Improving the Scope of Deformable Model Shape and Motion Estimation

Cristian Sminchisescu, Dimitris Metaxas, Sven Dickinson

Improving the Scope of Deformable Model Shape and Motion Estimation

Cristian Sminchisescu
INRIA Rhône-Alpes
655 avenue de l'Europe
38330 Montbonnot, France
Cristian.Sminchisescu@inrialpes.fr

Dimitris Metaxas
University of Pennsylvania
200 South 33rd Street
Philadelphia, USA
dnm@graphics.cis.upenn.edu

Sven Dickinson
University of Toronto
6 King's College Road
Toronto, Canada
sven@cs.toronto.edu

Abstract

Previous approaches to deformable model shape estimation and tracking have assumed a fixed class of shapes representation (e.g., deformable superquadrics), initialized prior to tracking. Since the shape coverage of the model is fixed, such approaches do not directly accommodate incremental representation discovery during tracking. As a result, model shape coverage is decoupled from tracking, thereby limiting both processes in terms of scope and robustness. We present a novel deformable model framework that accommodates the incremental incorporation during tracking of new geometric primitives (lines, in addition to points) that are not explicitly captured in the initial deformable model but that are moving consistently with its image motion. As these new features are detected via consistency checks, they are added to the model, providing incremental soft constraints on the estimation of its rigid parameters. The consistency checks are based on trilinear relationships between geometric primitives. Consequently, we not only increase both model scope and, ultimately, its higher-level shape coverage, but improve tracking robustness and accuracy, by directly employing the new features in both forward prediction and reconstruction. Our new formulation is a step towards automating model shape estimation and tracking, since it requires significantly reduced initial model hand-crafting. We demonstrate our approach on two separate image-based tracking domains, each involving complex 3D object shape and motion.

Keywords: *Deformable models, geometric constraints, object tracking, model grouping, bundle adjustement.*

1 Introduction

1.1 Approach

Deformable models offer a powerful, data-driven framework for recovering the shape and motion of an object through numerical integration of a dynamical system that encodes the fitting error. The process is a physics-based interpretation of an optimization process, in which the initial model shape and pose are supplied by an initialization process that aligns the model with the object of interest in the initial image. The model subsequently moves under the action of forces derived from image measurements which, in turn, are converted (through a Jacobian matrix) into generalized forces that act on the model's rigid and non-rigid degrees of freedom.

The quest for robustness and accuracy during model estimation naturally leads to the search for geometrically and photometrically derived image forces (cues), for model parameterizations, and for ways to combine them in a consistent manner. Although previous approaches have investigated the use of particular combinations of cues in the context of deformable models, they assumed *a fixed and known* initial model *shape representation*.

In this paper, we present a novel two-step technique in which new model structure is revealed during tracking using geometric consistency checks. These checks combine heterogeneous information, involving an initial model's rigid parameters as well as independently tracked line features in the image. Our goal is to enhance the higher level model shape representation with additional deformable shapes, and enrich its basic discretization power through the incremental addition of line features.

In our approach, line features (not belonging to the initial deformable model) are tracked in an image sequence. Those features whose motion is consistent with the model are reconstructed, integrated into the initial (high-level) model representation, and subsequently used to increase tracking robustness. *The technique consequently attempts to bridge the gap between model-based, top-down estimation techniques and classical bottom-up, feature-based reconstruction techniques, by relaxing some constraints on each side. Complete model representation is no longer fully known a-priori, while knowledge of feature motion is acquired during reconstruction, once their identity as being part of the model is established.* Our final model is a mixture of lines, points and parameterized shapes, and generalizes the shape coverage of previously-used deformable models. We show how a deformable model framework naturally accommodates rigid and non-rigid high-level shape representations and their associated constraints,

and various model discretizations such as points and lines, by combining the simplicity of linear reconstruction methods with the refinement and bias removal of non-linear estimation methods.

We present experiments involving complex objects and motion in monocular video sequences and show that the proposed method is able to recover model structure accurately and efficiently. We also quantitatively show how the novel recovered structure significantly improves the accuracy and speed of the deformable models estimation process, providing important additional constraints especially during difficult to track, towards camera motion.

1.2 Relation to Previous Work

In the area of physics-based deformable models, various formulations have been proposed ([25, 13, 17, 16]). As powerful as these techniques are, they typically assume the model representation is fixed and known a-priori, sometimes imposing a heavy burden on the model initialization/recovery process ([6, 5]). Furthermore, a representational "gap" exists between the coarse, parametric shapes used to model the objects and the point-based discretizations used to bind them to the image; no obvious method exists to bridge this gap through the inclusion of other basic geometric primitives, e.g., lines. On the other hand, it has been recognized ([19]) that the rather different parameterizations of various features encountered in different approaches or applications usually lead to difficult problems when attempting to integrate those features within a particular representation or optimization procedure.

