Generating 3D Virtual Populations from Pictures of a Few Individuals

WonSook Lee[1], Pierre Beylot[1], David Sankoff[2], and Nadia Magnenat-Thalmann[1]

[1] Miralab, Centre Universitaire d'Informatique, University of Geneva, 24, rue Général Dufour, CH 1211, Geneva 4, Switzerland, {wslee,beylot,thalmann}@cui.unige.ch, http://miralabwww.unige.ch/

[2] Centre de recherches mathématiques, Université de Montréal, CP 6128 Montréal H3C Québec, sankoff@ere.umontreal.ca

Abstract. This paper describes a method for cloning faces from two orthogonal pictures and for generating populations from a small number of these clones. An efficient method for reconstructing 3D heads suitable for animation from pictures starts with the extraction of feature points from the orthogonal picture sets. Data from several such heads serve to statistically infer the parameters of the multivariate probability distribution characterizing a hypothetical population of heads. A previously constructed, animation-ready generic model is transformed to each individualized head based on the features either extracted from the orthogonal pictures or determined by a sample point from the multivariate distribution. Using projections of the 3D heads, 2D texture images are obtained for individuals reconstructed from pictures, which are then fitted to the clone, a fully automated procedure resulting in 360° texture mapping. For heads generated through population sampling, a texture morphing algorithm generates new texture mappings.

1 Introduction

Animators agree that the most difficult subjects to model and animate realistically are humans and particularly human faces. The explanation resides in the universally shared (with some cultural differences) processes and criteria not only for recognizing people in general, but also for identifying individuals, expressions of emotion and other facial communicative signals, based on the covariation of many partially correlated shape, texture and movement parameters within narrowly constrained ranges. There are now a number of computer animation technologies for the construction of 3D virtual clones of individuals, or for the creation of new virtual actors [1–5, 7–9, 11, 13]. It is now of increasing interest in many applications to be able to construct virtual groups, crowds or populations of distinct individuals having some predefined general characteristics. Simple approaches, such as the random mix and match of features, do not

F. Dehne, A. Gupta, J.-R. Sack, and R. Tamassia (Eds.): WADS'99, LNCS 1663, pp. 134-144, 1999

take into account local and global structural correlations among facial sizes and structures, distances between features, their dimensions, shapes, and dispositions, and skin complexion and textures. In Section 2 of this paper, we describe an efficient and robust method for individualized face modeling, followed by techniques for generating animation-ready populations in a structurally principled way. The idea is to infer the statistical characteristics of a population from pairs of photographs (front and side views) of a relatively small input sample of individuals. The characteristics measured are the determinants of shape in a procedure for reconstructing individual heads through deformations of a generic head. The hypothetical population from which the input sample was drawn can then be represented by a small number of eigenvectors and eigenvalues. Any number of other head shapes can then be rapidly obtained by random sampling from this inferred population followed by application of the deformation method. At the same time, the original photos for each individual are used to construct a 2D texture image which can then be applied either to the corresponding head, or together with texture images from the other photographed heads in random proportion, to the new heads output from the random sampling of the hypothetical population. This texture mapping is described in Section 3. The results are shown in Section 4.

2 A hypothetical population of head shapes

There are precedents for several aspects of our method, including modeling from photographs, with feature detection and generic head modification [2–4, 8]. Free Form Deformations have been used to create new heads from a generic model [4]. Statistical methods were used by DeCarlo et al. [6], randomly varying distances between facial points on a geometric human face model, guided by anthropometric statistics. In the present research, however, we combined and improved selected elements with a view to a fast, flexible and robust method for creating large numbers of model head shapes typical of a given population. Speed means not having to manually detect too many feature points per input sample head, a restricted number of heads in the sample, and then completely automated processing; flexibility requires a method applicable to any sort of population, with no requirement for anthropometric or other data bank; and robustness allows for a range of non-studio photographic input and statistical procedures that work for small samples without generating aberrant shapes.

