Tracking Rigid Motion using a Compact-Structure Constraint

Yaser Yacoob and **Larry Davis**
Computer Vision Laboratory
University of Maryland
College Park, MD 20742

Abstract

An approach for tracking the motion of a rigid object using parameterized flow models and a compact-structure constraint is proposed. While polynomial parameterized flow models have been shown to be effective in tracking the rigid motion of planar objects, these models are inappropriate for tracking moving objects that change appearance revealing their 3D structure. We extend these models by adding a structure-compactness constraint that accounts for image motion that deviates from a planar structure. The constraint is based on the assumption that object structure variations are limited with respect to planar object projection onto the image plane and therefore can be expressed as a direct constraint on the image motion. The performance of the algorithm is demonstrated on several long image sequences of rigidly moving objects.

1 Introduction

Tracking a moving object in an image sequence is a fundamental capability of a vision system. Tracking can be defined as the process of identifying the region, R^{t+1}, at time $t+1$ in image $I(x,y,t+1)$ that corresponds to a known region, R^t, at image $I(x,y,t)$ and estimating the transformation that maps R^t into R^{t+1}. Recent research on object tracking can be divided into the following categories:

- Tracking using a foreground-background segmentation process [8, 10, 15]. These approaches typically employ models that capture properties of the foreground object or the background (e.g., color, motion, shape, silhouette) and then apply them to each image to delineate the regions of interest. These approaches typically do not directly estimate a transformation between the tracked regions in consecutive images.
- Tracking using an optical-flow formulation that accounts for brightness movement [3, 13, 17]. This formulation employs the brightness constancy constraint to estimate the instantaneous change of the object region between consecutive images. This brightness constancy is typically coupled with models for prototypical motions of the object [17].
- Tracking using appearance-based recognition of object views [4, 6, 11], where the appearance of the object from prototypical views is learned off-line. Then, tracking is posed as a local search in the image for a region that best matches a familiar object appearance under a smooth motion constraint.
- Tracking using 3D motion and structure models [7, 12]. These approaches exploit information about the geometry of the object and the imaging parameters of the scene acquired through direct estimation or learning. Then, tracking is posed as estimation of the object motion while employing the structure information to account for changes in object views.

In this paper we propose a hybrid approach that combines the economy and effectiveness of parameterized flow models with a structure-compactness constraint that seeks to explain image motion differences by global structure variations with respect to a planar rigid structure. We show that structure variation can be posed as a constraint on the optical flow in the image. Then, we illustrate how this constraint can be used in tracking and motion estimation in image sequences of rigid objects in view-variant motions. We demonstrate our approach on several long image sequences of objects in motion and contrast the performance of our approach with parameterized flow tracking based on a planarity assumption.

2 Compact Object Tracking Model

2.1 Objective

Visual tracking marks an image region, R^{t+1} of image $I(x,y,t+1)$, as corresponding to region, R^t of image $I(x,y,t)$, based on a relation $\mathcal{T}$. The estimation of $\mathcal{T}$ typically employs models of brightness values, 2D or 3D motion, rigidity, imaging geometry or scene structure.

Figure 1 shows four frames from a long sequence of a rotating box and the enclosed regions that would be

produced by an ideal tracker. The top row shows two images that are 5 frames apart in which the box rotation reveals a new surface that was not visible in the initial image. The tracker should be able to mark the expanded region as being part of the box. In contrast, the bottom row shows two images that are 5 frames apart in which the box rotation conceals a side of the box. In this case, even though it might be predicted that the dotted area in Figure 1(d) should contain a transformed instance of part of the box in Figure 1(c), the tracker should realize that the surface segment has become occluded and adapt the tracked region to reflect its disappearance.

We make the following assumptions to design the tracker:

- the tracked object is moving rigidly,
- the global illumination is constant throughout the sequence (the object surface may change orientation relative to the camera; therefore, local illumination may vary over time),
- the object structure is compact (to be defined below),

Our algorithm starts with a given initial region and employs the above assumptions to compute the successive regions of the object, a planar object transformation that captures the relationship between consecutive regions, and additions/deletions of points to the object region between consecutive regions to capture appearance changes. The tracker is based on approximating a compact object by a 3D plane and depth variation of points from this plane. This approximation allows us to compute upper and lower bounds on the image flow of points in the object region, and therefore can be employed as a criterion for determining if points correspond to object points or non-object points. The advantage of this criterion is that it approximates simultaneously the 3D motion and structure and thus supports identifying corresponding and added/deleted regions.

In order to compute added/deleted regions we examine the object region boundary where appearance differences occur. Object motions that involve self-occlusions do not affect the tracked region. However, if holes exist (e.g., tracking a rotating doughnut) the additions/deletion analysis must also be applied over the interior of the region.

2.2 Preliminaries

The diameter, D, of a 3D rigid object, O, is the maximum distance between any two surface points P_1, P_2.

