ENERGY CONSTRAINTS ON DEFORMABLE MODELS: Recovering Shape and Non-Rigid Motion

Demetri Terzopoulos
Andrew Witkin
Michael Kass

Schlumberger Palo Alto Research
3340 Hillview Avenue, Palo Alto, CA 94304

Abstract

We propose a paradigm for shape and motion reconstruction based on dynamic energy constraints. Objects are modeled as deformable elastic bodies and constraints derived from image data are modeled as external forces applied to these bodies. The external constraint forces are designed to mold a deformable body into a configuration that satisfies the constraints, making the model consistent with the images. We present a particular shape model whose internal forces induce a preference for surface continuity and axial symmetry. We develop a constraint force for dynamic stereo images and present results for the recovery of shape and non-rigid motion from natural imagery.

1. Introduction

To reconstruct the shapes and motions of 3D objects from their images it is necessary to synthesize 3D models that simultaneously satisfy a bewildering variety of constraints. Some of these constraints derive from the immediate content of the image. Others reflect background knowledge of the image-forming process and of the shape and behavior of real-world objects. We propose a paradigm for shape and non-rigid motion reconstruction in which objects are modeled as deformable elastic bodies and constraints are modeled as dynamic external forces applied to the bodies. The external forces are designed to mold the deformable body into a configuration that satisfies the constraints. This minimum-energy configuration is computed by numerically solving the equations of motion for the deformable body.

In this paper, we consider the reconstruction of the shape and non-rigid motion of objects possessing rough axial symmetries. The input data consists of a temporal sequence of stereo image pairs. Several researchers have investigated motion-stereo fusion as a means of facilitating the recovery of 3D scene information [Nevatia, 1976; Regan and Beverley, 1979; Ballard and Kimball, 1983; Richards, 1985; Waxman and Sinha, 1986]. In our approach, an energy functional is defined which varies temporally according to the evolving stereo image pair. To reconstruct the shape and motion of a non-rigid object of interest, the dynamically deforming model maintains consistency with the image data by continually seeking lower energy states. An interesting feature of our procedure (though by no means a necessary consequence of our energy constraint methods in general) is that dynamic 3D object models are computed directly from image data without an intervening 2.5D surface representation.

Our deformable model of shape is governed by internal forces that imbue it with a preference for surface continuity as well as a preference for axial symmetry. In the latter regard, our model is close in spirit to generalized cylinder representations [Nevatia and Binford, 1977; Marr, 1977]. However, while generalized cylinders impose exact symmetries on any object they represent, our energy-based model is *symmetry-seeking:* It is capable of representing any shape, but those with axial symmetry have lower energy and hence are preferred.

Reconstruction is accomplished by applying image-derived forces to the symmetry-seeking deformable model. For each image we compute a local measure of the intensity gradient magnitude. After an appropriate linear transformation, the local minima of the resulting potential functions indicate locally highest contrast and are interpreted as silhouettes (occluding contours). By de-projecting the gradient of these image potentials through the binocular camera model, a time-varying force field is created in 3-space.

Given the camera parameters and the model's current state, only unoccluded points where the lines of sight from either the left or right eye graze its surface (occluding boundaries) are sensitive to the force field. The forces move boundary points laterally and in depth such that their binocular projections are consistent with both the left and right image silhouettes. Consistency is achieved when the projected boundary points rest at local minima of the image potentials. The shape over the remainder of the surface is determined by the model's internal continuity and symmetry forces. It has been observed that occluding boundaries present difficulties to standard stereo matching methods, largely because occluding contours in the two images correspond to different occluding boundary curves on smooth objects. Our method overcomes these difficulties by applying separate forces to points along the left and right boundary curves.

The work in this paper is an instance of a dynamic energy constraints paradigm which has been successfully applied to a variety of problems in graphics and modeling as well as vision: In [Terzopoulos, Platt, Barr, and Fleischer, 1987] energy constraints are applied to deformable curve, surface, and solid models to build and simulate objects made of rubber, cloth, and similar materials. In [Witkin, Fleischer, and Barr, 1987] energy constraints are applied to parameterized shape models such as cylinders or spheres to automatically dimension, assemble, and animate objects made of collections of such parts. In [Barzel and Barr, 1987] articulated objects are assembled and simulated with accurate Newtonian dynamics. In [Kass, Witkin, and Terzopoulos, 1987] image energy constraints are applied to deformable plane curves to interactively locate and track edges and other image features. In [Witkin, Terzopoulos, and Kass, 1986] a deformable sheet in image coordinates is subjected to forces derived from area correlation to perform stereo reconstruction in the style of the 2.5D sketch (Fig. 1). In [Terzopoulos, Witkin, and Kass, 1987] symmetry-seeking models are used in a limited way to perform object reconstruction from static monocular silhouettes (Fig. 1). In [Platt, 1987] a deformable space curve model is extended into a space-time surface, and used to recover rigid motion.

