A New Approach to Tracking 3D Objects in 2D Image Sequences

Michael Chan[1], Dimitri Metaxas[1] and Sven Dickinson[2]

[1]Dept. of Computer & Information Science
University of Pennsylvania
Philadelphia, PA 19104-6389

[2]Dept. of Computer Science
University of Toronto
Toronto, Ontario, Canada M5S 1A4

mchan@grip.cis.upenn.edu, dnm@central.cis.upenn.edu, sven@vis.toronto.edu

Abstract

We present a new technique for tracking 3D objects from 2D image sequences through the integration of qualitative and quantitative techniques. The deformable models are initialized based on a previously developed part-based qualitative shape segmentation system. Using a physics-based quantitative approach, objects are subsequently tracked without feature correspondence based on generalized forces computed from the stereo images. The automatic prediction of possible edge occlusion and disocclusion is performed using an extended Kalman filter. To cope with possible occlusion caused by a previously undetected object, we monitor the magnitude and direction of the computed image forces exerted on the models. Abrupt changes to these forces trigger scene re-segmentation and model re-initialization through the qualitative shape segmentation system. Tracking is subsequently continued using only local image forces. We demonstrate our technique in experiments involving image sequences from complex motions of 3D objects.

Introduction

Research in 3D model-based object tracking from image sequences is typified by approaches which attempt to recover the six degrees of freedom of an object in each frame, e.g., (Thompson & Mundy 1988; Verghese, Gale, & Dyer 1990; Lowe 1991; Gennery 1992). Once correspondences between image and model features are determined, changes in the positions of image features in successive frames are used to update the pose of the object. Although these techniques provide accurate pose of the object at each frame, they require an exact geometric specification of the object; they do not allow models to deform as they move. Recently, to cope with the challenges of nonrigidity, several researchers have adopted a physics-based approach to estimate the shapes and motions of nonrigid 3D objects from visual data to different levels of accuracy (Terzopoulos, Witkin, & Kass 1988; Huang 1990; Pentland & Horowitz 1991; Metaxas & Terzopoulos 1993). The 2D problem has received similar attention (Kass, Witkin, & Terzopoulos 1988; Duncan, Owen, & Anandan 1991; Szeliski & Terzopoulos 1991; Blake, Curwen, & Zisserman 1993).

In this paper, we develop a new approach to tracking shapes and motions of objects in 3D from 2D image sequences. Our method makes use of both the framework of qualitative shape segmentation (Dickinson, Pentland, & Rosenfeld 1992b; 1992a) and the physics-based framework for quantitative shape and motion estimation (Terzopoulos & Metaxas 1991; Metaxas & Terzopoulos 1993). To be able to track multiple objects, initialization of the models is performed in the first frame of the sequence based on a shape recovery process that uses recovered qualitative shapes [1] to constrain the fitting of deformable models to the data (Metaxas & Dickinson 1993). For successive frames, the qualitative shape recovery process can be avoided in favor of a physics-based model updating process requiring only a gradient computation in each frame. Assuming no occlusion and small deformations between frames, local forces derived from stereo images are sufficient to update the positions, orientations, and shapes of the models in 3D.

Kalman filtering techniques have been applied in the vision literature for the estimation of dynamic features (Deriche & Faugeras 1990) and rigid motion parameters (Dickmanns & Graefe 1988; Broida, Chandrashekhar, & Chellappa 1990) of objects from image sequences. We use a Kalman filter for the estimation of the object's shape and motion, which consequently allows the prediction of possible edge occlusion and disocclusion. The occurrence of these situations may be due to changes of an object's aspect from frame to frame or due to motions of other independently moving objects (situations where most tracking approaches based on feature correspondence may not work robustly). By predicting the occurrence of these situations in our approach, we can confidently determine which part of an object will be occluded and suppress their contributions to the net forces applied to

[1]We assume that objects are constructed from a finite set of volumetric part classes.

the model. In fact, an advantage of our technique is that we do not need to perform costly feature correspondence during 3D tracking.

