

Incremental Model-Based Estimation Using Geometric Consistency Constraints

Cristian Sminchisescu, Dimitris Metaxas, Sven Dickinson

▶ **To cite this version:**

Cristian Sminchisescu, Dimitris Metaxas, Sven Dickinson. Incremental Model-Based Estimation Using Geometric Consistency Constraints. [Research Report] RR-4209, INRIA. 2001. inria-00072413

HAL Id: inria-00072413
https://inria.hal.science/inria-00072413

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Incremental Model-Based Estimation Using Geometric Consistency Constraints

Cristian Sminchisescu, Dimitris Metaxas, Sven Dickinson

N° 4209

June 2001

THÈME 3



ISSN 0249-6399 ISRN INRIA/RR--4209--FR+ENG



Incremental Model-Based Estimation Using Geometric Consistency Constraints

Cristian Sminchisescu*, Dimitris Metaxas†, Sven Dickinson ‡

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet MOVI

Rapport de recherche n° 4209 — June 2001 — 17 pages

Abstract: We present a physics-based deformable model framework for the incremental object shape estimation and tracking in image sequences. The model is estimated by an optimization process that relates image-based cost functions to model motion via the Lagrangian dynamics equations. Although previous approaches have investigated various combinations of cues in the context of deformable model shape and motion estimation, they generally assume a fixed, known, model parameterization, along with a single model discretization in terms of points. Our technique for object shape estimation and tracking is based on the incremental fusing of point information and new line information. Assuming that a deformable model has been initialized to fit part of a complex object (e.g., a bicycle) new line features belonging to the object but excluded from the initial model parameterization are identified during tracking. The identification is based on a set of novel model-based geometric consistency checks relating separate, independent tracking processes at the image feature and model levels, respectively. Identified features are reconstructed and integrated into the model parameterization which results in more accurate object shape estimation and subsequently they are used to increase model coverage in the image, thereby increasing tracking robustness. We derive the forward transfer under the action of the Euclidean group and Jacobian matrices for the underlying line feature mapping. New corresponding image alignment and generalized forces are introduced as soft constraints and combined with the forces derived from model contours to incrementally improve its motion estimation. We demonstrate our approach on image sequences with complex object shape and motion.

Key-words: deformable models, geometric constraints, constrained optimization, bundle adjustement

* INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 Montbonnot, France, *e-mail: Cristian.Sminchisescu@inria.fr*
† University of Pennsylvania, 200 South 33rd Street Philadelphia, PA 19104-6389, USA, *e-mail: dnm@cis.upenn.edu*
‡ University of Toronto, Toronto, Ontario, Canada M5S 3G4, *e-mail: sven@cs.toronto.edu*

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Estimation Incrémentale Basée Sur un Modèle Utilisant des Contraintes Géométrique de Consistance

Résumé : On présente un approche basée sur la physique des modèles deformables pour l'estimation incrémentale de structure et suivi de mouvement dans des séquences d'images. Le modèle est estimée par un processus d'optimisation qui lie des fonctions de coût basée sur les images, avec le mouvement du modèle, grâce à des équations dynamiques Lagrangienes. Les approches précédants qui ont étudié une variété des combinaisons des modalités dans le contexte d'estimation de structure et du mouvement d'un modèle deformable, ont généralement assumé une seule représentation, fixe et connue, pour la paramétrisation du modèle, et une seule discrétisation avec des points. Notre technique pour l'estimation de structure et du mouvement est basse sur la fusion incrèmentale des informations issues des points et des nouvelles informations issues des lignes. En supposant qu'un modèle déformable à été aligné avec une partie d'un objet complexe (par exemple, un vélo), des nouvelle lignes appartenant á l'objet, mais non inclus dans la paramétrisation initiale du modèle, sont identifié pendant le suivi. L'identification est basée sur un nouvelle ensemble de tests de consistance, en lisant des processus différentes, independentes au niveau de l'image et du modèle. Les cibles identifiés sont reconstruites et intégrées dans la représentation du modèle, en résulte une estimation plus précise de la structure du modèle, et il sont par conséquent utilisés pour augmenter la complexité du modèle dans les images, en améliorant la robustesse du suivi. On dérive les équations de transfert pour le groupe Euclidienne et les matrices Jacobiennes pour les lignes correspondantes. De nouvelles forces d'alignement dans les images et de nouvelles forces généralisés sont introduites comme des contraintes faibles et combinées avec les forces dérives du contour pour améliorer incrementalement l'estimation de mouvement du modèle. On démontre notre approche sur des séquences d'images contenant des objets avec des structures et mouvements complexes.

