STRUCTURE FROM MOTION IN DYNAMIC SCENES WITH MULTIPLE MOTIONS

Engin Tola, Sebastian Knorr†, Evren İmre, A. Aydın Alatan, and Thomas Sikora†

Department of Electrical & Electronics Engineering, Middle East Technical University
Balgat, 06531 Ankara, Turkey
E-mail: {etola, alatan}@eee.metu.edu.tr, eimre@metu.edu.tr

† Department of Communication Systems, Technical University of Berlin
Einsteinufer 17, Berlin, Germany
E-mail: {knorr, sikora}@nue.tu-berlin.de

ABSTRACT

In this study, an algorithm is proposed to solve the multi-body structure from motion (SfM) problem for the single camera case. The algorithm uses the epipolar criterion to segment the features belonging to independently moving objects. Once the features are segmented, corresponding objects are reconstructed individually by applying a sequential algorithm, which uses the previous structure to estimate the pose of the current frame. A tracker is utilized to increase the baseline and improve the F-matrix estimation, which is beneficial for both segmentation and 3D structure estimation. The experimental results on synthetic and real data demonstrate that our approach efficiently deals with the multi-body SfM problem.

1. INTRODUCTION

Structure from motion in static scenes is an extensively studied problem with some well established solutions [6]. However, these solutions are not capable of dealing with dynamic scenes with many moving objects, which are often encountered in practice. A common strategy is to eliminate the dynamic elements of the scene, and perform an incomplete reconstruction. Inclusion of the dynamic elements into the reconstruction requires solving the multi-body SfM problem which can be stated as follows:

Given a set of frames with feature points, estimate the locations of the feature points in 3D world coordinates, camera motion and the motion of the independently moving objects (IMO).

The literature on the multi-body SfM problem is shaped by the observation that, when the feature set is segmented into partitions corresponding to the background and the individual objects, the problem can be decomposed into several static SfM problems. Hence, the literature is shaped by a *divide-and-conquer approach*. As some mature tools are available for the latter, first part of the problem i.e., segmentation of motions, is in the focus of most studies.

The solution techniques for the multi-body SfM problem can be examined in 4 categories. Optical flow based methods [1][7][8] assume a scene composed of planes of varying depths. In this case, a simple clustering of the optical flow values is sufficient to achieve the desired segmentation. Another set of methods use the *affinity matrix* [12], a structure which contains information about the similarity among the features. Eigen decomposition of the matrix reveals the identity of the features [14]. Statistical techniques also have a niche in this field. In [10], *sequential importance sampling* is used to construct the conditional probability density function (pdf) of the structure, motion and segmentation, given the features. Once the pdf is computed, it is possible to estimate the structure, motion and the segmentation optimally in ML sense.

Finally, it is possible to exploit the constraints derived from the epipolar geometry and the rigid body motion assumptions. The most common approach is to estimate the individual F-matrices for each motion, and to use the epipolar constraint for the classification [4][13][15]. However, in [12], different geometric constraints are available, and both the partitions and the models for each partition are determined after utilizing a Bayesian criterion. Yet another technique is presented in [3], which exploits the rank constraint on a matrix composed of feature trajectories, imposed by the rigid body motion assumption.

In this paper, both the segmentation and the reconstruction aspects of the multi-body SfM problem are studied. Segmentation by using the epipolar constraint is straightforward and is adopted into our algorithm. The 3D

reconstruction uses a two-frame triangulation. Both techniques enjoy a better reliability for the large baseline case. However, the solution of the correspondence problem is better facilitated by a small baseline. It is observed that the use of a tracker reduces the need for a compromise, providing a satisfactory solution to the correspondence problem, while providing a larger baseline.

The organization of this paper is as follows: In the next section, the proposed solution is outlined. In Section 3 and 4, segmentation and reconstruction stages are described, respectively. The experimental results are presented in Section 5. Finally, in Section 6, the paper is concluded by a discussion of results and the future work.

2. PROPOSED SOLUTION

In order to facilitate the solution, the following assumptions are made:

- **Motion:** There are two different motion types: camera and IMOs. These motions are assumed to be non-degenerate, slow and generally in one direction, such that the baseline is sufficiently large between the first and the last frame.
- **Scene:** The scene is assumed to have a 3D structure, and no significant changes in the environmental conditions, such as light, are allowed. Also, the scene should fulfill the rigid body assumption.
- **Camera:** Internal camera parameters are assumed to be known and constant throughout the sequence

The outline of our proposed solution is as follows: The first processing step is the construction of the trajectories by using a pyramidal Lucas-Kanade tracker [2]. This step effectively solves the correspondence problem for the large baseline case. As a next step, the F-matrix construction and trajectory segmentation is performed by the help of a robust algorithm. Finally, the 3D reconstructions of the independently moving object and the background are performed by the technique described in [16][9]. Finally, for clarity of presentation, the reconstructions are merged manually.

