A STRUCTURE-FROM-MOTION METHOD FOR 3-D RECONSTRUCTION OF MOVING OBJECTS FROM MULTIPLE-VIEW IMAGE SEQUENCES

Ha Vu Le

Department of Electrical and Computer Engineering
Vietnam National University, Hanoi
hvle@hn.vnn.vn

ABSTRACT

Solving the correspondence problem is the most essential task for multiview reconstruction techniques, yet finding unique correspondences between views is impossible at some points, due to such problems as occlusions and ambiguities. We have developed a closed-form solution through constructive geometry for a special case of the structure-from-motion (SfM) problem with four rigidly moving points. This solution allows the 3-D position of a point on a moving object to be computed without having to find the correspondence between its projections on the image planes of multiple views, given its projected 2-D motion vector on an image plane and 3-D information of three other points. With this method we do not have to depend entirely on multiview feature correspondences in reconstructing 3-D objects, hence easing those problems caused by occlusions and ambiguities.

1. INTRODUCTION

Multiview vision is the most commonly-used approach for 3-D structure reconstruction in computer vision applications. To recover the 3-D structure of a scene/object, correspondences between the views must be established. The correspondence problem is ambiguous in general even with various constraints. On the other hand, there may be no correspondences for parts of the views because of occlusions. It is often required to use additional cues in order to resolve the ambiguities or to fill in incomplete parts of the recovered 3-D structures. When reconstructing dynamic scenes from image sequences, motion can be used as a cue for 3-D structure recovery. Although extracting motion components from image sequences also requires finding correspondences between image frames, this correspondence problem is fairly easy to solve comparing to the correspondence problem in multiview vision.

In [1, 2], it was shown that 3-D positions and motions of a number of rigidly moving points can be recovered from a monocular image sequence. The disadvantage of SfM methods is that they are highly non-linear. A linear approach is the Tomasi-Kanade factorization [3], which is based on the orthographic projection. The Tomasi-Kanade factorization method proved to be robust and computationally inexpensive. However, the orthographic projection does not reflect the scaling effect and the position effect, hence limiting the applicability of the method. Some authors have discussed the use of motion in multiview reconstruction. Most of the proposed approaches [4, 5, 6] were to apply multiview and motion constraints to computations of 3-D information without explicitly addressing what multiview-based and motion-based computations could benefit from each other. Some others used motion to help solve the stereo correspondence problem [7, 8]. In this case the motion cue is not directly involved in recovering 3-D data, and therefore finds no use at points for which multiview correspondences do not exist due to occlusions.

We propose a solution for a simplified case of the SfM problem with four rigidly moving points, among them three points already have their 3-D information recovered. We developed this solution for the problem of reconstructing 3-D structures of moving objects from multiview image sequences, in which 3-D positions of some points can not be computed through multiview correspondences while their motions are observable from one of the views. It is obvious that we can apply the equations in [1, 2] to this particular case and the resulted equations will be much less complex than the original ones. However, those simplified equations are still non-linear and solving them is not a trivial task. Our solution is closed-form, obtained through constructive geometry under the perspective imaging model. Due to space limit, we are only able to present a brief description of the solution inn this paper.

2. PROBLEM DEFINITION

The following notations are used. M represents a point in 3-D space, whose position in the camera-centered 3-D coordinate system is represented by the vector $\mathbf{M} = [X\ Y\ Z]^T$, while m denotes the projection of M on the image plane

0-7803-8554-3/04/$20.00 ©2004 IEEE.

