3D FACE APPERANCE MODEL

B. Lading, R. Larsen

Informatics and Mathematical Modeling
Technical University of Denmark
Kgs. Lyngby, Denmark
{bl, rl}@imm.dtu.dk

and K. Åström

Department of Mathematics
Lund Institute of Technology
Lund, Sweden
{kalle}@maths.lth.se

ABSTRACT

We build a 3d face shape model, including inter- and intra-shape variations, derive the analytical jacobian of its resulting 2d rendered image, and show example of its fitting performance with light, pose, id, expression and texture variations.

1. INTRODUCTION

Building models which are able to capture the variations of faces in images is an extensive research field. Most work have been focusing on shape and appearance modeling based on data from 2d images. These approaches includes the eigenmode face model by turk [1] and the widely used asm and aam models originally by Cootes et al. [3, 2]. In resent years the 3d laser scanner and the structured light scanner have given researchers the ability to record and construct 3d shape models. Most notably, in this new field, being the work done by Blanz and Vetter [4, 5, 7].

Face models could be useful for expression and emotion recognition, facial conversation agents, animation in film and gaming industry, compressed data transfer for video conferencing, and for identification. Although these fields have been active for decades no robust scheme have yet been proposed.

2. DATA

We have collected 35 laser scans of one id with different facial muscle pertubations, and 24 laser scans of different ids with neutral expressions. All scans where aquired using a Minolta Vivid 900 laser scanner. Shapes aquired are edited, surrounding points are deleted, holes filled, and the polygonal mesh between lips are seperated. The final registered, smoothed, and reduced models contain 1595 vertices connected in 3049 triangles.

3. METHODS

In the following we will develop a 3D face model accounting for facial identity (ID), facial expression, lighting and pose, as well as a scheme for optimizing the resulting model image to a 2d face image.

3.1. Shape set surface registration

The expression model should account for an arbitrary expression. In order to train the model we have sampled extreme pertubations of one ID. We assume that most expressions can be generated from interpolating these extremes. However, the extreme pertubations makes surface registration a formidable task. Standard algorithms such as iterated closest point [8, 9] fails. Therefore, we have developed a semi-automated iterative registration procedure for sets of surface shapes. In addition to hole filling of all shapes the procedure requires annotation of a small set of landmarks on the shape set. In our example we use 13 landmark points.

The procedure has three steps. In the first step we construct a series of geodesic distance representations of each shape. In our example we use 3 geodesic representation. Each of these representations consists of an embedding of the geodesic distances of each surface point to 3 pre-selected landmarks in a 3D Euclidean space. We select a prototype shape. In each of the embedding spaces and the original Euclidean space - for all shapes - we apply a thin plate spline warp [10] to the prototype to align all landmarks followed by a closest surface point registration. A resistant mean registration for all surface points is calculated.

In the second step we extract a subset of these registered shapes and build a principal component based point distribution model (PDM) [6]; $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi b}$.

In the third step the crude PDM is fitted to the remaining shapes; $\hat{\mathbf{b}} = [\mathbf{\Phi}^T \mathbf{\Phi}]^{-1} \mathbf{\Phi}^T (\mathbf{x} - \bar{\mathbf{x}}_0)$, and closest surface point registration is carried out. Steps 2 and 3 are iterated until convergence.

In Figs. 1a and 1b we see the prototype shape and a unregistered shape. Figs. 1c and 1d one of the geodesic dis-

tance embeddings are shown for the two shapes. In Figs. 1e and 1g the registration based of Euclidean space only is shown. The combined geodesic embedding registration is shown in Figs. 1f and 1h. The latter is obviously superior.
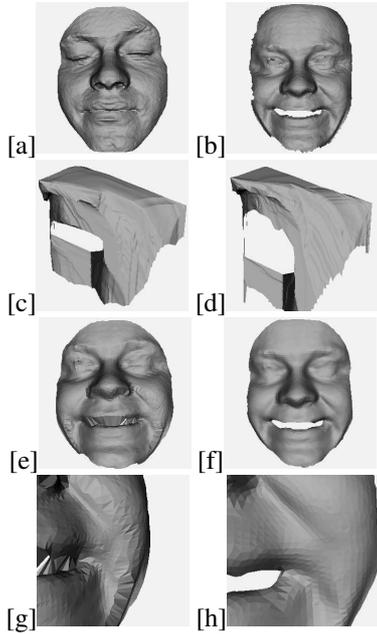

Fig. 1. Registration in different spaces. See text for explanation.

Based on the final PDM we can even - given suitable initialization - register new shapes that are not annotated. Two examples from the ID and the expression database is shown in Fig. 2.
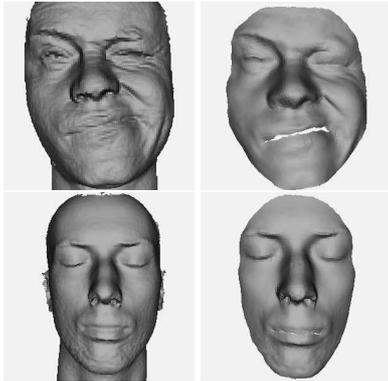

Fig. 2. Automatic registered shapes: **top:** Expression. **bottom:** ID.

