3D Stereo Reconstruction of Human Faces driven by Differential Constraints

Richard Lengagne[1], Pascal Fua[2], and Olivier Monga[1]

[1] INRIA Rocquencourt, Domaine de Voluceau BP105
78153 Le Chesnay, FRANCE
{Richard.Lengagne, Olivier.Monga}@inria.fr
[2] Computer Graphics Lab, EPFL
CH-1015 Lausanne, SWITZERLAND
fua@lig.di.epfl.ch

Abstract. Conventional stereo algorithms often fail in accurately reconstructing a 3D object because the image data do not provide enough information about the geometry of the object. We propose a way to incorporate a priori information in a reconstruction process from a sequence of calibrated face images. A 3D mesh modeling the face is iteratively deformed in order to minimize an energy function in a snake-like process. Differential information about the object shape is used to generate an anisotropic mesh that can both fulfill the compacity and the accuracy requirements. Moreover, in areas where the stereo information is not reliable enough to accurately recover the surface shape, because of inappropriate texture or bad lighting conditions, we propose to incorporate some geometric constraints related to the differential properties of the surface. These constraints can be intuitive or can refer to some predefined geometric properties of the object to be reconstructed. They can be applied to scalar fields such as curvature values, or structural features such as crestlines.

Category: Head and Face Modelling Techniques

1 Introduction

3D face reconstruction is currently receiving a lot of attention in the Computer Vision and Computer Graphics communities. It is a thriving research field with many applications such as virtual reality, animation, face recognition, etc... In all these cases, the recovered model must be compact and accurate, especially around significant areas like the nose, the mouth, the orbits, etc... These areas can often be characterized in terms of their differential properties. Several attempts to deal with that problem have been made. In [DF94], the differential properties of the surface are inferred from a disparity map and used to modify the shape of a correlation window. In [LFM96], crest line extraction is performed on a 3D model and used to improve the reconstruction around sharp ridges. These methods improve the accuracy of the reconstruction but do not suffice if the initial 3D model is not reliable. For instance, it is well known that bad lighting conditions or lack of texture can make correlation-based stereo fail. Consequently, the image information alone is not always sufficient to recover 3D shape. In [FB96], constraints on the depth of a given set of points on a surface mesh are applied in order to improve terrain reconstruction. In [LMF97], curvature information and structural features such as crest lines are extracted from the 3D model or interactively specified in order to generate an anisotropic surface mesh that reflects the geometric properties of the object. In this paper, we propose a further step towards incorporating a priori information in the reconstruction process from sets of sequences of calibrated face images. Differential information is used to constrain the topology of a mesh modeling the surface and the parameters of an analytical surface model, through the specification of low(high)-curvature areas, or structural features. Mathematically, this is achieved by constrained mesh optimization. We show preliminary results of this ongoing work, whose goal is to build 3D face models using entirely passive techniques.

2 The reconstruction process

2.1 An energy minimization scheme

Our reconstruction process is based on the iterative deformation of a 3D triangular mesh (i.e. a collection of vertices, triangular faces and edges) modeling the face in order to minimize an energy function E. The reconstruction process is thus treated as a snake-like process ([KWT88],[FL95],[LFM96]).
The initial mesh is computed by fitting a generic animation mesh to 3-D points derived from a correlation-based disparity map ([FM98]). It is then refined by minimizing an energy function that is the weighted sum of two terms: one stereo term E_{ext}, whose minimization makes the model fit to the image data (see [FL95] or [LFM96] for more details), and one regularization term E_{int}. This process is based on correlation; consequently, in many well-known cases (lack of texture, lighting problems,...), it will fail in accurately recovering the 3D shape.
This optimization uses a finite-element scheme. The depth Z of each surface point is expressed as a piecewise polynomial function of the two other coordinates X and Y. This polynomial is of degree 5, which guarantees that the surface is piecewise C^1 (see [Neuen95], [ZT88]). The parameters of the optimization process are the depths of each vertex, as well as the 5 first and second-order partial derivatives of the depth with respect to X and Y. To compute the initial values of the partial derivatives, we have locally approximated the surface by a quadric and set the partial derivatives of the surface to the partial derivatives of the corresponding quadric.