The remainder of the paper is organized as follows: Sec-

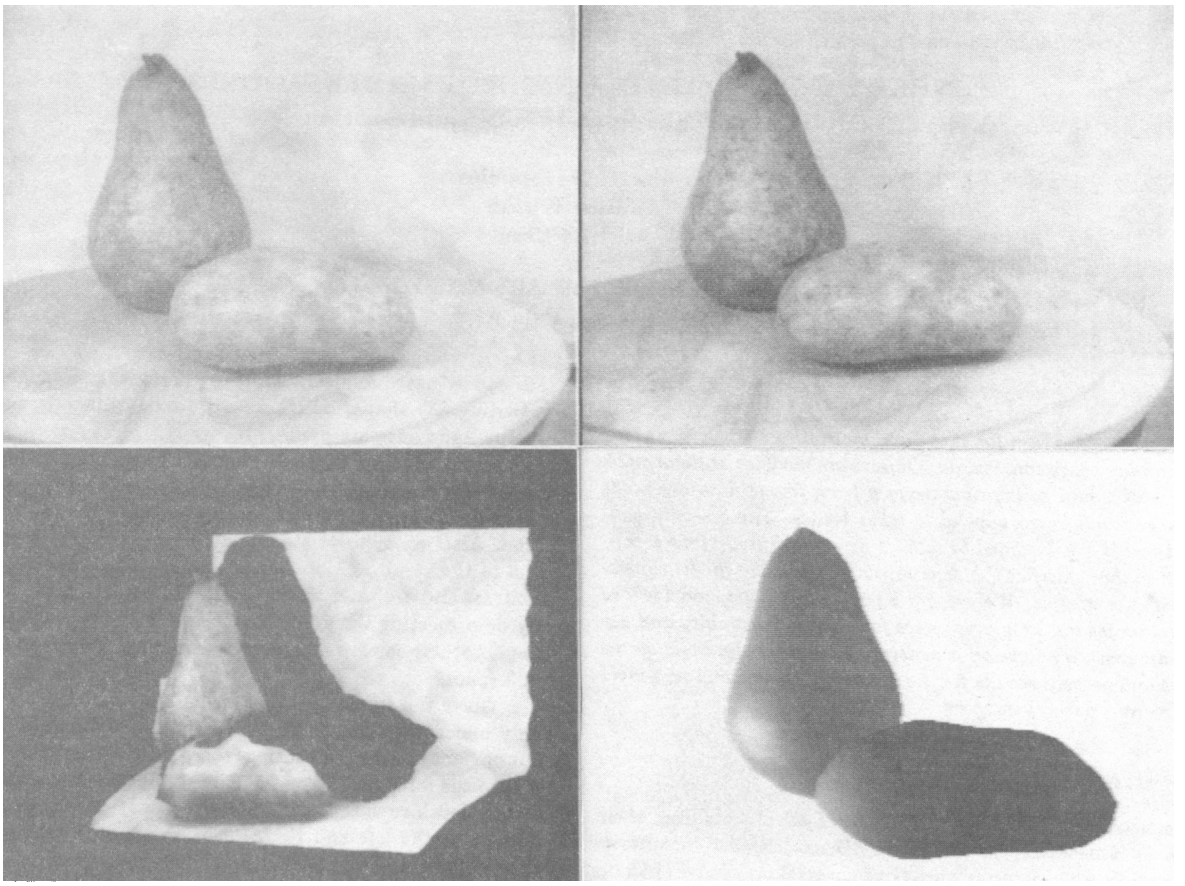

Figure 1. Reconstructions of a still life scene. Stereo images (top). 2.5D reconstruction of stereopair using a deformable sheet disparity model (bottom left). 3D reconstruction of objects in left image using a symmetry-seeking deformable model (bottom right).

tion 2 describes the geometry and dynamics of the deformable symmetry-seeking model. Section 3 describes the stereo-motion image force. Section 4 discusses the implementation and Section 5 presents results.

