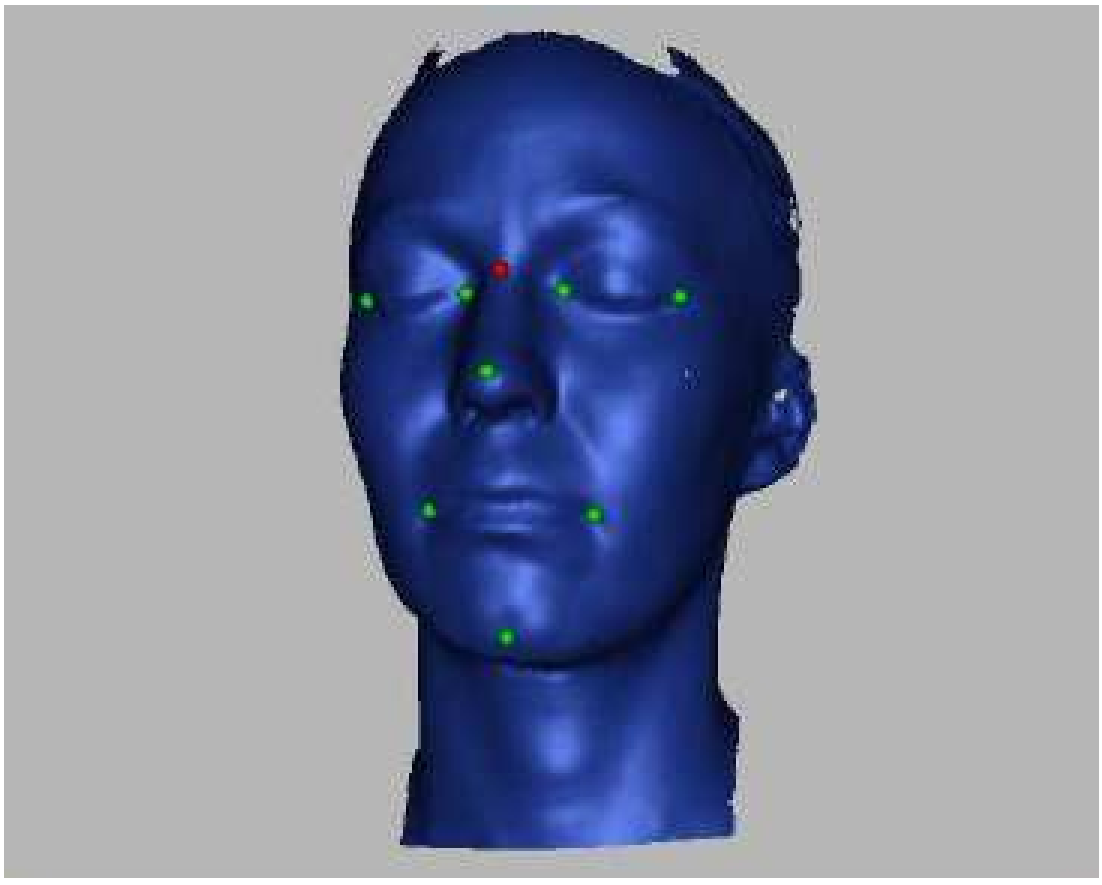
DTU Technical Report: ARTTS

Title: **Face pose tracking and recognition and 3D cameras**

Author: **Rasmus Larsen**

Project: **ARTTS**

Date: **February 10th, 2006**



Contents

Introduction

This technical report focuses on the state-of-the-art of face tracking and face pose estimation with a special emphasis on the added value of 3D cameras to these applications.

These applications receive a lot of attention and are the focus of a series of international scientific workshops, e.g.

