

# Dynamic Depth Recovery from Multiple Synchronized Video Streams<sup>1</sup>

Hai Tao<sup>†</sup>, Harpreet S. Sawhney<sup>‡</sup>, and Rakesh Kumar<sup>‡</sup>

<sup>†</sup> Department of Computer Engineering  
University of California at Santa Cruz  
Santa Cruz, CA 95064  
tao@soe.ucsc.edu

<sup>‡</sup> Sarnoff Corporation  
201 Washington Road  
Princeton, NJ 08543  
{hsawhney, rkumar}@sarnoff.com

## Abstract

*This paper addresses the problem of extracting depth information of non-rigid dynamic 3D scenes from multiple synchronized video streams. Three main issues are discussed in this context: (i) temporally consistent depth estimation, (ii) sharp depth discontinuity estimation around object boundaries, and (iii) enforcement of the global visibility constraint. We present a framework in which the scene is modeled as a collection of 3D piecewise planar surface patches induced by color based image segmentation. This representation is continuously estimated using an incremental formulation in which the 3D geometric, motion, and global visibility constraints are enforced over space and time. The proposed algorithm optimizes a cost function that incorporates the spatial color consistency constraint and a smooth scene motion model.*

## 1 Introduction

The problem of recovering depth information using images captured simultaneously from multiple viewpoints has been extensively studied in the past. In recent years, with advances in computing and imaging technologies, capturing multiple synchronized high quality video streams has become easier, and the problem of recovering depth maps of dynamic scenes using synchronized capture has received increasing attention [Vedula99, Zhang99, Carceroni01]. This problem is termed *dynamic depth recovery* in this paper. It can be considered as an extension of the traditional stereo computation problem where the depth solution should make images consistent not only across multiple views, but also across different time instants.

A straightforward approach for dynamic depth recovery is to apply a standard stereo estimation algorithm at each time instant. A comprehensive survey on early stereo algorithms can be found in [Dhond89] while a short list of newer algorithms consists of [Belhumeur96, Roy98, Boykov99, Kutulakos99, Szeliski99, Lhuillier00, Zhang00, Tao00]. The principle underlying these algorithms is to find a depth solution that optimizes an image match measure across views. We call this measure *spatial match measure*.

However, this straightforward solution ignores two new constraints present in multi-view image sequences.

The first constraint encodes the geometric relationship between the 3D motion of a scene point and its 2D projections in multiple synchronized images. This relationship, which is called the *scene flow constraint*, has been investigated in [Vedula99, Zhang99]. By applying this constraint, temporal 2D image correspondences can be used to infer 3D scene motion and therefore, constrain the depth information over time. To successfully apply scene flow constraint directly in depth estimation, the accuracy of the optical flow is crucial since the effect of unreliable flow at object boundaries and in untextured regions will propagate into the final depth map. More reliable results may be obtained by estimating parametric motion models in local image regions [Zhang99].

The second constraint arises from the observation that objects in the scene usually deform or move smoothly. Applying this constraint helps to obtain temporally consistent depth solutions and to eliminate ambiguities that usually can not be easily resolved at any single time instant. Rigid [Zhang92, Adiv85, Young90, Hanna93] and non-rigid parametric motion models [Zhang99] have been employed in previous work. Local parametric motion models can also be estimated for scenes with arbitrary non-rigid motions [Carceroni01].

This paper proposes a dynamic depth recovery method in which a scene representation that consists of piecewise planar surface patches is estimated within an incremental formulation. Such a representation can be derived based on color segmentation of input images. The proposed formulation integrates constraints on geometry, motion, and visibility over both space and time. More specifically, the scene surface corresponding to each homogeneous color region in the image is modeled as a 3D plane. The motion of this plane is described using a constant velocity model. The spatial match measure and the scene flow constraint for this representation is investigated. Based on the observation that the spatial match measure depends only on the out-of-plane motion of each planar surface, a dynamic depth recovery algorithm is developed. The proposed method enforces the motion model without explicitly using the scene flow

---

<sup>1</sup> This work was performed while Hai Tao was employed by the Sarnoff Corporation. This material is based upon the work supported by the Air Force Research Laboratory under Contract number F30602-00-C-0143.

constraint and therefore avoids the propagation of the errors in optical flow computation to depth estimation. The global matching framework proposed in [Tao00] is adopted for enforcing the visibility constraint and also for initializing and refining the depth solution. As a result, the visibility constraint is enforced both across different views and over different time instants.