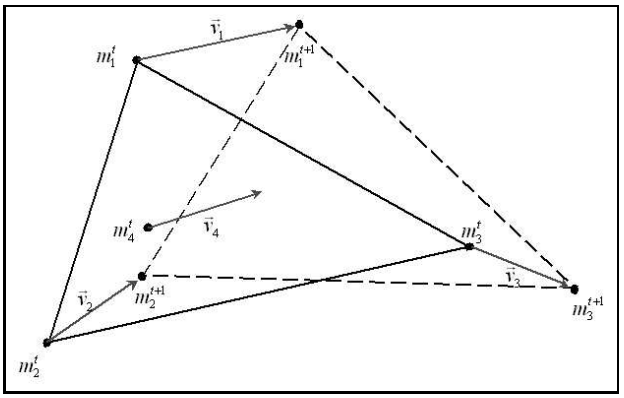

Fig. 1. *Images of four points and their corresponding 2-D motion vectors between two frames t and $t+1$.*

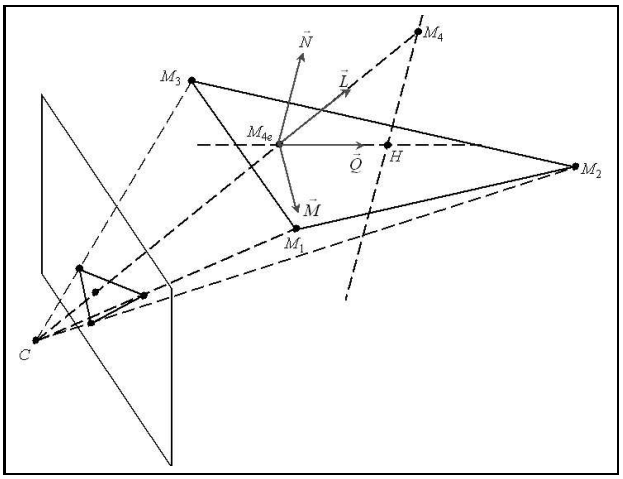

Fig. 2. *Construction of the solution.*

of a camera whose *optical center* is located at point C. The image plane is the plane $Z = -f$ of the 3-D coordinate system which has its origin at C, where f is the focal length of the camera. The 2-D coordination of m in the image coordinate system is represented by the vector $\mathbf{m} = [x\ y]^T$. Let c denote the intersection between axis Z and the image plane, c is then called the *image center*, whose image coordination is (x_0, y_0).

The problem is stated as follows: given three rigidly moving points; M_1, M_2, and M_3; whose positions in the camera-centered 3-D coordinate system have been computed, and their projections in two image frames t and $t+1$; m_1^k, m_2^k, and m_3^k $(k \in \{t, t+1\})$ (Fig. 1); the question is how to compute the 3-D position of a fourth point, M_4, given its image m_4^t and its projected 2-D motion vector $\mathbf{v}_4$.

3. BRIEF DESCRIPTION OF THE SOLUTION

Let M_{4e} be the intersection between the optical ray going through M_4 and the plane defined by three points M_1, M_2, and M_3, and let H be the perpendicular projection of M_4 onto the plane (M_1, M_2, M_3). The line that goes through M_{4e} and H is defined by point M_{4e} and a vector $\mathbf{Q}$, which can be computed from the unit normal vector $\mathbf{N}$ of the plane and the unit normal vector $\mathbf{L}$ of the optical ray going through M_4 and M_{4e}.

Consider the positions of four points moving under a rigid motion at time t and time $t+1$. Superscripts will be added to the symbols used above to indicate the time stamps. Thus, point H^t lies on a line going through M_{4e}^t and having the unit normal vector $\mathbf{Q}^t$. Since the 2-D motion vector $\mathbf{v}_4$ at m_4^t is known, the position of m_4^{t+1} is $\mathbf{m}_4^{t+1} = \mathbf{m}_4^t + \mathbf{v}_4$. Therefore, the above results can also be applied to points at time $t+1$; i.e., point H^{t+1} lies on a line going through M_{4e}^{t+1} and having the unit normal vector $\mathbf{Q}^{t+1}$. Note that H^t and H^{t+1} are a same point with respect to the moving plane defined by three moving points M_1, M_2, and M_3, while in general M_{4e}^t and M_{4e}^{t+1} are different points with respect to that plane, except when the motion of point M_4 between t and $t+1$ is in the direction parallel to the optical ray (C, M_4^t). It means the position of H can be determined by finding the intersection of two lines. Also, the distance h from M_4 to the plane (M_1, M_2, M_3) can be computed from positions of M_{4e} and H, and vector $\mathbf{L}$. That completes the solution.