1. IEEE Workshop on Face Recognition Grand Challenge Experiments
2. International Conference on Audio- and Video-Based Person Authentication
3. International Conference on Automated Face and Gesture Recognition and authentication

A recent paper

> Kevin W. Bowyer, Kyong Chang, Patrick Flynn (2006*) A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition.* Computer Vision and Image Understanding Vol 101, pp 1–15

summarizes the conclusions of the first of the workshops mentioned above. From this paper the following statement summarizes the idea of using 3D depth information for face tracking and recognition

> *We are particularly interested in 3D face recognition because it is commonly thought that the use of **3D sensing** has the potential for greater recognition **accuracy** than 2D. For example, one paper states—''Because we are working in 3D, we overcome limitations due to viewpoint and lighting variations'' [13]. Another paper describing a different approach to 3D face recognition states—''Range images have the advantage of capturing shape variation irrespective of illumination variabilities'' [12]. Similarly, a third paper states—''Depth and curvature*

features have several advantages over more traditional intensity-based features. Specifically, curvature descriptors: (1) have the potential for higher accuracy in describing surface-based events, (2) are better suited to describe properties of the face in a areas such as the cheeks, forehead, and chin, and (3) are viewpoint invariant'' [11].

With respect to the choice of 3D sensor the paper concludes

An ideal 3D sensor for face recognition applications would combine at least the following properties: (1) image acquisition time similar to that of a typical 2D camera, (2) a large depth of field; e.g, a meter or more in which there is essentially no loss in accuracy of -depth resolution, (3) robust operation under a range of ''normal'' lighting conditions, (4) no eye safety issues arising from projected light, (5) dense sampling of depth values; perhaps 1000 · 1000, and (6) depth resolution of better than 1 mm. Evaluated by these criteria, we do not know of any currently available 3D sensor that could be considered as ideal for use in face recognition.

The ARTTS consortium believes that the Swiss ranger provides a large step in the right direction. Furthermore, the combination of the swiss ranger with high resolution 2D cameras in a calibrated multimedia setup will greatly enhance the usability according to the statement above.

State-of-the art 3D face pose and tracking

In general the approaches to face tracking and pose estimation can be categorized according to the following table

		data	
		2D images	3d image
model	2d model	cf. 1 below	
	3d model	cf. 2 below	ARTTS

All approaches implicitly or explicitly make use of a (generative) model of a face.

1. in 2D this model may a patch of intensities, e.g. a rectangular (or more general polygonal) region-of-interest normalized with respect to size and distance between facial features. In [2] a rectangular patch is defined in terms of eyes position and distance. In [6] an elliptic patch aligned to match eyes and mouth is used. Other approaches build models based on more general warps of faces, e.g. by matching a set of landmarks [7]. In both cases [6,7] subspace methods can be used to extract significant

modes of variation. Such subspace methods range from simple linear principal components, non-linear methods based on manifold approximations [8,9] of more structured linear (tensor) methods incorporating various types of variations such as pose and illumination [6]. These 2D models and used to segment 2D images using various types of optimization techniques. Often particular model parameters can be related (dy design or statistically) to pose and illumination variation. It is quite possible in some cases to reach 4-30 Hz speed on standard PC hardware[7]. In an appendix 2d face segmentation using the algorithm in[7] is illustrated

Figure 1: Minolta VIVID 910 3D laser/camera scanner (www.minolta.com) and scanned face

Figure 2: 3DMD multi camera/structures light scanner (from www.3dmd.com)

2. By modelling the face in 3D a number of advantages is achieved. First of all pose and illumination parameters may directly be separated from the biological/expression face variation. Such models can be build from 3D scanning images obtained from 3D scanning devices. The most commonly used technologies are based on laser/camera triangulation (as for the Minolta VIVID 9xx scanner series, cf. Figure 1) or from triangulation using multiple camera and sequences of random (structured) light patterns (as for the 3dmd [10] products

www.3dmd.com, cf. Figure 2). The Minolta scanner scans slowly (~30s /scan), the 3dmd is faster but heavy computations are necessary for the 3d image reconstruction. Also solutions using multiple cameras (calibrated and non-calibrated) are used. However, in the later case the stereo problem or the structure for motion problem will also have to be solved. Regularization is necessary in these cases to arrive at dense depth maps. In the appendix figure 8 & 9 illustrate how 3D models can be used to separated the pose and illumination parameters (cf. DTU papers [14,15]). These models are based on scans from the Minolta VIVID 910 instrument. In general , the matching of the 3D model to 2D data is an ill posed problem. Often stochastic relaxation algorithms including particle filtering, Gibbs sampling, RANSAC must be applied in order to reach the optimum. Particularly, occlusion is difficult to handle.