Definition 1: A view of a 3D rigid object, O, is called *D'-compact* if

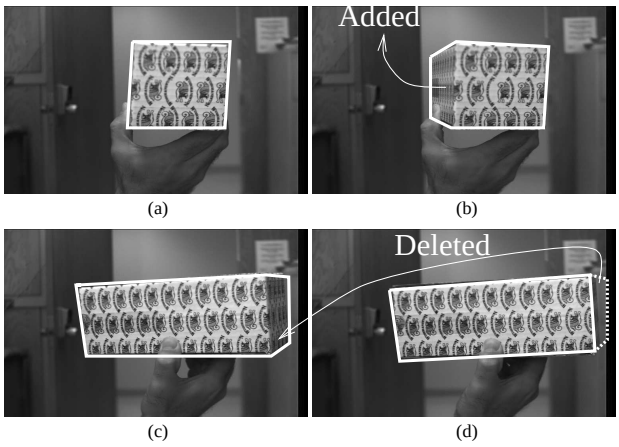

Figure 1: Illustration of object region addition/deletion for a rotating box. The upper and lower two images are 5 frames apart.

1. D is the diameter of O.
2. R is the greatest distance between the camera center and any visible point on O's surface, and
3. $\frac{D}{R} \leq D'$

Our goal is to track an object through any sequence of D'-compact views. The notion of a D'-compact view is clearly related to the typical qualitative assumption that an object is "far" from the camera. This has been employed in various ways to overcome the non-linearity of perspective projection (i.e., D' is assumed to be very small or close to zero). However, we consider here scene geometry with strong perspective in which $0 < D' \leq 1$.

Definition 2: Let L be a plane passing through O. A view of O is *${D_L}'$-plane-compact* if for every visible surface point P of O, if

1. d is the distance of P from the camera center,
2. d' is the distance of P from L along the line of sight through P ($d' = ||P\ P_L||$ where P_L is the intersection point of the line of sight through P with L),

then $\frac{d'}{d+d'} \leq {D_L}'$.

Our tracking algorithm will approximate the visible surface of O with a plane L and then track the plane as long as views remain ${D_L}'$-plane-compact (in fact the plane will be slowly adapted over time so that the sequence of D'-compact views remain ${D_L}'$-plane-compact). Typically, for views that are D' and ${D_L}'$-plane-compact, ${D_L}' < D'$

Consider a ${D_L}'$-plane-compact view of an object. The quantity $\frac{||P||}{||P^L||} = \frac{||P^L|| \pm ||P\ P^L||}{||P^L||}$ is the ratio between the actual point distance from the camera and

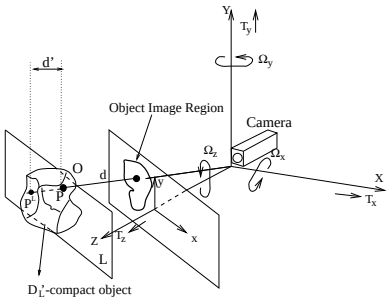

Figure 2: The motion and geometry of the camera.

the planar point distance from the camera. This ratio satisfies $1 \pm \frac{||P,P^L||}{||P^L||} \leq 1 \pm {D_L}'$ and therefore $||P|| \leq (1 \pm {D_L}')||P^L||$ which is a property that will be exploited in the flow model.

In the following sections we will show that the ${D_L}'$-plane-compactness is a valuable property that enables object region tracking based on a traditional brightness constancy assumption and error minimization of point motion with respect to the ${D_L}'$-plane-compactness constraint.

2.3 Motion and Structure Model

We employ the standard conventions [14] for representing the spatio-temporal variation of the optical flow as the camera moves through a static scene (equally applicable to object motion with a stationary camera). Assume a camera moving in a static scene with instantaneous 3D translational velocity (T_x, T_y, T_z) and rotational velocity $(\Omega_x, \Omega_y, \Omega_z)$ relative to an external coordinate system fixed with respect to the camera. A texture element P in the scene with instantaneous coordinates (X, Y, Z) will induce an optical flow vector (u, v) where u and v are the horizontal and vertical instantaneous velocities

$$\begin{aligned} u &= \Omega_x xy - \Omega_y(1+x^2) + \Omega_z y - (T_x - T_z x)/Z \\ v &= \Omega_x(1+y^2) - \Omega_y xy - \Omega_z x - (T_y - T_z y)/Z \end{aligned} \quad (1)$$