## 2 Multi-view depth recovery of dynamic scenes

Figure 1 illustrates the configuration of a multi-camera video capture system. Without loss of generality, we will first develop the formulation for the motion of a single planar structure. The formulation can then be directly applied to handle a piecewise planar scene description. We assume that both the intrinsic and extrinsic parameters of the cameras are known. The camera extrinsic parameters, namely the positions and orientations of the cameras, are represented with respect to the camera coordinate system of a *reference* view. The planar scene surfaces are also represented in this coordinate system. The calibration matrices of the  $M + 1$  cameras are  $K_v, v = 0, \dots, M$ . The rotation and translation pair  $(R_0, T_0) = (\mathbf{I}, 0)$  represents the camera pose for the reference camera 0 and the pair  $(R_v, T_v)$  represents the pose of the *inspection* cameras  $v \in \{1, \dots, M\}$ . For time instant  $t$ , only the depth map in the reference view is estimated. In our representation, each of the  $S$  homogeneous color segments in the reference view corresponds to a planar surface in the world. Therefore, the depth map can be derived from the plane parameters  $\Psi_{s,t,0} = [n_{s,t,0}, d_{s,t,0}]$ ,  $s \in [1, \dots, S]$ , where  $n_{s,t,0}$  is the normal vector of the plane surface in segment  $s$  at time instant  $t$  and  $d_{s,t,0}$  is the distance of that plane to the reference camera center.

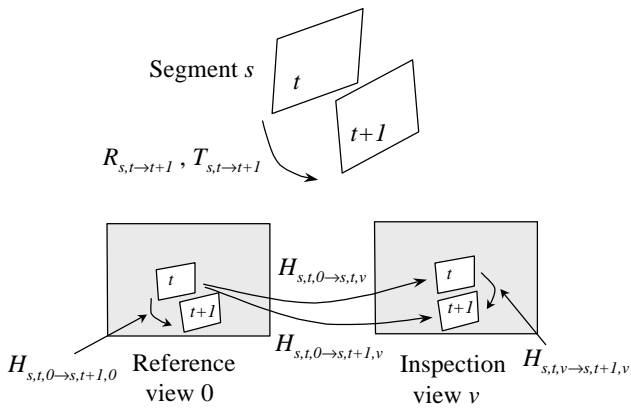


Figure 1. Dynamic depth recovery: synchronized image sequences of a dynamic scene are captured from multiple viewpoints.

At any given time instant, relating multiple views of a scene plane leads to the spatial matching constraint. Matched image points on the same scene plane across these views

over-constrain the plane's geometry. On the other hand, relating multiple views of a scene plane over time leads to the planar scene flow constraint. Matched image points on the same scene plane over time over-constrain the plane's 3D motion and in turn its geometry. The goal of the dynamic depth recovery algorithm is to find a piecewise planar depth map that optimizes a spatial match measure and a temporal match measure subject to a motion smoothness prior.

### 2.1 Spatial match measure

Adopting the global matching framework [Tao00], a match measure is computed as a function of the difference between the real inspection views and the inspection views predicted based on the reference image and its depth map. Prediction is done by depth based forward warping in which the global visibility is handled using z-buffering. The warping function for each segment is a homography determined by its plane parameters. More specifically, at time instant  $t$ , for the reference view  $I_{t,0}$ , and the inspection view  $I_{t,v}$ ,  $v = 1, \dots, M$ , the global match measure is

$$E_t^1(\Psi_{1,t,0}, \dots, \Psi_{S,t,0}) = \sum_{s=1}^S \sum_{v=1}^M E(I_{s,t,0}, I_{t,v}), \quad (1)$$

where  $E(I_{s,t,0}, I_{t,v}) = g(I_{t,v}, \text{fwarmp}(I_{s,t,0}, H_{s,t,0 \rightarrow s,t,v}))$ , and  $g()$  is the sum of squared differences function. It should be noticed that only pixels visible in the inspection views are considered. For segment  $s$ , the forward warping function  $\text{fwarmp}()$  warps the image of the segment  $I_{s,t,0}$  to inspection view  $v$  through a homography  $H_{s,t,0 \rightarrow s,t,v}$  induced by the plane model. This homography can be computed as

$$H_{s,t,0 \rightarrow s,t,v} = K_v \left[ R_{0 \rightarrow v} + \frac{T_{0 \rightarrow v} n_{s,t,0}^T}{d_{s,t,0}} \right] K_0^{-1}, \quad (2)$$

where  $R_{0 \rightarrow v}$  and  $T_{0 \rightarrow v}$  are the relative rotation and translation between view 0 and view  $v$ . This homography is determined only by the normal vector of the plane and the distance of the plane from the camera center. Therefore, it is only affected by out-of-plane motions since an in-plane motion leaves the plane normal and the distance unchanged.