This solution can not be applied if point M_4 is moving in the direction parallel to the optical ray going through it. A simplified case happens when the optical ray (C, M_4) is perpendicular to the plane (M_1, M_2, M_3). In that case we would not have to go through all the computations to find the position of H because H and M_{4e} are a same point. Given the position of H with respect to M_1, M_2, and M_3, the position of M_4 can be decided whenever the optical ray going through it is not perpendicular to the plane defined by M_1, M_2, and M_3.

Like other SfM methods, this technique is error-proned when the displacements are very small between two consecutive frames. It can be overcome by using the displacements over more than two frames in an image sequence.

4. EXPERIMENTAL RESULTS

In the first experiment a four-camera setup is used. The test object is a polyhedron whose edges are 40mm in length (Fig. 4). The polyhedron was slowly moved and the cameras simultaneously captured its images at each position. Stereo technique is used first to reconstruct the 3-D object at each position. Each stereo pair is composed of two adjacent cameras, thus three stereo pairs are available. The length of each edge of each reconstructed polyhedron at each position from each stereo pair is computed and compared to the actual length, the difference between them is used as

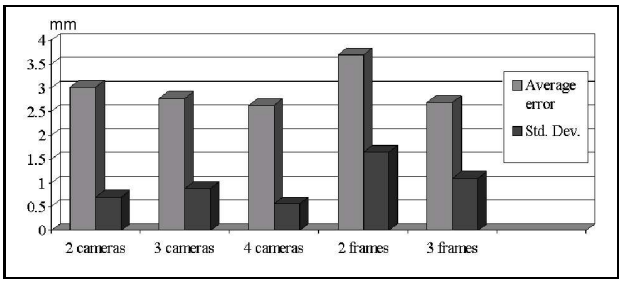

Fig. 3. *Performance of techniques used to reconstruct the polyhedron in 3-D.*

the error measure. The average error and standard deviation of errors for edges of the reconstructed polyhedron at all positions from all stereo pairs are shown in Fig. 3. Because increasing number of cameras is a mean to improve performance of 3-D structure recovery techniques, we also apply three-camera and four-camera reconstructions, using each set of three adjacent cameras, then all four cameras. In this experiment, we want to compare the performance of our SfM technique with the traditional multiview reconstructions. For each stereo-reconstructed polyhedron at each position, three of its vertices are retained and the rest are recomputed using these three 3-D vertices and the 2-D motion vectors between two consecutive frames of a view. Since the displacements between any two consecutive frames are small, errors of this case are high comparing to the multiview reconstructions above. However, when we use three consecutive frames to reconstruct each polyhedron instead of using only two consecutive frames, errors drop sharply and are comparable to the errors of the multiview cases with three and four cameras.

The next experiment is more realistic, in which a toy car was used as test object. Stereo reconstruction is used first for some matched feature points between two views to generate a 3-D surface mesh of the object (Fig. 7), then motion vectors between two consecutive frames of a view are used to refine the 3-D mesh (Fig. 8). Among optical flow techniques available for this step of computing motion vectors between frames, we chose the correlation-based approach, because of its tendency toward accuracy. The tradeoff is high computational cost. For every point of a view whose motion vector has been determined, the four-point solution can be applied to compute its depth: choose the three known 3-D points by picking a triangle of the 3-D mesh whose 2-D projection encloses or is the closest to the point under consideration. The refined 3-D surface mesh of the object shows a significant improvement in details over the stereo-reconstructed surface (Fig. 9). We can even apply this method to objects moving non-rigidly by assuming local rigidity.

5. SUMMARY

We have developed a closed-form solution using constructive geometry for a special case of the SfM problem with four rigidly moving points. The solution allows the 3-D position of a point to be calculated from its projected 2-D motion and the 3-D positions of three other points without having to solve non-linear equations. Experiments with real image sequences shows the performance of this solution is reliable when applied to 3-D structure recovery problems if motions at points of interest in image sequences are significant enough. This SfM solution provides an effective way to use the motion cue in reconstructing 3-D scenes/objects from multiview image sequences.