3. SEGMENTATION

Trajectory segmentation is handled by geometric means. For each independent motion in the sequence, there exists a corresponding F-matrix, $\mathbf{F}_i$, which fulfills the epipolar constraint

$$\mathbf{x}_1^T \mathbf{F}_i \mathbf{x}_2 = 0\,, \qquad (1)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are corresponding points in two views. A RANSAC-based F-matrix estimation algorithm identifies the feature pairs belonging to the dominant motion and

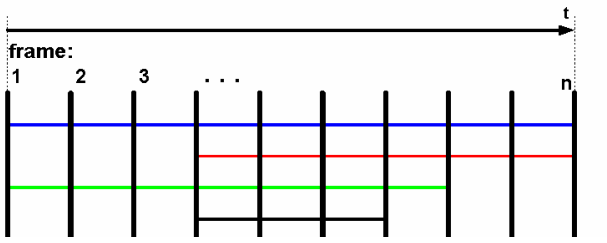


Figure 1. Trajectory classification: complete trajectory (blue), incomplete right (red), incomplete left (green), incomplete (black)

labels the rest as outliers. If the same procedure is repeated with the outliers, some of the outliers should satisfy the epipolar constraint according to this new F-matrix, which corresponds to the motion of the IMO. Hence, upon successive iteration of the procedure for all frames, the feature trajectories can be classified, either as background or IMO.

Figure 1 illustrates 4 different kinds of trajectories:

1. Complete trajectories, which are visible in all frames of the sequence,
2. Incomplete right trajectories, which appear within the sequence and continue to the last frame,
3. Incomplete left trajectories, which appear in the first frame and end within the sequence
4. Incomplete trajectories, which appear and end within the sequence

In the proposed approach, only the most reliable ones, i.e. the first three classes of trajectories, are utilized. Incomplete trajectories and trajectories, which are labeled as outliers after RANSAC and re-RANSAC, are removed.

Algorithm 1: Trajectory segmentation algorithm

1. Compute F-matrix corresponding to the first and the last frame of the sequence by using a RANSAC-based procedure and label the inliers as background trajectories.
2. Compute F-matrix on the outliers of step 1 by using again RANSAC algorithm and label the inliers as IMO trajectories.
3. Proceed one frame backwards according to the last frame. Estimate F-matrix between first and current frame for each motion using the labeled trajectories and classify new trajectories, i.e. incomplete left ones.
4. Proceed one frame forward according to the first frame. Estimate F-matrix between current and last frame using again the labeled trajectories and classify new trajectories, i.e. incomplete right ones.
5. Repeat Step 3 and 4 for all frames as long as the baseline is large enough.

4. RECONSTRUCTION

The *initial* 3D reconstruction for each object motion is achieved by a two-view reconstruction algorithm. The segmented trajectory information is used to form a correspondence set. Since the calibration information is assumed to be known, the F-matrix is decomposed into rotation and translation parameters by using the algorithm given in [11]. Following this step, the projection matrix, **P**, is formed and the initial structure is computed using polynomial triangulation proposed in [6].

Final stage is the update of the initial structure. This is accomplished by adding the features detected in the additional frames of the sequence to the initial structure sequentially [16][9]. For a complete reconstruction of the scene, this procedure is repeated for features belonging to each motion.

The pose of the new frame with respect to the current structure is obtained by utilizing the correspondences of the new view with a previously inserted view and the structure points. First, the epipolar geometry between the new view and a previously inserted view is obtained. As the next step, 2-D points, whose 3-D structure points are already calculated, are selected from the obtained correspondence set. Later, from this set of 3-D - 2-D points the P-matrix of this new frame is calculated by using a robust algorithm, similar to the one used in the computation of the F-matrix. Finally, new structure points are initialized by using the remaining points that do not have a 3-D point associated to them by using the calculated P-Matrix. The overall reconstruction is then refined using a global bundle adjustment algorithm [17].