Our approach also allows the detection of object occlusion due to a previously undetected object by monitoring changes to the image forces exerted on the models. In such an ambiguous situation, we invoke the qualitative shape segmentation module for scene re-segmentation and model re-initialization. Tracking can then be continued by using only local images forces. Our technique is robust and can handle scenes with complex motions and occlusion due to the triggering of the qualitative shape segmentation system when necessary.

Dynamic Deformable Models

This section reviews the formulation of the deformable model we adopted for object modeling and the physics-based framework of visual estimation (see (Metaxas & Terzopoulos 1993) for greater detail).

Geometry of Deformable Models

The positions of points on the model relative to an inertial frame of reference Φ in space are given by a vector-valued, time varying function $\mathbf{x}(\mathrm{u},t) = (x(\mathrm{u},t), y(\mathrm{u},t), z(\mathrm{u},t))^T$, where ^{T} denotes transposition and u are the model's material coordinates. We set up a noninertial, model-centered reference frame ϕ and express the position function as $\mathbf{x} = \mathbf{c} + \mathbf{R}\mathbf{p}$, where $\mathbf{c}(t)$ is the origin of ϕ at the center of the model and the rotation matrix $\mathbf{R}(t)$ gives the orientation of ϕ relative to Φ. Thus, $\mathbf{p}(\mathrm{u},t)$ gives the positions of points on the model relative to the model frame.

We further express $\mathbf{p} = \mathbf{s} + \mathbf{d}$, as the sum of a reference shape $\mathbf{s}(\mathrm{u},t)$ and a displacement $\mathbf{d}(\mathrm{u},t)$. We define the reference shape as: $\mathbf{s} = \mathbf{T}(\mathbf{e}(\mathrm{u}; a_0, a_1, \ldots); b_0, b_1, \ldots)$. Here, a geometric primitive $\mathbf{e}$, defined parametrically in u and parameterized by the variables $a_i(t)$, is subjected to the *global deformation* $\mathbf{T}$ which depends on the parameters $b_i(t)$. Although generally nonlinear, $\mathbf{e}$ and $\mathbf{T}$ are assumed to be differentiable (so that we may compute the Jacobian of $\mathbf{s}$) and $\mathbf{T}$ may be a composite sequence of primitive deformation functions $\mathbf{T}(\mathbf{e}) = \mathbf{T}_1(\mathbf{T}_2(\ldots \mathbf{T}_n(\mathbf{e})))$. We concatenate the global deformation parameters into the vector $\mathbf{q}_s = (a_0, a_1, \ldots, b_0, b_1, \ldots)^T$. To illustrate our approach in this paper, we will use as a reference shape a deformable superquadric ellipsoid that can also undergo parameterized tapering deformations, as defined in (Metaxas & Terzopoulos 1993).

Model Kinematics and Dynamics

The velocity of a 3D point on the model is given by

$$\dot{\mathbf{x}} = \mathbf{L}\dot{\mathbf{q}}, \tag{1}$$

where $\mathbf{L}$ is the Jacobian matrix that converts q-dimensional vectors to 3D vectors (Metaxas & Terzopoulos 1993). The vector $\mathbf{q}(\mathrm{t})$ represents the generalized coordinates of the model consisting of the translation, rotation, global and local deformations. To make the model dynamic, we assume that it is made of a simulated elastic material that has certain mass distribution. From Lagrangian mechanics, we obtain second-order equations of motion which take the form (see (Terzopoulos & Metaxas 1991) for derivations):

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{D}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{g}_q + \mathbf{f}_q, \quad \mathbf{f}_q = \int \mathbf{L}^T \mathbf{f} \, d\mathrm{u}, \tag{2}$$

where $\mathbf{f}_q$ are generalized external forces associated with the components of $\mathbf{q}$, and $\mathbf{f}(\mathrm{u},t)$ is the image force distribution applied to the model. Here $\mathbf{M}$ is the mass matrix, $\mathbf{D}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix and $\mathbf{g}_q$ is the vector of the generalized coriolis and centrifugal forces.