2.2 Adaptive meshes

The computation time can be very high if we keep a very large number of vertices. Therefore, we have to reduce the number of vertices in featureless areas and to keep many points in the most significant areas of the face. Furthermore, this has to be achieved with as much automation as possible. For instance, we would like to keep many points in the nose area, the orbits, the mouth, i.e. areas which are likely to act as landmarks in an animation or a recognition process. All these areas can be characterized by geometrical properties of the surface, especially differential properties. Indeed, areas like the nose ridge, the orbits, can be expressed in terms of high curvature areas, or crest lines, whereas the cheeks, the forehead (where we would like a small number of facets) can be described as low curvature areas. We have thus chosen to refine the 3D model according to the differential properties of the surface that can be easily inferred from the analytical expression of the surface or estimated by a local quadric approximation. As described in [LMF97], we generate an adaptive mesh governed by the principal curvatures and the principal curvature directions of the surface. Information about the computation of the differential properties can be found in [DoCar76].

The algorithm can be summarized as follows:

- compute on the initial mesh the principal curvatures k_{max} and k_{min} and the principal curvature directions $\boldsymbol{t}_{max}$ and $\boldsymbol{t}_{min}$.
- specify for each vertex of the initial mesh the three parameters (two scalar values h_1 and h_2 and an angle θ) of an ellipse centered on the vertex which governs the generation of a new mesh.
- optimize the new mesh by minimizing the energy function $E = \lambda_{ext}E_{ext} + \lambda_{int}E_{int}$.

The algorithm completely remeshes a 2D domain (which is taken here to be a frontal projection of the face) according to the values of h_1, h_2 and θ. These values govern the local topology of the new mesh in the vicinity of the old vertex they are attached to. As shown in figure 1, the angle θ determines in which direction the new facet in the remeshed surface will be "elongated". This direction will be given by $\boldsymbol{t}_{min}$. In other terms, the edges of the new facets will be longer in the minimum curvature direction than in the maximum curvature direction (those two directions are orthogonal). This is rather intuitive: for instance, in the case of the nose ridge, the minimum curvature direction lies along this ridge. We want to capture as many details as possible in the direction orthogonal to this ridge, since there is a high curvature variation in that direction. Consequently, it is natural to generate longer edges in the minimum curvature direction (i.e.along the ridge) than in the maximum curvature direction (i.e. across the ridge). The scalar values h_1 and h_2 determine the average lengths of the edges in those two directions. They are decreasing functions of k_{max} and k_{min}, since we want more facets in low curvature areas. Typically, they are chosen as inverses of a second order polynomial function. h_1 is determined by the minimum curvature and h_2 is determined by the maximum curvature. This procedure uses a mesh generation software developed for the Computational Field Simulations ([BCGHM96]). This scheme can also be used if we want to remesh the surface according to structural information such as crest lines than can be automatically detected ([LMF97]) or

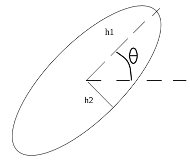

Fig. 1. The ellipse defining the local topology of the new mesh.

interactively specified. Let us recall that a crest line is defined as a set of zero-crossings of the derivative of the maximum curvature in the maximum curvature direction, i.e. the set of points such that $dk = \nabla k_{max}.\boldsymbol{t}_{max} = 0$. A crest line can thus be composed of maxima or minima of the maximum curvature. We typically threshold the result of crest line extraction according to the maximum curvature value in order to only keep significant lines. The angle θ is thus determined by the direction of the crest line, and h_1 and h_2 are fixed parameters.

