ERK 2000

Portorož, Slovenija, 21. - 23. september 2000

Zbornik devete Elektrotehniške in računalniške konference ERK 2000
Proceedings of the ninth Electrotechnical and Computer Science Conference ERK 2000

Zvezek B / Volume B

Računalništvo in informatika / Computer and Information Science
Umetna inteligenca / Artificial Inteligence
Robotika / Robotics
Razpoznavanje vzorcev / Pattern Recognition
Biomedicinska tehnika / Biomedical Engineering
Didaktika / Didactics
Študentski članki / Student Papers

Uredil / Edited by
Baldomir Zajc

Slovenska sekcija IEEE / Slovenian Section IEEE

Building Animated 3D Face Models From Range Data

Žiga Kranjec, Franc Solina
Faculty of Computer and Information Science
University of Ljubljana
Tržaška 25, 1001 Ljubljana, Slovenia
ziga.kranjec@ijs.si, franc.solina@fri.uni-lj.si

Abstract

This paper describes the issues involved in building functional 3D models of individuals with the help of coded light range sensor. A range image, acquired with the range sensor is triangulated, mirrored and zippered to construct a medium resolution face model. A low resolution face template is fitted to 3D data to create a low resolution model. Techniques used for fitting involve finding surface/line intersections and continuous piecewise linear transformations through barycentric coordinates. A texture and a parametric muscle model are applied to create expressions. Medium resolution results are good and low resolution results are acceptable for some applications.

Keywords: building models from sensor data, facial animation.

1 Introduction

Building 3D models of humans for use in animation manually can be a very time consuming process. This is particularly true for head and face, where attention of the observer is usually focused most. Popular approach is to modify templates from a library of several different face prototypes to create a wide variety of characters. For modeling particular individuals, this can require considerable artistic skills. Conformance to particular individual is achieved through a) shape and b) texture. By using a range sensor [3], precise shape of an individual can be acquired. 3D models constructed automatically from range data can bear very good resemblance to particular individual, especially if they are properly textured.

Such models do have several shortcomings:

- they can be relatively large - typical scans we used contained 10000-20000 vertices - and as such less suitable for interactive and real time use, such as games. Optimal face models vary in complexity in different areas, depending on curvature and intensity of articulation in that area. Common polygon reduction algorithms usually don't consider the latter.
- They can be difficult to animate effectively - sophisticated facial animation techniques often require various additional information about the model, such as regions of the face, elastic properties, etc. This information is difficult to construct automatically for 3D models, created directly from range data. Also, it is often useful to model certain features, such as eyes and hair separately, since they do not differ as much among individuals.

We tried to address these problems by modifying a generic face template from the library to conform to the data, acquired with range sensor. A template has known topology and structure and can be animated immediately after conformation.

2 Computer Facial Animation

Computer facial animation beginnings date back to early seventies. In recent years, much research has been done in the area, particularly for use in animation industry. For a good overview, see [2, 5, 10].

Techniques, commonly used for facial animation include:

- Direct parameterization - model modification is specified implicitly with handcrafted procedures, which take arguments (parameters). This technique can be difficult to extend and develop, since it requires much experimentation and interaction between parameters is complex and generally not very well known.

- Physics based modeling - models are assigned some physics-based properties and model and simulation is run. Often, spring based models are used.
- Interpolation - common, fast and powerful technique of relative vertex keys requires one neutral and one or more extreme poses; expressions are generated by adding weighted differences between extreme and neutral poses to the neutral pose. Convincing animation can be generated by using 5-6 poses [6]. Extreme poses can be conveniently generated with one of the former techniques and fine-tuned manually.

Latter techniques are more general and mesh/model independent. Sophisticated animation methods often use a combination of several techniques.

The actual animation technique and representation used for modelling of course depends on the speed/level-of-detail compromise. Note that it is possible to make effective character animation without great overall level of detail. Well done eyes, for example, add much efectiveness. Pelachaud *et al.* [10] describe the major parts of the head/face. Particularly important are: overall head shape, facial mask, mouth, eyes and hair.

For our use, we adopted a parametric pseudo-muscle face model developed by Waters [2] (figure 1). The model is computationally relatively simple and can create wide variety of very intense expressions. Muscle model is actually independent of the facial mesh and thus the same muscles can be used on different models, even the ones, made by use of range sensor. The model was augmented with several conformation parameters to adjust overall proportions of a model in a manner, similar to one described in [1].