### 2.2 Scene flow constraint

The scene flow constraint relates the 3D motion of a plane and the resulting 2D optical flow fields or image correspondences. Let the plane of segment  $s$  undergo a motion described by the rotation/translation pair  $(R_{s,t \rightarrow t+1}, T_{s,t \rightarrow t+1})$ . The image projections of the plane in the reference view between times  $t$  and  $t+1$  are related by the homography:

$$H_{s,t,0 \rightarrow s,t+1,0} = K_0 \left[ R_{s,t \rightarrow t+1} + \frac{T_{s,t \rightarrow t+1} n_{s,t,0}^T}{d_{s,t,0}} \right] K_0^{-1}. \quad (3)$$

Also, the homography between the reference view at time  $t$  and view  $v$  at time  $t+1$  is given by

$$H_{s,t,0 \rightarrow s,t+1,v} = H_{s,t,v \rightarrow s,t+1,v} H_{s,t,0 \rightarrow s,t,v}, \quad (4)$$

where the first term is the homography in view  $v$  for segment  $s$  between the two time instants, which can be computed using (3) in the camera coordinate system of view  $v$ . The second term is the spatial homography defined in the previous section.

The temporal match measure  $E_{t,s}^2$  of the segment  $s$  induced by the 3D motion of the plane can now be written as

$$E_{t,s}^2(R_{s,t \rightarrow t+1}, T_{s,t \rightarrow t+1}) = \sum_{v=0}^M g(\text{f Warp}(I_{s,t,0}, H_{s,t,0 \rightarrow s,t+1,v}), I_{t+1,v}). \quad (5)$$

This function computes the match measure between view 0 and  $v$  before and after the motion. Note that the temporal homography constraint for just one view is sufficient to solve for the planar motion parameters. Therefore the above error function over-constrains motion parameters of the plane. It is also clear that the temporal match measure is affected both by in-plane and out-of-plane motions.

### 2.3 Motion smoothness constraint

Various motion models such as a constant velocity model or a constant acceleration model can be employed to constrain the depth estimation process. We adopt a constant rotational/translational velocity model in the local coordinate system of each segment. This coordinate system is obtained by translating the origin of the reference camera coordinate system to the point on the plane that corresponds to the centroid of the image segment. Using an instantaneous motion formulation, in the Appendix, we show that this model induces constant velocities in the normal vector  $n$  and the plane distance  $\tilde{d}$  also. In other words, constant velocity model of for the motion of the plane results in a constant velocity model for its out-of-plane motion parameters. Deviations from this model are penalized through a motion smoothness cost function, which is integrated in the depth estimation process. For segment  $s$ , the cost function is:

$$E_o(\Psi_{s,1,0}, \dots, \Psi_{s,T,0}) = \sum_{t=2}^{T-1} \left| \frac{n_{s,t-1,0} + n_{s,t+1,0}}{|n_{s,t-1,0} + n_{s,t+1,0}|} - n_{s,t,0} \right| + \kappa \left| \frac{\tilde{d}_{s,t-1,0} + \tilde{d}_{s,t+1,0}}{2} - \tilde{d}_{s,t,0} \right| / \left( \left| \tilde{d}_{s,t+1,0} - \tilde{d}_{s,t-1,0} \right| + \delta \right), \quad (6)$$

where  $\kappa$  is the weight for the second term, and  $\delta$  is a small positive number for avoiding overflow in division. The in-plane motion smoothness measure  $E_i$  can also be formulated. However, since it will not be used in the proposed algorithm, the result will not be shown here.

### 2.4 Batch formulation and incremental estimation

With the above geometric constraints and the motion model defined, estimating the depth maps from time 1 to  $T$  is equivalent to finding the plane parameters that minimize the cost function

$$\mathbf{e} = \mathbf{a} \sum_{t=1}^T E_t^1(\mathbf{y}_{0,t,0}, \dots, \mathbf{y}_{s,t,0}) + \mathbf{b} \sum_{t=1}^{T-1} \sum_{s=1}^S E_t^2(R_{s,t \rightarrow t+1}, T_{s,t \rightarrow t+1}) + \sum_{s=1}^S \{ \gamma E_o(\mathbf{y}_{s,1,0}, \dots, \mathbf{y}_{s,T,0}) + \mathbf{I} E_i \} \quad (7)$$

where the constants  $\alpha, \beta, \gamma, \lambda \in [0,1]$  are the weights of each term. Ideally, all the terms in the above function should be utilized for finding the optimal depth maps. However, the second term, i.e., the scene flow constraint, relies on the homographies over time which in turn depends on the accuracy of flow or piecewise parametric motion estimation. It may become unreliable for untextured regions or segments with small spatial support. Therefore, for the algorithm in this paper, we do not use the second term and rely only on the spatial match measure and the temporal smoothness of motion to compute the depth maps. The in-plane motion smoothness term  $E_i$  is also dropped for the same reason. In this mode, motion estimates in terms of optical flow are still utilized but only for the purpose of establishing temporal correspondences between pixels or segments. This correspondence is subsequently employed to predict planar parameters using the motion smoothness constraint. Such a simplification lowers the requirement for the accuracy of the optical flow because as long as the corresponding pixels are in the correct segment, the errors in flow fields will not affect the prediction of the plane parameters. This advantage will be elaborated further in Section 3. The simplified cost function (7) is given by:

$$\varepsilon(\Psi_{s,1,0}, \dots, \Psi_{s,T,0}) = \alpha \sum_{t=1}^T E_t^1(\Psi_{0,t,0}, \dots, \Psi_{s,t,0}) + \gamma \sum_{s=1}^S E_o(\Psi_{s,1,0}, \dots, \Psi_{s,T,0}). \quad (8)$$

When the depth information before time  $t$  is given or already estimated, the depth at time  $t$  can be computed using an incremental formulation. The cost function consists of the spatial match measure at time  $t$  and a motion smoothness measure. More specifically, the cost function for the incremental formulation is given by:

$$\varepsilon_t = \alpha E_t^1(\Psi_{1,t,0}, \dots, \Psi_{s,t,0}) + \gamma \sum_{s=1}^S E(\Psi_{s,t,0}^-, \Psi_{s,t,0}), \quad (9)$$

where  $\psi_{s,t,0}^-$  is the predicted plane parameters based on the smooth motion model. The function  $E(\psi_{s,t,0}^-, \Psi_{s,t,0})$  represents the difference between the two planes  $\psi_{s,t,0}^-$  and  $\Psi_{s,t,0}$ , which can be computed as the average distance between points inside a segment.

### 3 Algorithm

The block diagram of an algorithm based on the incremental formulation is shown in Figure 2. We first briefly describe the main steps of the algorithm and then discuss the details of the algorithm in the following subsections. The three steps of the algorithm are:

**Step 1:** Predicting plane parameters  $\psi_{s,t,0}^-$  from the previous time instants. This step consists of three tasks: (1) for each segment, find the corresponding regions in the previous time instants, (2) find or estimate plane parameters in those regions, and (3) predict the plane parameters at current time instant.

**Step 2:** Initialization of the piecewise planar depth representation  $\psi_{s,t,0}^0$  in the reference view using the spatial match measure  $E_t^1$  only.

**Step 3:** Global depth hypothesis testing. For each segment  $s$ , choose either  $\psi_{s,t,0}^-$  or plane parameters in the initial depth estimation as its depth representation. Find a locally optimal solution for the cost function in (9) using local search. A greedy hypothesis testing algorithm similar to the one proposed in [Tao00] is adopted for this purpose.

### 3.1 Predicting plane parameters

In order to predict the plane parameters of a segment, its corresponding positions in the previous images need to be determined. Based on the depth information in those regions, the depth representation at the current time instant is predicted. We propose two complementary methods for finding this temporal image correspondence. The first method is based on temporal color segmentation and is good for large homogeneous color regions; the second method is based on optical flow and works well for textured regions.

#### 3.1.1 Segmentation based method

For large untextured scene surfaces, the corresponding color segments at different time instants tend to have large overlapping areas when their motions are relatively small. This property is exploited to track corresponding segments over time. The algorithm is illustrated in Figure 3a and is described as follows:

For any large segment  $s$  at time instant  $t-1$

- Project the segment to the next frame  $t$ , noted as  $S_f$
- At time instant  $t$ , find the segments that satisfy the following criteria:
  - 85% of their areas are covered by  $S_f$
  - Have similar color as the segment  $s$

For these segments at time instant  $t$ , their image correspondence at time instant  $t-1$  is image segment  $s$ .

There will be segments that have no correspondence in the previous frame according to the above method. For these segments, the optical flow based method is applied to find image correspondences.

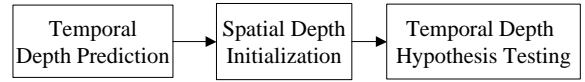


Figure 2. Block diagram of the proposed dynamic depth recovery algorithm.

#### 3.1.2 Optical flow based method

For any pixel in segment  $s$  at time instant  $t$ , its corresponding position in the previous frame can be found if the optical flow is available. This process can also be used to find its image correspondences in multiple previous frames by concatenating the flow fields (Figure 3b). Since optical flow is only used to find corresponding regions of a segment in the previous frames to fit a plane model, as long as the image correspondences are in the right region, error will not affect the resulting plane parameters. In addition, errors in optical flow only affect the temporal plane prediction, which is tested amongst a number of different hypotheses. The erroneous ones will in general not survive.