Figure 2 shows a stereo pair of a face and the initial model obtained by the deformation of an animation mask. Our purpose is thus to capture more details in significant areas of the face, while preserving a reasonable number of vertices. Figure 3 shows an optimized anisotropic mesh of the face governed by curvature information. The nose ridge is well recovered since we have extracted high curvature values in this area and the principal curvature directions have oriented the facets along the ridge. However, the mouth is not very well recovered. Figure 4 shows a map of automatically extracted crest lines and the optimized mesh governed by crest line information. The crest line extraction algorithm ([LMF97]) ensures that the crest line lies inside a facet. The mesh of the face model has edges on the nose ridge. This is why this ridge has not been detected at the exact location by the crest line extraction algorithm. However, the purpose of this extraction is to specify areas of interest on the face and refine them, so we do not need a very accurate crest line extraction. In the example, we show the automatic extraction of the nose ridge, orbits, some lines on the lips and other lines that are not as intuitive but that can also describe the face geometry, such as cheek or forehead lines. In this case, the mouth is better recovered than using curvature information. We show in Figure 5 that an interactive outline of some crests can also help the reconstruction of key areas such as the orbits. We have shown examples where mesh topologies driven by curvature, automatically extracted crest lines and manually specified crest lines have been generated separately. An optimal reconstruction algorithm would merge all these kinds of information. This is part of the software we are currently developing.

Governing the mesh topology by surface differential properties and running a correlation-based optimization algorithm on the adaptive mesh is sometimes not sufficient to accurately recover 3D shapes. For instance, in the above example, the shape of the eyes cannot be recovered accurately from stereo information alone because of specularity (see also fig. 6). Therefore, it seems necessary to incorporate in the reconstruction process extra information.

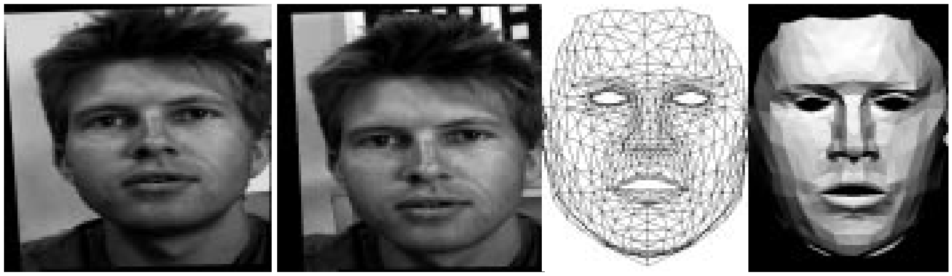

Fig. 2. A stereo pair of a face. The generic animation model fitted to correlation data: the mesh and a shaded view.

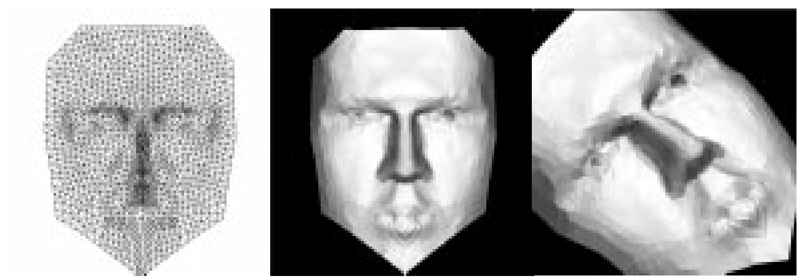

Fig. 3. The reconstructed surface using an anisotropic mesh governed by curvature information: the mesh and two shaded views.

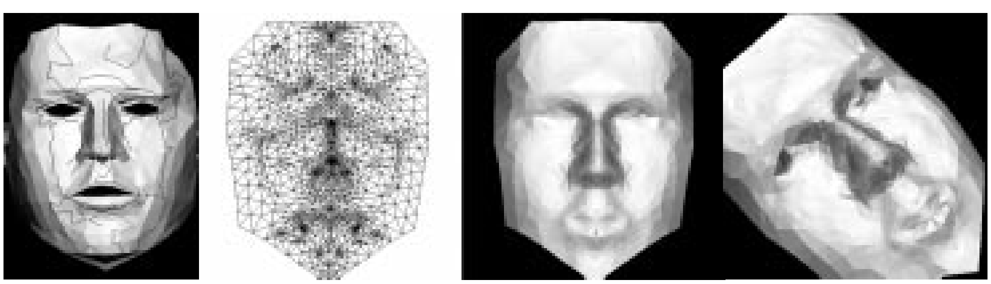

Fig. 4. Some crest lines automatically detected on the face model, and the reconstructed surface using an anisotropic mesh governed by crest line information: the mesh and two shaded views.