3D camera advantages

The use of combining 3D models with segmentation and tracking 3D images as is proposed in the ARTTS will address a series of shortcomings of the current state-of-the-art.

1. reliable curvature descriptors may be derived directly from the images and need not be (unreliably?) inferred from intensity patterns or prior models. Curvature descriptors carry information for tracking and recognition [5].
2. the 3D TOF camera supplies true 3D data almost instantaneously at normal a standard video rate (30 Hz at close range) which is a (vast) improvement over current 3D imaging equipment as described above.
3. Occlusions (one object moving in front of another) are easily detected by sharp discontinuities in the depth map.
4. To a large extent the 3D TOF camera will be invariant to external illumination changes due the sensor being active (ie containing it own light source. A light source that I n principle avoids shadows because of its alignment with optical axis of the system.

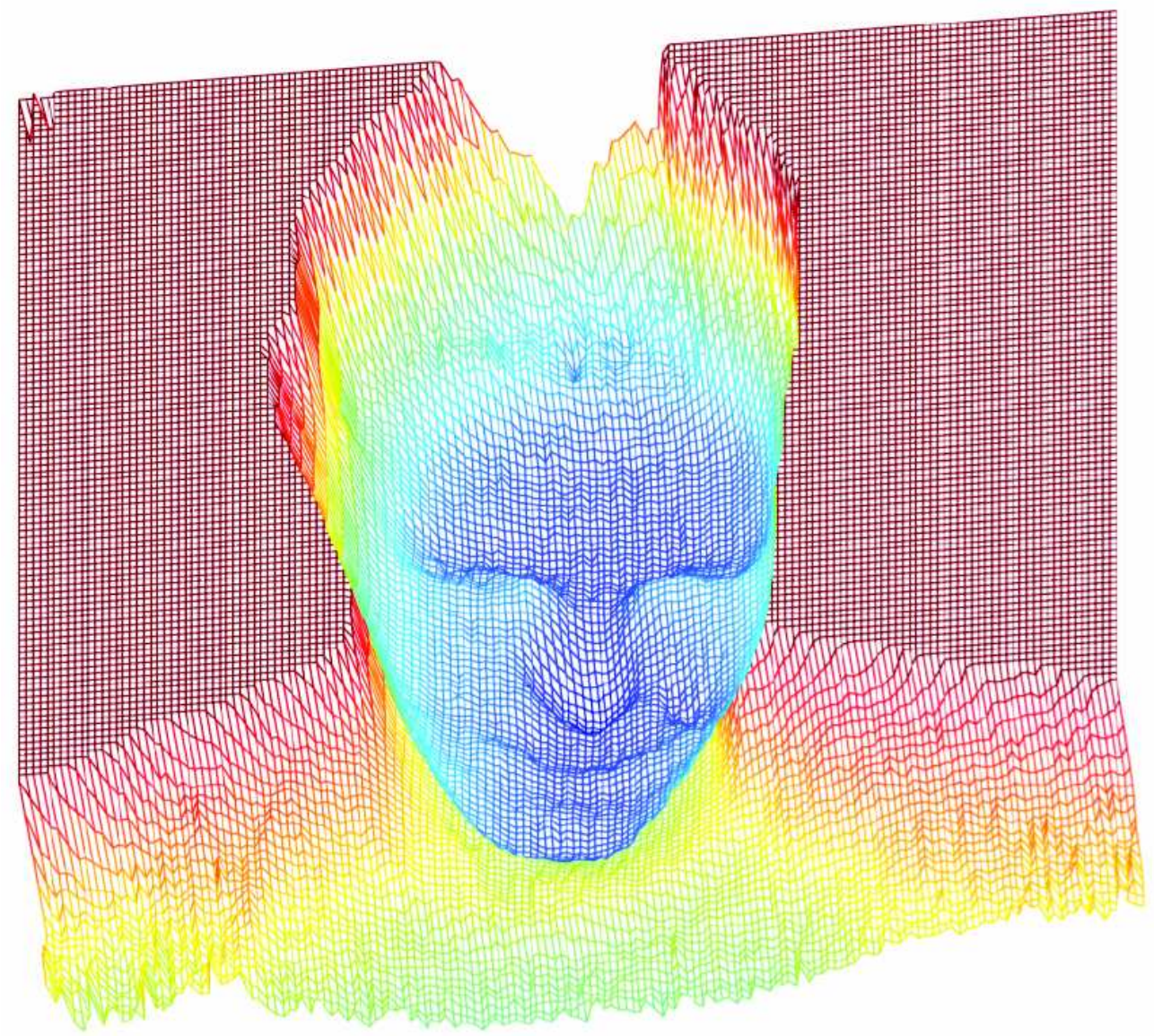


Figure 3: swiss ranger depth map

References

1. Wang J.J.1; Singh S. (2003) Video Analysis of Human Dynamics a Survey. Real-Time Imaging, Volume 9, Number 5, October 2003, pp. 321-346(26)

2. Zhiwei Zhu; Qiang Ji (2004) Real Time 3D Face Pose Tracking From an Uncalibrated Camera, The First IEEE Workshop on Face Processing in Video, Computer Vision and Pattern Recognition
3. Le Lu, Xiang-tian Dai, Gregory Hager (2004) A Particle Filter without Dynamics for Robust 3D Face Tracking. The First IEEE Workshop on Face Processing in Video, Computer Vision and Pattern Recognition
4. Fadi Dornaika and Jörgen Ahlberg (2003). FAST AND RELIABLE ACTIVE APPEARANCE MODEL SEARCH FOR 3D FACE TRACKING. Proceedings of Mirage 2003, INRIA Rocquencourt, France, March, 10-11 2003