In the area of constraint integration, particularly in a deformable model setting, contours and stereo ([26]), shading and stereo ([7]), contours and optical flow ([3, 4]), and shading ([18]) have been introduced, with good results. Beyond the particular choice of sources of information to use, the approaches differ in the way they fuse their component information. Some combine the information in a symmetric manner and weight them statistically (*soft* constraints). Others favor a particular hierarchical constraint satisfaction order with an exact policy, such that inconsistent contributions to the solution from constraints further down in the hierarchy are pruned away by constraints higher-up (*hard* constraints).

Finally, in the area of feature-based rigid motion and structure estimation, various approaches based on the type of correspondences (2-D to 2-D, 2-D to 3-D, or 3-D to 3-D) as well as the type of features (lines, points, or corners), have also been proposed (see [10] for a review). Several algorithms, commonly based on Kalman filtering, are also known ([2, 27]). Approaches to simultaneous structure and motion estimation, based on the inversion of the so-called forward model, have been presented. Many of these methods, such as [23, 1, 24], could be related to the classical relative orientation paper ([11]).

The approach we present here is based on a deformable model formulation similar to ([25]), in which we employ an articulated chain of deformable superquadrics with global deformations, such as tapering bending, shearing and twisting. For broader modeling coverage, additional local deformation functions could also be used. As in the standard formulation, we employ local contour-based forces derived from image gradients for basic model motion estimation. We diverge from previous approaches in that we assume our initial model is incomplete, and attempt to recover additional model structure through the use of geometric consistency checks embedded in a tracking framework. Specifically, we integrate constraints in the form of new line features discovered in the image as moving consistently with the model, with constraints in the form of point features derived from a discretization of the model.

The line feature consistency checks we use are based on the conditions present in separate, bottom-up structure and motion estimation under 2-D to 2-D line correspondences in the Euclidean calibrated case ([10, 15, 28]). A later body of work for calibrated, uncalibrated, and projective reconstruction cases ([22, 21, 8] addresses trilinear constraints between points and lines in 3 views, and presents linear, bottom-up, and model-free rigid reconstruction methods (see [9] for a comprehensive review).

However, unlike these approaches, we assume an incomplete deformable and Euclidean model and we don't attempt to solve for the rigid parameters. On the contrary, by knowing the estimated rigid parameters of the model, we only attempt to check whether the consistency conditions are indeed satisfied for arbitrarily independent tracked lines in at least three frames. Once consistent lines have been identified, they are reconstructed, and the model representation is enhanced to include them. Their prediction in subsequent frames under the action of the Euclidean group is derived, while their structure is continuously re-estimated together with the deformable structure and motion of the other model parts. Their contribution to the solution for the model's estimated rigid parameters is linearized by means of their corresponding Jacobian matrix and integrated as an additional soft constraint (see section 3.5).

2 Deformable Model Formulation

In the next two sections, we shall briefly review the deformable model formulation in terms of its underlying geometry and dynamics (see [14] for details).

2.1 Model Geometry

The reference shape of the model is defined over a domain Ω as $\mathbf{p} = \mathbf{T}(\mathbf{q_d}, \mathbf{u})$, where $\mathbf{T}$ defines a global deformation based on the parameters, $\mathbf{q_d}$ and $u \in \Omega$. Following [20], we employ a deformable superquadric ellipsoid, having global tapering and bending deformations, as a reference shape. The posi-

tion of a point on the model is expressed with respect to a world coordinate system Φ as $\mathbf{x} = \mathbf{D}(\mathbf{q_r}, \mathbf{p})$, where $\mathbf{D}$ represents a rigid displacement parameterized by $\mathbf{q_r}$. Finally, both rigid and non-rigid parameters are assembled in a vector, $\mathbf{q} = (\mathbf{q_r}, \mathbf{q_d})$, sometimes referred to as the model's generalized coordinates.

2.2 Dynamics and Generalized Forces

The velocity of points on the model is given by:

$$\dot{\mathbf{x}}(\mathbf{u}) = \mathbf{L}(\mathbf{q}, \mathbf{u})\dot{\mathbf{q}} \qquad (1)$$

where $\mathbf{L}$ is the model Jacobian matrix [25]. By omitting the mass term (since we want a model free of inertia), the Lagrangian dynamics equations become [25]:

$$\mathbf{D}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f_q}, \quad \mathbf{f_q} = \int \mathbf{L}^\top \mathbf{f}\mathbf{du} \qquad (2)$$

where $\mathbf{D} = \int \gamma \mathbf{L}^\top \mathbf{L}$, and $\mathbf{K} = \mathbf{diag}(k_{si})$, with k_{si} being the stiffness associated with the global parameter i.

