Motion and Structure Factorization and Segmentation of Long Multiple Motion Image Sequences *

Chris Debrunner[1] *and Narendra Ahuja*

Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801
[1] Currently at Martin Marietta Corporation, Denver CO 80201

Abstract. This paper presents a computer algorithm which, given a dense temporal sequence of intensity images of multiple moving objects, will separate the images into regions showing distinct objects, and for those objects which are rotating, will calculate the three-dimensional structure and motion. The method integrates the segmentation of trajectories into subsets corresponding to different objects with the determination of the motion and structure of the objects. Trajectories are partitioned into groups corresponding to the different objects by fitting the trajectories from each group to a hierarchy of increasingly complex motion models. This grouping algorithm uses an efficient motion estimation algorithm based on the factorization of a measurement matrix into motion and structure components. Experiments are reported using two real image sequences of 50 frames each to test the algorithm.

1 Introduction

This paper is concerned with three-dimensional structure and motion estimation for scenes containing multiple independently moving rigid objects. Our algorithm uses the image motion to separate the multiple objects from the background and from each other, and to calculate the three-dimensional structure and motion of each such object. The two-dimensional motion in the image sequence is represented by the image plane trajectories of feature points. The *motion* of each object, which describes the three-dimensional rotation and translation of the object between the images of the sequence, is computed from the object's feature trajectories. If the object on which a particular group of feature points lie is rotating, the relative three-dimensional positions of the feature points, called the *structure* of the object, can also be calculated.

Our algorithm is based on the following assumptions: (1) the objects in the scene are rigid, i.e., the three-dimensional distance between any pair of feature points on a particular object is constant over time, (2) the feature points are orthographically projected onto the image plane, and (3) the objects move with constant rotation per frame. This algorithm integrates the task of segmenting the images into distinctly moving objects with the task of estimating the motion and structure for each object. These tasks are performed using a hierarchy of increasingly complex motion models, and using an efficient and accurate factorization-based motion and structure estimation algorithm.

This paper makes use of an algorithm for factorization of a measurement matrix into separate motion and structure matrices as reported by the authors in [DA1]. Subsequently in [TK1], Tomasi and Kanade present a similar factorization-based method which allows arbitrary rotations, but does not have the capability to process trajectories starting and ending at arbitrary frames. Furthermore, it appears that some assumptions about the magnitude or smoothness of motion are

* Supported by DARPA and the NSF under grant IRI-89-02728, and the State of Illinois Department of Commerce and Community Affairs under grant 90-103.

still necessary to obtain feature trajectories. Kanade points out [Ka1] that with our assumption of constant rotation we are absorbing the trajectory noise primarily in the structure parameters whereas their algorithm absorbs them in both the motion and structure parameters.

Most previous motion-based image sequence segmentation algorithms use optical flow to segment the images based on consistency of image plane motion. Adiv in [Ad1] and Bergen *et al* in [BB1] instead segment on the basis of a fit to an affine model. Adiv further groups the resulting regions to fit a model of a planar surface undergoing 3-D motions in perspective projection. In [BB2] Boult and Brown show how Tomasi and Kanade's motion factorization method can be used to split the measurement matrix into parts consisting of independently moving rigid objects.

2 Structure and Motion Estimation

Our method relies heavily on the motion and structure estimation algorithm presented in [DA1], [De1], and [DA2]. The input to this algorithm is a set of trajectories of orthographically projected feature points lying on a single rigid object rotating around a fixed-direction axis and translating along an arbitrary path. If these constraints do not hold exactly the algorithm will produce structure and motion parameters which only approximately predict the input trajectories. Given a collection of trajectories (possibly all beginning and ending at different frames) for which the constraints do hold, our algorithm finds accurate estimates of the relative three-dimensional positions of the feature points at the start of the sequence and the angular and translational velocities of the object. The algorithm also produces a confidence number, in the form of an error between the predicted and the actual feature point image positions. Aside from SVDs, the algorithm is closed form and requires no iterative optimization.

In Section 4, the results of applying our motion and structure estimation algorithm to real image sequences are presented in terms of the rotational parameters ω and $\hat{\mathbf{n}}$ and the translational motion parameter $\hat{\mathbf{v}}$. The parameter ω represents the angular speed of rotation about the axis along the unit vector $\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is chosen such that it points toward the camera (ω is a signed quantity). Since we are assuming orthographic projection, the depth component of the translation cannot be recovered, so $\hat{\mathbf{v}}$ is a two vector describing the image plane projection of the translational motion. Although the motion and structure estimation algorithm can accommodate arbitrary motion, most of the objects in the experimental image sequences are moving with constant velocity and their translational velocity is given in terms of $\hat{\mathbf{v}}$.

3 Image Sequence Segmentation and Motion and Structure Estimation

The segmentation of the feature point trajectories into groups corresponding to the differently moving 3D objects and the estimation of the structure and motion of these objects are highly interrelated processes: if the correct segmentation is not known, the motion and structure of each object cannot be accurately computed, and if the 3D motion of each object is not accurately known, the trajectories cannot be segmented on the basis of their 3D motion. To circumvent this circular dependency, we integrate the segmentation and the motion and structure estimation steps into a single step, and we incrementally improve the segmentation and the motion and structure estimates as each new frame is received.