6. REFERENCES

[1] J. W. Roach and J. K. Aggarwal, "Determining the Movement of Objects from a Sequence of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, pp. 554–562, November 1980.

[2] O. D. Faugeras and S. Maybank, "Motion from Point Matches: Multiplicity of Solutions," *International Journal of Computer Vision*, vol. 4, no. 3, pp. 225–246, 1990.

[3] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: a Factorization Method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, November 1992.

[4] A. M. Waxman and J. H. Duncan, "Binocular Image Flows: Steps Toward Stereo-Motion Fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 11, pp. 715–729, November 1986.

[5] A. Mitiche, "A Computational Approach to the Fusion of Stereo and Kineopsis," in *Motion Understanding: Robot and Human Vision*, W. N. Martin and J. K. Aggarwal, Eds., pp. 81–95. Kluwer Academic Publishers, 1988.

[6] Z. Zhang and O. D. Faugeras, "3D Dynamic Scene Analysis," in *Springer Series in Information Sciences*, T. S. Huang, Ed., vol. 27. Springer-Verlag, 1992.

[7] L. Li and J. H. Duncan, "3-D Translational Motion and Structure from Binocular Image Flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 7, pp. 657–667, July 1993.

[8] P.-K. Ho and R. Chung, "Stereo-Motion with Stereo and Motion in Complement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 2, pp. 215–220, February 2000.

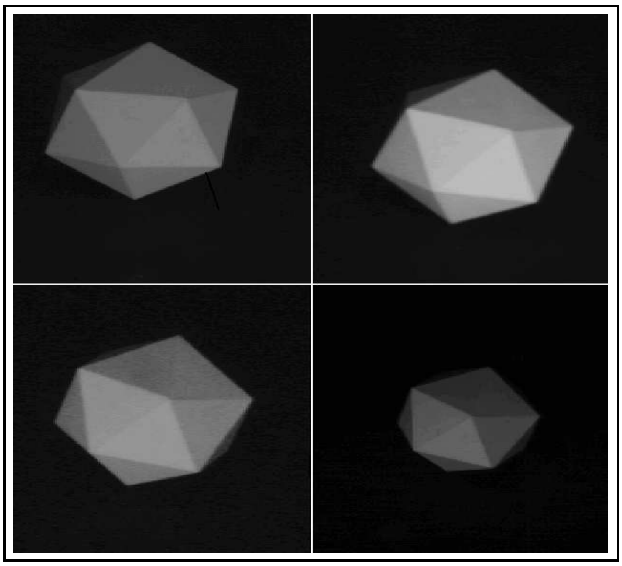

Fig. 4. *Images of the polyhedron captured by four cameras.*

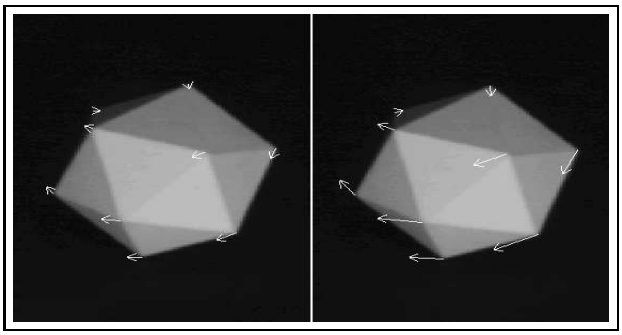

Fig. 5. *Motion vectors between the first frame and the last frame of a two-frame sequence (left) and of a three-frame sequence (right) of a view.*