2.1 Range data

The individual was scanned with the LRV range sensor [3], from a side angle of about 45 degrees (figure 2). This gave very good results for half a face, almost all relevant data was present. Range image histogram (figure 3) nicely shows two segments – segment in darker, farther-away area belongs to the right shoulder and was removed for our use. The rest of range image was converted to 3D triangular mesh and vertices lying outside specified box were discarded. This removed most of the noise, but left some in the area where left shoulder meets

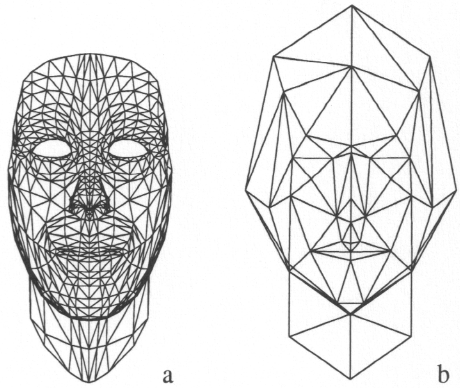

Figure 1: a) Model and b) hull

the neck, which had to be removed interactively. Proper frontal orientation of the mesh was specified by range sensor calibration.

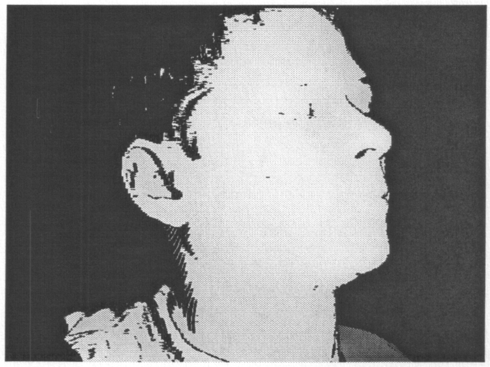

Figure 2: Range image acquired with LRV range sensor

The other half was created by mirroring and zippering both halves together [4]. When zippering, an additional line of vertices was extruded from the border vertices of the mesh and positioned exactly in the center between both halves. These vertices lie on the face profile, which is a very important face characteristic, and provide additional control for further adjustments. Both meshes were then zippered to this line.

Lee *et al.* [1] used data acquired with Cyberware range sensor, which is in cylindrical coordinates and from around the whole head. Cyberware sensor is slower and more expensive in terms of equipment cost than LRV sensor.

Texture from a portrait photograph was applied to the model. This ensured, that the light-

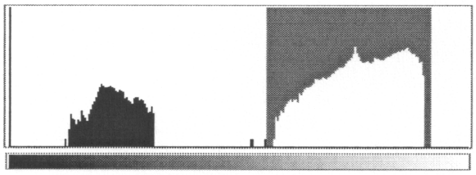

Figure 3: Range image histogram. Interesting data is in the upper part (highlighted)

ing on the texture was uniform, so that the actual light when rendering was provided by OpenGL. Quality portrait photographs are usually readily available. Since a photograph and 3D model belong to the same individual, texture maps very well onto a surface. We used *uv* coordinate mapping, which keeps texture attached to vertices, so only vertex positions need to be modified for animation. Texture coordinates were computed and attached to vertices data structure.

Figure 4: Muscle actions, applied to zippered symmetrical model (2735 vertices, 5225 faces)

A pseudo-muscle model was used to apply several expressions from the expression library. The resulting models (figure 4) can be used as key poses for interpolation. However, this works only for animation procedures, which use only 3D data. These models cannot be used as is for animation, which requires more structural information – opening a mouth, or moving the eyes, for example.

3 Fitting template to 3D data

A template model (figure 1a) was fitted to 3D model. For fitting, we used a method, somewhat similar to ray tracing to find locations on model surface in 3D space. Point P belonging to the template is moved to position P' in the direction r. Point P' is found by choosing a point from a mesh, which is closest to a line defined by P and r. Vector r defines, what kind of mapping is being used. We tried cylindrical, spherical and the direction of estimated normal of point P. Cylindrical mapping seemed to produce best results.

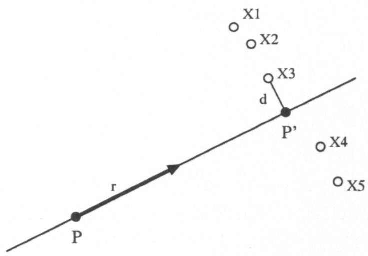

Figure 5: Finding intersection between points and line

This works well on a dense enough untriangulated mesh, but if triangulation information is present for a scanned mesh, intersection point on the appropriate triangle can be used. This greatly improves the result for meshes with less points.