The general segmentation paradigm is split and merge. Each group of trajectories (or *region*) in the segmentation has associated with it one of three *region motion models*, two of which describe rigid motion (the *translational* and *rotational* motion models), and the third (*unmodeled* motion) which accounts for all motions which do not fit the two rigid motion models and do not contain any local motion discontinuities. When none of these motion models accurately account

for the motion in the region, the region is split using a region growing technique. When splitting a region, a measure of motion consistency is computed in a small neighborhood around each trajectory in the region. If the motion is consistent for a particular trajectory, we assume that the trajectories in the neighborhood all arise from points on a single object. Thus the initial subregions for the split consist of groups of trajectories with locally consistent motion, and these are grown out to include the remaining trajectories.

Initially all the trajectories are in a single region. Processing then continues in a uniform fashion: the new point positions in each new frame are added to the trajectories of the existing regions, and then the regions are processed to make them compatible with the new data. The processing of the regions is broken into four steps: (1) if the new data does not fit the old region motion model, find a model which does fit the data or split the region, (2) add any newly visible points or ungrouped points to a compatible region, (3) merge adjacent regions with compatible motions, (4) remove outliers from the regions.

Compatibility among feature points is checked using the structure and rotational motion estimation algorithm or the translational motion estimation algorithm described in [De1]. A region's feature points are considered incompatible if the fit error returned by the appropriate motion estimation algorithm is above a threshold. We assume that the trajectory detection algorithm can produce trajectories accurate to the nearest pixel, and therefore we use a threshold (which we call the *error threshold*) of one half of a pixel per visible trajectory point per frame.The details of the four steps listed above may be found in [De1] or [DA3].

4 Experiments

Our algorithm was tested on two real image sequences of 50 frames: (1) the cylinder sequence, consisting of images of a cylinder rotating around a nearly vertical axis and a box moving right with respect to the cylinder and the background, and (2) the robot arm sequence, consisting of images of an Unimate® PUMA® Mark III robot arm with its second and third joints rotating in opposite directions. These sequences show the capabilities of the approach, and also demonstrate some inherent limitations of motion based segmentation and of monocular image sequence based motion estimation.

Trajectories were detected using the algorithm described in [De1] (using a method described in [BH1]), which found 2598 trajectories in the cylinder sequence and 202 trajectories in the robot arm sequence. These trajectories were input to the image sequence segmentation algorithm described in Section 3, which partitioned the trajectories into groups corresponding to different rigid objects and estimated the motion and structure parameters.

The segmentation for the cylinder sequence is shown in Fig. 1. The algorithm separated out the three image regions: the cylinder, the box, and the background. The cylinder is rotating, and thus its structure can be recovered from the image sequence. Fig. 2 shows a projection along the cylinder axis of the 3D point positions calculated from the 1456 points on the cylinder. The points lie very nearly on a cylindrical surface. Table 1 shows the estimated and the actual motion param-

Table 1. Comparison of the parameters estimated by the algorithm and the true parameter for the cylinder image sequence experiment.

Parameters	Estimated	Actual
ω	-0.022	-0.017
$\hat{\mathbf{n}}$	(0,.99,.12)	(0,.98,.19)
$\hat{\mathbf{v}}$	(.29,-.19)	(.14,0)

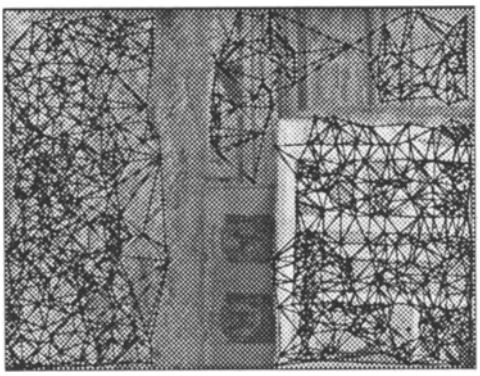

Fig. 1 The image sequence segmentation found for the cylinder sequence (the segmentation is superimposed on the last frame of the sequence).

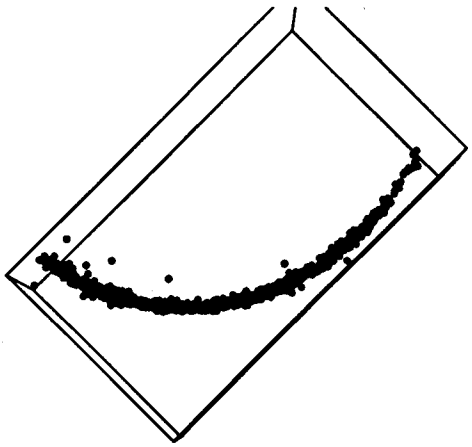

Fig. 2 An end-on view of the three-dimensional point positions calculated by our structure and motion estimation algorithm from point trajectories derived from cylinder image sequence.