5. Kevin W. Bowyer, Kyong Chang, Patrick Flynn (2006) A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. Computer Vision and Image Understanding Vol 101, pp 1–15
6. Terzopoulos, Demetri; Lee, Yuencheng; Vasilescu, M. Alex O. (2004). Model-based and image-based methods for facial image synthesis, analysis and recognition. Proceedings - Sixth IEEE International Conference on Automatic Face and Gesture Recognition FGR 2004
7. M. B. Stegmann, B. K. Ersbøll, R. Larsen, FAME - A Flexible Appearance Modelling Environment, IEEE Transactions on Medical Imaging, vol. 22(10), pp. 1319-1331, Institute of Electrical and Electronics Engineers (IEEE), 2003
8. Tenenbaum, J. B., Silva, V. de, & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290, 2319–2323.
9. Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290, 2323–2326.
10. 3DMD Systems. 3q qlonerator. <http://www.3q.com/offerings_prod.
11. G. Gordon, Face recognition based on depth and curvature features, Computer Vision and Pattern Recognition (CVPR) (June) (1992) 108–110.
12. C. Hesher, A. Srivastava, G. Erlebacher, A novel technique for face recognition using range imaging, in: Seventh International Symposium on Signal Processing and Its Applications, 2003, pp. 201–204.
13. G. Medioni, R. Waupotitsch, Face recognition and modeling in 3D, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003), October 2003, pp. 232–233.
14. K. Skoglund, R. Larsen, B. Lading, Building a 3-D Appearance Model of the Human Face, DSAGM, DIKU (Datalogisk Institut), 2003
15. R. Larsen, K. B. Hilger, K. Skoglund, S. Darkner, R. R. Paulsen, M. B. Stegmann, B. Lading, H. Thodberg, H. Eiriksson, Some Issues of Biological Shape Modelling with Applications, 13th Scandinavian Conference on Image Analysis (SCIA), Gothenburg, Sweden, vol. 2749, pp. 509-519, Springer, 2003