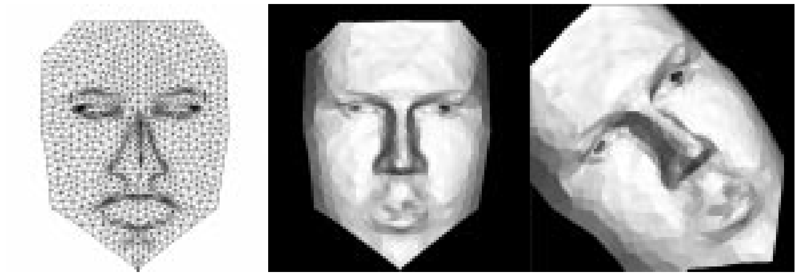

Fig. 5. The reconstructed surface using an anisotropic mesh governed by a priori knowledge about crest lines: the locations of the crest lines have been specified manually.

3 Incorporating a priori knowledge

When reconstructing an object, we have a rough idea about its shape, especially about typical features like crest lines, or about areas that can be labeled as "flat", "spherical", "cylindrical", etc... This kind of a priori knowledge can be of great interest where classical stereo methods fail. The a priori knowledge that a user can have about the shape he wants to reconstruct can be intuitive ("This region is flat, or spherical") or can rely on well-known geometric properties (anthropometric in case of face reconstruction). In any case, this a priori knowledge can very often be expressed in terms of differential properties. For instance, the knowledge "This area is flat" is obviously "translated" as: at each vertex, $k_{max} = k_{min} = 0$. "This area is spherical" means: at each vertex, $k_{max} = k_{min}$.

We can also express "structural" knowledge such as "There is a crest line here", and interactively outline the crest on the surface (or, ideally, on the images). If we restrict our problem to the crest lines that are sets of maxima of the maximum curvature in the maximum curvature direction, the corresponding constraint can be expressed as: $\forall i \in V, k_{max}(i) \geq k_{max}(j)$ and $k_{max}(i) \geq k_{max}(j')$ where V is the set of vertices lying on the crest line and j and j' are surface points such that the directions respectively defined by (i, j) and (i, j') are orthogonal to the crest line.

Incorporating a priori knowledge in the reconstruction process can be achieved using constrained optimization, since all the constraints are expressed in terms of the partial derivatives of the surface, which are the parameters of the optimization process. We use for that purpose a constrained optimization software especially designed for large systems [LZT96] (which is our case, since we have 6 parameters per vertex).

We have reconstructed one eye of the face shown in the previous section, using the a priori assumption that the eye is spherical. We first constrain the topology of the mesh by manually outlining the eyelid, therefore generating more facets on the tip of the eyelid. The initial eye surface is computed by interpolation, since there is no information on the animation model in this area. We then minimize

$E = \lambda_{ext}E_{ext} + \lambda_{int}E_{int}$ under the following constraints:

$$\forall i \in V, k_{max}(i) = k_{min}(i) \tag{1}$$

$$\forall (i,j) \in V^2, k_{max}(i) = k_{max}(j) \tag{2}$$

$$\forall (i,j) \in V^2, k_{min}(i) = k_{min}(j). \tag{3}$$

where i denotes the i-th vertex and V the set of vertices lying on the eye surface. We show in the results a fine isotropic mesh of the reconstructed eye surface after resampling the anisotropic mesh and using the polynomial surface approximation given by the finite element scheme. The surface has been rotated for visualization purposes, thereby inverting the signs of the curvatures.