2. The Symmetry-Seeking Deformable Model

Before we review the formulation of the symmetry-seeking model proposed in [Terzopoulos, Witkin, and Kass, 1987], here is an informal description:

Imagine a deformable sheet made of elastic material (a blending of a membrane and thin plate). Take this sheet and roll it into a tube. Next, pass through the tube a deformable wire spine made of the same material and at regularly spaced points along its length couple it to the tube with radially projecting Hookean springs. The spring strengths can be adjusted so as to maintain the spine in approximately axial position within the tube. Additional forces are introduced that coerce the tube into a quasi-symmetric shape around the wire. Extra control is provided through additional compression/expansion forces radiating from the spine. The rigidity of the spine and the tube can be controlled independently, and their natural rest metrics and curvatures can be prescribed in advance or modified dynamically. For instance, if the circumferential metric of the tube is set to zero, the tube will tend to shrink around the spine unless expansion forces prevail; the model will shorten or lengthen as the longitudinal metrics of the tube and spine are modified. Hence, a wide range of interesting behavior can be obtained by adjusting the control parameters of the model.

The spine is a deformable space curve defined by mapping a 1-dimensional parametric domain $s \in [0,1]$ into Euclidian 3-space: $\mathbf{v}(s) = (X(s), Y(s), Z(s))$. The tube is made from a deformable space sheet defined by mapping a 2-dimensional parametric domain $(x,y) \in [0,1]^2$ into 3-space: $\mathbf{v}(x,y) = (X(x,y), Y(x,y), Z(x,y))$. In this paper, the mapping functions represent 3-space positions (alternatively they may may represent displacements away from prescribed rest configuration in 3-space).

The mapping is governed by the minimum of an energy functional

$$\mathcal{E}(\mathbf{v}) = \int_{\Omega} E[\mathbf{v}(\mathbf{x})] + P[\mathbf{v}(\mathbf{x})]\, d\mathbf{x}, \tag{1}$$

where $\mathbf{x}$ is a point in the parametric domain Ω. Here, E is the internal potential energy density of deformation and P is a generalized potential function associated with an externally applied force field. In our deformable model, E is an instance of the controlled-continuity constraint kernels [Terzopoulos, 1986].

The deformation energy associated with the spine mapping $\mathbf{v}(s)$ is given by

$$\mathcal{E}_S(\mathbf{v}) = \int_0^1 w_1(s)|\mathbf{v}_s|^2 + w_2(s)|\mathbf{v}_{ss}|^2 + P_S(\mathbf{v})\, ds. \tag{2}$$

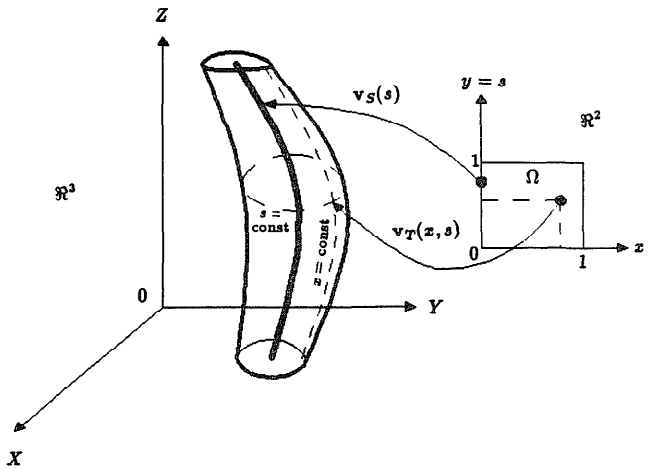

Figure 2. Parameterization of the the 3D model.

The weighting functions control the material properties: $w_1(s)$ determines the metric properties and tension along the spine, while $w_2(s)$ determines the curvature properties and rigidity of the spine.

The deformation energy of the sheet mapping $\mathbf{v}(x,y)$ is given by the functional

$$\mathcal{E}_T(\mathbf{v}) = \int_0^1 \int_0^1 w_{1,0}|\mathbf{v}_x|^2 + w_{0,1}|\mathbf{v}_y|^2 + w_{2,0}|\mathbf{v}_{xx}|^2 + 2w_{1,1}|\mathbf{v}_{xy}|^2 + w_{0,2}|\mathbf{v}_{yy}|^2 + P_T(\mathbf{v})\,dx\,dy. \tag{3}$$

The functions $w_{1,0}(x,y)$ and $w_{0,1}(x,y)$ determine the metric of the sheet along each parameter curve, while $w_{2,0}(x,y)$, $w_{1,1}(x,y)$, and $w_{0,2}(x,y)$ determine its natural curvature and rigidity.

The tube is formed by prescribing boundary conditions on two opposite edges of the sheet that "seam" these edges together. We seam the edge $x = 0$ to the edge $x = 1$, letting y span the length of the tube. The required periodic boundary conditions are

$$\mathbf{v}(0,y) = \mathbf{v}(1,y), \qquad \mathbf{v}_x(0,y) = \mathbf{v}_x(1,y). \tag{4}$$

To couple the two models together, we first identify $y \equiv s$, which brings into correspondence the spine parameter with the parameter along the tube (Fig. 2). We then distinguish the configuration vector function of the spine $\mathbf{v}_S$ from that of the tube $\mathbf{v}_T$.