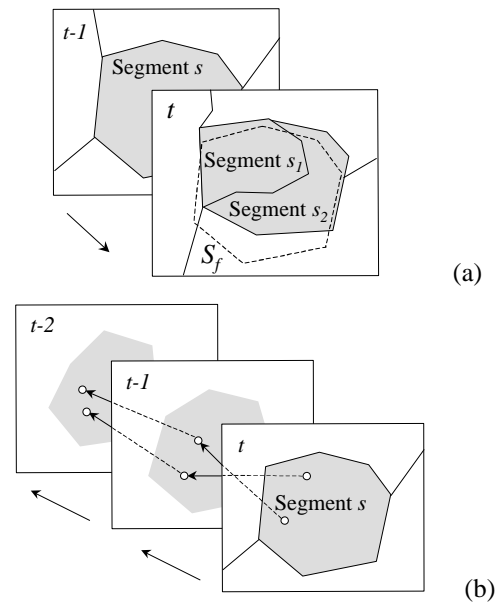


Figure 3. (a) Determine temporal image correspondence based on overlapping areas of color segments. Segment  $s_1$  and segment  $s_2$  at time  $t$  have corresponding segment  $s$  at time  $t-1$ . (b) Determine the corresponding pixels in the previous frames using optical flow.

### 3.1.3 Predicting plane parameters

The image correspondences of a segment in the previous frames are found either by the segmentation based method or by the optical flow based method. The associated plane parameters in the previous frames can then be obtained by using the plane parameters for the tracked segments from the former method, or by fitting planes to the depth values of tracked pixels from the latter method. Based on the plane parameters in the previous frames, the plane parameters at current time instant can be predicted using the smooth motion model. For example, to predict plane parameters from two previous frames  $t-1$  and  $t-2$ , if we denote the plane parameters in those two frames as  $[n_{s,t-1,0}, \tilde{d}_{s,t-1,0}]$  and  $[n_{s,t-2,0}, \tilde{d}_{s,t-2,0}]$ , the normal vector at time  $t$  is computed as the solution of

$$(n_{s,t-2,0} + n_{s,t-1,0}^-) / |n_{s,t-2,0} + n_{s,t-1,0}^-| = n_{s,t,0} \quad (10)$$

which is  $n_{s,t,0}^- = (2n_{s,t-1,0} n_{s,t-2,0}^T - I)n_{s,t-2,0}$ . The plane distance parameter is computed as  $\tilde{d}_{s,t,0}^- = 2\tilde{d}_{s,t-1,0} - \tilde{d}_{s,t-2,0}$ .

### 3.2 Initial depth estimation

The color segmentation based stereo algorithm [Tao00] has been implemented to initialize the depth solution at the current time instant. This algorithm minimizes the spatial match measure  $E_t^1$  and considers visibility implicitly in the global matching criterion. The four steps in this algorithm are (1) color segmentation, (2) correlation based initial depth computation, (3) plane fitting in each segment based on the initial depth, and (4) for each segment, comparison of the initial depth representation with depth hypotheses from its neighboring segments, and selection of the one that improves  $E_t^1$  as the solution. The estimated initial depth representation for segment  $s$  is denoted as  $\psi_{s,t,0}^0$ .

### 3.3 Temporal depth hypothesis testing

The hypotheses for each segment  $s$  are the predicted plane parameters  $\psi_{s,t,0}^-$ , the initial plane parameters  $\psi_{s,t,0}^0$ , and the initial plane parameters of the neighboring segments of  $s$ . With plane parameters of the other segments fixed, these hypotheses are compared using the cost function in (9). This process is summarized as follows.

- At time instant  $t$ , for each segment  $s$  in the reference view
  - Compute the cost function (9) for  $\psi_{s,t,0}^-$ ,  $\psi_{s,t,0}^0$ , and  $\psi_{k,t,0}^0$ ,  $k \in$  neighboring segments of  $s$
  - Set plane parameter  $\psi_{s,t,0}$  to be the one with the lowest cost value.