Here, $p = (x, y)$ are the image coordinates of (X, Y, Z) relative to a coordinate system in which the positive Z is aligned with the line of sight of the camera (see Figure 2). Assume that the object is a plane that satisfies the equation $Z = A + BX + CY$; Then, its perspective projection is given by $\frac{1}{Z} = \alpha + \beta x + \gamma y$ where $\alpha = \frac{1}{A}$, $\beta = \frac{-B}{A}$, and $\gamma = \frac{-C}{A}$. It is well known [1] that the flow measured over the projected region, R, of the plane can be modeled by an eight parameter model,

$$\begin{aligned} u(x,y) &= a_0 + a_1 x + a_2 y + a_6 x^2 + a_7 xy \\ v(x,y) &= a_3 + a_4 x + a_5 y + a_6 xy + a_7 y^2 \end{aligned} \quad (2)$$

where

$$\begin{aligned} a_0 &= -\Omega_y - \alpha T_x & a_4 &= -\Omega_z + \beta T_y \\ a_1 &= \alpha T_z + \beta T_x & a_5 &= \alpha T_z + \gamma T_y \\ a_2 &= \Omega_z + \gamma T_x & a_6 &= -\beta T_z - \Omega_y \\ a_3 &= \Omega_x - \alpha T_y & a_7 &= -\gamma T_z + \Omega_x \end{aligned}$$

These eight parameters are estimated by pooling the motion of the points in R into an overconstrained system. Consider now an image point $\mathbf{x} = (x_n, y_n) \in R$ that corresponds to an object point not on the plane, i.e. with actual depth Z_n that is unequal to the plane-constrained depth Z_p. The flow at $\mathbf{x}$ is predicted by the planar model in Equation 2 as (u_p, v_p). If the translation in depth is not zero then the plane-induced flow will typically be different from the actual flow. We seek to model how the actual flow (u_n, v_n) in the image is related to the plane-induced flow (u_p, v_p). From Equation 1 we know that the actual flow satisfies

$$\begin{aligned} u_n &= \Omega_x x_n y_n - \Omega_y(1+{x_n}^2) + \Omega_z y_n - \frac{(T_x - T_z x_n)}{Z_n} \\ v_n &= \Omega_x(1+{y_n}^2) - \Omega_y x_n y_n - \Omega_z x_n - \frac{(T_y - T_z y_n)}{Z_n} \end{aligned} \quad (3)$$

while the planar-induced flow is given by

$$\begin{aligned} u_p &= \Omega_x x_n y_n - \Omega_y(1+{x_n}^2) + \Omega_z y_n - \frac{(T_x - T_z x_n)}{Z_p} \\ v_p &= \Omega_x(1+{y_n}^2) - \Omega_y x_n y_n - \Omega_z x_n - \frac{(T_y - T_z y_n)}{Z_p} \end{aligned} \quad (4)$$

The difference between the actual and plane-induced flow is given by

$$\begin{aligned} u_n - u_p &= (T_x - T_z x_n)(\frac{1}{Z_p} - \frac{1}{Z_n}) = (T_x - T_z x_n)\frac{1}{Z_p}(1 - \frac{1}{r}) \\ v_n - v_p &= (T_y - T_z y_n)(\frac{1}{Z_p} - \frac{1}{Z_n}) = (T_y - T_z y_n)\frac{1}{Z_p}(1 - \frac{1}{r}). \end{aligned} \quad (5)$$

where $Z_n = rZ_p$. If $0 < r < 1$ then this distance to P is less than the planar distance, while if $1 < r$ the distance to P is greater than the planar distance. Equation 4 can be rewritten as

$$\begin{aligned} u_p &= {u_p}^{rot} - {u_p}^{trans} \\ v_p &= {v_p}^{rot} - {v_p}^{trans} \end{aligned} \quad (6)$$

where the ${u_p}^{rot}, {v_p}^{rot}$ denote the rotation-induced horizontal and vertical flow components and ${u_p}^{trans}, {v_p}^{trans}$ the translation-induced horizontal and vertical flow components of the planar motion. Computing ${u_p}^{trans}, {v_p}^{trans}$ requires recovering the motion parameters of the plane. We employ the closed form solution proposed by [16] to recover the actual motion parameters from the estimated parameters $a_0, ..., a_7$. This solution recovers two dual solutions for the motion parameters. However, one of these solutions leads to negative depth and is thus eliminated.

Equation 5 describes the range of image flows as an object point deviates from planarity. To track an object through a sequence of $D_L{}'$-planar-compact views we must determine for each point in the tracked region (and its dilated boundary) if its 3D rigid motion model and deviation from the planar constraint correspond to the image flow.

Determining the value of $D_L{}'$ is critical to Equation 5, and it is directly related to the compactness of the object views. Consider two positive numbers $r_{min} < 1$, and $1 < r_{max}$ satisfying $\forall \quad i \quad r_{min} {Z_p}^i < {Z_p}^i < r_{max} {Z_p}^i$. Let $r = \max(1-r_{min}, r_{max}-1)$. If we assume that the object views are r-planar-compact then we can set $D_L{}' = 1 \pm r$ and employ Equation 5 to track the region and adapt it to appearance changes.

