INCREMENTAL SHAPE RECONSTRUCTION USING STEREO IMAGE SEQUENCES

Tai Jing Moyung and Paul W. Fieguth

Department of Systems Design Engineering
University of Waterloo, Ontario, Canada, N2L 3G1

ABSTRACT

The limitations of estimating structure from either stereo or motion alone can be addressed by the use of stereo image sequences; however, many existing techniques for processing stereo sequences rely on having accurate feature correspondences initially. In this paper, we present an *iterative* feature matching and shape reconstruction algorithm. The proposed method simultaneously resolves matching ambiguities by multiple hypothesis testing and develops an incrementally accurate and dense representation of the reconstructed object. This approach uses minimal domain specific constraints and can be easily generalised. The method's potential is demonstrated in a space vision application.

1. INTRODUCTION

The automated reconstruction of three-dimensional objects in space has many useful applications such as satellite identification, grasping, docking, and fault diagnosis (Fig. 1). However, the space context presents considerable computer vision challenges in terms of extreme lighting conditions, specular reflection and hard shadows, all of which can lead to mistaken or inappropriate image registration. Recently there has been a growing interest in the use of stereo image *sequences* for the estimation of motion and structure in dynamic scenes [1, 2]. Many of these approaches, with a few exceptions [3, 4], either assume that accurate feature correspondences are already established, or that the brightness patterns of the images remain constant, which may not be the case under conditions such as those in Fig. 2.

This paper addresses the problem of robust feature matching and rigid shape reconstruction from stereo image sequences. Specifically, we develop an algorithm to find feature matches that are consistent with both stereo (epipolar) and three-dimensional motion constraints, without any reliance on the specific attributes of the extracted features. Furthermore, the algorithm robustly bootstraps to a solution: as correspondences are found, motion estimates are improved, further constraining the matching problem.

Research in this paper is funded in part by Natural Science and Engineering Research Council of Canada. Images are courtesy of Macdonald Dettwiler Space and Advanced Robotics Ltd.

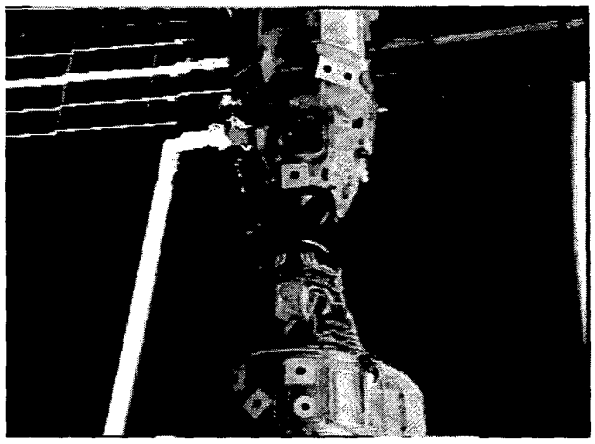


Fig. 1. Robotic arm grasping a micro-satellite in space for rendezvous and docking.

2. ROBUST FEATURE MATCHING

Consider a set of point features on a rigid body, $\{\boldsymbol{p}_i\}$, represented by their 3D position on some coordinate system. The 2D perspective projection of each point onto an image $I_s(f)$ at frame f from a viewpoint s can be defined as a nonlinear function of $\boldsymbol{p}_i$:

$$\boldsymbol{m}_i^s(f) = \boldsymbol{h}[s, \boldsymbol{p}_i(f)]. \tag{1}$$

The problem of 3D shape reconstruction from multiple images is to recover the depth information lost during projection and to determine $\{\boldsymbol{p}_i\}$ given a set of images $\{I_s(f)\}$, which vary temporally with f and/or spatially with s.

