Moving Camera Registration for Multiple Camera Setups in Dynamic Scenes

Evren Imre
h.imre@surrey.ac.uk
Jean-Yves Guillemaut
j.guillemaut@surrey.ac.uk
Adrian Hilton
a.hilton@surrey.ac.uk

Center for Vision, Speech and Signal Processing
University of Surrey
Guildford, UK

Problem Definition

This paper describes a method to register a moving (principal) camera, given a set of fully calibrated static cameras (witnesses) viewing a dynamic scene, a common scenario in broadcasting and film production. Our ultimate aim is to equip the existing free-viewpoint video algorithms with the ability to exploit any available moving cameras in generic dynamic scenes, and to facilitate 3D content production by augmented reality and stereoscopic rendering.

Proposed Method

Overview: The algorithm utilizes the witness cameras to build a 3D reference structure, by solving a multiview triangulation problem, and then, computes a pose measurement for the principal camera at each time instant with respect to this structure via a P3P solver. These measurements are filtered by an unscented Kalman filter (Figure 1).

Building a reference model: In order to solve the multiview triangulation problem, first, wide baseline guided matching [3] is performed on SIFT features [5] for all available image pairs. Then, the correspondences belonging to the same 3D scene feature are clustered together. In each cluster, the correspondences are triangulated [3] to obtain a set of 3D measurements for the actual position of the associated scene feature. The measurement covariance is estimated by the unscented transformation (UT) [4]. The measurements are fused into a position estimate by using a Kalman filter. The resulting scene feature is characterized by its position, position covariance, and a collection of 2D SIFT descriptors belonging to the members of the cluster.

Computing the pose measurements: The P3P problem is solved by applying the Finsterwalder's method [2] on the set of 3D-2D correspondences between the principal camera, and the reference structure.The similarity metric employed to assess a possible 3D-2D correspondence is the maximum similarity score between the descriptors associated with 3D scene feature, and that of the 2D image feature. The covariance of the pose estimate is obtained via the UT.

Jitter removal: In order to remove any jitter in the instantaneous pose measurements, and to exploit the temporal coherence of the camera motion, an unscented Kalman filter (UKF) [4] is utilized. The UKF employs a constant translational and rotational velocity model.

Results and Conclusion

The behaviour of the algorithm is studied on two dynamic indoor sequences, with a 7-witness, 1-principal camera setup. The recovered position trajectory for the sequence *Ball* is illustrated in Figure 2. In order to assess the quality of the estimated pose trajectory, the principal camera is then used in a number of applications, all of which require an accurate pose estimate (Figure 3). The main conclusions are:

- The performance of the algorithm is essentially independent of the number of witness cameras, as long as a reasonable overlap with the field-of-view with the principal camera is maintained.
- The algorithm is remarkably robust to occlusions by the dynamic elements in the scene (up to 50-60%).
- The estimated pose trajectory is sufficiently accurate for the intended applications (Figure 3 and the supplementary material).

The main limitation of the algorithm is the known internal calibration assumption, which can be alleviated by employing a P4P solver.

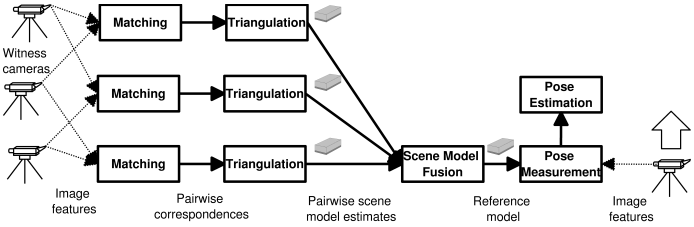

Figure 1: Overview of the proposed algorithm, for a 3 witness-1 principal camera setup.

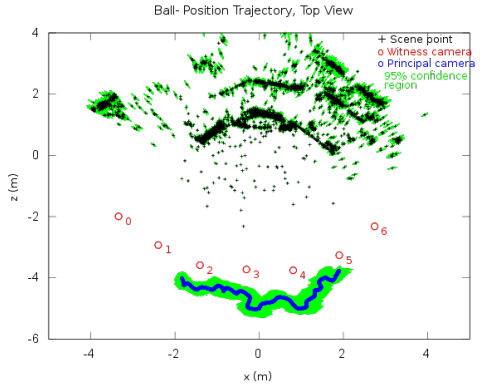

Figure 2: The reference model, the recovered camera position trajectory, and the confidence regions, as estimated by KF and UKF, respectively

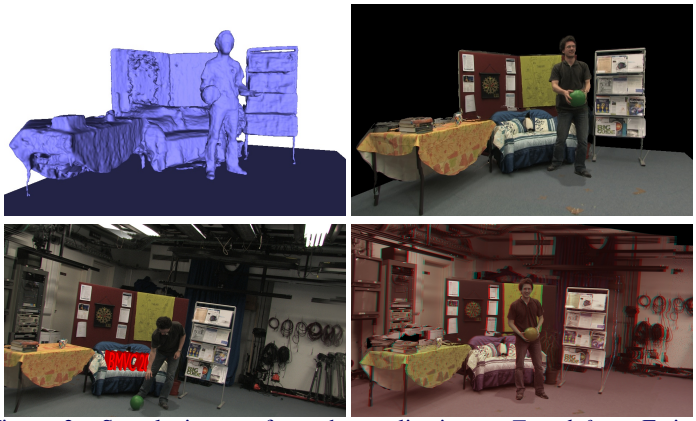

Figure 3: Sample images from the applications. *Top, left:* Estimated scene model [1]. *Top, right:* Free-viewpoint video. *Bottom, left:* Scene augmentation. *Bottom, right:* Stereoscopic rendering, in red/cyan anaglyph format.

[1] J.-Y. Guillemaut, J. Kilner, and A. Hilton. Robust graph-cut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes. In *Proc. ICCV*, 2009.

[2] R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nolle. Analysis and the solutions of the three point perspective pose estimation problem. In *Proc. CVPR*, pages 592–598, 1991.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. 2nd edition, 2003.

[4] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proc. IEEE*, 92(3):401–422, March 2004.

[5] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.