2.4 Brightness Constancy Constraint

Define

$$\mathcal{M}(\mathbf{x}) = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \end{bmatrix}^T$$

such that $\mathbf{F}(u,v) = \mathbf{F}(\mathbf{x};\mathbf{A}) = \mathcal{M}(\mathbf{x})\mathbf{A}$ represents the planar flow model described above (where $\mathbf{x} = (x,y)$). The brightness constancy assumption for the object region states that

$$I(\mathbf{x},t) = I(\mathbf{x} - \mathcal{M}(\mathbf{x})\mathbf{A}, t+1), \quad \forall \mathbf{x} \in R, \tag{7}$$

where I is the image brightness function and t represents time. Taking the Taylor series expansion of the right hand side, simplifying, and dropping terms above first order gives

$$\nabla I \cdot (\mathcal{M}(\mathbf{x})\mathbf{A}) + I_t = 0, \quad \forall \mathbf{x} \in R, \tag{8}$$

where $\nabla I = [I_x, I_y]$ and the subscripts indicate partial derivatives of image brightness with respect to the spatial dimensions and time. To estimate the parameters $\mathbf{A}$ we minimize the error term

$$E = \sum_{\mathbf{x} \in R} \rho(\nabla I \cdot (\mathcal{M}(\mathbf{x})\mathbf{A}) + I_t, \sigma), \tag{9}$$

for some error norm ρ where σ is a scale parameter (see Geman-McClure [9]).

The error term in Equation 9 divides the points in the region R, as well as the rest of the image into those conforming to the planar motion model (i.e., small error) and those that violate it (large error). Points with large errors can result from brightness constancy violations, imaging noise or non-planarity. In the following we propose a criterion for determining if the violation is due to non-planarity that doe not violate the planar-compactness of views.

The $D_L{}'$-planar-compactness of object views can be used to estimate the error of Equation 8 at a point p^n that has an error greater than a maximum error E^{max}. From Equation 8,

$$E(p^n) = abs(u_p I_x + v_p I_y + I_t) > E^{max} \tag{10}$$

where (u^p, v^p) have been computed from the planar model. Equation 5 provides bounds for the actual flow of p^n based on the $D_L{}'$-planar-compactness assumption. Specifically, for a $D_L{}'$-compact view, the range of the actual flow is bounded by $({u_n}^1, {v_n}^1)$ and $({u_n}^2, {v_n}^2)$ (from Equations 5 and 6)

$$\begin{aligned} {u_n}^1 &= u_p + {u_p}^{trans}(1 - \frac{1}{1 - D_L{}'}) \\ {v_n}^1 &= u_p + {v_p}^{trans}(1 - \frac{1}{1 - D_L{}'}) \end{aligned} \tag{11}$$

$$\begin{aligned} {u_n}^2 &= u_p + {u_p}^{trans}(1 - \frac{1}{1 + D_L{}'}) \\ {v_n}^2 &= u_p + {v_p}^{trans}(1 - \frac{1}{1 + D_L{}'}) \end{aligned} \tag{12}$$

and corresponding errors,

$$E^1 = {u_n}^1 {I_x}^n + {v_n}^1 {I_y}^n + {I_t}^n \tag{13}$$

$$E^2 = {u_n}^2 {I_x}^n + {v_n}^2 {I_y}^n + {I_t}^n \tag{14}$$

The values of E^1 and E^2 can be compared to the value of E^{max} resulting in the following cases

1. If ($E^1 > E^{max}$ and $E^2 > E^{max}$) or if ($E^1 < -E^{max}$ and $E^2 < -E^{max}$) then the $D_L{}'$-planar compactness of the object views fails to explain the point brightness motion in accordance with the estimated planar motion and the allowed structure variation.
2. If ($E^1 > 0$ and $E^2 > 0$) and ($E^1 < E^{max}$ or $E^2 < E^{max}$) then either $(1 - D_L{}', 1 + D_L{}')$ will provide a minimum error at $\min(E^1, E^2)$ and therefore the point can be attached to the object region.
3. If ($E^1 < 0$ and $E^2 < 0$) and ($E^1 > -E^{max}$ or $E^2 > -E^{max}$) then either $(1 - D_L{}', 1 + D_L{}')$ will provide a minimum error at $\max(E^1, E^2)$ and therefore the point can be attached to the object region.
4. If ($0 < E^1$ and $0 > E^2$) or ($0 > E^1$ and $0 < E^2$) then the minimum error is achieved at the zero crossing along the line defined by the points $(1 - D_L{}', E^1)$ and $(1 + D_L{}', E^2)$ and therefore the point can be attached to the object region.

The $D_L{}'$-planar-compactness assumption provides a basis for adding/deleting points in the image to the region of interest while relying on the structure variation model to predict and explain brightness movement in consecutive images.

3 Computation Details

The computational aspects of the proposed tracking approach follow, generally, the algorithm proposed by Black and Anandan [2] for planar motion estimation.