Multiple Object Tracking

This section describes our new approach for tracking multiple objects in the presence of occlusion. It is based on the intelligent use of a qualitative shape segmentation system (Metaxas & Dickinson 1993) and techniques for quantitative shape and motion estimation (Metaxas & Terzopoulos 1993). The deformable models are first initialized based on the qualitative segmentation system. Objects are subsequently tracked using a physics-based approach by applying image forces simultaneously derived from the stereo images. We can handle partial edge occlusion and disocclusion due to the object's own motion or occlusion by another object by predicting their occurrences using an extended Kalman filter. We handle more complex cases of object occlusion due to a previously undetected object by monitoring changes to the image forces exerted on the models. These changes trigger the use of the qualitative shape segmentation system for scene re-segmentation and model re-initialization, and tracking is subsequently continued using local image forces only.

Qualitative Shape Recovery and Model Initialization

We employ the methodology developed in (Metaxas & Dickinson 1993) to initialize our deformable models. We start by assuming that objects are constructed from a finite set of volumetric part classes (Dickinson, Pentland, & Rosenfeld 1992b; 1992a). The parts, in turn, are mapped to a set of viewer-centered aspects. During the qualitative shape recovery process, the system first segments the image into parts using an aspect matching paradigm. Each recovered qualitative part defines: 1) the relevant non-occluded contour data belonging to the part, 2) a mapping between the image faces in their projected aspects and the 3D surfaces on the quantitative models, and 3) a qualitative orientation that is exploited during model fitting. Based on these constraints, we assign forces from monocular image data points to the corresponding points on the 3D model. The model is then fitted dynamically to the image data under the influence of the image forces. In the following sections, we will discuss how we handle sequences of stereo images taken under non-parallel geometry without requiring the continuous use of qualitative constraints.

Short Range Forces from Image Potentials

For each frame in the image sequence, we create an image potential such that the "valleys" of this potential correspond to the locations in the image where there are sharp changes in intensity or edge features. If we denote the intensity image by $I(x, y)$, the image potential can be computed as follows (Terzopoulos, Witkin, & Kass 1988):

$$\Pi(x, y) = -\beta \left| \nabla (G_\sigma * I)(x, y) \right| \tag{3}$$

where σ determines the width of the Gaussian function G_σ, $*$ denotes the convolution operation, and β determines the "steepness" of the potential surface. The corresponding 2D force field induced by this potential is given by:

$$\mathbf{f}(x, y) = -\nabla \Pi(x, y). \tag{4}$$

The model's degrees of freedom respond to the 2D force field through a process which first projects the model's nodes into the image. As the projected nodes are attracted to the valleys of the potential surface, the model's degrees of freedom are updated to reflect this motion. The mapping of 2D image forces to generalized forces acting on the model requires the derivation a Jacobian matrix.

Jacobian Computation for Perspective Projection

Let $\mathbf{x} = (x, y, z)^T$ denotes the location of a point j with respect to the world coordinate frame. Then we can write

$$\mathbf{x} = \mathbf{c}_c + \mathbf{R}_c \mathbf{x}_c, \tag{5}$$

where $\mathbf{c}_c$ and $\mathbf{R}_c$ are respectively the translation and rotation of the camera frame with respect to the world coordinate frame, and $\mathbf{x}_c = (x_c, y_c, z_c)^T$ is the position of the point j with respect to the camera coordinate frame.

Under perspective projection, the point $\mathbf{x}_c$ projects into an image point $\mathbf{x}_p = (x_p, x_p)^T$ based on the formulas:

$$x_p = \frac{x_c}{z_c} f, \quad y_p = \frac{y_c}{z_c} f, \tag{6}$$

where f is the focal length of the camera.