The spine is coerced into an axial position within the tube by introducing the interaction potential energy functional

$$\mathcal{E}_a(\mathbf{v}_S,\mathbf{v}_T) = a\int_0^1 |\overline{\mathbf{v}}_T - \mathbf{v}_S|^2\,ds, \tag{5}$$

where

$$\overline{\mathbf{v}}_T(s) = \int_0^1 \mathbf{v}_T(x,s)\,dx \tag{6}$$

and a is the strength of the interaction.

To make the tube prefer symmetry with respect to the spine, we first define the radial vector anywhere on the tube as $\mathbf{r}(x,s) = \mathbf{v}_T(x,s) - \mathbf{v}_S(x,s)$, the unit radial vector as $\hat{\mathbf{r}}(x,s) = \mathbf{r}(x,s)/|\mathbf{r}(x,s)|$, and $\bar{r}(s) = \int_0^1 |\mathbf{r}(x,s)|\,dx$. The potential energy functional is then given by

$$\mathcal{E}_b(\mathbf{v}_S,\mathbf{v}_T) = b\int_0^1\int_0^1 \big(\mathbf{r}(x,s) - \bar{r}(s)\big)\cdot\hat{\mathbf{r}}(x,s)\,dx\,ds, \tag{7}$$

where b is the strength of the symmetrizing force.

Finally, we want to provide control over the expansion or shrinkage of the tube around the spine. This is accomplished by introducing the functional

$$\mathcal{E}_c(\mathbf{v}_S,\mathbf{v}_T) = \int_0^1 c(s)\int_0^1 \mathbf{r}(x,s)\cdot\hat{\mathbf{r}}(x,s)\,dx\,ds. \tag{8}$$

Here, $c(s)$ is the strength of the radial force; the tube will inflate if $c > 0$ and deflate if $c < 0$.

In particular, an end of the tube can be cinched shut by setting an endpoint factor $c(0)$ or $c(1)$ to be a large positive value.

The potential energy of deformation of the model is then obtained by combining the potential energy of deformation of the spine and tube models with the three coupling energies:

$$\mathcal{E}(\mathbf{v}_S,\mathbf{v}_T) = \frac{1}{2}\big(\mathcal{E}_S(\mathbf{v}_S) + \mathcal{E}_T(\mathbf{v}_T) + \mathcal{E}_a(\mathbf{v}_S,\mathbf{v}_T) + \mathcal{E}_b(\mathbf{v}_S,\mathbf{v}_T) + \mathcal{E}_c(\mathbf{v}_S,\mathbf{v}_T)\big). \tag{9}$$

The variational principle involves the minimization of (9) within a space of suitably differentiable deformations. The associated Euler–Lagrange equations are given in [Terzopoulos, Witkin and Kass, 1987].

3. Dynamic Stereo Forces

The symmetry-seeking model has the freedom to deform and to undergo translations and rotations in 3-space. The model is coupled to the dynamically evolving stereopair via a coupling energy term. The energy term is designed to impart forces that dictate the model's deformations and motions such that it remains maximally consistent with an object of interest in the dynamic stereopair.

Our goal in the present paper is to match the deformable model to an object's occluding contours in the time-varying left and right images $I_L(\eta_L,\xi_L)$ and $I_R(\eta_R,\xi_R)$. We assume that the object is imaged in front of a contrasting background, so that we can formulate a simple force field of attraction towards strong intensity gradients which, by assumption, will include the occluding contours. Then, the occluding boundaries of the deformable tube are made sensitive to this force field. We shall show in the next section that in spite of its simplicity this force field nonetheless yields interesting results.

To couple the model to the image potential function, we stereoscopically project the material points of the tube into the left and right image planes through binocular imaging equations. The points sense the image potential at the projected locations. The material points of spine are projected as well, but in our current implementation this is done simply for display purposes—the spine experiences no image forces.

Although it is possible to use a general binocular camera model (see, e.g., [Duda and Hart, 1973, Sec. 10.6]), its parameters need not be known with great accuracy for our approach to work. Consequently, we have found it convenient to employ a simplified perspective stereoprojection with eye vergence at infinity. Specifically, letting $\Pi_L[\mathbf{v}_T(x,s)]$ and $\Pi_R[\mathbf{v}_T(x,s)]$ denote the stereoprojection of the tube material point 3-space coordinates $(X_T(x,s), Y_T(x,s), Z_T(x,s))$ into the image planes (η_L,ξ_L) and

(η_R, ξ_R) respectively, we employ

$$\Pi_L : (\eta_L(x,s), \xi_L(x,s)) = (X_T(x,s) + \alpha Z_T(x,s), Y_T(x,s)),$$
$$\Pi_R : (\eta_R(x,s), \xi_R(x,s)) = (X_T(x,s) + \alpha Z_T(x,s), Y_T(x,s)), \quad (10)$$

where α is a constant.