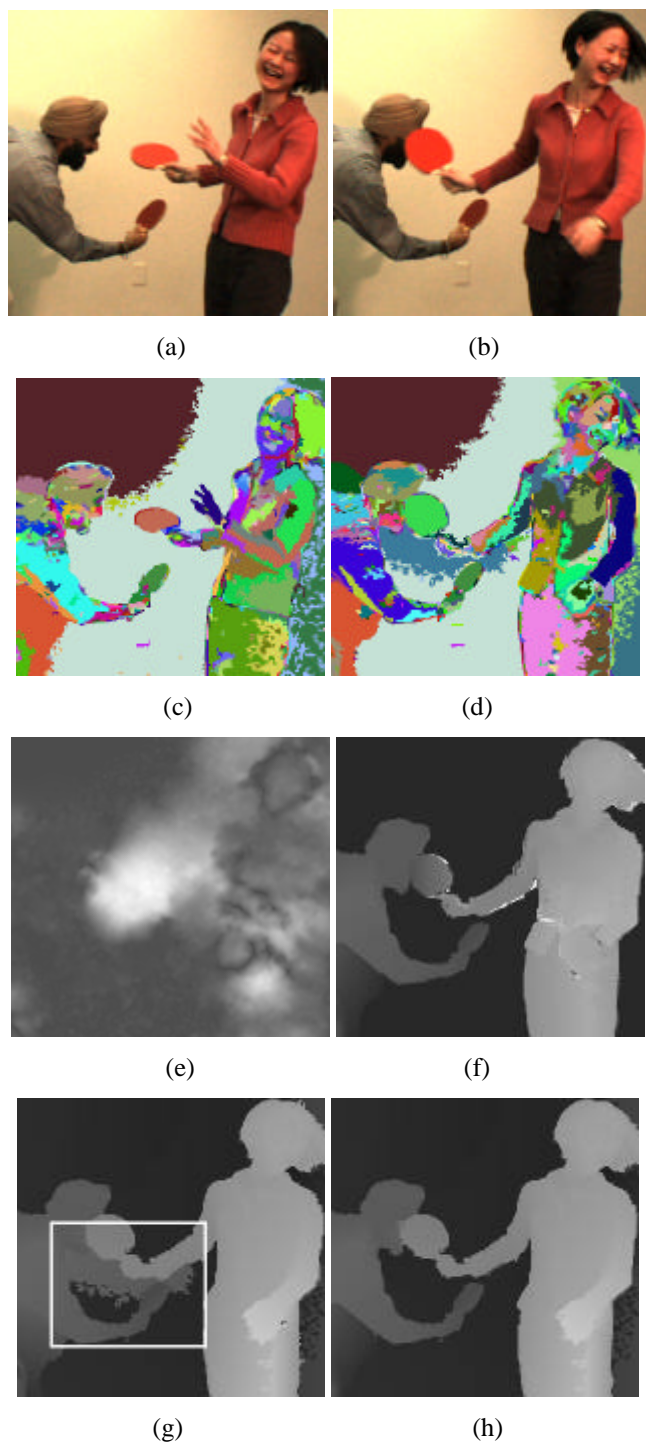


Figure 4. The dynamic depth recovery algorithm. (a,b) Original video frames 267 and 270. (c,d) Color segmentation in frames 267 and 270. Tracked segments are painted using the same color. (e) The magnitude of the optical flow computed in frame 270. (f) The predicted depth map in frame 270. (g) The initial depth map computed using the color segmentation based stereo algorithm in frame 270. The depth value of the area below the PingPong paddle is wrong. (h) By combining (f) and (g), the final depth map is computed. The errors in the area below the Ping Pong paddle are corrected.

## 4 Experimental results

The proposed method has been implemented and tested on synchronized video streams collected from an eight-camera video capture system. All cameras fixate on subjects about nine feet away, with an approximate angular separation of 10 degrees from each other. The videos are captured at 640x480 resolution in RGB color at 30Hz. The intrinsic parameters and poses of all the cameras are computed using the camera calibration algorithm of [Zhang00].

Figure 4 shows the main steps of the proposed dynamic depth recovery algorithm. Figures 4a and 4b are two frames captured from the reference view at different time instants. The color segmentation and the segment tracking results are shown in Figures 4c and 4d. Corresponding segments tracked using the method described in Section 3.1.1 are painted with the same color. For segments that are not tracked successfully, optical flow based method is used to perform pixel level tracking. The optical flow between the current frame and the previous frame is computed using a direct method [Bergen92]. The magnitude of the flow field is shown in Figure 4e. Using the image correspondences computed from the segmentation based method or the optical flow based method, the depth representation at the current time instant is predicted (Figure 4f). This predicted depth map is combined with the initial depth map computed using the color segmentation based stereo algorithm (Figure 4g) to derive the final depth estimation (Figure 4h). Note that the depth error in the highlighted area in the figure below the Ping Pong paddle in the initial depth map is corrected using the predicted depth map.

Figure 5 shows a time sequence of the dynamic depth recovery results. The depth maps are temporally consistent and sharp depth boundaries such as the accurate contours around fingers are preserved. Compared to the single frame algorithm, improvements in temporal consistency have been observed from the resulting depth sequences. Unfortunately, the reduction in temporal scintillation cannot be easily shown in print. Figure 6 shows the depth estimation results of another sequence. The reader is particularly directed to observe the sharp delineation of thin structures, like the fingers, from the background surface.