By taking the derivative of (6) with respect to time, we arrive at the following matrix equation:

$$\begin{pmatrix} \dot{x}_p \\ \dot{y}_p \end{pmatrix} = \begin{bmatrix} f/z_c & 0 & -x_c/z_c^2 f \\ 0 & f/z_c & -y_c/z_c^2 f \end{bmatrix} \begin{pmatrix} \dot{x}_c \\ \dot{y}_c \\ \dot{z}_c \end{pmatrix}. \tag{7}$$

Based on (5) and (1) we get

$$\dot{\mathbf{x}}_c = \mathbf{R}_c^{-1} \dot{\mathbf{x}} = \mathbf{R}_c^{-1} \mathbf{L} \dot{\mathbf{q}}. \tag{8}$$

Rewriting (7) in compact form using (8), we get

$$\mathbf{x}_p = \begin{bmatrix} f/z_c & 0 & -x_c/z_c^2 f \\ 0 & f/z_c & -y_c/z_c^2 f \end{bmatrix} \mathbf{R}_c^{-1} \mathbf{L} \dot{\mathbf{q}} = \mathbf{L}_p \dot{\mathbf{q}}. \tag{9}$$

By replacing the Jacobian matrix in (2) by $\mathbf{L}_p$, two dimensional forces $\mathbf{f}$ derived from image data can be appropriately converted into generalized forces $\mathbf{f}_q$ measured in the world coordinate frame.

Forces from Stereo Images

By computing generalized forces in the world coordinate frame, the 2D image forces in a pair of stereo images can be simultaneously transformed into generalized forces $\mathbf{f}_q$ measured in a common world coordinate frame. Measurements from two different views are sufficient to determine the scale and depth parameters of the model. If we define as *active* nodes those model nodes on which image forces are exerted, then the generalized forces are computed by summing the image forces exerted on all the active nodes of the discretized model. More precisely, if we denote the position of the jth active node on the model surface by $\mathbf{x}_j$, then the generalized force on the model can be computed as follows:

$$\begin{aligned} \mathbf{f}_q &= \sum_{j \in \mathcal{A}_L} \mathbf{L}_{p_L}^T (\mathbf{f}_L(\mathbf{P}(\mathbf{R}_{c_L}^{-1}(\mathbf{x}_j - \mathbf{c}_{c_L})))) \\ &+ \sum_{j \in \mathcal{A}_R} \mathbf{L}_{p_R}^T (\mathbf{f}_R(\mathbf{P}(\mathbf{R}_{c_R}^{-1}(\mathbf{x}_j - \mathbf{c}_{c_R})))), \end{aligned} \tag{10}$$

where $\mathcal{A}$ is the set of indices of active nodes. Here the subscripts L and R denote dependence on the left and right images respectively and $\mathbf{P}(x, y, z) = (\frac{x}{z} f, \frac{y}{z} f)$ describes the perspective projection equation.

Determining Active Model Nodes

When our measurements are 2D images, as opposed to 3D range data, only a subset of the nodes on the model surface are selected to respond to forces. From a given viewpoint, we can compute this active subset of model nodes based on the model's shape and orientation. In particular, a model node is made active if at least one of the following conditions is true:

1. it lies on the occluding contour of the model from that viewpoint, [2]
2. the local surface curvature at the node is sufficiently large and the node is visible.

(Note that it is possible that a model node is active with respect to one view, but not to another.). Instead of calculating analytically the positions of the active nodes on the model surface, we "loop" over all the nodes on the discretized model surface and check if one of the above two conditions is true. Condition 1 is true if: $|\mathbf{i}_j \cdot \mathbf{n}_j| < \tau$, where $\mathbf{n}_j$ is the unit normal at the jth model node, $\mathbf{i}_j$ is the unit vector from the

[2] See also (Terzopoulos, Witkin, & Kass 1988).

focal point to that node on the model, and τ is a small threshold. Condition 2 is true if

$$\exists k \in K_j \text{ s.t. } |\mathbf{n}_k \cdot \mathbf{n}_j| > \kappa \;\&\; \exists k \in K_j \text{ s.t. } \mathbf{n}_k \cdot \mathbf{i}_k < 0, \tag{11}$$

where K_j is a set of indices of the nodes adjacent to the jth nodes on the model surface. κ in (11) is a threshold to determine if the angle between adjacent normal vectors is sufficiently large.