When the estimation of the model is based on image contours, the above formulation has to be extended to include: 1) the perspective projection Jacobian, $\mathbf{L_p}$, computed in terms of the model Jacobian and the perspective transform Jacobian evaluated at points on the model discretization, and 2) the definition of local forces $\mathbf{f_p}$ derived from image potentials (see [14] for derivations). Consequently, the generalized forces with respect to edges are defined as:

$$\mathbf{f_q}^{\mathbf{e}} = \int \mathbf{L_p}^\top \mathbf{f_p}\mathbf{du} \qquad (3)$$

3 Line Feature Formulation

The formulation of the tracking process in the previous sections is based on a physics-based deformable framework, where the model moves under the action of local forces derived from image potentials. In this section, we extend that formulation by integrating new line features into the model – line features that are initially not part of the model, but for which evidence exists in the image. We begin by tracking a minimal model in a sequence of images and attempt, incrementally over time, to: 1) identify line features moving consistently with the model, and 2) augment the model with these consistent features to improve its tracking. At each iteration, the model not only increases in scope but covers more of the image features, resulting in more robust tracking of the object.

Our approach consequently attempts to relate two processes at the image and model levels, respectively. At the image level, we employ purely image-based techniques to track lines in a sequence of frames. We call such lines *image-tracked lines* (ITL). At the model level, we use contour based

gradient forces to estimate the model (rigid and non-rigid) parameters. First, we decide whether an ITL represents a line belonging to the object (but not present in its model) by means of two geometric consistency checks, derived from tracking the ITL in at least three successive frames. The lines that pass this test are called *consistent image-tracked lines* (CITL). Second, for CITLs detected in the previous step, we attempt to robustly recover their structure in terms of an underlying parameterization in a model-centered frame. Third, we predict how a CITL will appear in a subsequent image based on the current estimate of the model's rigid motion. We call this prediction a *model-predicted line* (MPL). Finally, we use the error between a CITL and a MPL to define additional image alignment forces. They are subsequently linearized by means of their corresponding Jacobian matrices and combined into the model estimation procedure as soft constraints. The initially reconstructed line representation is subsequently re-estimated, together with the entire model structure and motion, within the deformable model framework.

3.1 Line parameterization

In the following formulation, we denote a line in 3-D by lower-case letters, $l_i(i = 1..n)$, and its corresponding projected line (or segment) in the image plane by capital letters, $L_i(i = 1..n)$. A 3-D line is parameterized by a unit vector $\mathbf{v}$, representing one of two possible directions on the 3-D line l, and a vector $\mathbf{d}$ terminating on l and perpendicular to it (see Fig. 1a). This line representation forms a 6-dimensional parameter space with 4 degrees of freedom. Consequently, the underlying relation between the line primitives can be identified as a 4-dimensional manifold embedded in the abstract parameter space, and any line can be identified with (actually two) points on this manifold ([12, 24]).

(a) Line Pameterization   (b) Line Motion in Three Frames

Figure 1: Line Representation and Motion

The 3-D line, l, and the optical center of the camera determine a plane (line's interpretation plane) with normal, $\mathbf{N} =$

(N_x, N_y, N_z). This plane intersects the image plane, defined by the equation $z = f$ (f being the camera focal length) at line L. The equation of line L in the image plane can be consequently written as:

$$N_x X + N_y Y + N_z f = 0 \qquad (4)$$

The above relation, used in an inverse manner, yields an immediate representation for the normal of the plane containing a 3-D line in space by knowing the image plane equation of its projected (observed) line, or segment. In vector terms, given a plane, $\mathbf{P} = (N_x, N_y, N_z, 0) = (\mathbf{N}, 0)$, and any point belonging to the interpretation plane of a given line, $\mathbf{X} = (X, Y, Z, 1)$, we can write the plane equation as $\mathbf{P}^\top \mathbf{X} = \mathbf{0}$.

3.2 Model-Based Consistency Checks

The standard formulations for recovering motion and structure using 2-D to 2-D line correspondences (e.g., [10, 28, 15]) rely on three frames and at least six line correspondences (although no formal proof is yet available, [10]) for uniquely recovering the structure and motion of a rigid object (within a scale factor for translation and structure parameters). Furthermore, for the two view case, the resulting system of equations is underconstrained, so one would expect infinitely many solutions.

Consider now the motion of a 3-D line, l, in three successive frames ($l_i, i = 1, 2, 3$, with direction support $\mathbf{v}_i, i = 1, 2, 3$) and it's corresponding image projected lines ($L_i, i = 1, 2, 3$). The motion between frames 1 and 2 is described by the translation and rotation, t_{12} and R_{12}, and for frames 1 and 3, by t_{13} and R_{13}, respectively. The corresponding normals for the interpretation planes, P_1, P_2, P_3, determined by the line L and the center of projection in the three frames, are $\mathbf{N_1}, \mathbf{N_2}, \mathbf{N_3}$, respectively (see Fig. 1b).