The estimated parameters of the planar models are then used to determine how points in the image are moving with respect to the model. We focus on points within a small band b (in the following experiments $b = 9$ pixels) on either inside or outside of the boundary of the object region. This choice is motivated by the observation that object motion leads to most drastic appearance changes at the boundary of the object region while interior regions change more gradually; therefore, point addition/deletions are most important at the boundary. We consider those points with an error larger than E^{max} in this region. Using the $D_L{}'$-planar-compactness assumption the bound on allowable image flow are computed and used to calculate E^1 and E^2. Then, as described in the previous section, if the point complies with the planar model plus $D_L{}'$-planar-compactness it is added to the object region, otherwise, it becomes a non-object point.

The above procedures are accompanied by a coarse-to-fine computation that allows us to deal with large motions. The estimated region and the motion parameters at the coarsest level are used to warp the image using the planar and $D_L{}'$-planar-compactness models. Then, the new image is used at the finer level to estimate the residual motion and the process is repeated.

4 Experimental Results

In this section we show experiments of tracking objects based on the planar-compactness assumption. In these experiments we compare the performance of a planar tracker to the proposed planar-compactness tracker. It is assumed that an initial rectangle around the object is manually chosen at the first frame. Similar parameters were used for all sequences except for the compactness parameter. That was set to $D_L{}' = 0.5$ for the hat and $D_L{}' = 0.8$ for the box due to its short distance from the camera. The objects are hand carried and moved; therefore, the hand usually satisfies the same motion and planar-compactness of the object and is tracked as well.

Figure 3 shows a few frames from a 150 frames sequence of a hand-held box that is rotated about 180° in depth. The views of the box vary considerably over the sequence. The effects of perspective in this scene is clearly visible (see frames 50 and 100). The second row shows the tracking of the box using the planar model. The accuracy of the tracking quickly degrades as the planarity of the object is violated. In contrast, the third row shows the planar-compactness tracking. Here, the object is well tracked over the sequence with occasional boundary points tagging along (for a few frames at a time). Figure 4 provides two sets of images (two frames each) that show the added and deleted points during tracking (the first row shows added points and the second row shows deleted points, see also Figure 1). The frames showing added points demonstrate that many points are added at the left boundary of the box where a new surface is becoming visible. The frames showing deleted points indicate that many points are deleted at the right boundary of the box where the surface is gradually disappearing. Notice that in all frames points are sparsely added/deleted around the boundary due to large flow errors resulting from the derivative estimates that involve both object and boundary neighborhoods. Figure 6 shows several of the eight motion parameters that are computed using the planar and planar-compactness tracking. The columns show (left to right) the parameters a_0 and a_3, divergence $(a_1 + a_5)$ and out-plane rotations a_6 and a_7.

Figure 5 shows a few frames from a 300 frames sequence of a hand-held hat that is rotated in depth. Despite the complexity of the hat structure and appearance all its parts are tracked quite well for the duration of the sequence.

Finally, Figure 7 illustrates a complex experiment with two objects rotating in depth around their center. The objects are a box and a gourd that are placed to create neighboring regions in the image. The induced flow due the objects rotations is qualitatively similar but quantitatively different due to depth variation of each object points. In this experiment track each object separately during its motion while minimizing the overlapping of regions due to similary in motion. It should be noticed that at times (i.e., instantaneously) the flow created by the objects' motions may not reflect significant structure difference (e.g., very small translation in depth), which naturally leads to merging points of the two objects. Also, since the box is farther away, its structure information is less pronounced. In Figure 7 the edge of the gourd is not tracked well around frame 110 since the motion was quite large (20 pixels/frame) so that the differential framework is ineffective.

5 Discussion

Tracking of rigidly moving objects is a fundamental vision problem that has been studied extensively. Most of the proposed approaches have involved a-priori acquisition of scene-specific information about the object structure, scene geometry or appearance. Our objective has been to minimize a-priori knowledge about the tracked object and scene since most real-world objects exhibit rich structural and surface texture variations