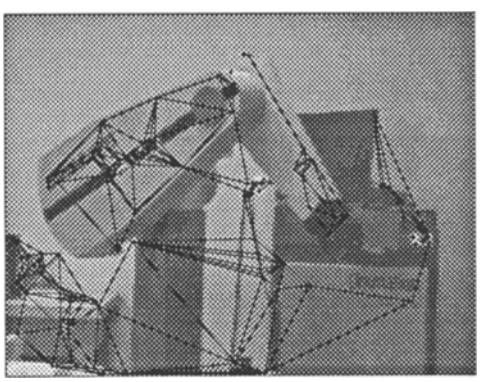

Fig. 3 The image sequence segmentation found for the robot arm sequence (the segmentation is superimposed on the last frame of the sequence).

Table 2. Comparison of the estimated and the true parameter values for the second (larger) segment of the robot arm.

Parameters	Estimated	Actual
ω	.0133	.0131
$\hat{\mathbf{n}}$	(-.67,-.01,.74)	(-.62,.02,.79)
$\hat{\mathbf{v}}$	(.02,-.07)	(0,0)

Table 3. Comparison of the estimated and the true parameter values for the third (smaller) segment of the robot arm.

Parameters	Estimated	Actual
ω	-.0127	-.0131
$\hat{\mathbf{n}}$	(-.58,.06,.81)	(-.62,.02,.79)

eters for the cylinder. The error in the ω estimate is large because the cylinder is rotating around an axis nearly parallel to the image plane and, as pointed out in [WH1], a rotation about an axis parallel to the image plane is inherently difficult to distinguish from translation parallel to the image plane and perpendicular to the rotation axis (this also explains the error in $\hat{\mathbf{v}}$). Note that the predicted trajectory point positions still differ from the actual positions by an average of less than the error threshold of 0.5 pixel. The accuracy of the motion and structure estimation algorithm for less ambiguous motion is illustrated in the experiments on the robot arm sequence.

The image sequence segmentation for the robot arm sequence is shown in Fig. 3. Note that several stationary feature points (only two visible in Fig. 3) on the background are grouped with

the second segment of the arm. This occurs because any stationary point lying on the projection of a rotation axis with no translational motion will fit the motion parameters of the rotating object. Thus these points are grouped incorrectly due to an inherent limitation of segmenting an image sequence on the basis of motion alone. The remaining points are grouped correctly into three image regions: the second and the third segments of the robot arm, and the background. The two robot arm segments are rotating and their three-dimensional structure was recovered by the motion and structure estimation algorithm.Only a small number of feature points were associated with the robot arm segments making it difficult to illustrate the structure on paper, but the estimated motion parameters of the second and third robot arm segments are shown in Table 2 and Table 3, respectively. Note that all the motion parameters were very accurately determined.

5 Conclusions

The main features of our method are: (1) motion and structure estimation and segmentation processes are integrated, (2) frames are processed sequentially with continual update of motion and structure estimates and segmentation, (3) the motion and structure estimation algorithm factors the trajectory data into separate motion and structure matrices, (4) aside from SVDs, the motion and structure estimation algorithm is closed form with no nonlinear iterative optimization required, (5) the motion and structure estimation algorithm provides a confidence measure for evaluating any particular segmentation.

References

[Ad1]Adiv, G.: Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects. IEEE Transactions on PAMI **7** (1985) 384-401

[BB1]Bergen, J., Burt, P., Hingorani, R., Peleg, S.: Multiple Component Image Motion: Motion Estimation. Proc. of the 3^d ICCV, Osaka, Japan (December 1990) 27-32

[BB2]Boult, T., Brown, L.: Factorization-based Segmentation of Motions. Proc. of the IEEE Motion Workshop, Princeton NJ (October 1991) 21-28

[BH1]Blostein, S., Huang, T.: Detecting Small, Moving Objects in Image Sequences using Sequential Hypothesis Testing. IEEE Trans. on Signal Proc. **39** (July 1991) 1611-1629

[DA1]Debrunner, C., Ahuja, N.: A Direct Data Approximation Based Motion Estimation Algorithm. Proc. of the 10th ICPR, Atlantic City, NJ (June 1990) 384-389

[DA2]Debrunner, C., Ahuja, N.: Estimation of Structure and Motion from Extended Point Trajectories. (submitted)

[DA3]Debrunner, C., Ahuja, N.: Motion and Structure Factorization and Segmentation of Long Multiple Motion Image Sequences. (submitted)

[De1]Debrunner, C.: Structure and Motion from Long Image Sequences. Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL (August 1990)

[Ka1]Kanade, T.: personal communication (October 1991)

[KA1]Kung, S., Arun, K., Rao, D.: State-Space and Singular-Value Decomposition-Based Approximation Methods for the Harmonic Retrieval Problem. J.of the Optical Society of America (December 1983) 1799-1811

[TK1] Tomasi, C., Kanade, T.: Factoring Image Sequences into Shape and Motion. Proc. of the IEEE Motion Workshop, Princeton NJ (October 1991) 21-28

[WH1]Weng, J., Huang, T., Ahuja, N.: Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation. IEEE Transactions on PAMI **11** (1989) 451-476