The following two relations can be derived either geometrically or algebraically, from the quantities presented above ([10, 15, 28]):

$$\mathbf{N_1} \cdot (R_{12}^{-1}\mathbf{N_2} \times R_{13}^{-1}\mathbf{N_3}) = 0 \qquad (5)$$

$$-t_{12} \cdot (R_{12}\mathbf{N_1}) = \frac{\|\mathbf{N_2} \times R_{12}\mathbf{N_1}\|}{\|\mathbf{N_2} \times R_{23}^{-1}\mathbf{N_3}\|} \cdot R_{23}^{-1}t_{23} \cdot R_{23}^{-1}\mathbf{N_3} \quad (6)$$

In this model-based formulation, we are not interested in solving for the rotation and translation but, given a model with *known* motion and some *independent* ITLs, *checking* whether those lines are moving consistently with our model (CITLs). More specifically, given an ITL in three frames (that is, knowing $\mathbf{N_1}$, $\mathbf{N_2}$ and $\mathbf{N_3}$) as well as the motion of the model (that is, R_{12}, t_{12} and R_{13}, t_{13}), we use equations (5) and (6) to check whether the motion of the line is consistent with the model's motion. If so, we hypothesize that the line is part of the object and therefore should be added to the model. One

can prove that the above two relations are necessary and sufficient conditions for consistency. In practice, the above two relations are never satisfied exactly, due to slight errors in the image tracked lines and errors in the estimated model rigid parameters, so a threshold must be chosen, based on the covariance of both model rigid parameters estimation and the image tracked lines, respectively.

3.3 Robust Unbiased Model-Based Structure Recovery

Once a moving image line has been assigned to the model through the above consistency check, the next step is to recover its structure, i.e., the vector pair $(\mathbf{v}, \mathbf{d})$ in a model-centered coordinate system. In order to increase the robustness of the recovery process, we can use as many line correspondences in as many frames (at least two) as are available. The process can be formulated as follows: all interpretation planes for the line correspondences in a camera frame are transformed to a common, model-centered coordinate frame. Each line, l_i, having the interpretation plane, $\mathbf{P_i}$, is subject to the displacement, $\mathbf{D_c^{-1}D_i^{-1}}$, where:

$$\mathbf{D_c} = \begin{bmatrix} R_c & t_c \\ 0 & 1 \end{bmatrix}$$

is the displacement corresponding to the camera, and $\mathbf{D_i}$ is the displacement corresponding to the model (in the world coordinate system) in frame i. Then, the equation of the plane in the object-centered frame is: $\mathbf{P_i}^\top \cdot \mathbf{D_c^{-1}D_i^{-1}} \cdot \mathbf{X} = 0$. By stacking together the equations for corresponding lines, we obtain:

$$\mathbf{A} \cdot \mathbf{X} = \begin{bmatrix} \mathbf{P_1^\top} \cdot \mathbf{D_c^{-1}D_1^{-1}} \\ \mathbf{P_2^\top} \cdot \mathbf{D_c^{-1}D_1^{-1}} \\ \cdots \\ \cdots \\ \cdots \\ \mathbf{P_k^\top} \cdot \mathbf{D_c^{-1}D_k^{-1}} \end{bmatrix} \cdot \mathbf{X} = \mathbf{0}$$

Since we are looking for the line intersection of all the above planes, the above [k x 4] matrix $\mathbf{A}$ should have rank 2. Any point $\mathbf{p}$ on the intersecting line can be written as a linear combination of the singular vectors corresponding to the 2 smallest singular values of the matrix $\mathbf{A}$:

$$\mathbf{p} = a \cdot \mathbf{X_{s1}} + b \cdot \mathbf{X_{s2}} \qquad (7)$$

The corresponding line parameterization can be subsequently recovered as:

$$\mathbf{v} = \mathbf{X_{s1}} - \mathbf{X_{s2}} \qquad \mathbf{d} = (\mathbf{I} - \frac{\mathbf{v} \cdot \mathbf{v}^\top}{\|\mathbf{v}\|^2}) \cdot \mathbf{X_{s1}} \quad (8)$$

As mentioned before, a 3D line has 4 intrinsic degrees of freedom while a projected image line has just 2. Measurements are collected in 3 frames, so this will determine 3x2-4=2 independent relations.

The stability of the reconstruction can be verified in terms of the ratio of the 2nd and 3rd singular values of $\mathbf{A}$ (remember that in the noise free case the last two singular values should be zero), being satisfactory when this ratio is high. The above linear method, although robust, might be prone to bias in the initial line parameter estimates, due to fixed rigid displacements. However, we are working in a deformable model framework, in which *both* model structure *and* model motion are jointly estimated. Consequently, any such initial linear bias is automatically removed during estimation in subsequent frames.

3.4 Forward Model Line Prediction

Once consistent lines (CITL) have been identified, recovered, and effectively added to the model, we use these new lines to improve the tracking of the enhanced model by imposing further constraints on its alignment with the data. The approach we follow is to define alignment forces between a CITL in the image and the projection (MPL) of its corresponding (new) model line, as predicted under the forward motion of the model.