2.1 Feature detection

To reconstruct a photographically realistic head, ready for animation, we detect corresponding feature points on both of two orthogonal images – front and side – and from these deduce their 3D positions. This information is to be used to modify a generic model through a geometrical deformation. Feature detection is processed in a semiautomatic way (manual intervention is sparing and efficient) using the structured snake method [4] with some anchor functionality. Figure 1

depicts an orthogonal pair of images, with feature points highlighted. The two images are normalized so that the front and side views of the head have the same height, as measured by certain predetermined feature points. The two 2D sets of position coordinates, from front and side views, i.e., the (x, y) and the (z, y) planes, are combined to give a single set of 3D points. Outside a studio, it would be rare to obtain perfectly aligned and orthogonal views. This leads to difficulties in determining the (x, y, z) coordinates of a point from the (x, y) on the front image and the (y, z) on the side image. Taking the average of the two y measurements often results in unnatural face shapes. Thus we rely mainly on the front y coordinate, using the side y only when we do not have the front one. This convention is very effective when applied to almost orthogonal pairs of images. In addition, for asymmetrical faces, this convention allows for retention of the asymmetry with regard to the most salient features, even though a single side image is used in reconstructing both the right and left aspects of the face. A global transformation re-situates the 3D feature points (about 160 of them) in the space containing a generic head. A part of feature points are detected by manual intervention and others by snake method [4]. We are now in a position to deform the generic head (which has a far more detailed construction than just 160 feature points) so that it becomes a model for the photographed head. However, the deformation process will be identical for these heads as for the heads generated by our statistical sampling procedure, so we will describe this procedure first.

2.2 Constructing and sampling from the hypothetical population

Our approach to generating populations is based on biostatistical notions of morphological variation within a community. The underlying hypothesis is that if we determine a large number of facial measurements, these will be approximately distributed in the population according to a multivariate normal distribution, where most of the variation can be located in a few orthogonal dimensions. These dimensions can be inferred by principal component analysis [21] applied to measurements of relatively few input heads. A random sample from the reduced distribution over the space spanned by the principal components yields the facial measurements of a new head, typical of the population.

Inference The 160 feature points are divided into predetermined subsets according to the head region where they are defined (mouth, eye, nose, cheek, and etc.). Two sets of pre-specified 3D vectors representing distances between feature points are calculated for each head. The first set reflects the overall shape parameters and consists of distances between a central point situated on the nose and a number of regional "landmarks", each of them belonging to a different region. The second set represents local relationships and corresponds to distances between the landmarks and the other point in the same region. Denote by n the total number of measurements represented by all the distance vectors. The measurements for the H heads are each standardized to $Normal[0, 1]$ and entered

into an $H \times n$ matrix M. The principal components of variation are found using standard procedures, involving the decomposition $XLX^t = MM^t$, where X is orthonormal and L contains the eigenvalues in decreasing order. Only those dimensions of X with non-negligible eigenvalues, i.e. the principal components, are retained. This ensures that we are considering correlations among the measurements for which there is strong, consistent evidence, and neglecting fluctuations due to small sample size.

Sampling For each head in the population being constructed, independent samples are drawn from a $N[0,1]$ distribution for each principal component. The i-th component is multiplied by L_i, where L_i is the i-th eigenvalue in L, and the feature point distance vectors are then constructed by inverting the transformation to X from the original measurement coordinates in n-dimensional space, and then inverting the standardization process for each measurement. It is then straightforward to position all the new feature points, starting with the central point on the nose. Sampling from the principal component space is a rapid method for generating any number of feature point sets.

2.3 Modification of a generic model

We have a certain set of 3D feature points, which has about 160 points. The problem is how to deform a generic model, which has more than a thousand points interconnected through a triangulation pattern, to make an individualized smooth surface. One solution is to use the 3D feature points as a set of control points for a deformation. Then the deformation of a surface can be seen as an interpolation of the displacements of the control points. The particular deformation we use is Dirichlet Free-Form Deformations (DFFD)[19] in which the position of each surface point is interpolated from that of a number of control points through the Sibson natural neighbors coordinate system [15]. The latter is determined by the structure of the generic head. This is a rapid but rough method. It does not attempt to locate all points on the head exactly, in contrast to automated scanning methods, for example a laser scanner, which create enormous numbers of points but do not precisely identify the control points necessary to compare heads and to link up with adptive mesh technology[5]. Considering the input data (pictures from only two views), the result is quite respectable. More important, it greatly limits the size of the data set associated with an individual head, and hence processing time, as is necessary in animation technology. The imprecision in head shape can be almost compensated for by the automatic texture mapping procedures described in the next section

3 Texture mapping

Texture mapping imbues the face with realistic complexion, tint and shading, and in so doing, it also disguises the approximate nature of shape modeling determined by feature point identification only. Texture data is captured from