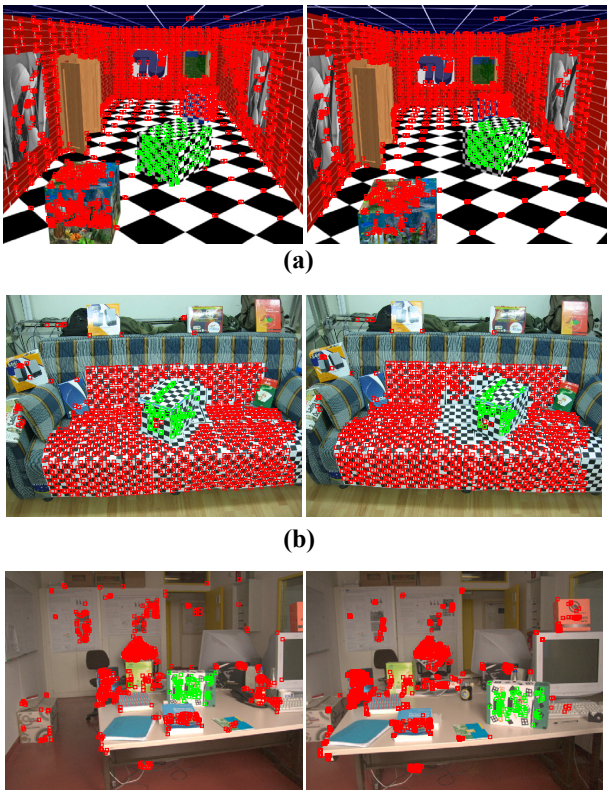
5. EXPERIMENTAL RESULTS

The algorithm is tested on both synthetic and real data. For the real data, the camera is calibrated manually.

In Figure 2, segmentation results are presented. The segmentation step generally functions well, despite the existence of occasional erroneous classifications. However, the presence of outliers does not pose a serious problem as they are removed in the 3D reconstruction step. It is observed that for a successful segmentation, a large baseline and reasonable number of inliers are essential.

Figure 3 depicts the reconstruction results for the "TUB-Room"-sequence and the "sofa"-sequence. While the background reconstruction is usually accurate, the IMO often contains an insufficient number of features.

Due to sparse features on the walls, the reconstructed background of the "desk"-sequence is not shown. Moreover, the reconstruction of the IMO would fail because all features are located on a planar surface.



(a)

(b)

(c)

Figure 2: Segmentation results. Red squares indicate the background features, and the green ones the IMO **a.** TUB-Room **b.** Sofa **c.** Desk

6. CONCLUSION AND FUTURE WORK

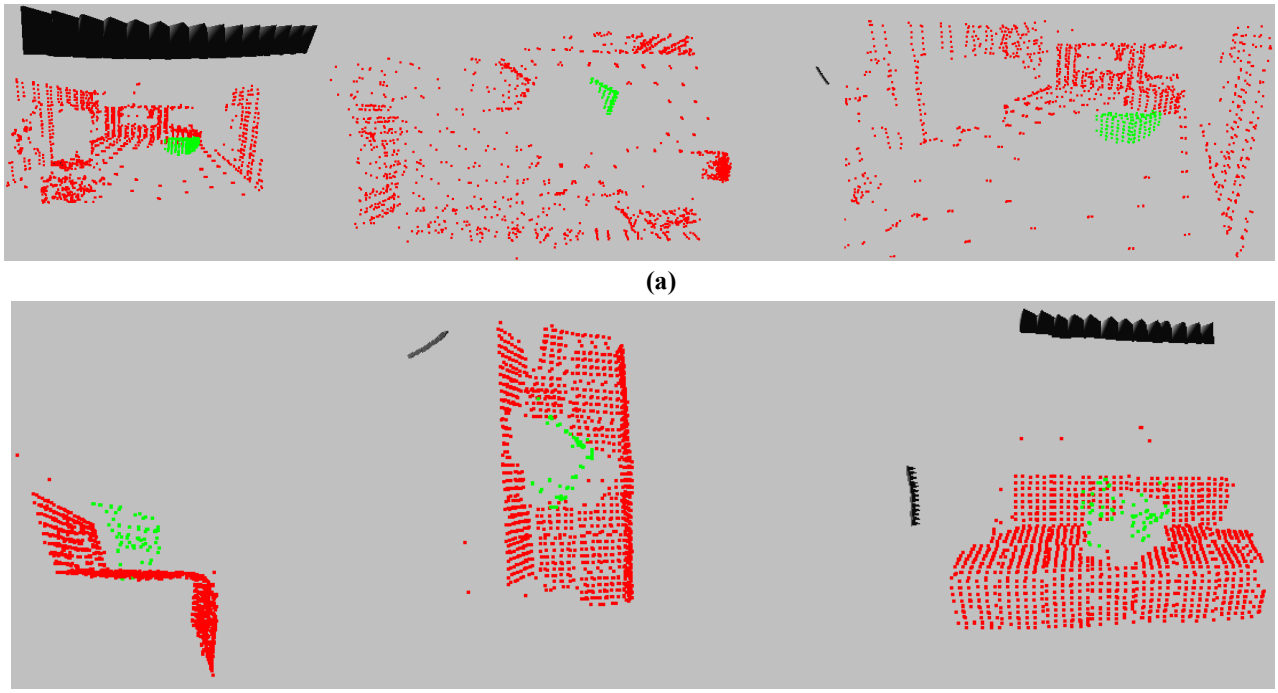
In this paper, an algorithm for the reconstruction of dynamic scenes is proposed. The algorithm utilizes the epipolar constraint to partition the feature set into independent motions. The F-matrix estimate is improved by increasing the baseline through the use of a tracker.