(a) Forward Line Transfer   (b) Alignment forces
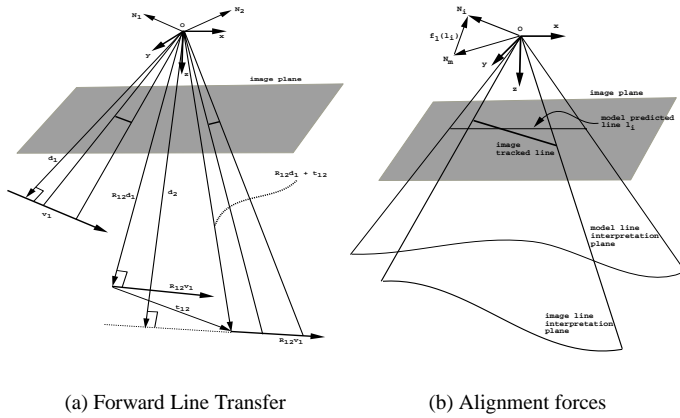
Figure 2: Lines transfer and alignment

Consider the two frame case, as illustrated in Fig. 2a. Given $(\mathbf{N_1}, \mathbf{v_1}, \mathbf{d_1})$, one can obtain $(\mathbf{N_2}, \mathbf{v_2}, \mathbf{d_2})$ by geometric means as follows:

$$\mathbf{v_2} = R_{12}\mathbf{v_1} \qquad (9)$$

$$\mathbf{d_2} = (R_{12}\mathbf{d_1} + \mathbf{t_{12}}) - \mathbf{v_2}((R_{12}\mathbf{d_1} + \mathbf{t_{12}}) \cdot \mathbf{v_2}) \qquad (10)$$

$$\mathbf{N_2} = \frac{\mathbf{v_2} \times \mathbf{d_2}}{\| \mathbf{v_2} \times \mathbf{d_2} \|} \qquad (11)$$

As mentioned in the introductory section, $\mathbf{N_2}$ fully identifies the MPL in the second frame.

The Jacobian matrix associated with the line parameterization is derived analytically (it is omitted here, due to lack of space). Given a line representation, $\mathbf{u_l} = (\mathbf{v}^\top, \mathbf{d}^\top)^\top$, and the

rigid parameterization of the model $\mathbf{q_r}$, the Jacobian matrix of the line parameters with respect to the model parameters is a [6x7] matrix:

$$\mathbf{L_{u_l}} = \frac{\partial \mathbf{u_l}}{\partial \mathbf{q_r}} \qquad (12)$$

3.5 Image Forces and Soft Constraints

Given an MPL, identified by equation (4), and a CITL, identified by its endpoints $\mathbf{P_1} = (X_1, Y_1)$ and $\mathbf{P_2} = (X_2, Y_2)$, we define 2-D forces to compensate for their alignment error. Since the CITL and MPL each define an interpretation plane, we define the line alignment forces in terms of the alignment of their corresponding interpretation planes (see Fig. 2b). Given $\mathbf{N_i}$, the normal of the CITL interpretation plane, and $\mathbf{N_m}$, the normal of the MPL interpretation plane, the corresponding alignment force can be written as:

$$\mathbf{f_l}(\mathbf{u_{l_i}}) = k \parallel \mathbf{N_i} - \mathbf{N_m} \parallel \qquad (13)$$

Notice that a final transformation maps a line representation to an observable "normal" on which the line alignment is actually performed. The corresponding transformation is given by equation (11), and involves the computation of another [3x6] Jacobian matrix:

$$\mathbf{L_N} = \frac{\partial \mathbf{N}}{\partial \mathbf{u_l}} \qquad (14)$$

The Jacobian $\mathbf{L_l} = \frac{\partial \mathbf{N}}{\partial \mathbf{q_r}}$ corresponding to the mapping between the line representation in the model frame and the observation (the normal of the corresponding line interpretation plane) is computed via the chain rule in terms of Jacobians (12) and (14).

The normal alignment forces map the observation error corresponding to the line, l_i, to the model's parameter space by means of the corresponding line Jacobian, $\mathbf{L_l}(\mathbf{u_{l_i}})$. As a result, the combined generalized forces (due to image gradient and line alignment) acting on the model's degrees of freedom sum over both point and line forces mapped by their corresponding Jacobians:

$$\mathbf{f_q} = \sum_i \mathbf{L_p}(\mathbf{u_i})^\top \mathbf{f_p}(\mathbf{u_i}) + \sum_{l_j} \mathbf{L_l}(\mathbf{u_{l_j}})^\top \mathbf{f_l}(\mathbf{u_{l_j}}) \qquad (15)$$