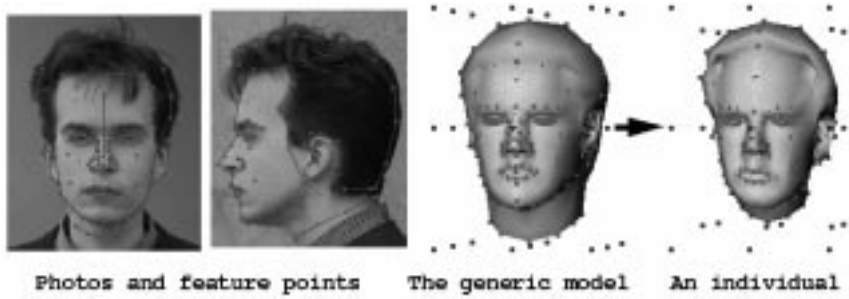


Fig. 1. Modification of a generic head according to feature points detected on pictures. Points on a 3D head are control points for DFFD.

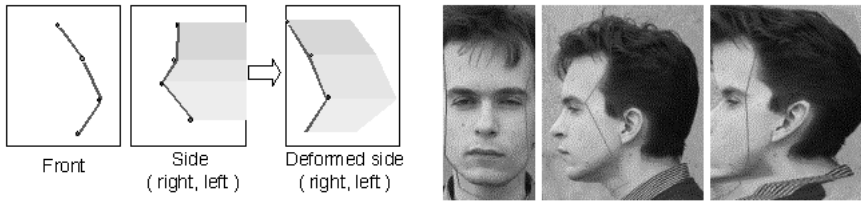
the two photographic views and a single composite texture image is produced by joining them together, with the help of a "multiresolution" smoothing technique. The set of feature points identified as in Section 2.1 helps us to project all points on a 3D head to the 2D image. The triangulation of (or inherited from) the generic head defines texture regions in the image which can then be mapped back to the surface of the head.

3.1 Texture generation

We first connect the two pictures along predefined feature lines, i.e. connecting predetermined feature points which are passed from feature detection process, using geometrical deformations and, to avoid visible boundary effects, a multiresolution technique. This process is fully automatic.

Image deformation We privilege the front view, since it provides the highest resolution for facial features. The side view is deformed to join the front view along certain defined feature points lines on the left hand side and, flipped over, on the right hand side. The feature lines are indicated on the front image in Figure 2 by thick lines. A corresponding feature line is defined for the side image. We deform the side image so that the feature line lines up with the one on the front view. Image pixels on the right side of the feature line are transformed in the same way as the line itself. To get the right part of the image, we deform the side image according to the right-hand feature line on the front image. For the left part of the image, we flip the side image and deform it according to the left-hand feature line on the front image. The resulting three images are illustrated in Figure 3 (a). A piecewise linear transformation is depicted, based on piecewise feature lines, but smoother deformations are easily produced using higher degree feature curves. This geometrical deformation guarantees feature points matching between front and side images, while a simple blending on some overlapping area or conventional cylindrical projection of front and side views [1–3] creates unexpected holes in the final texture image.

Multiresolution image mosaic No matter how carefully the picture-taking environment is controlled, in practice boundaries are always visible between the three segments of the texture image, as in Figure 3 (a). To correct this, the three images resulting from the deformation are merged using multiresolution [10]. Figure 3 (b) shows how this technique is effective in removing the boundaries between the three images.



(a) (b)

Fig. 2. (a) Thick lines are feature lines. (b) Feature lines on three images.

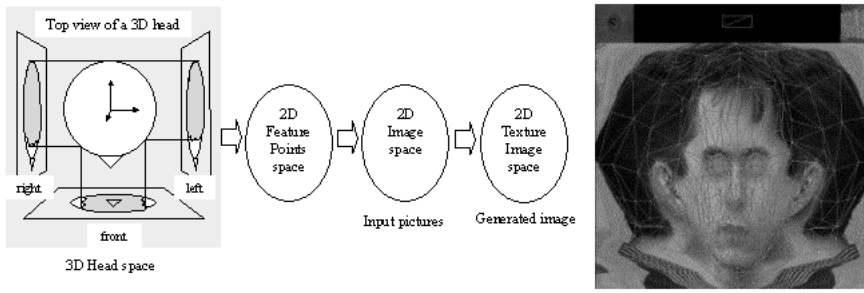
(a) (b)

Fig. 3. Combining the texture images generated from the three (front, right and left) images without multiresolution techniques, in (a) and with the technique in (b).