Appendix

2D AAM face model

A face model is build form at set of training images annotated using a set of landmarks as shown in Figure 4

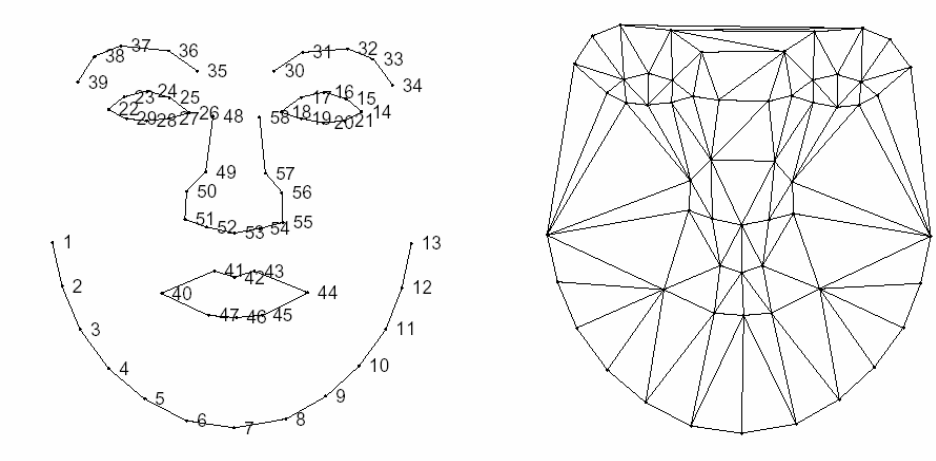


Figure 4: landmark positions and the convex hull of landmarks, which is the patch that is modelled

The variation of the landmark position and modelled by a few significant modes of variation as illustrated in Figure 5. In Figure 6 the most important intensity variation are shown. Figure 7 shows the segmentation in progress.

