ON MATCHING DEFORMABLE MODELS TO IMAGES

Demetri Terzopoulos

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

A crucial goal of machine vision is the computation of mathematical models describing the shapes of objects appearing in images. In this paper we formulate a broadly applicable class of models which may be viewed as deformable bodies composed of an abstract elastic material. The models are matched to images through the action of external forces which deform the bodies from their prescribed natural states and impart motions toward desired equilibrium configurations. Forces are typically derived from images, but they can also be applied interactively by the user (say, by means of a pointing device) or abstracted from physical phenomena (such as gravity).

A prior example of a deformable model is the thin plate model used for surface reconstruction from sparse visual data [1]. From its naturally planar state at depth $Z = 0$, the thin plate surface is deformed to an equilibrium configuration $Z = Z(x, y)$ by localized forces due to ideal springs attached to depth constraints. The model is characterized by a variational principle involving the minimization of a deformation energy functional.

This paper focuses on the formulation of variational principles that characterize a more sophisticated class of deformable models, based on an extension of the controlled-continuity constraint functionals proposed in [2]. The generalized models have d deformational degrees of freedom, each with a parametric dimensionality of order p. The degrees of freedom may, for example, denote Euclidean n-space positions or displacements of the model's points. For the case of surface reconstruction [1, 2], the deformable model is embedded in $n = 3$-space (X, Y, Z), $d = 1$ (the depth displacement Z) and $p = 2$ (the image coordinates (x, y)).

The matching to images of deformable models of different dimensionalities will be illustrated with three new applications [3]. The simplest of these is "intensity-edge-seeking" deformable contours embedded in images ($p = 1$, $d = 2$, $n = 2$) [4]. The second is signal matching; specifically correspondence matching of stereo images or motion sequences ($p = 2$, $d = 2$, $n = 3$) [5]. The third and most complex is the direct reconstruction of 3D objects from monocular images using a composite 3D model, a deformable tubular shell ($p = 2$, $d = 3$, $n = 3$) surrounding a deformable spine ($p = 1$, $d = 3$, $n = 3$) [6].

Generalized Deformable Models

Let $\mathbf{x} = (x^1, \ldots, x^p) \in \Re^p$ be a point in parameter space. Let Ω be a subset in $\Re^p$ with boundary $\partial\Omega$. A model is given by the image of the set Ω under the regular mapping $\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), \ldots, v_d(\mathbf{x})) : \Omega \subset \Re^p \mapsto \Re^d$. Let $\mathcal{H} \subset \Re^d$ be a space of admissible deformations. A general deformable model of order $q > 0$ minimizes in $\mathcal{H}$ the deformation energy functional $\mathcal{E}_q : \mathbf{v} \in \mathcal{H} \mapsto \Re$ where

$$\mathcal{E}_q(\mathbf{v}) = \sum_{m=1}^{q} \sum_{|j|=m} \frac{m!}{j_1! \ldots j_p!} \int_\Omega w_j(\mathbf{x}) \left| \frac{\partial^m \mathbf{v}(\mathbf{x})}{\partial x_1^{j_1} \ldots \partial x_p^{j_p}} \right|^2 d\mathbf{x} + \int_\Omega P[\mathbf{v}(\mathbf{x})]\, d\mathbf{x}. \tag{1}$$

Here, $j = (j_1, \ldots, j_p)$ is a multi-index with $|j| = j_1 + \ldots + j_p$, and $P : \Re^d \mapsto \Re$ is a generalized potential function associated with the externally applied force field.

In addition to the potential, the model is controlled by the vector $\mathbf{w}(\mathbf{x})$ of distributed parameter functions $w_j(\mathbf{x})$. For instance, a discontinuity of order $k < q$ is permitted to occur at $\mathbf{x}_0$ in the limit as $w_j(\mathbf{x}_0) \to 0$ for $|j| > k$ (see [2]). Elastic properties including natural rest states (natural lengths and curvatures) can be prescribed through suitable choices of $\mathbf{w}(\mathbf{x})$. A $q = 2$ order deformation model is employed in the applications described below.

Application 1: Image Contour Models

Our first goal is to design a deformable contour model which is attracted to salient features in an image $I(x,y)$. Visualize the model as a "snake" that has an affinity for extended intensity maxima, minima, or edges, and which can be guided across the image by user-controlled forces (Fig. 1) [3, 4].

To formulate the model we set $p = 1$ and parameterize the contour by $\mathbf{x} = s$ where $s \in [0,1]$. The model is embedded in $n = 2$ dimensional image space (x,y). A $d = 2$-dimensional mapping $\mathbf{v}(s) = (x(s), y(s))$ will be used, whose components denote image coordinates. The energy of deformation of the contour is given by

$$\mathcal{E}_2(\mathbf{v}) = \int_\Omega w_1(s)|\mathbf{v}_s|^2 + w_2(s)|\mathbf{v}_{ss}|^2 + P(\mathbf{v})\, ds. \tag{2}$$

Here, $w_1(s)$ controls "tension" while $w_2(s)$ controls "rigidity." Setting $w_1(s_0) = w_2(s_0) = 0$ permits a position discontinuity and setting $w_2(s_0) = 0$ permits a tangent discontinuity to occur at s_0.