3.2 Texture fitting

To find suitable coordinates on the combined image for every point on a head, we first project an individualized 3D head onto three planes as shown in Figure 4 (a). We are guided by the feature lines of Section 3.1 to decide to which plane a point on a 3D head is to be projected. This helps us find texture coordinates and the mapping of points on the integrated texture image. The final texture fitting on a texture image is shown in Figure 4 (b). This results in smoothly connected images inside triangles of texture coordinate points, which are accurately positioned. Eyes and teeth are added automatically, using predefined coordinates and transformations related to the texture image size. The triangles in Figure 4 (b) are projections of triangular faces on a 3D head. Since our generic model is endowed with a triangular mesh, the texture mapping benefits from an

efficient triangulation of the texture image containing finer triangles over the highly curved and/or highly articulated regions of the face and larger triangles elsewhere, as in the generic model.



(a) (b)

Fig. 4. Texture fitting giving a texture coordinate on an image for each point on a head. (b) Texture coordinates overlaid on a texture image.

3.3 Textures for statistically generated heads.

For head shapes newly created through sampling in the principal component space, we create texture by combining textures of some of the input sample heads in various, possibly random, proportions. Because of the common triangulation inherited from the generic head, each point on the surface of a new head can be identified with a point on the texture image of each of the input sample heads. The position of each pixel in the triangle that contains it can be written in barycentric coordinates, and it can then be identified with corresponding pixels (i.e. with the same coordinates, in the corresponding triangle) in each of the contributing texture images. The color value of the pixel is the sum of the values of the corresponding pixels in the contributing texture images, weighted by the given proportions. Smoothing of the image pixels is achieved through bilinear interpolation among four neighboring pixels.

4 Results

4.1 Cloning

Figure 5 shows several views of the head reconstructed from the two pictures in Figure 1.

Other examples covering wide range of age and ethnic group are shown in Figure 7. Every face in this paper is modified from the SAME generic model shown in Figure 1. How individualized the representations are depends on how many feature points are identified on the input pictures. We routinely use about 160 feature points including many on the eyes, nose and lips. Some points are allotted to face and hair outlines, but our generic model currently does not have many points on the "hairdo", so it is not easy to vary hair length, for example.

Fig. 5. Several views of a reconstructed head.

4.2 Real-time animation

The predefined regions [20] of the generic model are associated with animation information, which can be directly transferred to the heads constructed from it by geometrical modification in Section 2. Figure 6 shows several expressions on a head reconstructed from three pictures, one for front and others for side. This extension with input up to four images (two for front and two for side) is a generalized method separating shape and texture input sources.

Fig. 6. Examples of reconstructed heads and several expressions.

4.3 Creating a population

We used seven orthogonal photo sets as inputs. Figure 7 shows the input photos and output reconstruction. Included are four Caucasians of various ages, an Indian, an Asian, and an African. There are three adult females, three adult males and a child. The creation of a population is illustrated in Figure 8. The eight faces are drawn from the many dozens we generated from the $H = 7$ sets of orthogonal photos in Figure 8 according to the steps in Section 2. All faces are animation-ready. The size of the texture image for each person is 256×256, which has less than 10 KB. The total amount of data for the heads in OpenInventor format is small considering their realistic appearance. The size of Inventor format (corresponding to VRML format) is about 200 KB. The texture image is stored in JPEG format and is from $5 \sim 50$ KB in size, depending on the quality

Fig. 7. Examples of reconstructed heads from pictures. These are ready for immediate animation in a virtual world.

of pictures; all examples shown in this paper have size less than 45 KB. The number of points on a head is 1257, where 192 of them are for teeth and 282 of them are for the eyes. This leaves only 783 points for the individualization of the rest of the facial surface.

Fig. 8. Virtual faces created from seven faces reconstructed from images.

5 Conclusion

We have introduced a suite of methods for the generation of large populations of realistic faces, enabled for immediate real-time animation, from just a few pairs of orthogonal pictures. One key to our technique is the efficient reconstruction of animation-ready individualized faces fitted with seamless textures. This involves shape acquisition through the modification of a generic model and texture fitting through geometric deformation of an orthogonal pair of texture images, followed by multiresolution procedures. This technique was robust enough to allow one

operator to clone some 70 individuals in five days in public demonstration of a computer fair. The procedure is universal, applicable to men and women, adults and children, and different races, all using the same generic model. To generate a population from a small number of heads such as those produced by the reconstruction technique, the first step is to characterize the shape in more detail using vectors between feature points and to calculate the correlation matrix of these measurements. Principal component analysis is then applied to discover the statistical structure of this input sample, namely a representation of the data in terms of a reduced number of significant (and independent) dimensions of variability. Each point in this space, for example one chosen at random according to the probability distribution inferred from the input, determines all the feature points and other characteristics of a new head. The representation of a population as a probability distribution has great potential for allotting variation among face shapes into gender, age, race and residual components with eventual feedback to more realistic and efficient modeling.