Two traditional approaches of 3D reconstruction are structure-from-stereo [5] and structure-from-motion [6]. In stereo vision, structure is reconstructed from feature matches between the left and right camera images by finding the set of point pairs $\{\boldsymbol{m}_i^L(f), \boldsymbol{m}_i^R(f)\}$. Since the baseline in a typical stereo system is relatively large, geometric distortion, occlusion and change in reflection lead to difficulties in feature matching. Therefore most of the literature in this area [5] focuses on developing constraints and techniques to establish accurate correspondences in a *single* stereo im-

0-7803-6297-7/00/$10.00 © 2000 IEEE

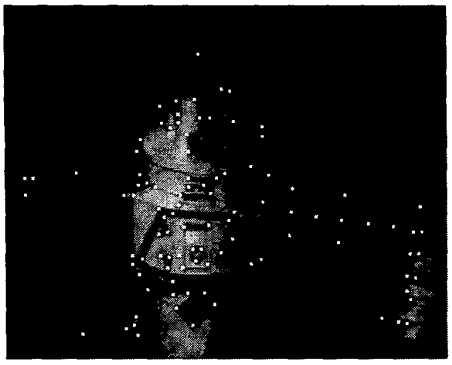


Fig. 2. A mock-up model of a satellite. Detected corner features are shown as white dots. Notice how specular reflection and strong shadows introduce false features.

age pair. Structure from motion uses temporally varying monocular images and requires finding the set of point matches $\{\boldsymbol{m}_i^s(f), \boldsymbol{m}_i^s(f+1)\}$. It takes advantage of the small baseline between frames to establish feature matches by tracking or optical flow. However, the reconstructed shape based on a few closely sampled image frames is often noise sensitive because of the small baseline. Batch processing of a long sequence can improve accuracy, but reconstruction is possible only after all images are available.

Due to noise, occlusion and errors in feature extraction, ambiguities often arise in establishing feature correspondences. The key to accurate 3D reconstruction is the ability to resolve matching ambiguities and to reject outliers. Statistical data association techniques and multiple hypothesis trees have been used in the past for establishing motion correspondences [7, 8] in monocular image sequences. In this paper, we use a similar approach, but include ambiguous stereo matching candidates in the list of hypotheses to be tested at each time step. Hypotheses for stereo and motion correspondences are created and deleted based on four sets of constraints: frame-to-frame motion constraints $\{\boldsymbol{m}_i^L(f), \boldsymbol{m}_i^L(f+1)\}$ and $\{\boldsymbol{m}_i^R(f), \boldsymbol{m}_i^R(f+1)\}$, and view-to-view epipolar constraints $\{\boldsymbol{m}_i^L(f), \boldsymbol{m}_i^R(f)\}$ and $\{\boldsymbol{m}_i^L(f+1), \boldsymbol{m}_i^R(f+1)\}$. Together, these constraints provide robust outlier rejection.

3. DYNAMIC ESTIMATION

Shape reconstruction in the proposed algorithm is an iterative process between motion estimation and reconstruction from feature matches, as illustrated in Fig. 3. We assume that the motion of the object with respect to the cameras is constant or varying slowly, and follows a smooth trajectory. At system start-up, stereo matching candidates are identified and the 2D feature points forming these candidates are tracked independently using a standard Kalman filter. The

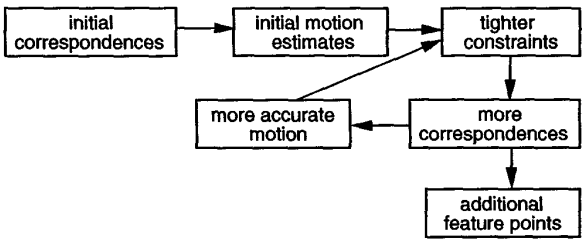


Fig. 3. Iterative motion estimation and reconstruction.