Mots-clés : modèles deformables, contraintes géométriques, optimisation contraint, ajustement des faisceaux

Contents

1 Introduction **4**
1.1 Approach 4
1.2 Relation to Previous Work 4

2 Deformable Models Formulation **5**
2.1 Model Geometry 6
2.2 Dynamics and Generalized Forces 6

3 Line Features Formulation **6**
3.1 Line parameterization 7
3.2 Model-Based Consistency Checks 7
3.3 Robust Model-Based Structure Recovery 9
3.4 Forward Model Line Prediction 10
3.5 Image Forces 11

4 Experiments **12**

5 Conclusions **14**

1 Introduction

1.1 Approach

A deformable model framework estimates structure and motion by 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 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 particular combinations of cues in the context of deformable models, they assumed *a fixed and known* model representation with points serving as the basic level of model discretization. In this paper we present a novel two-step technique in which new model structure is revealed during tracking using novel geometric consistency checks. These checks combine heterogeneous information, involving an initial model's parameters as well as independently tracked line features in the image. Our goal is to enhance the basic discretization power of the model through the incremental addition of line features, while still maintaining higher-level shape abstraction in terms of parameterized shapes, such as deformable superquadrics.

In our approach, line features (not belonging to the 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 structure is no longer fully known a-priori, while knowledge of feature motion is known during reconstruction, once their identity as being part of the model is established. Our final model is a mixture of lines and parameterized shapes and generalizes the shape coverage of previously used deformable models.

1.2 Relation to Previous Work

We briefly review the relation of our approach to relevant work in object modeling and tracking based on deformable models, constraint integration, and structure and motion estimation from feature correspondences.

In the area of physics-based deformable models, various formulations have been proposed ([11, 13, 16, 15]). 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 ([18]) 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, shading and stereo ([7]), contours and optical flow ([3, 4]), and shading ([17]) 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 [8] for a review). Several algorithms, commonly based on Kalman filtering, are also known ([2, 22]). 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 [19, 1, 21], could be related to the classical relative orientation paper ([9]).

The approach we present here is based on a deformable model formulation similar to ([13]), in which we employ an articulated chain of deformable superquadrics with global deformations, such as tapering and bending. 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 novel 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 structure and motion estimation under 2-D to 2-D line correspondences ([8]). However, unlike these approaches, we don't attempt to solve for the rigid parameters. On the contrary, by knowing the rigid parameters of the model, we only attempt to check whether the consistency conditions are indeed satisfied for arbitrarily 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, and they are used as additional cues in the form of soft constraints to increase the robustness of the model's motion estimation. Their contribution to the solution for the model's estimated rigid parameters is linearized by means of their corresponding Jacobian matrix (see section 3.5).

2 Deformable Models Formulation

In the next two sections we shall briefly review the deformable model formulation in terms of its underlying geometry and dynamics (see [12] 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 position 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:

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

where $\mathbf{L}$ is the model Jacobian matrix [13]. By omitting the mass term (since we want a model free of inertia), the Lagrangian dynamics equations become [13]:

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

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, $\mathbf{L_{pp}}$, 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 [12] for derivations). Consequently, the generalized forces with respect to edges are defined as:

$$\mathbf{f_q}^{\mathbf{e}} = \int \mathbf{L}_{\mathbf{pp}}^\top \mathbf{f_p} \mathbf{du} \tag{3}$$

3 Line Features 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 to incremental over time: 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.

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 **v**, representing one of two possible directions on the 3-D line l, and a vector **d** terminating on l and perpendicular to it (see Fig. 1). 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 ([10, 21]).

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:

$$N_x X + N_y Y + N_z f = 0 \tag{4}$$

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., [8, 23, 14]) rely on three frames and at least six line correspondences (although no formal proof is yet available, [8]) 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 under-constrained, so one would expect infinitely many solutions.