We associate with the contour a metric function $L(s)$ which prescribes the natural arc length of the snake (measured in the image) as a function of s. This is accomplished by defining $w_1(s) = \sqrt{x_s^2 + y_s^2} - L(s)$. The natural curvature $C(s)$ is prescribed by defining $w_2(s) = \kappa(s) - C(s)$ where $\kappa(s)$ is curvature along the curve.

The contour will have an affinity for darkness or brightness if $P[\mathbf{v}(s)] = \pm[G_\sigma * I(\mathbf{v}(s))]$, depending on the sign, and for intensity changes if $P[\mathbf{v}(s)] = -|\nabla[G_\sigma * I(\mathbf{v}(s))]|$, where $G_\sigma * I$ denotes the image convolved with a (Gaussian) smoothing filter whose characteristic width is σ.

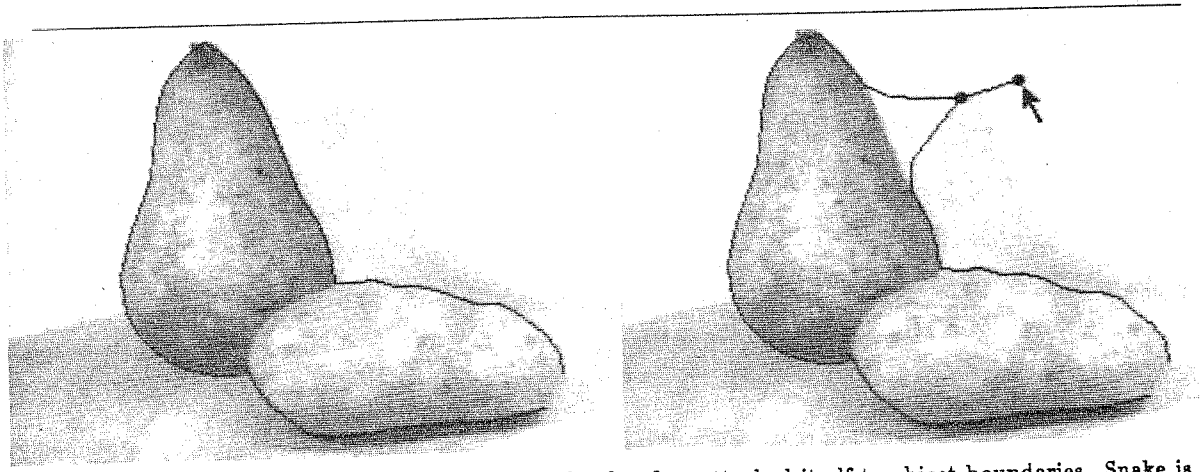

Figure 1. Dark contour is an edge-seeking snake that has attached itself to object boundaries. Snake is deformed (right) by a user-controlled spring force.

Application 2: Signal Matching; Stereo and Motion Correspondence

Given a set of similar signals that have been deformed with respect to one another, the signal matching problem is to recover the deformation. Important signal matching problems include stereo and motion correspondence matching and a variety of registration problems including template matching. A general treatment of this problem for an arbitrary number of multidimensional signals is given in [5]. Here, we consider the special case of matching two 2D images $I_1(x,y)$ and $I_2(x,y)$, which applies to stereo and motion correspondence problems. Fig. 2 shows the result of matching a deformable model for the disparity field to a stereopair.

In this particular case, $p = 2$, $d = 2$, and $n = 3$. The components of the mapping $\mathbf{v}(x,y) = (v_1(x,y), v_2(x,y))$, with $(x,y) \in \Omega = [0,1] \times [0,1]$, represent the displacement (disparity) in the horizontal x and vertical y directions respectively. The deformation energy is given by

$$\mathcal{E}_2(\mathbf{v}) = \int\int_\Omega w_1(x,y)(|\mathbf{v}_x|^2 + |\mathbf{v}_y|^2) + w_2(x,y)(|\mathbf{v}_{xx}|^2 + 2|\mathbf{v}_{xy}|^2 + |\mathbf{v}_{yy}|^2) + P(\mathbf{v})\, dx\, dy. \tag{3}$$