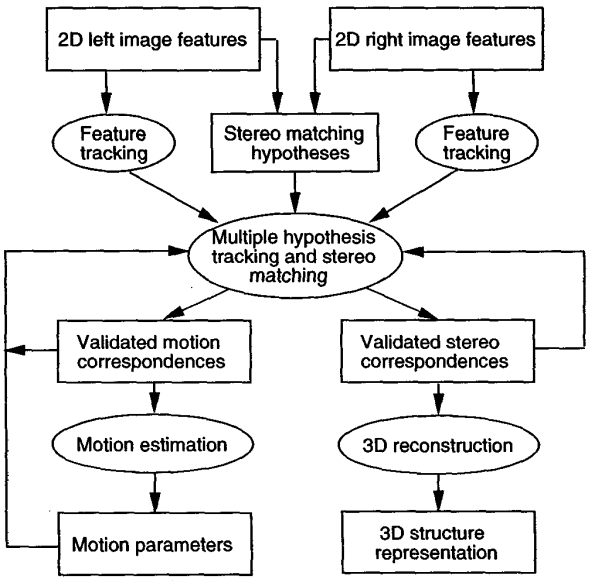


Fig. 4. Flow chart of the incremental reconstruction algorithm at each time step.

dynamic model at this time consists of only 2D motion and each feature point is not constrained by object rigidity; predictions are made using second order polynomial extrapolation. Some of the initial stereo matching candidates may be unambiguous. If their underlying 2D feature points also have unambiguous temporal matches across the next frame, two sets of 3D points, $\{\boldsymbol{p}_i(f)\}$ and $\{\boldsymbol{p}_i(f+1)\}$, can be reconstructed. These two sets of points provide an initial estimate of the rigid motion.

Let $R(f)$ be a 3×3 orthogonal rotation matrix and $\boldsymbol{t}(f) = [t_x\ t_y\ t_z]^T$ a translation vector. The positions of the reconstructed 3D points at each frame f can be modelled as

$$\boldsymbol{p}_i(f+1) = R(f)\boldsymbol{p}_i(f) + \boldsymbol{t}(f). \tag{2}$$

The twelve unknown parameters of $\hat{R}(f)$ and $\hat{\boldsymbol{t}}(f)$ are estimated uniquely using linear techniques from a minimum of four corresponding pairs of $\{\boldsymbol{p}_i(f), \boldsymbol{p}_i(f+1)\}$. As a result, a 3D motion dynamic model can now be applied to improve

tracking accuracy and to resolve matching ambiguities in future frames using rigidity constraints.

An extended Kalman filter is used to predict the location of 2D feature points and to update the current position of the reconstructed 3D points using the following dynamic and measurement model:

$$\boldsymbol{p}_i(f+1) = \hat{R}(f)\boldsymbol{p}_i(f) + \hat{\boldsymbol{t}}(f), \tag{3}$$

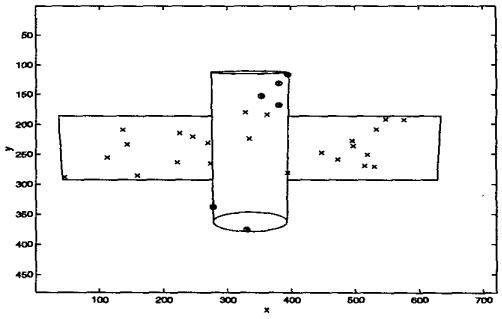
$$\boldsymbol{z}_i(f) = \begin{bmatrix} \boldsymbol{h}[L, \boldsymbol{p}_i(f)] \\ \boldsymbol{h}[R, \boldsymbol{p}_i(f)] \end{bmatrix} + \boldsymbol{v}(f). \tag{4}$$

$\boldsymbol{z}_i(f)$ is the measurement vector and $\boldsymbol{v}$ is zero-mean Gaussian white noise, with covariance Q, representing the measurement error. The accuracy of the motion estimate used in the above model can be improved as the number of reconstructed 3D points increases and their position estimates improve. Fig. 4 presents a simplified flow chart of the overall algorithm.