3.2. ID and expression 3D shape model

The face expression is derived as principal component point distribution model around a prototype neutral face expres-

sion shape,

$$x = \bar{x}_{\mathrm{ID}} + \Phi b.$$

Here $\bar{x}_{\mathrm{ID}}$ is a neutral face from the ID database, and the principal modes of variation Φ are estimated from the single ID face expression database. Φb expresses the facial animation.
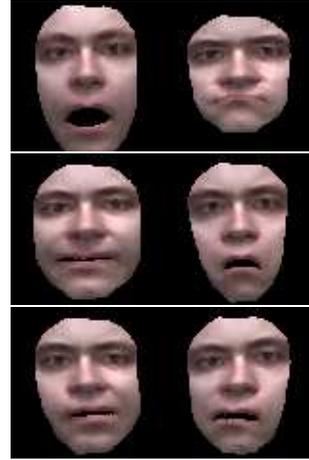
Fig. 3. Most important variations in the expression dataset.

In Fig. 3 the three most dominant modes of variation accounting for 84 % of the training set variability are show. Each row shows ± 2 standard deviations of each mode of variation.

In Fig. 4 the same modes of variations are applied to a different ID from the ID database. We see that the facial expressions are transferred to another ID in a visually meaningful way.
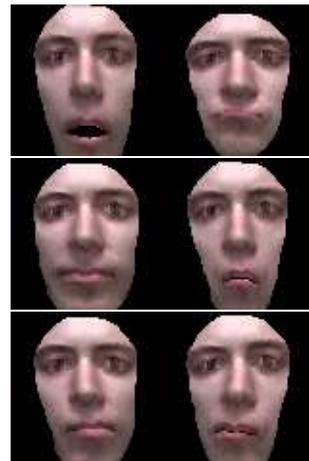
Fig. 4. Shape variations from figure 3 applied to another ID (The ID is the same as the id shown the bottom of figure 2).

3.3. Fitting the model

The complete model includes parameters describing facial deformation, id, pose and lights. With the positioning of one directional light source and the Phong illumination model [11] to add light effects to our scene, we have 7 light parameters, 7 pose parameters, and 8 shape parameters, collected in a single parameter vector $\mathbf{p}$. The mapped texture of the model is optimized in a seperate fashion. That is when the pose and the lights are approximately correct we optimize the texture by the use of a texture pca model, with 10 modes of variation.

Defining the error vector, $\mathbf{r}(\mathbf{p})$, as

$$\mathbf{r} = \mathbf{I}_{model}(\mathbf{p}) - \mathbf{I}_{sample}$$

we may calculate the estimated parameter displacement by use of modified Newton optimization ($\delta\mathbf{p} = -\left(\mathbf{J}^T\mathbf{J} + \epsilon\mathbf{I}\right)^{-1}\mathbf{J}^T\mathbf{r}$). The Jacobian may be approximated numerically by perturbing in parameters space, but since this approach is computationally expensive, and since we have complete knowledge of the rendered model, we derive an analytical expression for the elements of the Jacobian.

Writing our image of the model as a product of an image describing the lights, $\mathbf{I}_l$, and an image of the mapped texture values, $\hat{\mathbf{I}}_m$, we have

$$
\begin{aligned}
\frac{\partial \mathbf{r}}{\partial p_i} &= \frac{\partial}{\partial p_i}\left(\mathbf{I}_l(\mathbf{p})\hat{\mathbf{I}}_m(\mathbf{p}) - \mathbf{I}_{sample}\right) = \frac{\partial \mathbf{I}_l}{\partial p_i}\hat{\mathbf{I}}_m + \mathbf{I}_l\frac{\partial \hat{\mathbf{I}}_m}{\partial p_i} \\
&= \frac{\partial \mathbf{I}_l}{\partial p_i}\hat{\mathbf{I}}_m + \mathbf{I}_l\left(\frac{\partial \hat{\mathbf{I}}_m}{\partial \mathbf{s}}\right)^T\frac{\partial \mathbf{s}}{\partial p_i}.
\end{aligned}
$$

(we also include a self-shadowing map, but variations in this map are not included in the Jacobian).

When fitting the model, a good initial pose and light position need to be applied for the fitting to converge to an acceptable minima. So far all fitting is done in a handheld fashion. A standard optimization is as follows; first we let the system converge in the pose regime, then we apply fitting for the id, next the Phong-lighting parameters and position of the light source, and finally the texture and expression. If result is not satisfactory we iterate. Optimization may be done in different imagebands (I,R,G,B,H,S,V) and different combinations of these, but a superior imageband combination have yet to be found. Better optimization results are sometimes observed when the optimization is done on a hierachy of Gaussian blurred images. It is noted that texture optimization is only applied if the sample image is in ambient light.

Fig. 5. Results examples for the optimization of pose, light position and light parameters.
Left column shows the original sample image, center column is the model, and the right column shows the model overlayed the sample image.