Tracking and Prediction

We incorporate into our dynamic deformable model formulation a Kalman filter by treating their differential equations of motion (2) as system models. Based on the use of the corresponding extended Kalman filter, we perform tracking by updating the model's generalized coordinates $\mathbf{q}$ according to the following equation

$$\dot{\hat{\mathbf{u}}} = \mathbf{F}\hat{\mathbf{u}} + \mathbf{g} + \mathbf{P}\mathbf{H}^T\mathbf{V}^{-1}(\mathbf{z} - \mathbf{h}(\hat{\mathbf{u}})), \tag{12}$$

where $\mathbf{u} = (\dot{\mathbf{q}}^T, \mathbf{q}^T)^T$ and matrices $\mathbf{F}, \mathbf{H}, \mathbf{g}, \mathbf{P}, \mathbf{V}$ are associated with the model dynamics, the error in the given data and the measurement noise statistics (Metaxas & Terzopoulos 1993). Since we are measuring local short range forces directly from the image potential we create, the term $\mathbf{z} - \mathbf{h}(\hat{\mathbf{u}})$ represents the 2D image forces. Using the above Kalman filter, we can predict at every step the expected location of the data in the next image frame, based on the magnitude of the estimated parameter derivatives $\dot{\mathbf{q}}$.

Self Occlusion and Disocclusion

As an object rotates in space, or as the viewpoint of the observer changes substantially, certain faces of the object will become occluded or disoccluded by itself (a *visual event*). Hence, the corresponding line segment or edge feature in the image will appear or disappear over time. By using the Kalman filter to predict the position and orientation of the model in the next time frame, we can quantitatively predict the occurrence of a visual event. In other words, we can determine by using our active node determination approach, which subset of the model nodes will be active in the next image frame, and suppress their contributions to the net forces applied to the model. For stereo images, this prediction can be performed independently to the left and right images. In this case, two sets of active model nodes are maintained at any particular moment.

Tracking Multiple Objects with Occlusion

Our framework for tracking objects based on image potentials can be easily extended to deal with multiple independently moving objects and multi-part objects. The complication here is that object parts may occlude one another in different ways. By tracking objects in 3D using stereo images, we can predict the 3D positions of the nodes on each model based on the current estimates of their respective model parameters and their rate of change. Active nodes on each model will be made "inactive" if they are predicted to be occluded by surfaces of other models. This visibility checking is performed for each node on a model and against every surface of the other models in the scene. In practice, much of this checking can be avoided based on approximate estimates of each object's size and 3D location. We demonstrate in the experiments section that we are able to track all the objects in a scene even when some object parts become partially occluded.

There are also two more cases of object occlusion in case of multiple independently moving objects. The first case occurs when another moving object that was not previously present in the scene occludes the object being tracked. The second is due to an error from the qualitative segmentation system which did not detect an object during the model initialization step. Our system can handle both situations by monitoring the local forces exerted on the model. If no force or forces of unusual magnitude and direction are exerted on some of the predicted active nodes of the currently tracked model, the event signals the possibility that we have lost track of the object. In such a situation, we apply the qualitative segmentation system to resolve the ambiguity. After proper re-initialization of our models, we continue tracking using local image forces based on our physics-based technique.

Experiments

We demonstrate our approach in a series of tracking experiments involving real stereo image sequences. All images are 256×256 pixels and all the examples run at interactive rates on a SGI R4000 Crimson workstation, including real-time 3D graphics. In the first experiment, we consider a sequence of stereo images (16 frames) of two independently moving objects. The objects move towards each other along 2 different paths which are approximately linear and the paths' relative angle is about 20 degrees. The baseline of the two cameras is 100mm, they are both at a declination angle of 30 degrees from the horizon, and their relative rotation is 8 degrees. Fig. 1(a) shows the first pair of stereo images. The initial pose and shape of the objects are recovered using techniques mentioned before and they are subsequently tracked based on image forces only. Figs. 1(b-g) show snapshots of the two objects being tracked with the wire-frame models overlaid on the image potential. They demonstrate that our technique is able to continue the tracking even when one of the blocks becomes partially occluded and then disoccluded. Note that those active model nodes which are temporarily occluded are automatically identified and made inactive. Figs. 2(a-d) show the relative positions of the recovered models at four different instants.