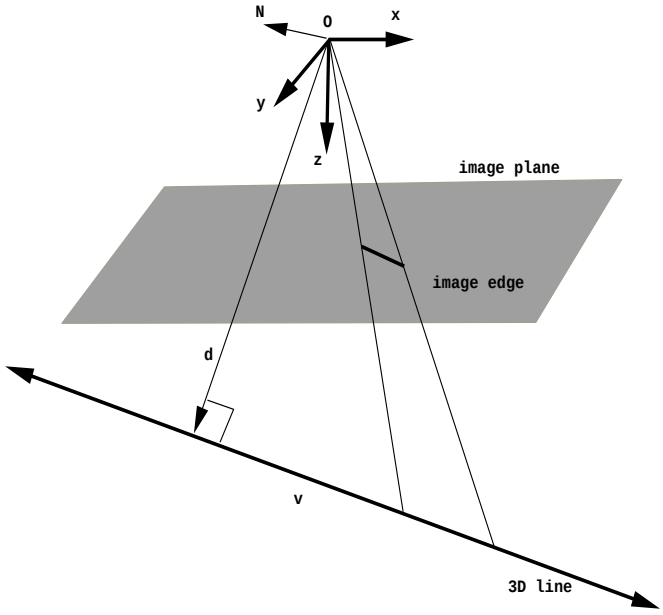


Figure 1: Line Parameterization

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. 2).

The following two relations can be derived either geometrically or algebraically, from the quantities presented above:

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

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

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

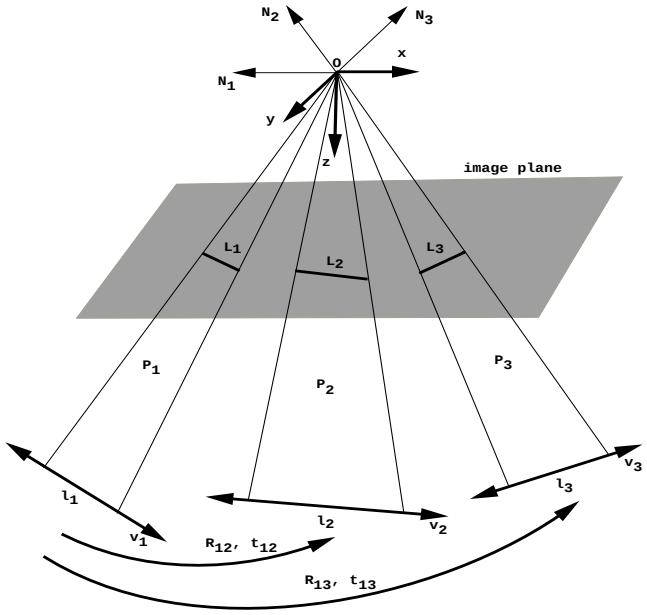


Figure 2: Line Motion in Three Frames

the model. One can prove that the above two relations are necessary and sufficient conditions for consistency.[1] 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 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:

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

[1] 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.

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: $\mathbf{P_i}^\top \cdot \mathbf{D_c^{-1}} \mathbf{D_i^{-1}} \cdot \mathbf{X} = \mathbf{0}$. By stacking together the equations for corresponding lines, we obtain:

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

Since we are looking for the line intersection of all the above planes, the above system of equations 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 $\mathbf{A}$:

$$\mathbf{p} = a \cdot \mathbf{X_{s1}} + b \cdot \mathbf{X_{s2}} \tag{7}$$

The corresponding line parameterization can be subsequently recovered as:

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

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.