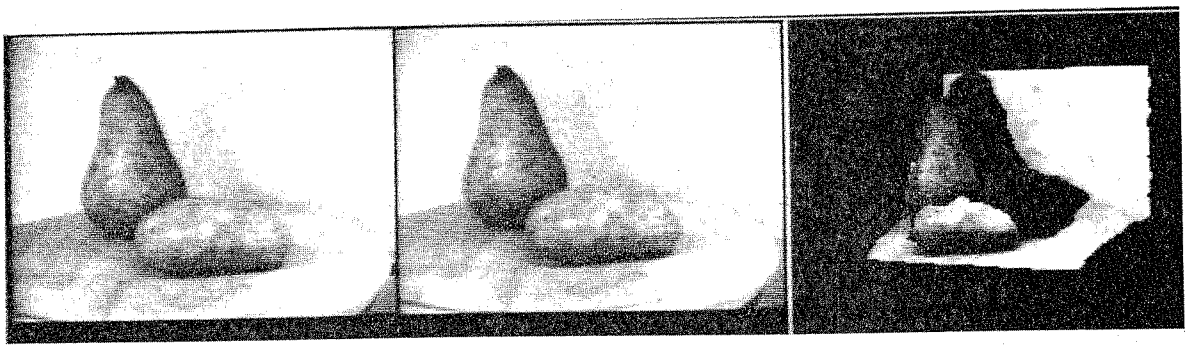

Figure 2. Stereopair and final state of the piecewise continuous disparity model rendered as a 3D surface with superimposed (left) image intensities.

The parameter functions w_1 and w_2 control the tension and rigidity of the deformation model [2] and discontinuity reconstruction requires that they be estimated during matching (see Fig. 2).

The generalized potential function is $P[\mathbf{v}(x,y)] = K_\sigma[I_1(\mathbf{v}(x,y)), I_2(x,y)]$, a measure of the local dissimilarity of the two images around (x,y). In [5] we employ normalized, windowed cross-correlation of the images smoothed by a filter with characteristic width σ.

Application 3: 3D Object Reconstruction

Finally, we consider the problem of 3D object reconstruction from monocular images. Here, we are interested in the restricted case of objects having quasi axial symmetry, with relatively little surface texture, and in general position before a contrasting background [3, 6]. Our approach is to match a 3D deformable model, having compatible symmetry, to an object's occluding contours in the image $I(\eta,\xi)$. The model has two components, a shell and a spine. Both components are $q = 2$ order functionals with $d = 3$ deformational degrees of freedom representing the $n = 3$-space positions (X,Y,Z) of material points.

Fig. 3 illustrates the reconstruction of a crook-necked squash from its image. After the user initializes the spine's projection in the image, the shell grows due to internal expansion forces. As the model deforms, its configuration in 3-space is dynamically projected onto the image (using computer graphics techniques), where its boundaries are subject to image-based forces; specifically, they are attracted by and lock onto significant image gradients. The mutually inverse processes of computer vision and computer graphics are coupled together to solve the reconstruction problem [3]; visual processing constrains and guides the reconstruction process, while graphical processing is necessary to apply the constraints and to provide the user with visual feedback.

The spine is a deformable space curve. For a configuration $\mathbf{v}(s) = (X(s), Y(s), Z(s))$ its deformation energy is given by a functional like (2), but where $\mathbf{v}(s)$ has dimensionality $d = 3$. The natural arc length of the spine is prescribed by $w_1(s) = \sqrt{X_s^2 + Y_s^2 + Z_s^2} - L(s)$ and the natural curvature by $w_2(s) = \kappa(s) - C(s)$.

The shell is composed of deformable sheet material. The deformation energy of the configuration $\mathbf{v}(x,y) = (X(x,y), Y(x,y), Z(x,y))$ is given by the $p = 2$ parameter functional

$$\mathcal{E}_2(\mathbf{v}) = \int\int\int_{\Omega_s} w_{1,0}|\mathbf{v}_x|^2 + w_{0,1}|\mathbf{v}_y|^2 + w_{2,0}|\mathbf{v}_{xx}|^2 + 2w_{1,1}|\mathbf{v}_{xy}|^2 + w_{0,2}|\mathbf{v}_{yy}|^2 + P(\mathbf{v})\,dx\,dy. \quad (4)$$

The natural metric of the sheet along each parameter curve is prescribed by the functions $w_{1,0}(x,y) = (X_x^2+Y_x^2+Z_x^2)^{1/2} - L_{1,0}(x,y)$ and $w_{0,1}(x,y) = (X_y^2+Y_y^2+Z_y^2)^{1/2} - L_{0,1}(x,y)$. Analogous expressions for $w_{2,0}(x,y)$, $w_{1,1}(x,y)$, and $w_{0,2}(x,y)$ determine its natural curvatures.