In the second experiment, we consider a sequence of stereo images (24 frames) of a scene containing multiple objects, including a two-part object. Fig. 3 shows the initial stereo images of the multi-object scene. The baseline of the stereo cameras is 150 mm, the cameras

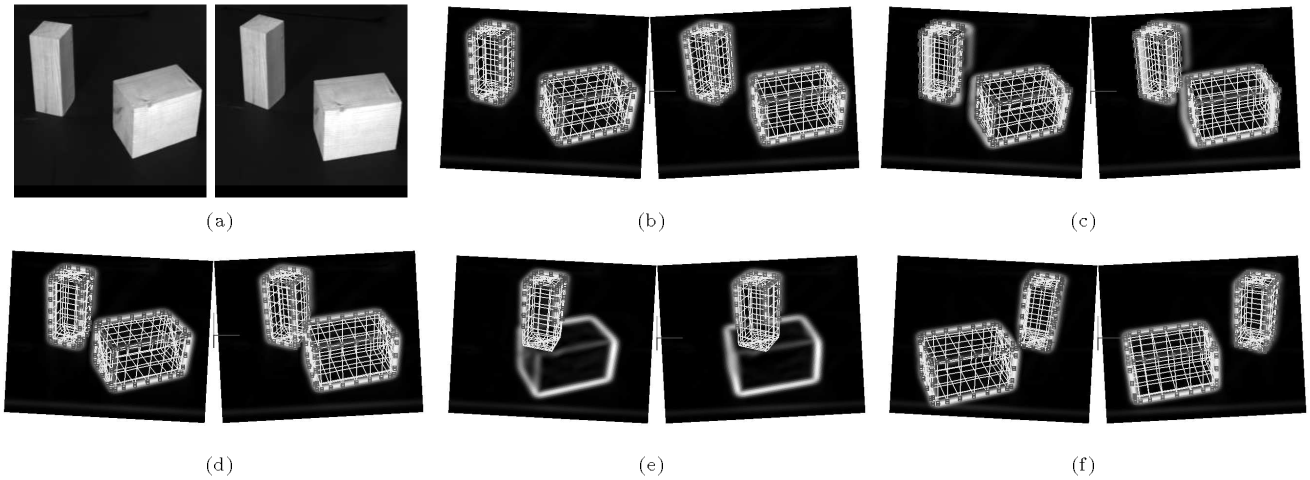

Figure 1: Tracking two independently moving blocks in a sequence of stereo images: (a) initialized models, (b) coming of a new frame, (c) beginning of the occlusion, (d) taller block partially occluded, (e) taller block becomes disoccluded, (f) no more occlusion. Note that only the active model nodes are marked, while the occluded ones are not.

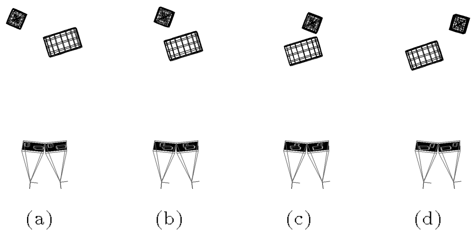

Figure 2: Recovered models of the two moving blocks in 3D over time from a top view (with the stereo arrangement).