Consider the two frame case, as illustrated in Fig. 3. 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:

$$\mathbf{v_2} = R_{12} \mathbf{v_1} \tag{9}$$

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

$$\mathbf{N_2} = \frac{\mathbf{v_2} \times \mathbf{d_2}}{\| \mathbf{v_2} \times \mathbf{d_2} \|} \tag{11}$$

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

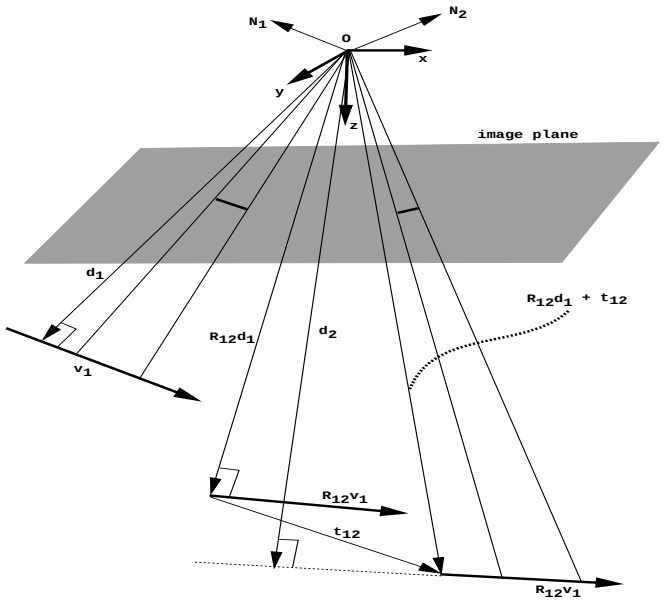


Figure 3: Forward Line Transfer

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:

$$\mathbf{L_l} = \frac{\partial \mathbf{u_l}}{\partial \mathbf{q_r}} \tag{12}$$

3.5 Image Forces

Given an MPL, identified by the equation (4), and an 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. 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:

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

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

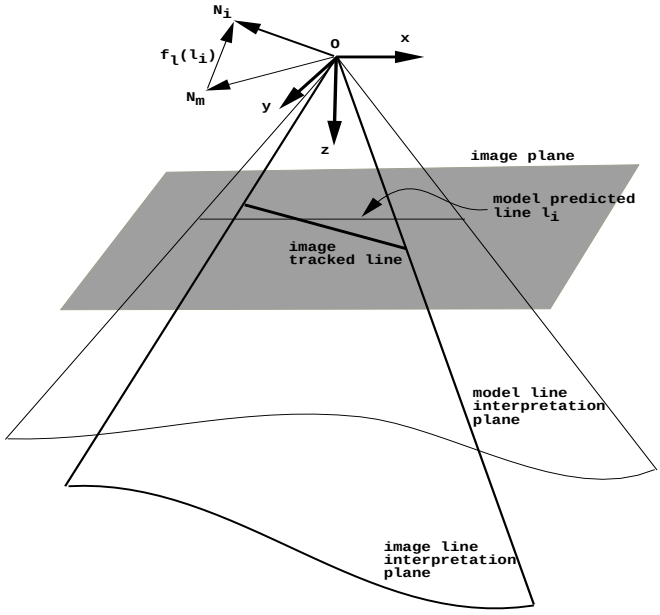


Figure 4: Alignment Forces

(11), and involves the computation of another [3x6] Jacobian matrix:

$$\mathbf{L_{pN}} = \frac{\partial \mathbf{p_N}}{\partial \mathbf{u_l}} \tag{14}$$

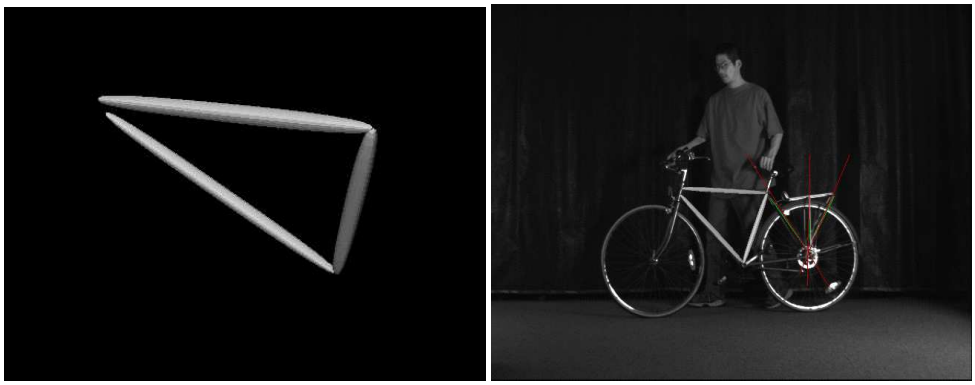
The Jacobian $\mathbf{L_{pl}}$ corresponding to the mapping between the line representation in the model frame and the observation (the normal of the corresponding line interpretation plane) is computed via the chain rule in terms of Jacobians (12) and (14).