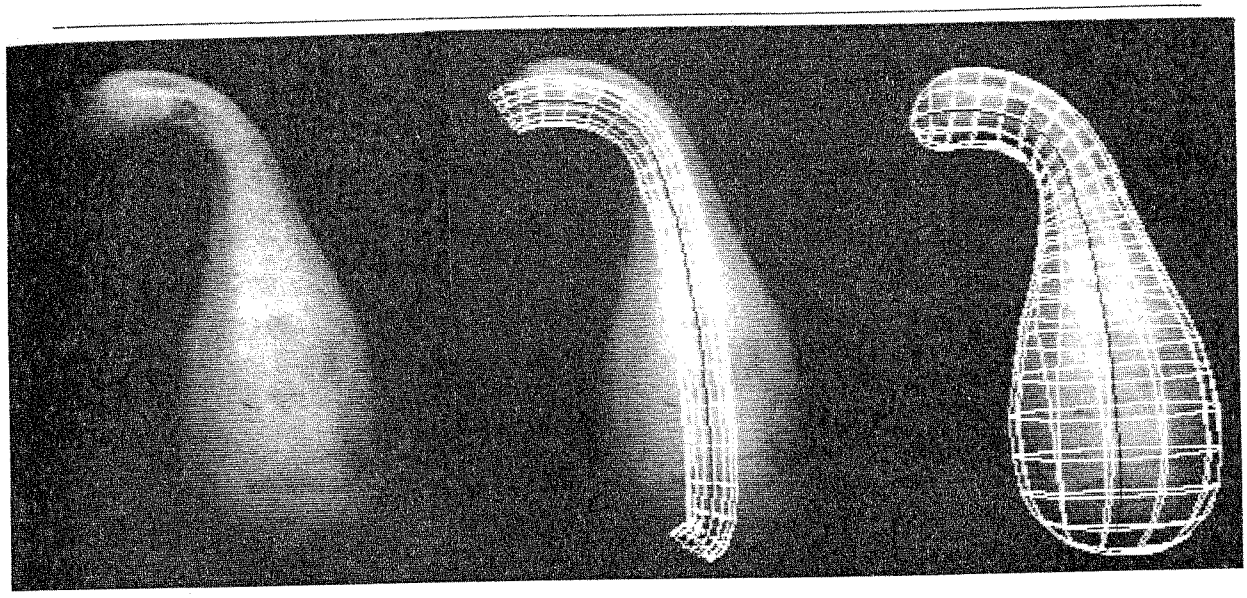

Figure 3. Reconstruction of a 3D model. Left to right: squash image; user initialized spine and shell; reconstructed squash.

A tubular shell is created by prescribing boundary conditions on two opposite edges of the sheet that effectively "seam" these edges together. Two forces of interaction are introduced between the shell and the spine: the first coerces the spine into a central position within the shell, while the second predisposes the shell to radial symmetry around the spine. The two open ends of the shell are cinched shut by the introduction of end-constricting force terms, thereby creating a sausage-like surface. These internal forces are formulated in [6].

Finally, the generalized potential $P[\mathbf{v}(x,y)] = -\gamma(x,y)|\nabla[G_\sigma * I(\Pi[\mathbf{v}(x,y)])]|$ imparts on the shell boundary an affinity for steep image intensity changes. The weighting function $\gamma(x,y)$ is nonzero only for material points (x,y) near occluding boundaries of the shell, which are selected by setting $\gamma(x,y) = 1 - |\mathbf{i}\cdot\mathbf{n}(x,y)|$ if this dot product is small (< 0.05), where $\mathbf{n}(x,y)$ is the unit normal of the shell at (x,y) and $\mathbf{i}$ is the unit normal of the image plane. $\Pi[\mathbf{v}(x,y)]$ denotes a projection of the material point 3-space coordinates $(X(x,y), Y(x,y), Z(x,y))$ into the image plane (η, ξ).

Acknowledgement and References

Thanks to A. Witkin, J. Platt, M. Kass, A. Barr, and K. Fleischer for their contributions to this research.

1. **Terzopoulos, D.**, "Multilevel computational processes for visual surface reconstruction," *Computer Vision, Graphics, and Image Processing*, **24**, 1983, 52–96.
2. **Terzopoulos, D.**, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **PAMI-8**, 1986, 413–424.
3. **Witkin, A., Kass, M., Terzopoulos, D., and Barr, A.**, "Linking perception and graphics: Modeling with dynamic constraints," *Images and Understanding*, H. Barlow, C. Blakemore, and M. Weston-Smith (ed.), Cambridge University Press, 1986, to appear.
4. **Kass, M., Witkin, A., and Terzopoulos, D.**, "Snakes: planar splines with potential wells," , 1986, in preparation.
5. **Witkin, A., Terzopoulos, D., and Kass, M.**, "Signal matching through scale space," *Proc. National Conf. on Artificial Intelligence, AAAI-86*, Philadelphia, PA, August, 1986, 714–719.
6. **Terzopoulos, D., Witkin, A., and Kass, M.**, "Symmetry-seeking models and 3D object reconstruction from images," , 1986, in preparation.