The coupling between the force field and the tube is through the external potential function P_T (see eq. 3). We define

$$P_T[\mathbf{v}_T(x,s)] = -\beta_L(x,s)\left|\nabla\left(G_\sigma * I_L(\Pi_L[\mathbf{v}_T(x,s)])\right)\right| - \beta_R(x,s)\left|\nabla\left(G_\sigma * I_R(\Pi_R[\mathbf{v}_T(x,s)])\right)\right|, \quad (11)$$

which imparts on the tube boundary an affinity for steep image intensity changes. Here, $G_\sigma * I$ denotes the image convolved with a (Gaussian) smoothing filter whose characteristic width is σ.

When partial occlusions occur between multiple objects (e.g., Fig. 1 in which the potato partially occludes the pear) only the unoccluded surface patches should be sensitive to image forces. We use 3D ray casting from each viewpoint to test surface patches for visibility in each image. Hence, the weighting functions $\beta_L(x,s)$ and $\beta_R(x,s)$ are non-zero only for visible material points (x,s) near occluding boundaries of the tube. Occluding boundary points are selected in the left image by setting

$$\beta_L(x,s) = 1 - \left|\mathbf{i}_L \cdot \mathbf{n}(x,s)\right| \quad (12)$$

if the dot product is small (< 0.05), and $\beta_L(x,s) = 0$ otherwise, where $\mathbf{n}(x,s)$ is the unit normal of the tube at $\mathbf{i}_L(x,s)$ is the unit normal to the left image plane. The analogous weighting function is used for the right image.

4. Implementation

In our implementation, the time evolution of the model is governed by an initial value problem involving the equations

$$\gamma\frac{\partial \mathbf{v}_S}{\partial t} + \frac{\partial \mathcal{E}(\mathbf{v}_S, \mathbf{v}_T)}{\partial \mathbf{v}_S} = 0, \qquad \gamma\frac{\partial \mathbf{v}_T}{\partial t} + \frac{\partial \mathcal{E}(\mathbf{v}_S, \mathbf{v}_T)}{\partial \mathbf{v}_T} = 0, \quad (13)$$

where γ is a damping factor. These first-order equations describe the motion of massless material in a viscous medium. The formulation of a second-order dynamic system incorporating mass density is also straightforward [Terzopoulos, 1987], but (13) has served well for the time being.

The components of the energy gradient $\partial\mathcal{E}/\partial\mathbf{v}_S$ are approximated using standard finite difference expressions on a linear array of N_s nodes, while a $N_x \times N_s$ array is used to similarly approximate the components of $\partial\mathcal{E}/\partial\mathbf{v}_T$. The external force components $\nabla P_T(\mathbf{v}_T)$ are computed numerically in the image domains (η, ξ) using bilinear interpolation between centrally differenced pixel values.

We use an iterative procedure of the alternating direction implicit (ADI) type [Press, Flannery, Teukolsky, and Vetterling, 1986] to solve the discrete equations of motion. This efficient procedure exploits the fact that we have a rectangular grid of nodes. Each time step of the ADI procedure involves (i) a sweep in the x direction solving N_s independent systems of algebraic equations in N_x unknowns, followed by (ii) a sweep in the s direction solving N_x independent systems in N_s unknowns. The ADI method is independently applied to each of the three tube position components (X_T, Y_T, Z_T).

The spine gives rise to an additional system of equations in N_s unknowns for each of its position components (X_S, Y_S, Z_S).

As a consequence of the controlled-continuity deformation model, each of the unidimensional systems of equations has a pentadiagonal matrix of coefficients, and it can be solved efficiently (linear-order in the number of unknowns) using direct solution methods. We employ a normalized Cholesky decomposition step followed by a forward-back resolution step. See [Terzopoulos, 1987] for a derivation of the pentadiagonal matrix and for a discussion of the algorithm and [Kass, Witkin, and Terzopoulos, 1987] for its application to "snakes."

Resolution, an inexpensive step, is performed at every ADI iteration as the applied forces change. Matrix decomposition is somewhat more expensive, but it is required only when the material properties of the model are altered (e.g., to increase rigidity or to introduce discontinuities). Currently, we perform only an initial decomposition because we have not yet experimented with the variation of material properties during solution.