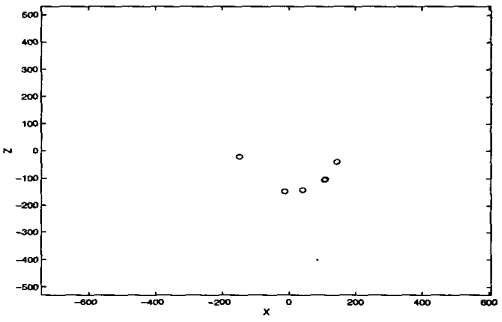
4. RESULTS

Simulation results of the incremental algorithm are illustrated in Fig. 5. A simple model consisting of a cylinder and two planar surfaces is constructed to imitate a satellite. Thirty points on the surface are randomly generated as the extracted features and a simulation experiment is conducted with synthetic motion and stereo set up. Issues such as occlusion and feature extraction errors are ignored. The cameras have coplanar and parallel optical axes perpendicular to the stereo baseline. The object model is rotated 5 degrees around the camera's optical axis between consecutive frames. Fig. 5(a) shows the shape of the satellite and the point features as seen in the left camera image. The six points that are reconstructed after frame 1 are shown in Fig. 5(b), where the vertical axis represents the recovered depth. It is clear why unambiguous stereo matches are found for these six points, because there are no other points in Fig. 5(a) close to their horizontal epipolar lines. As the algorithm learns about the motion and more constraints are available, more feature matches, and consequently more 3D points, can be constructed, as shown in Fig. 5(c), 5(d).

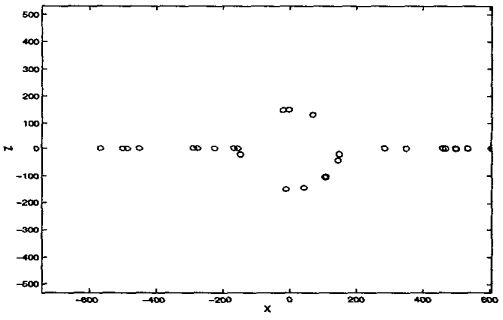
The second experiment involves reconstructing a real object from a sequence of stereo images captured in a laboratory environment. Twenty of the most significant point features are extracted from each image as described in [9]. Fig. 6 shows the set of point features in the images at frame 1. Only features left with one stereo matching candidate are reconstructed, and the reconstructed points at several frames are presented in Fig. 7. In order to demonstrate the results meaningfully, the estimated set of points $\{\boldsymbol{p}_i(f)\}$ are projected back onto their respective images. Because of the small number of features that we have chosen to extract and a high proportion of which being false features, many points disappear from frame to frame and are not present



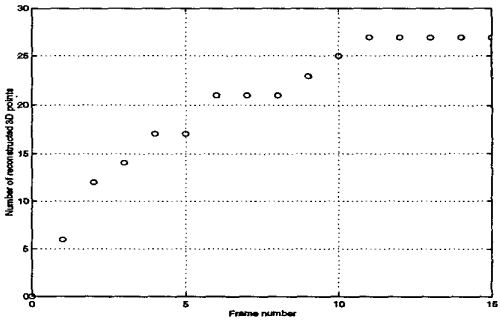
(a) Image of satellite shape and selected features(×). Points reconstructed in frame 1 are identified by (o).



(b) Reconstructed points after frame 1.



(c) Reconstructed points after frame 15.



(d) Number of points reconstructed vs. frame number.

Fig. 5. Results of simulation experiment.

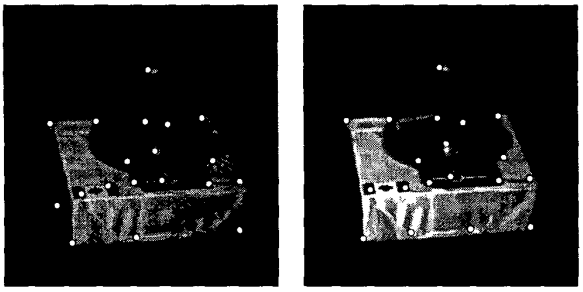


Fig. 6. The left and right images of a real stereo sequence at frame 1. The circles represent the extracted features.