4 Experiments

The experiments consist of two sequences, each containing of 4 seconds of video (200 frames recorded at 50 fps) of a moving bike and of a space robotics end-effector grapple fixture. Both sequences involve significant translational and rotational motion in the camera frame. Part of the bike frame is modeled and tracked by means of a model consisting of 3 parts (Fig. 3a), while the grapple fixture consists of only a square parallelepiped (Fig. 4a). Newly recovered models
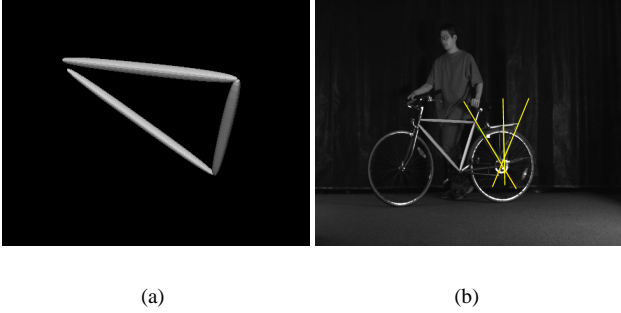
(a)   (b)

Figure 3: Initial and Reconstructed Model Tracking
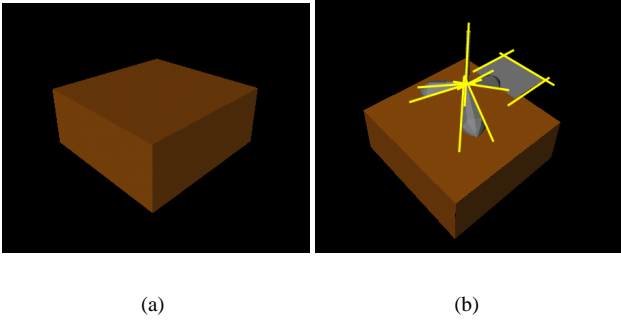

(a)   (b)

Figure 4: Initial and reconstructed model, with new superquadrics parts fitted

are displayed in Fig. 3b and Fig. 4b, in which additional deformable shapes have been fitted based on 3-D reconstructed model lines, their covariance, and their corresponding image contours (an algorithm for such higher-level shape-recovery and grouping is beyond the scope of this paper, and will be described in a subsequent paper).

The models have been manually initialized in the first frame of the sequence, and are displayed in subsequent frames aligned (overlaid and rendered in flat-shaded grey for the bike and brown wireframe for the grapple fixture, respectively) with the part of the object they model (for tracking results, see Fig. 6). Additional lines, not part of the initial models, are tracked through the sequence by means of an independent gradient-based line tracker, involving interest point-based tracking, followed by line fitting and re-normalization in each frame. Lines that are determined to be CITLs are displayed in green. The model reconstructed/predicted lines are displayed in yellow. The consistency check is evaluated, using a threshold $\tau = 0.05$, based on the lines present in frames (20, 40, 60) for the bike sequence. For the space shuttle sequences, lines are incrementally tracked and recovered (some are not visible initially) and the consistency checks are performed several times for different lines, in the frames (20, 40, 60), (120, 140, 160) and (160, 175, 190).

It is important that the estimation of the models' rigid motion is accurate as it affects both the validity of consistency checks and the quality of the line reconstruction. In practice, since the models are initialized manually, there is always uncertainty associated with their initialization. For this reason, we do not attempt to check and reconstruct lines immediately after model initialization, but rather allow a slight delay (around 20 frames in our experiments), such that the model locks onto the data.

Any initial *bias* in the linear reconstruction is subsequently eliminated through the re-estimation of the line representation jointly with all the rigid and non-rigid model parameters in the deformable model framework.

The least-squares-based reconstruction for the bike sequence is based on 12 frames within the interval, 20-60, while for the grapple fixture, we use 12 frames in 20-60, 120-160, 160-190, respectively. In our experiments, this provided us with sufficiently important motion and sufficient additional lines for accurate, constrained reconstruction. The stability of the reconstruction is checked, as explained in §3.3 by the ratio of the 2nd and 3rd singular values associated with the **A** matrix. We thus employ a principled criteria for selecting sufficient lines and corresponding informative displacements for the reconstruction. In both sequences, the frames prior to new line reconstruction are tracked using the prior model along with contour-based image forces, while the rest of the sequence (once the CITLs have been reconstructed) is tracked using the enhanced model (consisting of both the initial model *and* the reconstructed lines, and using *both* contour and alignment forces between CITLs and MPLs).

Although no lines are reconstructed until frame 60 in the bike sequence, their re-projection is displayed over the entire sequence (note that since they are reconstructed in a model-centered coordinate frame, they can be transferred backwards, since the inter-frame motion of the model has already been estimated). For the grapple fixture sequence, we avoided plotting the lines over the entire sequence in order to show how they are incrementally tracked and recovered. Notice that for both sequences, the reconstructed lines' projections are correct.

