Monocular SLAM for User Viewpoint Tracking in Virtual Reality

Raúl Mur-Artal and Juan D. Tardós

Abstract— **One challenge in Virtual Reality (VR) is to track the user viewpoint so that the virtual world is projected into the user screen consistently with its current position and orientation. This task can be accomplished using just one camera, e.g. attached to a VR headset, using Simultaneous Localisation and Mapping (SLAM) techniques. We have recently presented ORB-SLAM which processing in real-time a video sequence is able to create a map of the environment that is used for camera localisation. The system works in indoor, outdoor, small and large environments in real-time, even in the presence of some dynamic elements, and being able to close loops and relocalise the camera from different viewpoints. Thanks to the exceptional camera localisation capabilities of the system it provides a good solution for user viewpoint tracking for VR applications.**

I. FEATURE-BASED MONOCULAR SLAM

Based on previous excellent works [2], [6], [5], [1] we have developed ORB-SLAM [3], an open-source[1] monocular SLAM system that efficiently uses the same features for all system tasks, which run in parallel threads: tracking, local mapping and loop closing. The current implementation uses ORB features [4], which allows to perform tracking at 30 fps, and detect loops and relocalise the camera in real-time with good invariance to viewpoint and illumination.

The tracking localises the camera with every frame, being able to perform global relocalisation based on bags of words if the tracking is lost (e.g. due to occlusions or abrupt movements). Tracking find matches between the ORB in the frame and map points in a local map and optimizes the camera pose minimizing the reprojection error. The local map is retrieved from a covisibility graph of keyframes and allows to operate in large scale making the computational complexity independent of global map size.

The local mapping processes new keyframes, triangulates new points and performs local bundle adjustment to optimize the map minimizing the reprojection error. We use a *survival of the fittest* strategy, with rapid keyframe spawning, culling later the redundant keyframes. This enhances tracking robustness, especially under fast exploration motions, while the map is maintained compact, enhancing lifelong operation.

The loop closing detects loops and performs a global pose graph optimization to correct the translation, rotation and scale drift enforcing the global consistency of the map.

We evaluated this system in more than 20 public datasets achieving an accuracy in camera localisation typically below the 1% of the map dimensions, being clearly more robust and accurate than other state-of-the-art monocular systems.

II. APPLICATION TO VIRTUAL REALITY

Using the camera localisation capabilities of our monocular SLAM it is possible to track the viewpoint of a user in a virtual world, for example attaching a camera to a VR headset. The pipeline would consist in two steps:

Workspace map creation: Firstly the user creates a map of the environment where the user will run the VR application. The user will move the camera around the environment while ORB-SLAM processes the sequence in real-time. An interface shows if the tracking is working successfully and a view of the keyframes and points of the reconstruction. Once the user is satisfied with the result, the system can run a full bundle adjustment to optimize the whole map achieving the maximum accuracy.

Viewpoint tracking: During the VR operation, the mapping and loop closing threads can be disabled. When the VR application starts, ORB-SLAM will localise the user in the map, and continue tracking his *position and orientation*. If at some point the tracking fails, it will try to relocalise the user in the map.

Compared with inertial sensors or visual odometry, our approach provides *zero-drift tracking* as the estimation is not incremental, but based on the map. The typical error for a room-size environment is around 1-3cm. Noteworthy, it is possible to create the map with one camera, and localise in it a different one or even track several users in the same map. The only requirement to achieve good performance is a reasonably lit room without large untextured areas such as big plain walls. We plan to improve the agility and robustness of the tracking using information from an Inertial Measurement Unit (IMU).

This work was supported by the Dirección General de Investigación of Spain under Project DPI2012-32168 and the Ministerio de Educación Scholarship FPU13/04175.

The authors are with the Insitituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, María de Luna 1, 50018 Zaragoza, Spain. (e-mail: {raulmur,tardos}@unizar.es)

[1] http://webdiis.unizar.es/~raulmur/orbslam

REFERENCES

[1] D. Gálvez-López and J. D. Tardós. Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, 28(5):1188–1197, 2012.

[2] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007.

[3] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics (Under Review), ArXiv preprint: arXiv:1502.00956*, 2015.

[4] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: an efficient alternative to SIFT or SURF. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[5] H. Strasdat, A. J. Davison, J. M. M. Montiel, and K. Konolige. Double window optimisation for constant time visual SLAM. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[6] H. Strasdat, J. M. M. Montiel, and A. J. Davison. Scale drift-aware large scale monocular SLAM. In *Robotics: Science and Systems (RSS)*, 2010.