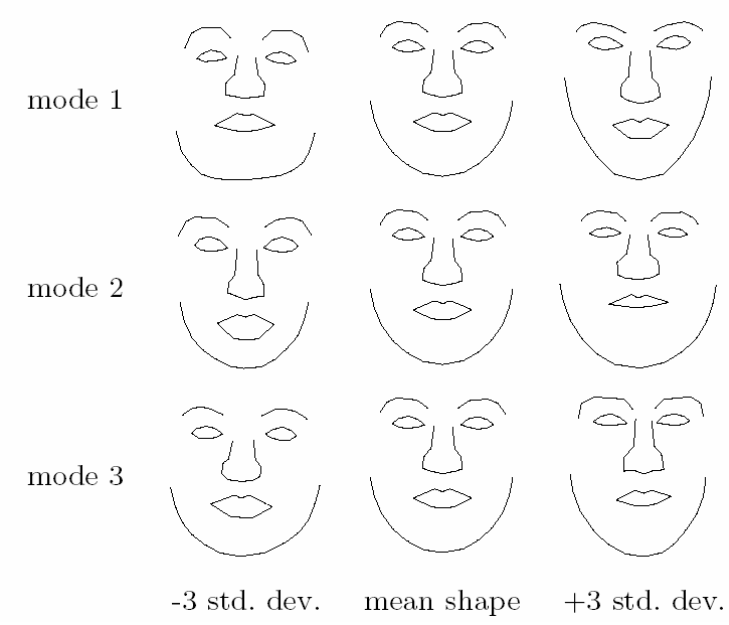


Figure 5: 3 most important facial modes of shape variation

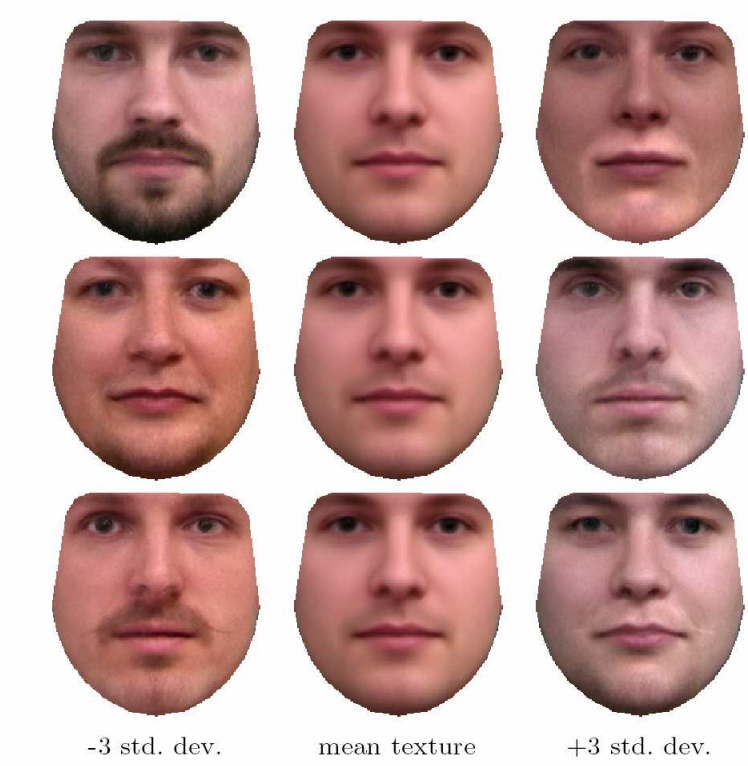


Figure 6: 3 most important facial modes of intensity variation

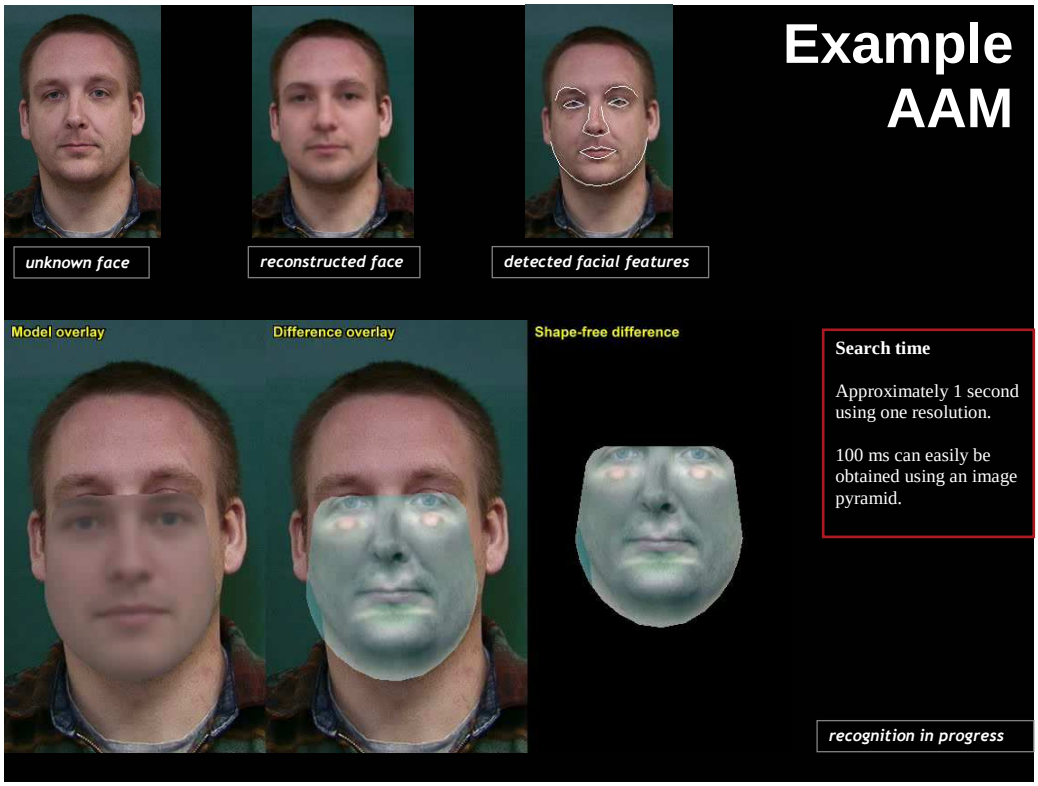