## 5 Discussions

Based on a piecewise planar scene representation, we study the three constraints applicable to the problem of dynamic depth recovery. The observation that constant velocity motion of a plane causes constant changes in out-of-plane motion parameters enables a simple algorithm that enforces motion smoothness across time without using the 2D optical flow field in the estimation of 3D homographies and the resulting planar velocities. The optical flow is only used for finding temporal correspondences between frames for the purpose of predicting plane parameters. Since in-plane motion does not change the predicted plane parameters, more errors in the optical flow field can be tolerated.

An algorithm based on the global matching framework is proposed. Experimental results show that temporal consistency in depth estimation is achieved and sharp depth boundaries are preserved. We are further exploring how the planar constraint on segments can be combined with non-parametric depths and with other smoothness constraints on the resulting surface shapes.

We plan to evaluate the quality of the generated depth maps by creating synthesized videos from novel viewpoints and assessing the spatial and temporal quality of the videos.

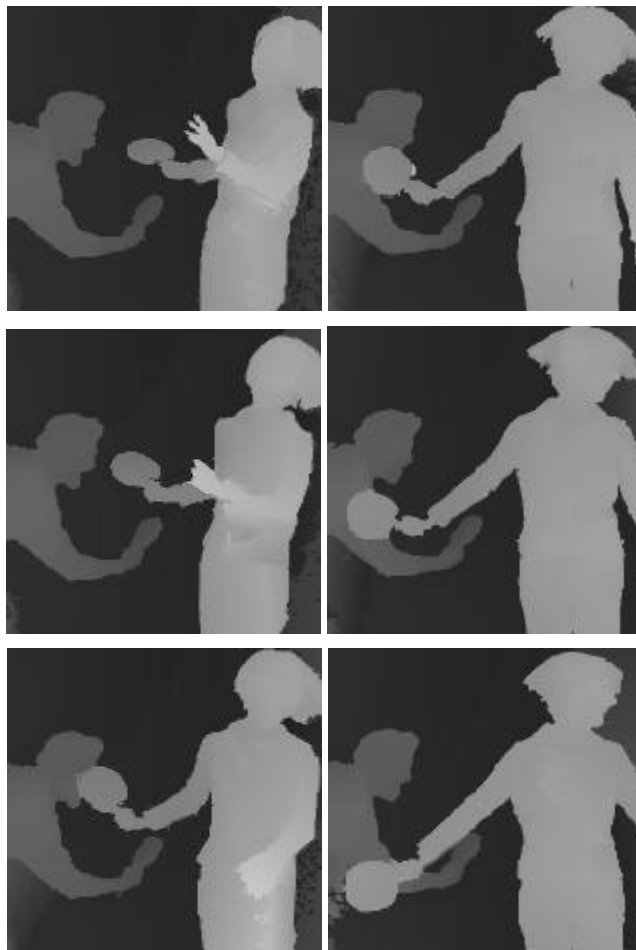


Figure 5. A depth map sequence estimated using the proposed dynamic depth recovery algorithm. The left column shows the first three depth maps and the right column shows later frames in the sequence. Sharp depth boundaries such as the finger contours in the first image are preserved even though the optical flow is blurry.

## Appendix

If a planar scene undergoes motion described by:  $P' = \mathbf{w} \times P + \mathbf{t}$ , where  $\mathbf{w}$  and  $\mathbf{t}$  are the rotation and translation respectively, then the out-of-plane motion of a plane,  $n^T P = d$ , is given by:  $n' = \mathbf{w} \times n$ , and  $d' = n'^T \mathbf{t}$ . It follows that for constant velocities,  $\mathbf{w}$  and  $\mathbf{t}$ , the induced velocities for the plane parameters  $n$  and  $d$  are constant too.