We find that for larger grid sizes and increasingly rigid material the ADI method evolves solutions faster than the successive over-relaxation (SOR) method that we employed previously [Terzopoulos, Witkin, and Kass, 1987]. This is attributable to the fact that the direct solution of each unidimensional system in the ADI method "immediately" distributes to all nodes along two perpendicular parametric grid lines the effects of forces acting on their common node.

5. Results

The reconstruction method was applied to a stereo motion sequence consisting of 40 video fields portraying the 3D motion of a human finger. The imaging apparatus was a beam-splitting stereo adaptor mounted on a CCD camera. An initial axis was specified on the first stereopair by the user, and the shell initialized to a cylinder around the axis (Fig. 3) The system's equations of motion were solved to equilibrium on the initial frame (requiring about 40 iterations), thus reconstructing the shape of the object in proper depth. The initial shape is rendered from several viewpoints in Fig. 3. The equilibrium solution then evolved over the remaining frames of the stereo sequence (using 20 iterations per frame), producing a dynamic 3D reconstruction of the finger's shape and motion. Fig. 4 shows six representative frames of the sequence along with the corresponding reconstructed shapes.

6. Conclusion

Our results illustrate the usefulness of dynamic energy constraints applied to deformable models as a means of recovering object shape and non-rigid motion.

A shortcoming of our current system is due to the fact that silhouette information alone, even with stereo and motion, provides limited information about objects. With large portions of the object's surface left unspecified, the symmetry-seeking material tends to make the reconstructed shape more symmetric than the actual one. Also, it is difficult to detect rotations around the object's axis only from silhouette information.

However, a key advantage of the energy constraints approach is the ease with which additional constraints can be integrated into the solution. A focus of our current work is the formulation and implementation of energy constraints that exploit shading and texture information over the entire visible surface, as well as constraints that make more effective use of motion information.

Our approach suggests energy constraint mechanisms for bringing higher-level knowledge to bear on the reconstruction problem. This remains a topic for future research. For the time being, the system is interactive; the user supplies an initial condition by instantiating a cylindrical surface about an approximate axis. We are investigating the use of scale-space continuation