Figure 7: optimizing model parameters to fit to image data yields a segmentation/recognition/pose estimate

3D AAM face model

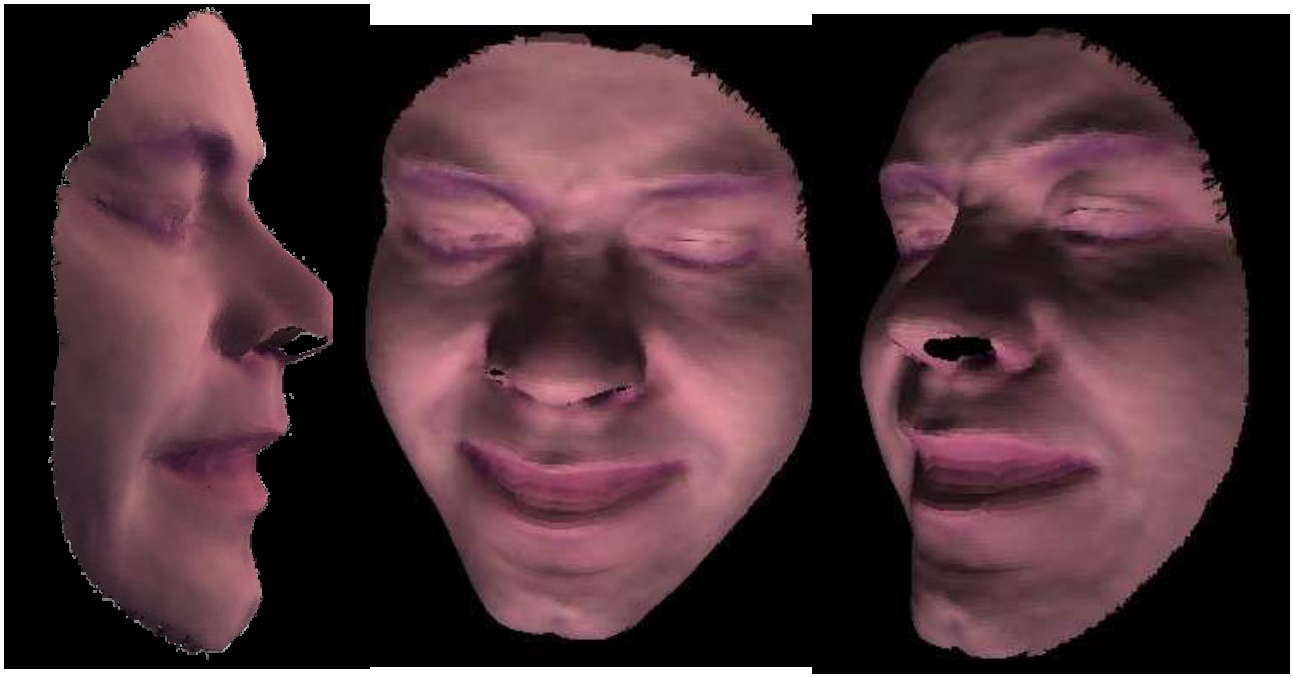


Figure 8: using a 3D model pose variation can be separated

Figure 9: using a 3D model illumination variation can be separated