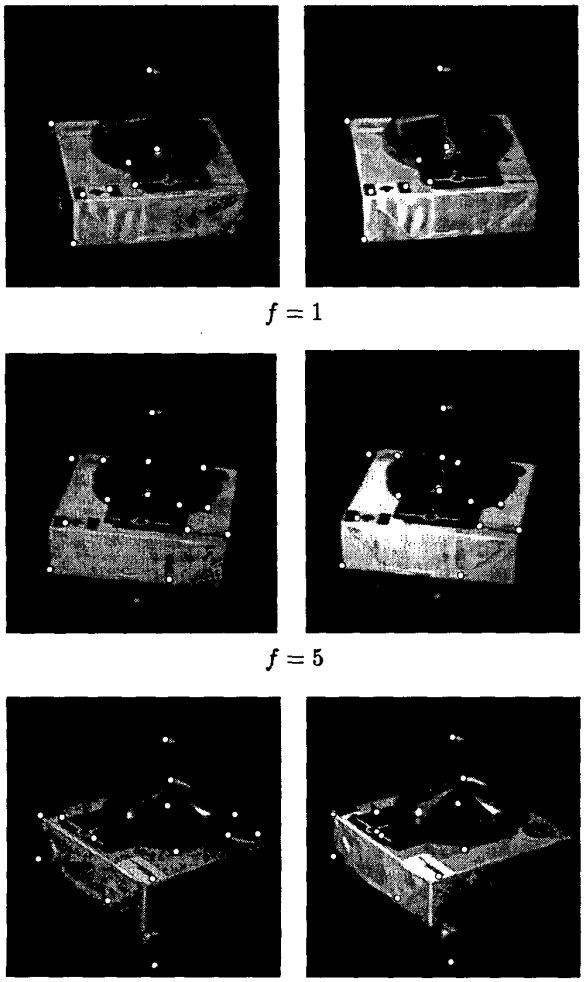


Fig. 7. Reconstructed 3D points projected back onto the left and right images at frames 1, 5, and 15.

long enough for them to be tracked with confidence. As a result, the total number of reconstructed points does not increase significantly. However, we have successfully reconstructed some of the feature points that have newly appeared in the sequence.

5. CONCLUSION

A feature-based, incremental 3D reconstruction algorithm using stereo image sequences has been presented. Its effectiveness has been demonstrated using both synthetically generated data and a real stereo sequence. Possible future work include the investigation of robust motion estimation techniques, full 3D reconstruction from complete 360° view information, and extending the framework to use other features such as edges and lines segments.

6. REFERENCES

[1] G. J. Young and R. Chellappa, "3-D motion estimation using sequence of noisy stereo images: Models, estimation, and uniqueness results," *IEEE Trans. PAMI*, vol. 12, no. 8, pp. 735–59, Aug. 1990.

[2] G. Stein and A. Shashua, "Direct estimation of motion and extended scene structure for a moving stereo rig," in *Proc. IEEE CVPR*, 1998.

[3] J. Yi and J. Oh, "Recursive resolving algorithm for multiple stereo and motion matches," *Image and Vision Computing*, vol. 15, no. 3, pp. 181–96, Mar. 1997.

[4] W. Liao and J. K. Aggarwal, "Cooperative matching paradigm for the analysis of stereo image sequences," *Int. J. of Imaging Systems and Technology*, vol. 9, no. 3, pp. 192–200, 1998.

[5] U. R. Dhond and J. K. Aggarwal, "Structure from stereo — a review," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489–1510, 1989.

[6] T. S. Huang and A. N. Netravali, "Motion and structure from feature correspondences: A review," *Proc. IEEE*, vol. 82, no. 2, pp. 252–268, Feb. 1994.

[7] I. J. Cox, "A review of statistical data association techniques for motion correspondence," *Int. J. Computer Vision*, vol. 10, no. 1, pp. 53–66, 1993.

[8] I. J. Cox and S. L. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm ... ," *IEEE Trans. PAMI*, vol. 18, no. 2, pp. 138–50, Feb. 1996.

[9] C. Tomasi and T. Kanade, "Detection and tracking of point features," Tech. Rep. CMU-CS-91-132, Carnegie Mellon University, Apr. 1991.