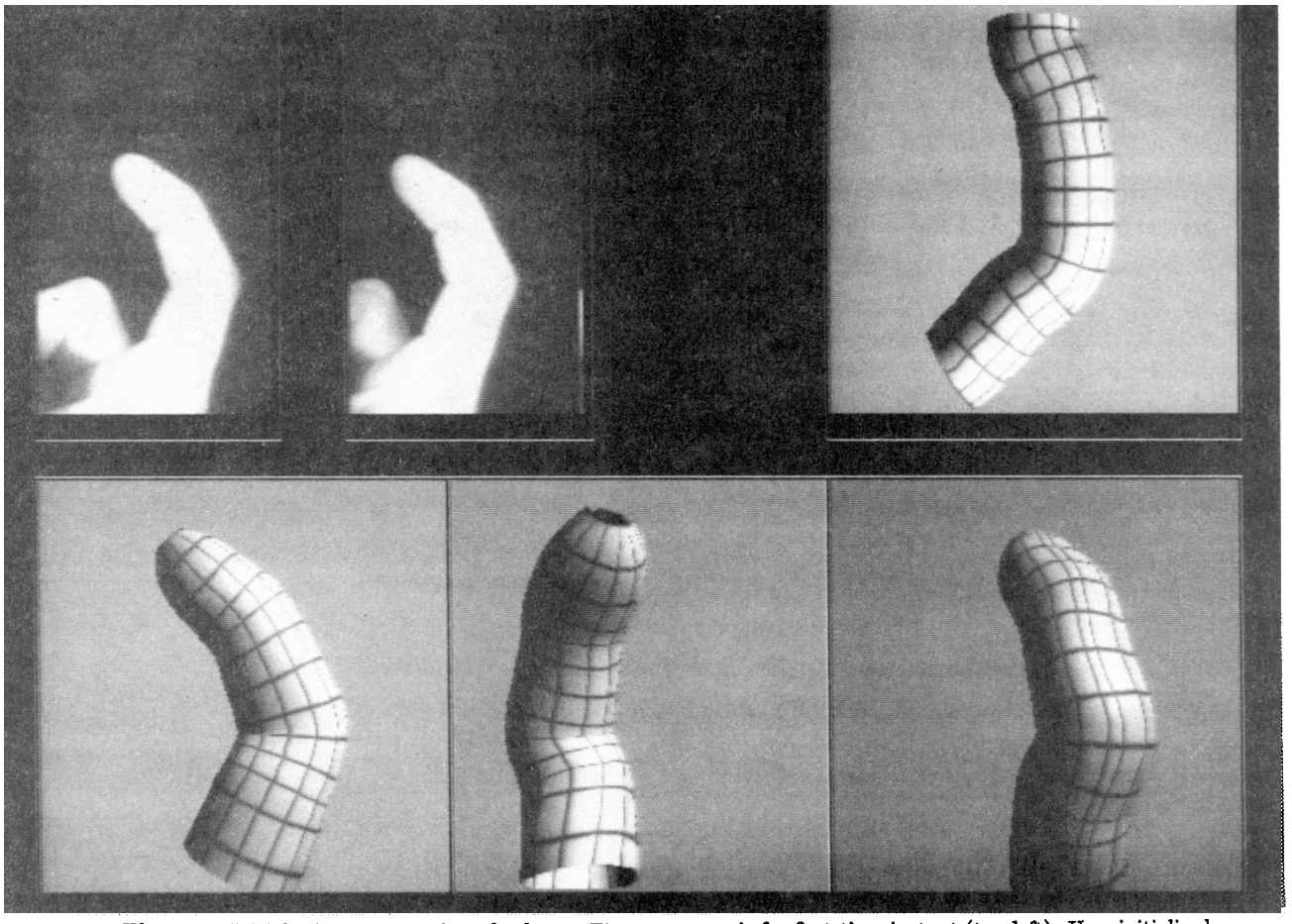

Figure 3. Initial 3D reconstruction of a finger. Finger stereopair for first time instant (top left). User-initialized cylinder (top right). Initial reconstructed shape from three viewpoints (bottom).

methods [Witkin, Terzopoulos, and Kass, 1986] to partially automate the initialization. We anticipate the ability to incorporate analytic camera models of greater sophistication into the energy functional and to automatically solve for the camera parameters as part of the minimization procedure.

Acknowledgments

Kurt Fleischer provided generous assistance in picture creation. John Platt contributed to the development of our ideas. Marty Tenenbaum provided suggestive interpretations of some of the figures.

References

Ballard, D.H., and Kimball, O.A., [1983], "Rigid body motion from depth and optical flow," *Computer Vision, Graphics, and Image Processing*, **22**, 95–115.

Barzel, R., and Barr, A., [1987], "Dynamic constraints," *to appear.*

Duda, R.O., and Hart, P.E., [1973], *Pattern Classification and Scene Analysis*, Wiley, New York.

Kass, M., Witkin, A., and Terzopoulos, D., [1987], "Snakes: Active contour models," *Proc. First Int. Conf. on Computer Vision*, London, UK, to appear in *Int. J. of Computer Vision*, **1***(4), 1987.*

Marr, D., [1977], "Analysis of occluding contour," *Proc. R. Soc. Lond. B*, **197**, 441–475.

Nevatia, R., [1976], "Depth measurement from motion stereo," *Computer Vision, Graphics, and Image Processing*, **9**, 203–214.

Nevatia, R., and Binford, T.O., [1977], "Description and recognition of curved objects," *Artificial Intelligence*, **8**, 77–98.

Platt, J., [1987], "An elastic model for interpreting 3D structure from motion of a curve," *to appear.*

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., [1986], *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK.

Regan, D., and Beverley, K.I., [1979], "Binocular and monocular stimuli for motion in depth: changing disparity and changing size feed the same motion in depth stage," *Vision Res.*, **19**, 1331–1342.

Richards, W., [1985], "Structure from stereo and motion," *J. Opt. Soc. Am. A*, **2**, 343–349.

Terzopoulos, D., [1986], "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **PAMI-8**, 413–424.

Terzopoulos, D., [1987], "Matching deformable models to images: Direct and iterative solutions," *Topical Meeting on Machine Vision, Technical Digest Series, Vol. 12., Optical Society of America*, Washington, DC, 160–167.

Terzopoulos, D., Platt, J., Barr, A., and Fleischer, K., [1987], "Elastically deformable models," *Proc. ACM SIGGRAPH-87 Conf.*

Terzopoulos, D., Witkin, A., and Kass, M., [1987], "Symmetry-seeking models for 3D object reconstruction," *Proc. First Int. Conf. on Computer Vision*, London, UK, to appear in *Int. J. of Computer Vision*, **1***(3), 1987.*

Waxman, A.M., and Sinha, S.S., [1986], "Dynamic stereo: Passive ranging to moving objects from relative image flows," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **PAMI-8**, 406–412.