The tracking results emphasize increased stability, as the line features are reconstructed and integrated into the tracking process as additional cues, especially during difficult to track towards camera motion of the grapple fixture. The number of iterations in the integration of motion equation (2) decreases with the integration of additional line cues (see Fig. 5 for plots). The average per-node error decreased from 0.9 (initially) to 0.4 pixels (frame 60) in the bike tracking sequence, and from 1.2 pixels (initially) to 0.8 (frame 60), 0.6 (frame 160) and 0.2 (frame 190) in the grapple fixture sequence, respectively, reflecting improved accuracy due to the integration of line soft constraints in the deformable model estimation process.
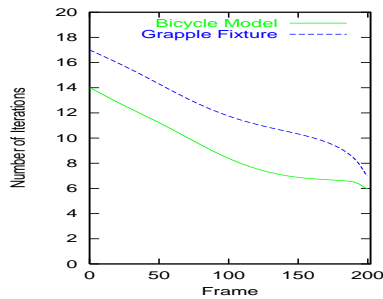
6

Figure 5: Number of iterations per frame (Bezier interpolation) decreases due to the addition of line constraints.

5 Conclusions

We have presented a framework for incremental model acquisition and tracking using deformable models. We have relaxed the constraint that the representation of the model has to be fully known a-priori, and have enhanced its basic discretization structure in terms of points and lines while preserving higher-level shape information (in terms of parametric shapes). This has allowed us to formulate geometric constraints for feature consistency, reconstruction, and tracking via cue integration in a flexible manner through the integration of top-down, unbiased deformable model-based estimation techniques and bottom-up, linear feature reconstruction techniques. Our initial experimental results are promising and show that the proposed framework is able to deal with objects with complex shape and motion, and to track and incrementally recover structure with improved accuracy. Subsequent work will include more rigorous quantitative evaluation of the reconstruction and tracking improvement results, as well as grouping and high-level model abstraction.

Acknowledgments

References

[1] A.Azarbayejani and A.Pentland. Recursive Estimation of motion, structure and focal length, IEEE PAMI, 1995.

[2] J.Crowley, P.Stelmaszyk, T.Skordas, P.Pugget. Measurements and integration of 3-D structures by tracking edge lines, IJCV, July 1992.

[3] D. DeCarlo and D.Metaxas. The integration of optical flow and deformable models with applications to human face shape and motion estimation, CVPR 1996, pp.231-238.

[4] D. DeCarlo and D.Metaxas. Combining Information in Deformable Models Using Hard Constraints, CVPR 1999.

[5] S.Dickinson and D.Metaxas. Integrating qualitative and quantitative shape recovery. IJCV, 13(3):1-20, 1994.

[6] S.Dickinson, A.Pentland, A.Rosenfeld. Shape recovery using distributed aspect matching. PAMI, 14(2):174-198, 1992.

[7] P.Fua and G.Leclerc. Taking Advantage of Image-Based and Geometry-based constraints to recover 3-D surfaces, CVIU, July 1996.

[8] R.Hartley. Lines and points in three views and the trifocal tensor, IJCV, 22(2):125-140, 1997.

[9] R.Hartley and A.Zisserman. Multiple View Geometry in Computer Vision, Cambridge University Press, 2000.

[10] T.S.Huang, A.N.Netravali. Motion and Structure from feature correspondences: A review, Proc. of the IEEE, 82(2), 1994.

[11] K.B.Horn. Relative Orientation, IJCV, January 1990.

[12] K.Kanantani. Statistical optimization for geometric computation: Theory and Practice, Elsevier, 1996.

[13] M.Kass, A.Witkin and D.Terzopoulos. Snakes:Active contour models. IJCV, 1(4):321:331, 1988.

[14] D.Metaxas. Physics-Based Deformable Models, Kluwer, 1997.

[15] A.Mitiche and J.Aggarwal. Line-based computation of structure and motion using angular invariance, IEEE Workshop on Motion, 1986.

[16] A.Pentland and B.Horowitz. Recovery of non-rigid motion and structure. PAMI, 13(7):730-742, 1991.

[17] A.Pentland and S.Sclaroff. Closed form solutions for physically-based shape modeling and recognition, PAMI, July 1991.

[18] D.Samaras and D.Metaxas. Incorporating illumination constraints in deformable models, CVPR '98, 322-329.

[19] S.Seitz. Implicit Scene Reconstruction from Probability Density Functions, Proc. DARPA Image understanding Workshop, CA 1998.

[20] F.Solina and R.Bajcsy. Recovery of Parametric models from range images: the case of superquadrics with local and global deformations. PAMI, 12(2):131-146, 1990.

[21] A.Shashua. Algebraic functions for recognition, PAMI, 17(8):779-789, 1995.

[22] M.Spestakis and J.Aloimonos. A multi-frame approach to visual motion perception, IJCV, 16(3):245-255, 1991.

[23] R.Szeliski and S.Kang. Recovery 3-D shape and motion from image streams using non-linear least-squares, DEC TR 93/3.

[24] C.Taylor and D.Kriegman. Structure and Motion from Line Segments in Multiple Images, PAMI, (11), 1996.

[25] D.Terzopoulos and D.Metaxas. Dynamic 3D models with local and global deformations:Deformable superquadrics, PAMI, 13(7):703-714, 1991.

[26] D.Terzopoulos, A Witkin and M.Kass. Constraints on deformable models: Recovering 3-D shape and non-rigid motion, Artificial Intelligence, 36(1):91-123,1988.