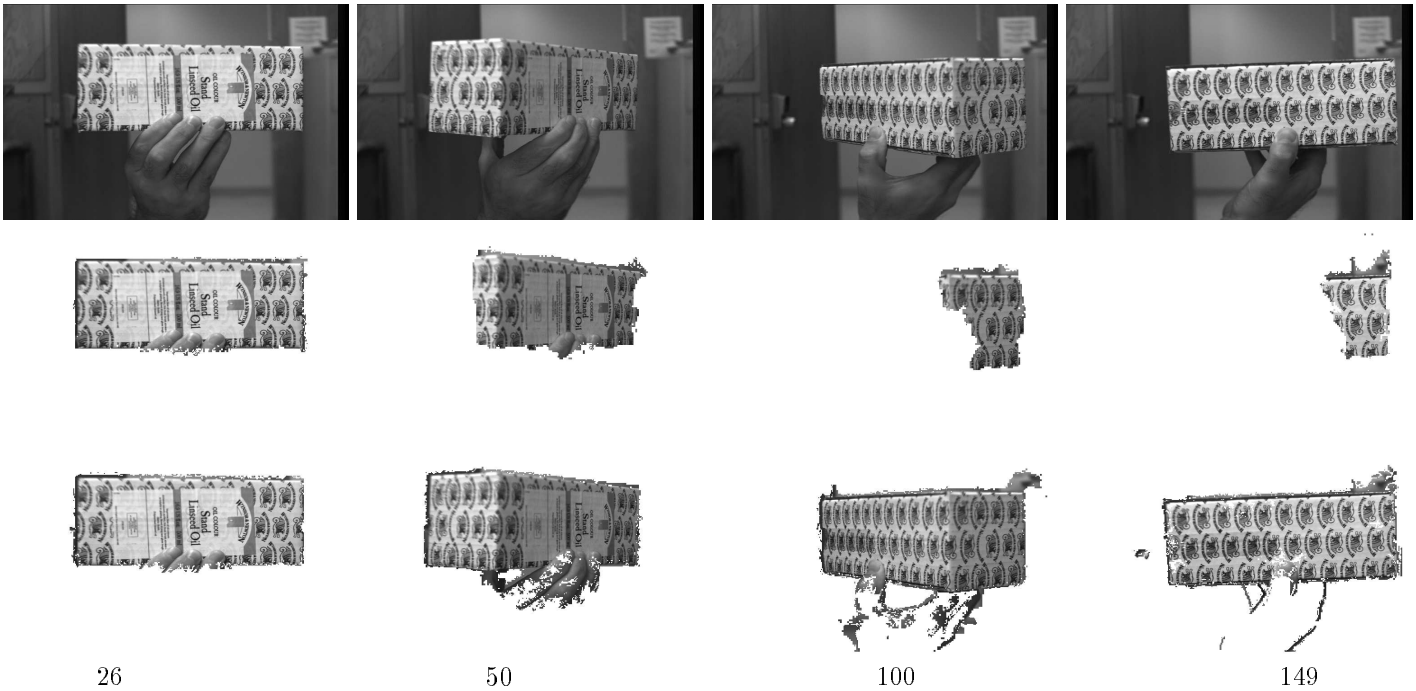

Figure 3: Selected frames from an image sequence of a rotating box, and the planar tracking (middle row) and planar+structure-compactness tracking (bottom row)

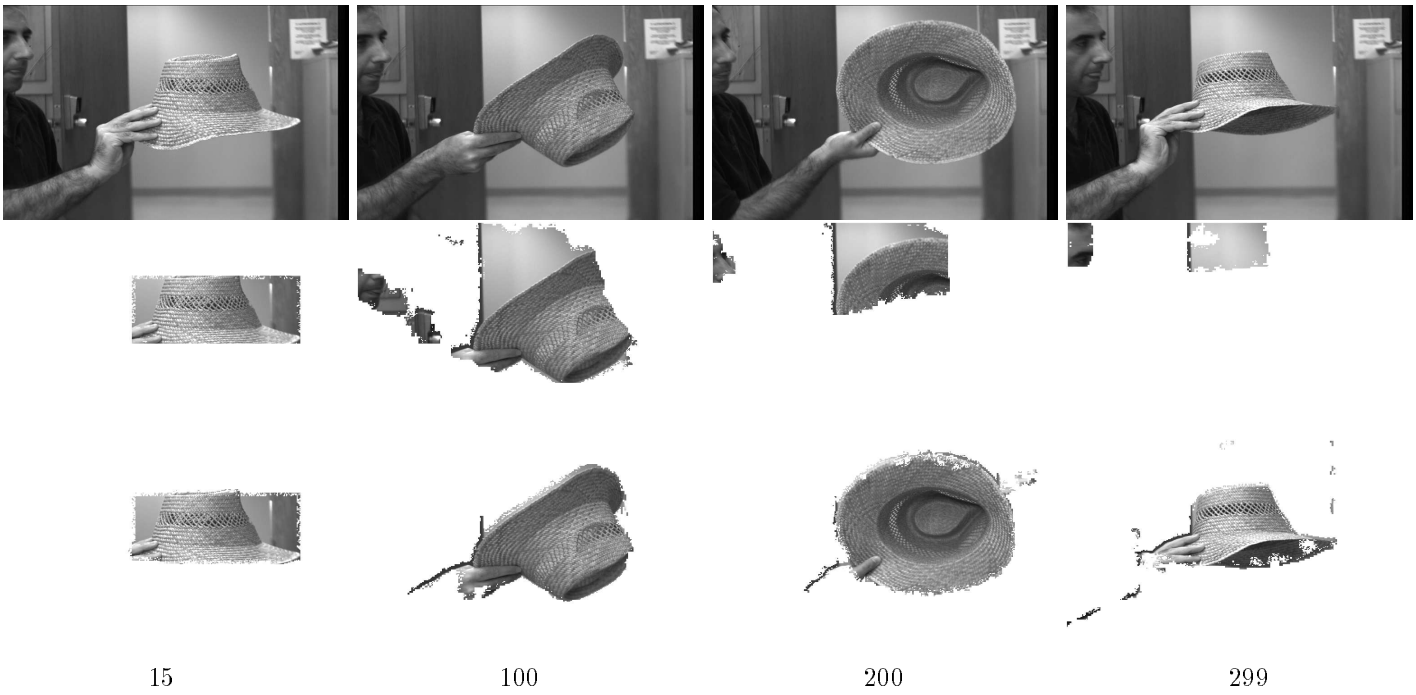

Figure 5: Selected frames from an image sequence of a rotating hat, and the planar tracking (middle row) and planar+structure-compactness tracking (bottom row)