Witkin, A., Fleischer, K., and Barr, A., [1987], "Energy constraints on parameterized models," *Proc. ACM SIGGRAPH-87 Conf.*

Witkin, A., Kass, M., Terzopoulos, D., and Barr, A., [1987], "Linking perception and graphics: Modeling with dynamic con-

straints," *Images and Understanding*, H. Barlow, C. Blakemore, and M. Weston-Smith (ed.), Cambridge University Press, to appear.

Witkin, A., Terzopoulos, D., and Kass, M., [1986], "Signal matching through scale space," *Proc. National Conf. on Artificial Intelligence, AAAI-86*, Philadelphia, PA, 714–719, to appear in *Int. J. of Computer Vision*, 1*(2), 1987.*

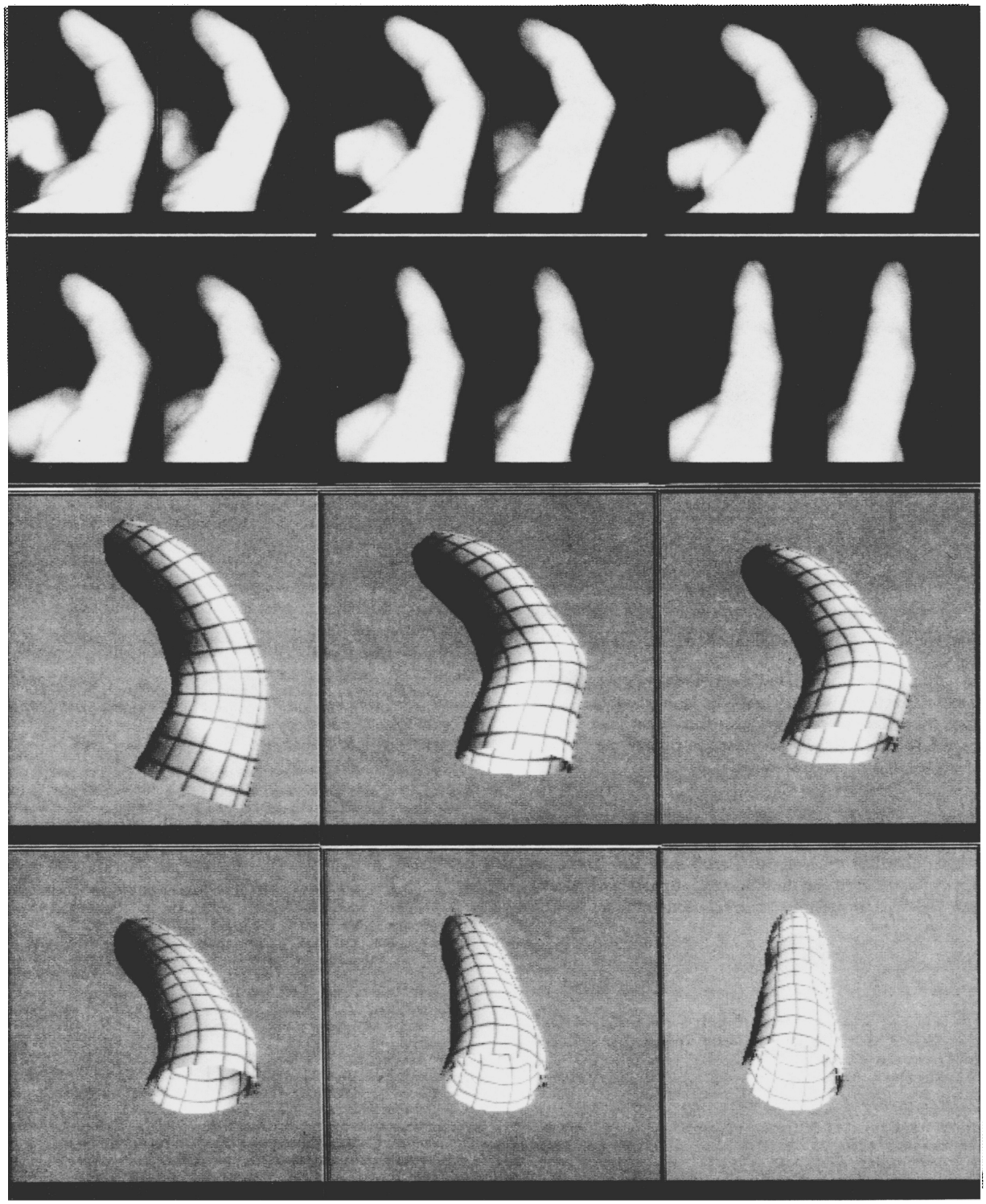

Figure 4. Evolution of the reconstructed 3D model through time. Six frames of the stereo sequence are shown (top) along with the evolving shape and motion of the model (bottom).