4. RESULTS

Figure 5 shows the optimization results for an image sequence with differences in head pose (since the pose difference between image samples is large, we optimized the samples independently). We observe that the pose and the lights are fitted nicely, even for the extreme rotations. It is noted that the subject in this sequence is identical to the subject from which all expression data samples was obtained.

In figure 6 we have optimized the lights, id, expression and texture for three different ids. In the first optimized image (Fig 6a) we observe that the ID model seem inadequate to model the mouth, and it relies on the expression model to fit this region, which in turn causes the model to get stuck in a non-optimal minima. The sample image is diffusively lit and texture optimization is therefore applied.
In the second image (Fig 6b) we get better results, but still not convincing.
The last optimization is without the ID model (not needed since the subject of sample image is the subject of the expression data set), and show a good approximation of the shape and lights on both chin and cheeks. In the last image texture optimization is not done since shadowing is extensive.
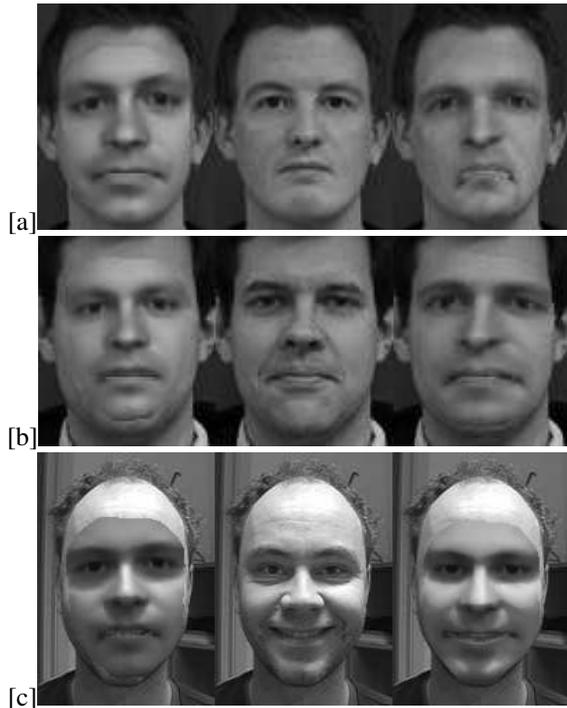
Fig. 6. Some results on the optimization of ID, texture, lights and expression.
Left column shows the initial state, center column is the sample image and the last column shows the optimized model image overlayed the sample image.

5. DISCUSSION

We have presented a technique for the registration of 3d shapes, a shape model representing facial expressions and an ID model, and we described an algorithm used to optimize these models on 2d face images, including the light environment and the face pose.

Future developments will focus on searches for different and better minimas (these may be tracked down by following some of the eigenvectors of the Hessian, see eg. [12]), a new texture optimization taking into account the cast shadows (we have a good estimate of the shadow map, and need to substract this from the sample image), and a technique to learn the texture to be mapped for the mouth region and the eyes (from a set of optimized images, we may build a new model for the texture which applies to these vertices).

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] M. Turk and A. Pentland (1991b). *Face recognition using eigenfaces.* In Proceedings of Computer Vision and Pattern Recognition, pages 586–591. IEEE.

[2] T.F. Cootes, G.J. Edwards and C.J. Taylor, *Active Appearance Models.* Proc. European Conference on Computer Vision, vol 2, p 484-98, 1998.

[3] T. Cootes, D. Cooper , C. Taylor and J. Graham *Active shape models— their training and application.* Computer Vision andImag Understanding 1995;61(1):38–59.

[4] V. Blanz and T. Vetter, *A Morphable Model for the Synthesis of 3D Faces.* Siggraph 1999, Computer Graphics Proceedings, p. 187-194 (1999)

[5] V. Blanz, C. Basso, T. Poggio and T. Vetter, *Reanimating faces in Images and Video.* Eurographics, vol 22, no. 3, 2003.

[6] T. J. Hutton, B. F. Buxton and P. Hammond, *Dense Surface Point Distribution Models of the Human Face.* IEEE CVPR, 2001.

[7] S. Romdhani and T. Vetter, *Efficient, Robust and Accurate Fitting of a 3D Morphable Model*, Procedings of the ninth IEEE International conference on Computer VISION, 2003.

[8] XXBL CORRECT? XXBL Besl, P. J. and McKay, N. D. (1992). *A method for registration of 3-d shapes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):239256.

[9] XXBL CORRECT? XXBL

[10] Bookstein FL: *Principal warps: thin-plate splines and the decomposition of deformations.* IEEE Transactions on Pattern Analysis and Machine Intelligence 1989;11:56785.

[11] Phong, Bui Tuong: *Illumination of Computer-Generated Images*, Department of Computer Science, University of Utah, UTEC-CSs-73-129, July 1973

[12] C. Sminchisescu and B. Triggs, *Building Roadmaps of Minima and Transitions in Visual Models*, ECCV 2002.