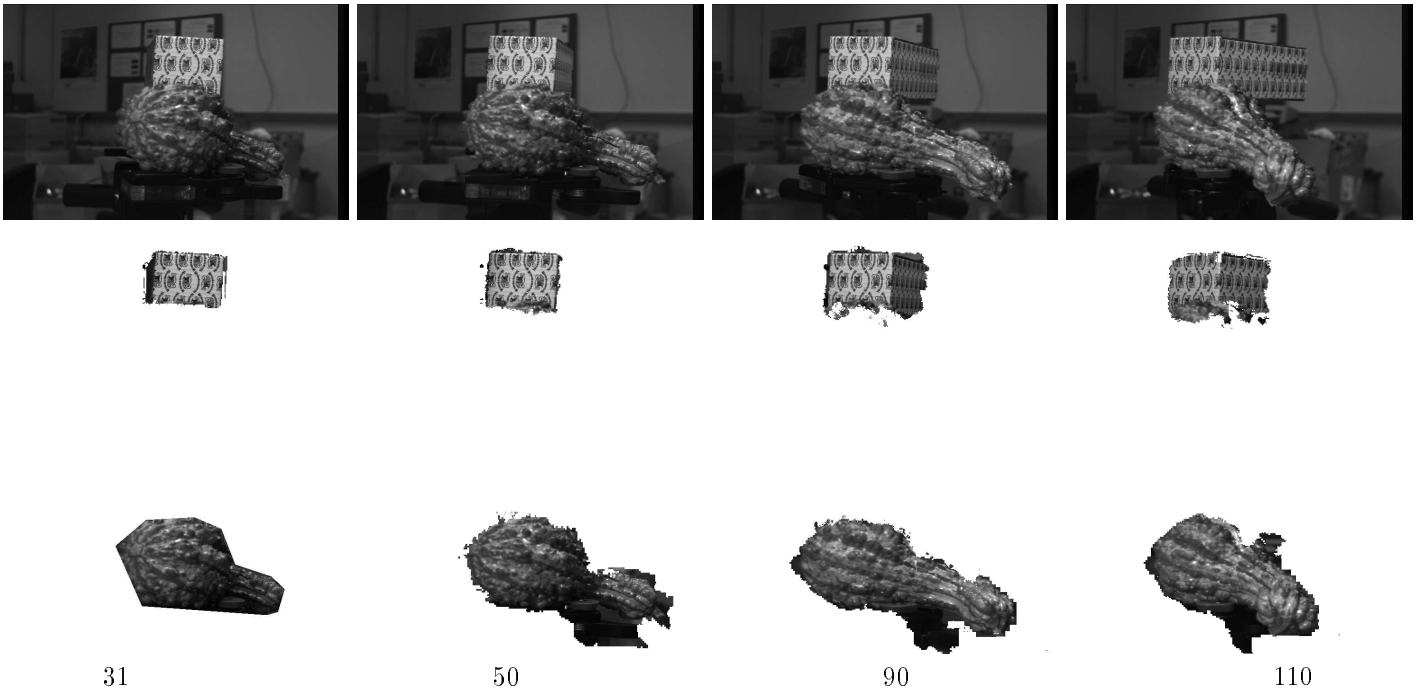

Figure 7: Selected frames from an image sequence of simultaneous rotation of a box and gourd, with difference in depth and the planar tracking (middle row) and planar+structure-compactness tracking (bottom row)

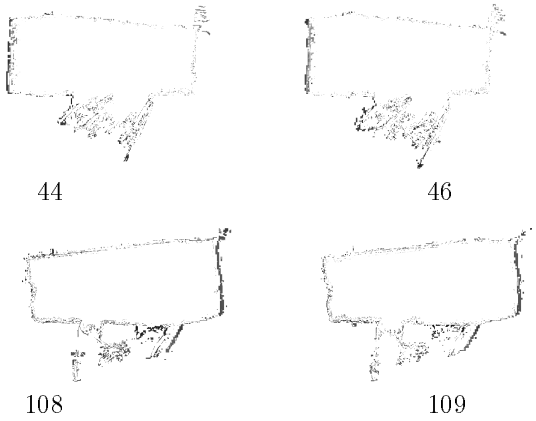

Figure 4: Selected frames illustrating the addition/deletion of points for the rotating box boundary (first and second rows, respectively).