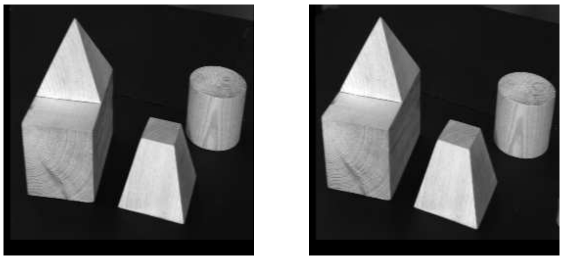

Figure 3: Initial stereo images of the multi-object scene.

are at a declination angle of 30 degrees from the horizon, and their relative rotation is 12 degrees. The cameras are rotated around the scene at a constant rate. Fig. 4(a) shows the initialized models using the same technique as before. Fig. 4(b) shows image potentials at an intermediate time frame where the aspects of some parts have changed and some parts have become partially occluded. Figs. 4(c-f) show that each object is still successfully tracked under these circumstances with the individual part models overlaid on the image potentials in Fig. 4(b).

In the last experiment we demonstrate the applicability of our technique in case of object occlusion by another undetected object. We use the same sequence of stereo images as in the first experiment where there are two independently moving blocks, but we do not assume prior detection of one of the blocks this time. Fig. 5 shows the instant at which some of the image forces exerted on the active model nodes of the moving block exceed a threshold. It then triggers the qualitative segmentation system to resegment the scene and correctly group edges belonging to each of the blocks. After the models are reinitialized, tracking continues as before.

Conclusion

We have presented a new integrated approach to object tracking in 3D from 2D stereo image sequences. After initializing our deformable models based on a part-based qualitative segmentation system, we subsequently track the objects using our physics-based approach. We further used a Kalman filter for estimating the object's shape and motion which allowed the prediction of possible visual events; thus we were able to determine where on the model, image forces can be exerted. We also demonstrated that our approach can deal with object occlusion from other independently moving objects by predicting each object's motion in the scene. Occlusion due to previously unidentified objects can also be detected by monitoring changes to the image forces exerted on the models. Based on these changes, the qualitative shape segmentation system is invoked for scene re-segmentation and model reinitialization. We are currently extending our system to handle objects composed of more complex primitives than the ones we have assumed.

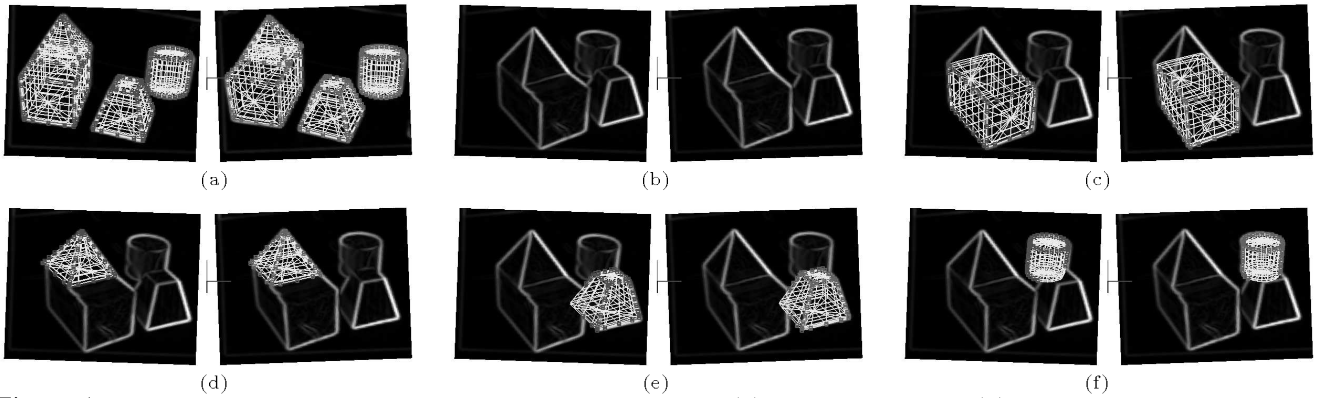

Figure 4: Tracking multiple objects in a sequence of stereo images (a) initialized models, (b) image potentials of an intermediate frame (both occlusions and visual events have occurred) (c-f) each object part correctly tracked with part models overlaid on the image potentials in (b). Note that only the active model nodes are marked, while the occluded ones are not.