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, $\mathbf{L_{pl}}(\mathbf{u_{l_i}})$. 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:

$$\mathbf{f_q} = \sum\nolimits_i \mathbf{L_{pp}}(\mathbf{u_i})^\top \mathbf{f_p}(\mathbf{u_i}) + \sum\nolimits_{l_j} \mathbf{L_{pl}}(\mathbf{u_{l_j}})^\top \mathbf{f_l}(\mathbf{u_{l_j}}) \tag{15}$$

4 Experiments

The experiment consists of a sequence of 5 seconds of video (425 frames recorded at 85 fps) of a moving bike involving both translational and rotational motion in the camera frame. Part of the



(a) Initial frame model used during tracking (b) Model and reconstructed lines, frame 200

Figure 5: Incremental model acquisition

bike frame is modeled and tracked by means of a model consisting of 3 parts (see Fig. 5 a). The model has been manually initialized in the first frame of the sequence, and is displayed in subsequent frames aligned (overlaid and rendered in grey) with the part of the bike object it models (tracking results). Additional lines, not part of the initial model, are tracked through the sequence by means of an independent gradient-based line tracker, involving interest point-based tracking followed by line fitting and renormalization in each frame. Lines that are determined to be CITLs are displayed in green. The model reconstructed/predicted lines are displayed in red (see Fig. 5 b and Fig. 6). The consistency check is evaluated, using a threshold $\tau = 0.05$, based on the lines present in frames 140, 170, 200. It is important that the estimation of the model rigid motion is accurate as it affects both the validity of consistency checks, and the quality of line reconstruction. In practice, the model being initialized manually, there is always an uncertainty associated with its initialization. This is the reason we do not attempt to check and reconstruct lines immediately after model initialization but leave a slight delay, such that the model locks into the data.

The least-squares based reconstruction is based on 12 frames within the interval, 140-200. This allows us to have a sufficiently important motion and sufficient lines for accurate, constrained, reconstruction. The sequence, 80-200, is tracked using only the frame model and contour based image forces, while the rest of the sequence (once the CITL have been reconstructed) is tracked using the enhanced model (consisting of both the initial three-part frame *and* the reconstructed lines, and using *both* contour and alignment forces between CITLs and MPLs).

Although no lines are reconstructed until frame 200, their reprojection is displayed over the entire sequence (note they are reconstructed in a model-centered coordinate frame, so they could be transferred backwards, as the interframe motion of the model has already been estimated). Notice that 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. On average, the number of iterations per time-step in the integration of motion equation (2) decreased from 12 (between frames 80-200) to 7 (between frames 200-425) while the average per-node error decreased from 0.9 to 0.4 pixels.

5 Conclusions

We have presented a contribution to incremental model acquisition and tracking. We have relaxed the constraint that the structure of the model has to be fully known a-priori, and have enhanced the basic discretization of the model's 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, model-based techniques and bottom-up, feature reconstruction techniques. Our initial experimental results are promising and provide an initial validation of our formulation. 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

The authors would like to thank Bill Triggs for many helpful discussions and guidance, Doug DeCarlo for explanations and discussions on deformable models and Alex Telea for implementation assistance.

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 '99

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

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

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

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

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

[13] D.Metaxas and D.Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. PAMI, 15(6):580-591, 1993.

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

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

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



(a) frame 80 (b) frame 110 (c) frame 140



(d) frame 170 (e) frame 200 (f) frame 230



(g) frame 260 (h) frame 290 (i) frame 320



(j) frame 350 (k) frame 370 (l) frame 390

Figure 6: Model Tracking (the model frame in grey) with CITLs (green) and MPLs (red)

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

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

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

[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] C.Taylor and D.Kriegman. Structure and Motion from Line Segments in Multiple Images, PAMI, November 1996.

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

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



Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399