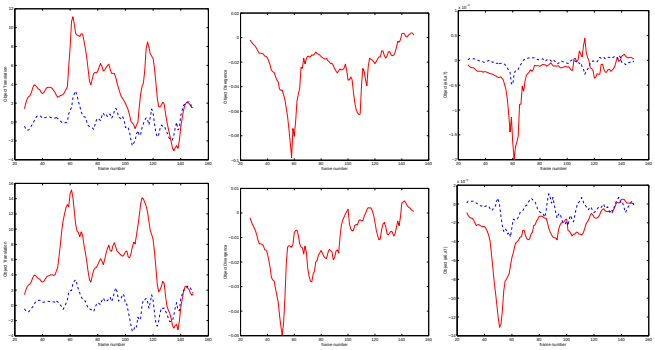

Figure 6: Graphs for the rotating box parameters a_0, a_3 (left, solid and dashed, respectively), divergence (center) and yaw and pitch a_6, a_7 (solid and dashed, respectively) for planar and planar+structure-compactness tracking (upper and lower rows, respectively).

that are quite challenging to model and recover.

We proposed a structural-compactness constraint for object tracking. This compactness was defined with respect to a 3D plane which made it possible to exploit the power of parameterized flow models. It was shown that this compactness translates into a constraint that makes it possible to model the resulting variation of optical flow relative to the 3D planar structure without computing exact structure or 3D motion. The constraint on the range of optical flow can be used

to determine if image brightness motion is consistent with a single object model, thereby facilitating addition/deletion of points to the object region over time.

Although the basis of computation is relative to a 3D planar model the object need not have an actual dominant plane. The error minimization of the planar motion model converges to a virtual plane (i.e., a solution always exists due to the error minimization formulation). In unreported animation of the accuracy of the flow estimation (which cannot be shown in printed form) we found that for the tested objects the motion of a virtual plane was very well estimated.

The definition of planar-compactness combines two independent components: the intrinsic object structure and the imaging geometry (specifically distance between the object and camera). It is well known that object motion will appear planar if the distance to the object is much larger than the compactness parameter D (see Section 2.2). In this case the object will be zero-planar-compact. If the object is planar then even at small distances it will be zero-planar-compact. Therefore, it is only when the object is non-planar but close to the camera that the planar-compactness parameter, ${D_L}'$, becomes important.

Our experiments indicate that an accurate compactness parameter is not necessary. Since the compactness parameter is dependent on the distance of the object from the camera (an unknown quantity in the reported experiments) a conservative value of ${D_L}'$ was used. If a rough depth estimate of the object is available (e.g., from stereo), then a tighter compactness parameter could be used. In the experiments we demonstrated that no prior modeling or knowledge of the object is needed beyond the compactness parameter. As a result, a wide variety of 3-D shapes can be tracked.

References

[1] Adiv G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *PAMI,* 7(4), 1985, 384-401.

[2] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVIU*, 63(1), 1996, 75-104.

[3] M. Black and Y. Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motions. *IJCV*, 25(1), 1997, 23-48.

[4] M. Black and A. Jepson. EigenTracking: Robust matching and tracking of articulated objects using a view-based representation. *IJCV*, 26(1), 1998, 63-84.

[5] A. Blake and A. Zisserman. *Visual Reconstruction* MIT Press, 1987.

[6] M. La Cascia, J. Isidoro and S. Sclaroff. Head tracking via robust registration in texture map images. *CVPR*, 1998, 508-414.

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

[8] P. Fieguth and D. Terzopoulos. Color-based tracking of heads and other mobile objects at video frame rates. *CVPR*, 1997, 21-27.

[9] S. Geman and D.E. McClure. *Statistical Methods for Tomographic Image Reconstruction.* Bulletin of the International Statistical Institute, LII-4:5-21, 1987.

[10] I. Haritaoglu, D. Harwood, and L. Davis, W4S: A real time system for detecting and tracking people in 2.5D. *ECCV,* 1998, 877-892.

[11] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 20(10), 1998.

[12] T. Jebara and A. Pentland. Parameterized structure from motion for 3D adaptive feedback tracking of faces. *CVPR*, 1997, 144-150.

[13] S. X. Ju, M. Black, and Y. Yacoob. Cardboard people: A parameterized model of articulated image motion. in *Proc. Int. Conference on Face and Gesture*, Vermont, 1996, 561-567.

[14] H.C. Longuet-Higgins and K. Prazdny, The interpretation of a moving retinal image. *Proc. Royal Society of London, B*, 208, 1980, 385-397.

[15] N. Oliver, A. Pentland and F. Berard. LAFTER: Lips and face real time tracker. *CVPR*, 1997, 123-129.

[16] M. Subbarao and A.M. Waxman, Closed form solutions to image flow equations for planar surfaces in motion. *CVGIP*, 36, 1986, 208-288.

[17] Authors Learned temporal models of image motion. *ICCV-98*, India, 1998, 446-453.