations of the algorithms are far from being optimised and the initial estimate delivered by the face detector is not adopted to the needs of the AAM-Search (it convergates in a local minimum if the initial estimate is too rough), one could say, that the results in this paper show the minimum potential of Active Appearance Models for face description. The possibility to model faces with a variability in shape and grey-level appearance makes this approach interesting for different interpretation tasks. Solving them could be possible, using the appearance parameters for classification. Further investigations should show which parameters are useful for which classification task (gender, age, pose, expression, identity). The possibility to compute the shape and grey-level vectors out of the appearance parameters could be used, some interpretation tasks are more related to the shape, some more to the grey-level distribution.
An approach to improve the quality of the model could be to use the similarity measurement to examine, which images can already be described by the model. Face images, for which the similarity to the model built synthetic image is too low, could be joined to the training data to improve the model. Also there are different approaches to make the search algorithm more effective and robust [8, 2], which deserve closer attention.

References

[1] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In $5th$ *European Machine Vision Conference*, volume 2, pages 484–492. Springer, 1998.

[2] T. Cootes, G. Edwards, and C. Taylor. A comparative evaluation of active appearance model algorithms. In $9th$ *British Machine Vision Conference*, volume 2, pages 680–689. BMVA Press, 1998.

[3] T. Cootes, G. Edwards, and C. Taylor. Advances in active appearance models. In $7th$ *International Conference on Computer Vision*, pages 137–142, 1999.

[4] T. Cootes and C. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, 2000.

[5] A. Lanitis, C. Taylor, and T. Cootes. Automatic face identification system using flexible appearance models. *IVC*, 13(5):393–401, June 1995.

[6] A. Lanitis, C. J. Taylor, and T. F. Cootes. A unified approach to coding and interpreting face images. In *ICCV*, pages 368–373, 1995.

[7] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[8] I. Matthews and S. Baker. Active appearance models revisited. Technical report, Carnegie Mellon University Robotics Institute, 2003.

[9] A. Störmer. Untersuchung der Möglichkeiten zur Beschreibung von Gesichtern mit einer minimalen Anzahl von Appearance-Parametern. Diploma Thesis (in german) , Ilmenau Technical University, Department of Neuroinformatics, 2004.

[10] T. Maurer and C. von der Malsburg. Tracking and Learning graphs and pose on image sequences of faces. In $2nd$ *International Conference on Automatic Face and Gesture Recognition*, pages 176–181. IEEE Computer Society Press, 1997.

[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.