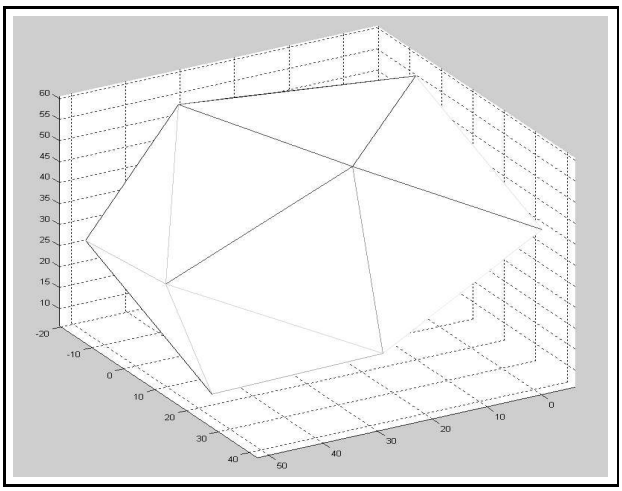

Fig. 6. *A reconstructed polyhedron shown in 3-D space.*

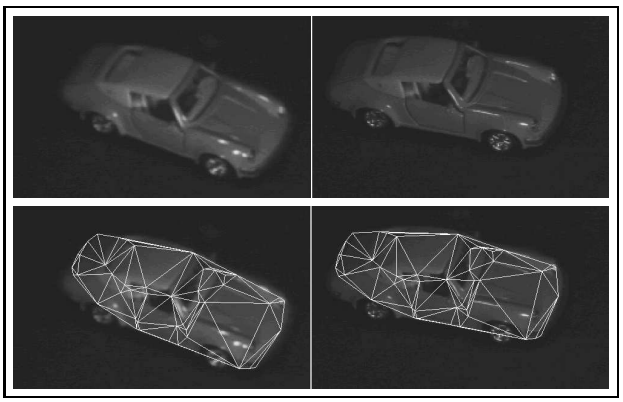

Fig. 7. *Images from two views of a stereo pair (top) and the triangular meshes whose vertices are corresponding feature points in two views (bottom).*

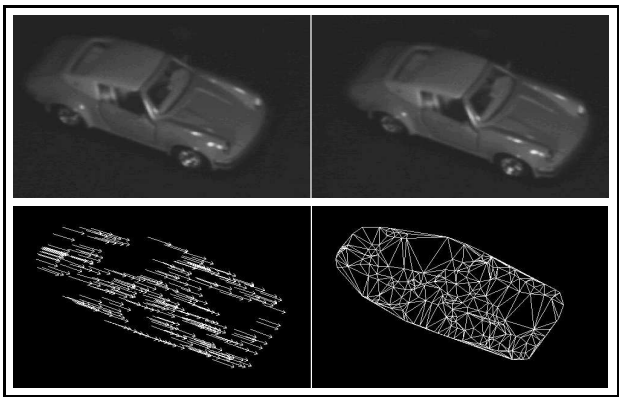

Fig. 8. *Top: two consecutive frames of a view. Bottom: motion vectors between two given frames (left) and the refined mesh for the first frame (right).*

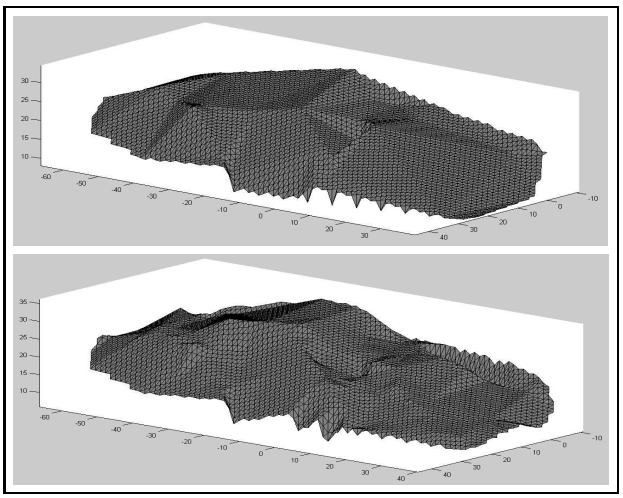

Fig. 9. *The reconstructed 3-D surface from stereo images (top) and the 3-D surface after being refined by applying the SfM solution (bottom).*