The experiments indicate that the proposed solution has an accurate performance, given enough features. For the background, this is usually not a significant problem. However, in practice, IMOs are often significantly smaller than the background, hence, contain less features.

In order to increase the number of features guided matching along the epipolar lines can be incorporated. This idea can be taken a step further by applying dense matching techniques.

Extending the algorithm to handle the multiple moving objects case is trivial. The set of outliers can be searched for more F-matrices, as long as the estimated F-matrices are reliable, or enough features are left in the outlier set.

Finally, the algorithm can be enhanced with the ability to select the key frames adaptively, to improve the 3D reconstruction and F-matrix estimation.



(a)

(b)

Figure 3: 3D reconstruction results. Red dots indicate the background features, and the green ones the IMO **a.** TUB -Room **b.** Sofa

7. ACKNOWLEDGEMENT

The work presented was developed within 3DTV, a European Network of Excellence (http://www.3dtv-research.org), funded under the European Commission IST FP6 programme.

8. REFERENCES

[1] E.H. Adelson, "Layered Representation for Vision and Video", Proc. of the IEEE WS on Representation of Visual Scenes, Cambirdge, MA, pp. 3, June 1995

[2] J.-Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker", Intel Corp., Microprocessor Research Labs, 2000, http://www.intel.com/research/mrl/research/opencv/

[3] J.P. Costeira, T. Kanade, "A Multibody Factorization Method for Independently Moving Objects", IJCV' 98

[4] W. Fitzgibbon, A. Zisserman, "Multibody Structure and Motion: 3D Reconstruction of Independently Moving Objects", ECCV 2000

[5] M.A. Fischler and R.C. Bolles "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, Volume 24 Number 6, June 1981

[6] R. Hartley, A. Zisserman, Multiple view geometry, Cambridge University Press, UK, 2003

[7] M. Irani, P. Anandan, "A Unified Approach to Moving Object Detection in 2D and 3D Scenes", IEEE Trans. on PAMI Vol. 20, Issue 6, pp. 577-589, June 1998

[8] M. Lourakis, A. Argyros, S. Orphanoudakis, "Independent 3D motion detection using residual parallax normal flow fields", Proceedings of ICCV98, Bombay, 1998

[9] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, "Visual modeling with a hand-held camera", IJCV 59(3), 207-232, 2004

[10] G. Qian, R. Chellappa, Q. Zheng, "Bayesian Algorithms for Simultaneous Structure from Motion Estimation of Multiple Independently Moving Objects", IEEE Trans. on IP, Vol. 14, Issue 1, Jan 2005

[11] A. M., Tekalp, Digital Video Processing, Prentice Hall, 1995

[12] P.H S. Torr, "Geometric Motion Segmentation and Model Selection", Phil. Trans. Royal Society of London A 356, 1740,1321–1340, 1998

[13] R. Vidal, S. Soatto, Y. Ma, S. Sastry, "Two-view Multibody Structure from Motion", IJCV 2002

[14] Y. Weiss, "Segmentation Using Eigenvectors: A Unifying View", Proceedings of ICCV99, pp. 975-982, 1999

[15] L. Wolf, A. Shashua, "Two-body segmentation from two perspective views", Proc. of CVPR 2001, Vol. 1, 2001

[16] E. Tola, "Multiview 3D Reconstruction of a scene containing independently moving objects", MS Thesis, Middle East Technical University Library, Turkey, 2005

[17] M.I.A. Lourakis and A. A. Argyros, "The Design and Implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm" ,http://www.ics.forth.gr/~lourakis/sba/, 2004