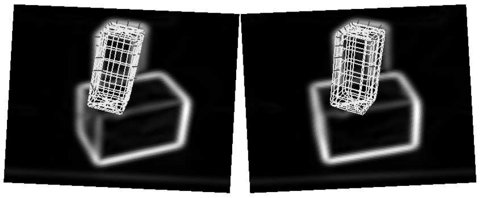

Figure 5: Unpredicted object occlusion: no knowledge of the 2nd block is assumed. It is detected by monitoring the forces exerted on the active nodes of the displayed model.

References

Blake, A.; Curwen, R.; and Zisserman, A. 1993. Affine-Invariant Contour Tracking with Automatic Control of Spatiotemporal Scale. In *Proc. IEEE 4th International Conference on Computer Vision*, 502–507.

Broida, T. J.; Chandrashekhar, S.; and Chellappa, R. 1990. Recursive 3-D Motion Estimation from a Monocular Image Sequence. *IEEE Transactions on Aerospace and Electronic Systems* 26(4):639–656.

Deriche, R., and Faugeras, O. 1990. Tracking Line Segments. *Image and Vision Computing* 8(4):261–270.

Dickinson, S.; Pentland, A.; and Rosenfeld, A. 1992a. From Volumes to Views: An Approach to 3D Object Recognition. *Computer Vision, Graphics, and Image Processing: Image Understanding* 55(2):130–154.

Dickinson, S.; Pentland, A.; and Rosenfeld, A. 1992b. Shape Recovery Using Distributed Aspect Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2):174–198.

Dickmanns, E. D., and Graefe, V. 1988. Applications of Dynamic Monocular Machine Vision. *Machine Vision and Applications* 1:241–261.

Duncan, J. S.; Owen, R. L.; and Anandan, P. 1991. Measurement of Nonrigid Motion Using Contour Shape Descriptors. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 318–324.

Gennery, D. 1992. Visual Tracking of Known Three-Dimensional Objects. *International Journal of Computer Vision* 7(3):243–270.

Huang, T. S. 1990. Modeling, Analysis and Visualization of Nonrigid Object Motion. In *Proc. IEEE 10th International Conference on Pattern Recognition*, volume 1, 361–364.

Kass, M.; Witkin, A.; and Terzopoulos, D. 1988. Snakes: Active Contour Models. *International Journal of Computer Vision* 1(4):321–331.

Lowe, D. 1991. Fitting Parameterized Three-Dimensional Models to Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(5):441–450.

Metaxas, D., and Dickinson, S. 1993. Integration of Quantitative and Qualitative Techniques for Deformable Model Fitting from Orthographic, Perspective, and Stereo Projections. In *Proc. IEEE 4th International Conference on Computer Vision*, 641–649.

Metaxas, D., and Terzopoulos, D. 1993. Shape and Nonrigid Motion Estimation Through Physics-Based Synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(6):580–591.

Pentland, A., and Horowitz, B. 1991. Recovery of Nonrigid Motion and Structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(7):730–742.

Szeliski, R., and Terzopoulos, D. 1991. Physically-Based and Probabilistic Modeling for Computer Vision. In *Proc. SPIE Geometric Methods in Computer Vision*, volume 1570, 140–152. Society of Photo-Optical Instrumentation Engineers.

Terzopoulos, D., and Metaxas, D. 1991. Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(7):703–714.

Terzopoulos, D.; Witkin, A.; and Kass, M. 1988. Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motion. *Artificial Intelligence* 36(1):91–123.

Thompson, D., and Mundy, J. 1988. Motion-Based Motion Analysis: Motion from Motion. *Robotics Research: The Forth International Symposium* 299–309.

Verghese, G.; Gale, K.; and Dyer, C. 1990. *Real-time, Parallel Tracking of Three-Dimensional Objects from Spatiotemporal Sequences.* New York: Springer-Verlag. 310–339.