## References

- [Adiv85] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 4, July 1985.
- [Belhumeur96] P. N. Belhumeur, "A Bayesian-approach to binocular stereopsis," *Int. Journal of Computer Vision*, vol. 19, no. 3, pp. 237-260, August 1996.
- [Bergen92] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proc. European Conference on Computer Vision (ECCV'92)*, pp. 237-252, 1992.
- [Boykov99] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," in *Proc. Int. Conf. on Computer Vision (ICCV'99)*, Sept. 1999.
- [Carceroni01] R. L. Carceroni and K. N. Kutulakos, "Scene capture by surfel sampling: from multi-view streams to non-rigid 3D motion, shape and reflectance," in *Proc. Int. Conf. on Computer Vision (ICCV'01)*, July 2001.
- [Dhond89] U. R. Dhond and J. K. Aggarwal, "Structure from stereo: a review," *IEEE Transactions on System, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489-1510, 1989.
- [Hanna93] K. J. Hanna and Neil E. Okamoto, "Combining stereo and motion analysis for direct estimation of scene structure," in *Proc. Int. Conf. on Computer Vision*, pp. 357-265, 1993.
- [Kutulakos99] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," in *Proc. Int. Conf. on Computer Vision (ICCV'99)*, pp. 307-314, 1999.
- [Lhuillier00] M. Lhuillier and L. Quan, "Edge-constrained joint view triangulation for image interpolation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'00)*, June 2000.
- [Roy98] S. Roy and I. J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem", in *Proc. Int. Conf. on Computer Vision (ICCV'98)*, Bombay, India, January 1998.
- [Shan2001] Z. Zhang and Y. Shan, "A progressive scheme for stereo matching," *Lecture Notes in Computer Science 2018*, Springer-Verlag, pp. 68-85, March 2001.
- [Szeliski99] R. Szeliski, "A multi-view approach to motion and stereo," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'99)*, pp. I-157-163, June 1999.
- [Tao00] H. Tao and H. S. Sawhney, "Global matching criterion and color segmentation based Stereo," in *Proc. Workshop on the Application of Computer Vision (WACV2000)*, pp. 246-253, December 2000.
- [Vedula99] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade, "Three-dimensional scene flow," in *Proc. Int. Conf. on Computer Vision*, pp. II-722 - 729, Sept. 1999.
- [Young90] G. Young and R. Chellappa, "3-D motion estimation using a sequence of noisy stereo images: models, estimation and uniqueness results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp.735-759, 1990.
- [Zhang92] Z. Zhang and O. Faugeras, *3D Dynamic Scene Analysis: A Stereo Based Approach*, Springer, Berlin, Heidelberg, 1992.
- [Zhang99] Y. Zhang and C. Kambhamettu, "Integrated 3D scene flow and structure recovery from multiview image sequences," in *Proc. IEEE Conf. on Computer Vision and Pattern*

*Recognition (CVPR'00)*, pp. II-674-681, South Carolina, June 2000.

- [Zhang00] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330-1334, 2000.

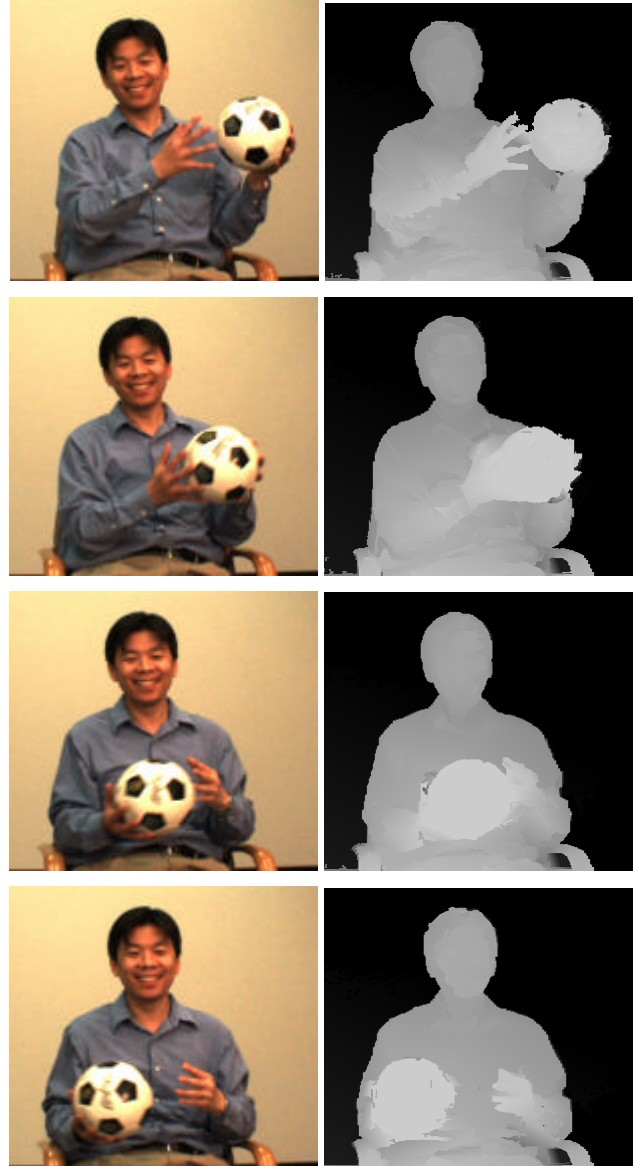


Figure 6. The depth map sequence estimated using the proposed dynamic depth recovery algorithm. The left column shows the original reference images and the right column shows the estimate depth maps.