[27] T.Vieville and O.Faugeras. Feed-forward recovery of motion and structure from a sequence of 2-D line matches, ICCV, December 1990.

[28] B.Yen and T.S.Huang. Determining 3-D Motion and Structure of a Rigid Body Using Straight line Correspondences, NATO ASI Series, 1983.
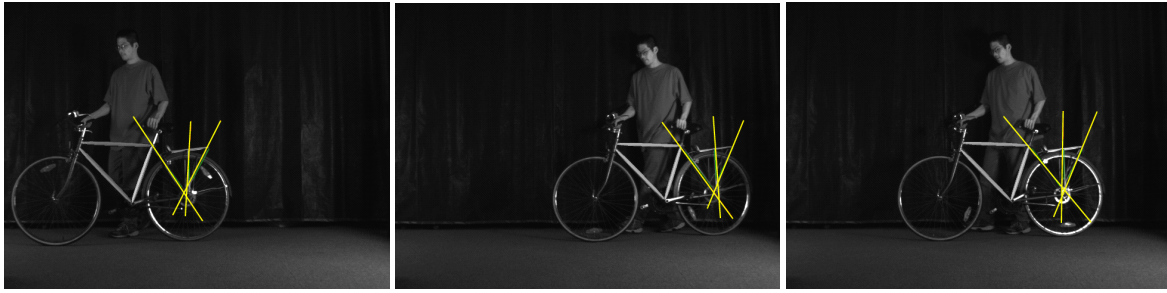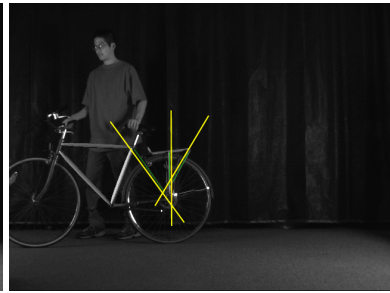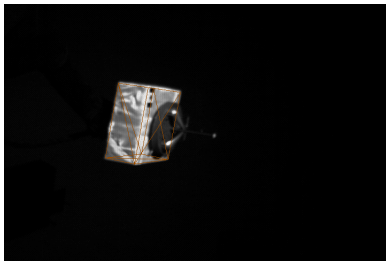
(a) frame 0

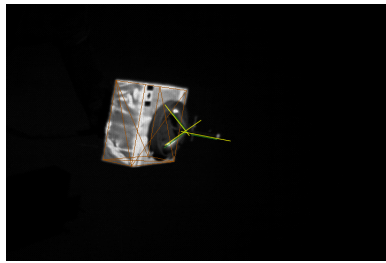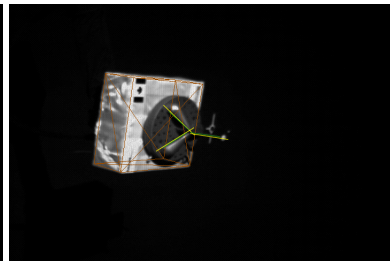(b) frame 40

(c) frame 80
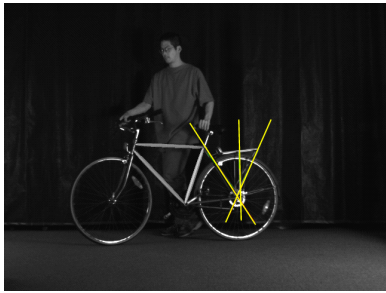
(d) frame 130
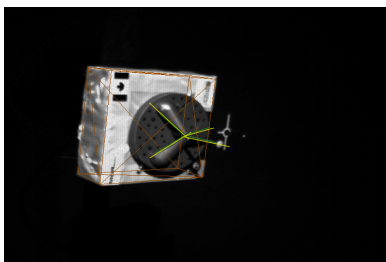
(e) frame 170

(f) frame 200
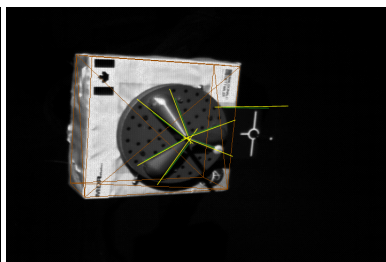
(g) frame 0, no reconstructed lines
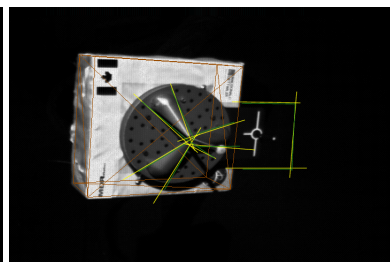
(h) frame 40, 3 reconstructed lines

(i) frame 80, 3 reconstructed lines

(j) frame 120, 4 reconstructed lines

(k) frame 160, 7 reconstructed lines

(l) frame 190, 10 reconstructed lines

Figure 6: Model Tracking (the bike model in grey flat-shaded and the grapple fixture model in brown wireframe) with CITLs (green) and MPLs (yellow)