It is important to note, that while this method sometimes gives good results, it doesn't differentiate among facial features of different importance. It would seem reasonable to adjust more important features first and detailed shape later. Research of facial landmarks and relations among them is a subject of *anthropometry* [8].

We constructed a hull (figure 1b), consisting of points on the original model (figure 1a), which are anthropometric landmarks. We fitted the hull by using the method described above. The displacement of points on the model, which are not present in the hull is computed by using barycentric coordinates relative to the triangles of the hull. This preserves first-order continuity of the deformation. This is somewhat similar to deformation model used by Arai *et al.* [9] to model faces from photographs.

This technique achieved overall general shape conformation. Details did not work out particulary well. It needs further refinement in terms of locating facial landmarks and final surface conformation.

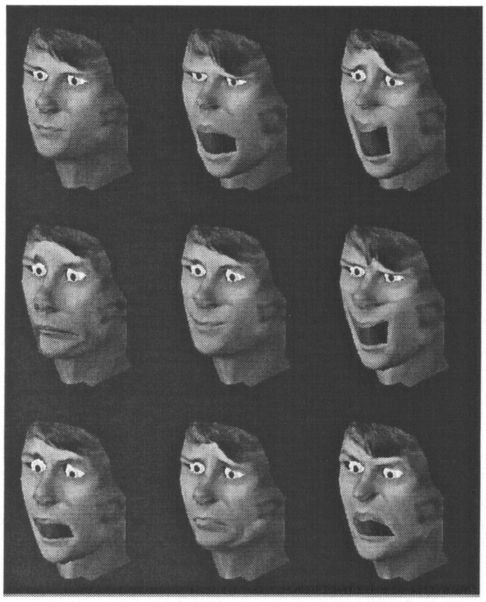

Figure 6: Conformed template (512 vertices, 875 faces)

The results in figure 6 are low polygon count models, which can be animated and rendered quickly. Eyes were added as library objects. Since template vertices were tagged, it was possible to use wider range of animation procedures, such as opening a mouth. Conformation to individual is visible, but there are problems in some areas. Some areas should really be smoother. One way to approach this problem would be to apply spring dynamics to a fitted model. Anthropometrically important vertices would be marked as immobile in the process.

4 Summary

Results look better, if rendering size is small. One area of use for such models is interactive games, where low polygon count is important.

Range sensor currently uses one projector and one monochrome camera for depth and texture acquisition. It would make sense to use another set of lights and frontal RGB camera of possibly higher resolution for texture acquisition. This would automate texture mapping, while giving similar results. The whole sensor setup could probably be conveniently built into a photo-cabin.

References

[1] Y. Lee, D. Terzopoulos, K. Waters. Realistic Modeling for Facial Animation. *Computer Graphics (SIGGRAPH '95)*, 20(4):56-62, August 1995. `http://www.cs.toronto.edu/~dt/papers/siggraph95b/siggraph95b.ps.gz`

[2] F. Parke and K. Waters. Computer Facial Animation. *A K Peters, 1996.* ISBN 1-56881-014-8. `http://crl.research.compaq.com/publications/books/waters/waters_book.html`

[3] D. Skočaj. Coded Light Range Sensor. `http://www.lrv.fri.uni-lj.si/~danijels/3d/rsensor.html`

[4] G. Turk and M. Levoy. Zippered Polygon Meshes from Range Images. *Computer Graphics (SIGGRAPH '94)*, 1994. `http://www-graphics.stanford.edu/papers/zipper/`.

[5] UCSC Perceptual Science Laboratory. *Facial Animation.* University of California - Santa Cruz. `http://mambo.ucsc.edu/psl/fan.html`

[6] B. Raitt. *Character Modelling.* Protozoa. `http://www.spiraloid.com/tutorial/morph_applied.html`

[7] G. Roth. Building Models from Sensor Data: An Application Shared by The Computer Vision and Computer Graphics Community. *NATO Advanced Research Workshop, 29-31 August 1999*, Ljubljana, Slovenia

[8] L. Farkas. *Anthropometry of Head and Face.* Raven Press, 1994

[9] T. Kurihara and K. Arai. A transformation model for modelling and animation of the human face from photographs. *Computer Animation '91*, Springer-Verlag, Tokyo, 1991

[10] C. Pelachaud, N.I. Badler, M.L. Viaud. *Final Report to the NSF of the Standards for Facial Animation.* University of Pennsylvania, School of Engineering and Applied Science, Computer and Information Science Department, Philadelphia, PA 19104-6389, August 1994, `http://www.cis.upenn.edu/~hms/pelachaud/workshop_face/workshop_face.html`