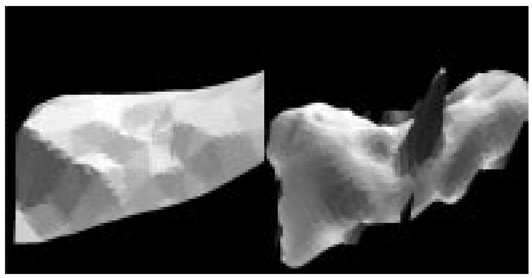

Fig. 6. The initial eye surface (left) and the reconstructed eye with the classical optimization algorithm (right). This area corresponds to the black spot in the eye in fig. 5.

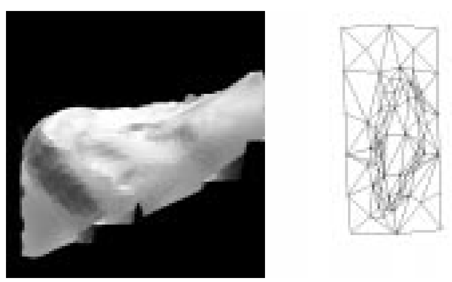

Fig. 7. The reconstructed eye after incorporating curvature-based constraints (left) and resampling the anisotropic mesh (right).

4 Conclusion

We have proposed a way of palliating the lack of information extracted from stereo images during a 3D reconstruction task. We interactively reconstruct from stereo a complex 3D object like a face using a priori information about its differential properties. Our purpose is to develop an interactive image-based modeling software that takes into account some a priori knowledge that a user can have about the differential properties of the object to reconstruct.

References

[BCGHM96] H. Borouchaki, M. J. Castro-Diaz, P. L. George, F. Hecht and B. Mohammadi: *Anisotropic Adaptive Mesh Generation in Two Dimensions for CFD.* Proceedings of the 5th International Conference on Numerical Grid in Computational Field Simulations, 1996, Mississippi State University, USA.

[DF94] F. Devernay and O.Faugeras: *Computing Differential Properties of 3-D Shapes from Stereoscopic Images without 3-D Models.* Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), 1994, Seattle, USA.

[DoCar76] M. P. do Carmo: *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, 1976.

[FB96] P. Fua and C. Brechbuhler: *Imposing Hard Constraints on Soft Snakes.* Proceedings of the European Conference on Computer Vision (ECCV), 1996, Cambridge, U.K., II, p. 495-506.

[FL95] P. Fua and Y.G. Leclerc: *Object-Centered Surface Reconstruction: Combining Multi-image Stereo and Shading.* International Journal on Computer Vision, 1995, 16(1), p. 35-56.

[FM98] P. Fua and C. Miccio: *Animated Heads from Ordinary Images: A Least Squares Approach.* Technical Report, LIG, EPFL, 1998.

[KWT88] M. Kass, A. Witkin and D. Terzopoulos: Snakes : *Active Contour Models.* International Journal on Computer Vision, 1988, 1(4) p. 321-331.

[LFM96] R. Lengagne, P. Fua and O. Monga: *Using Crest Lines to Guide Surface Reconstruction From Stereo.* Proceedings of the 13th International Conference on Pattern Recognition (ICPR), 1996, Vienna, Austria.

[LMF97] R. Lengagne, O. Monga and P. Fua: *Using Differential Constraints to Reconstruct Complex Surfaces From Stereo.* Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), 1997, San Juan, Puerto Rico, USA.

[LZT96] C. Lawrence, J. L. Zhou and A. L. Tits: *User's Guide for CFSQP Version 2.4: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints.* Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742.

[Neuen95] W. Neuenschwander: *Elastic Deformable Contour and Surface Models for 2D and 3D Image segmentation.* PhD Thesis of the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, 1995.

[ZT88] O.C. Zienkiewicz and R.L Taylor: *The finite element method*, Mc Graw-Hill, 1988.