6 Acknowledgments

The authors would like to thank other members of MIRALab for their help, particularly Laurent Moccozet and Hyewon Seo. This project is funded by an European project eRENA and Swiss National Research Foundation.

References

1. Tsuneya Kurihara and Kiyoshi Arai, "A Transformation Method for Modeling and Animation of the Human Face from Photographs", In Proc. Computer Animation'91, Springer-Verlag Tokyo, pp. 45-58, 1991.
2. Takaaki Akimoto, Yasuhito Suenaga, and Richard S. Wallace, Automatic Creation of 3D Facial Models, IEEE Computer Graphics & Applications, Sep., 1993
3. Horace H.S. Ip, Lijin Yin, Constructing a 3D individual head model from two orthogonal views. The Visual Computer, Springer-Verlag, 12:254-266, 1996.
4. Lee W. S., Kalra92 P., Magnenat-Thalmann N, "Model Based Face Reconstruction for Animation", In Proc. Multimedia Modeling (MMM'97), World Scientific, Singapore, pp. 323-338, 1997.
5. Yuencheng Lee, Demetri Terzopoulos, and Keith Waters, "Realistic Modeling for Facial Animation", In Computer Graphics (Proc. SIGGRAPH'96), pp. 55-62, 1996.
6. Douglas DeCarlo, Dimitris Metaxas and Matthew Stone, "An Anthropometric Face Model using Variational Techniques", In Computer Graphics (Proc. SIGGRAPH'98), pp. 67-74, 1998.
7. Brian Guenter, Cindy Grimm, Daniel Wood, "Making Faces", In Computer Graphics (Proc. SIGGRAPH'98), pp. 55-66, 1998.
8. Frederic Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, David H. Salesin, Synthesizing "Realistic Facial Expressions from Photographs", In Computer Graphics (Proc. SIGGRAPH'98), pp. 75-84, 1998.
9. http://www.turing.gla.ac.uk/turing/copyrigh.htm
10. Peter J. Burt and Edward H. Andelson, "A Multiresolution Spline with Application to Image Mosaics", ACM Transactions on Graphics, 2(4):217-236, Oct., 1983.

11. Marc Proesmans, Luc Van Gool, "Reading between the lines - a method for extracting dynamic 3D with texture". In Proc. of VRST'97, pp. 95-102, 1997.
12. S.-Y. Lee, K.-Y. Chwa, S.-Y. Shin, G. Wolberg, "Image metamorphosis using Snakes and Free-Form deformations", In Computer Graphics (Proc. SIGGRAPH'95), pp. 439-448, 1995.
13. P. Fua, "Face Models from Uncalibrated Video Sequences", In Proc. CAPTECH'98, pp. 215-228, 1998.
14. Sederberg T. W., Parry S. R., "Free-Form Deformation of Solid Geometric Models", In Computer Graphics (Proc. SIGGRAPH'86), pp. 151-160, 1986.
15. Sibson R., "A Vector Identity for the Dirichlet Tessellation", Math. Proc. Cambridge Philos. Soc., 87, pp. 151-155, 1980.
16. Aurenhammer F., "Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure", ACM Computing Survey, 23, 3, September 1991.
17. Farin G., "Surface Over Dirichlet Tessellations", Computer Aided Geometric Design, 7, pp. 281-292, North-Holland, 1990.
18. DeRose T.D., "Composing Bezier Simplexes", ACM Transactions on Graphics, 7(3), pp. 198-221, 1988.
19. Moccozet L., Magnenat Thalmann N., "Dirichlet Free-Form Deformations and their Application to Hand Simulation", In Proc. Computer Animation'97, IEEE Computer Society, pp.93-102, 1997.
20. Kalra P, Mangili A, Magnenat-Thalmann N, Thalmann D, "Simulation of Facial Muscle Actions Based on Rational Free Form Deformations", Proc. Eurographics'92, pp. 59-69, NCC Blackwell,1992.
21. Kendall, M.G. and Stuart, A. Advanced Theory of Statistics, vol. 3. Griffin, 1976.