Decoupled Mapping and Localization for Augmented Reality on a Mobile Phone

Pierre Martin
Orange – IRISA, Inria

Eric Marchand
Université de Rennes 1, IRISA, Inria

Pascal Houlier
Orange

Isabelle Marchal
Orange

ABSTRACT

Using Simultaneous Localization And Mapping (SLAM) methods become more and more common in Augmented Reality (AR). To achieve real-time requirement and to cope with scale factor and the lack of absolute positioning issue, we propose to decouple the localization and the mapping step. We explain the benefits of this approach and how a SLAM strategy can still be used in a way that is meaningful for the end user.

Keywords: Augmented reality, simultaneous location and mapping, mobile phone

Index Terms:

1 INTRODUCTION

Simultaneous Localization And Mapping (SLAM) approaches are now more and more common in Augmented Reality (AR) applications. Since PTAM [1], which demonstrated the feasibility of a deterministic SLAM system for augmented reality on a PC, a few industrial frameworks relying on such methods (from *13th Lab* or *Metaio* for instance) have emerged. After a dedicated initialization protocol, they propose a way for the user to automatically reconstruct and track the environment and define a plane where augmented objects can be displayed.

For the typical end-user, two problems emerge with such scenario. Due to the lack of absolute localization, the first issue is the difficulty to propose context aware augmentation. Although it is possible to detect known objects during the tracking [2] and display information around them, it is nearly impossible to closely register and augment the whole space of a scene, like a room or a hallway. It then becomes obvious that complete map acquisition is required prior to decide where we want the augmentations to be placed, so that they are meaningful. The second problem lies in the ergonomic side of the application. Typically, an unbriefed end-user should not have to follow a complex protocol to be able to localize himself in the scene. The system then requires a quick and very robust re-localization in a known map without any prior information on the pose.

Even if it has been shown in [3] that is possible to tweak the original algorithm to use PTAM [1] on a mobile phone, its performances are greatly diminished. With the recent evolution in mobile hardware (multi-core CPU), it is now possible to run the original algorithm on such platforms, even if the provided cameras are not as good as the one used in [1]. Nevertheless, a clear decoupling of the mapping and tracking step is still relevant to save computational power for the end-user application.

We here show our approach of the decoupling and explicit our first attempt to improve the end-user experience during the localization of the system in a pre-learnt and thus known map.

e-mails: pierremartin35@gmail.com, marchand@irisa.fr, pascal.houlier@orange.com, isabelle.marchal@orange.com

2 APPLICATION SUITE

The proposed system takes the form of an application suite, where three roles intervene: the map-maker, the designer and the end-user. To explicit further each role, we take the example of the augmentation of a store front (Fig. 1). One basic application idea is to allow a user/consumer to interact with a store even when it is closed.

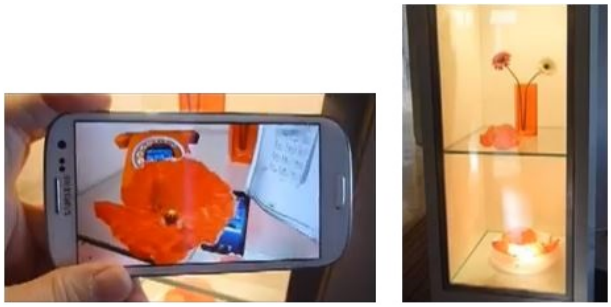

Figure 1: Augmentation of a store front. On the left the live representation of the augmentations (tablet, orange phone, white phone). On the right the real environment (see also video).

2.1 Mapping

This is the application used by the map-maker. Its role is to produce the map, a cloud of 3D points, which will be used as a reference frame for the augmentation design and for the localization.

A two frames initialization step is required to initialize the SLAM algorithm and localize the camera relatively to the first frame. The camera is then tracked while the map-maker tries to move in the target environment to acquire as much visual features as possible. A SLAM algorithm (similar to [1]) allows the map-maker to have a visual feedback while he extends the recognizable environment for the user. During this discovery step, a set of keyframes are automatically stored with their known poses to optimize the map by bundle adjustment and to provide a set of possible references for the localization. At this point, the map-maker has the possibility to insert keyframes manually to further increase the coverage of reference images.

As we will see in Section 3, at the end of this process, a more computational heavy task can occur offline, which is the computation and optimization of descriptors.

2.2 Augmentation Design

The augmentation designer has to place each augmentation (virtual objects) in the reference frame of the whole map. This allows for augmentations anywhere in the scene where a part of the map can be observed and tracked, which is more comfortable than just around a set of fixed patches.

To handle occlusions and real object interactions, a 3D model of the scene can be acquired by a multiple views structure from motion, and is registered with the point cloud produced in the Mapping step. From there, the task of the designer is quite easy. He has to put the virtual object anywhere he wants in regard of the textured 3D model of the scene. This step can be done offline with a modeling tool or online using the tracker to place augmentations directly in the real environment with the mobile phone.

2.3 Localization

This app is used by the end-user. It tracks the map, relocalizes itself (pose computation), from the stored keyframes, automatic or manual, and possibly with the help of previously computed descriptors, it obviously also displays the augmentations.

In this app, all map discovery and optimization features (such as bundle adjustment) are disabled, reducing the amount of computation needed.

Exceptionally, the system can still behave as a real SLAM system for short periods of time to allow the user to move a little bit outside of the designed environment but is not intended to create a large map and forgets rapidly new features.

3 Improving Re-localization

During previous experiments, it was established that the re-localization of the end-user from an unknown pose, should work anywhere and from any point of view in the augmented scene, in less than a few seconds.

If we look at the real space where the user would like to be able to locate itself, we can roughly evaluate the performance of the proposed re-localization algorithm, taking only a few significant orientations of the camera, see Fig. 2.

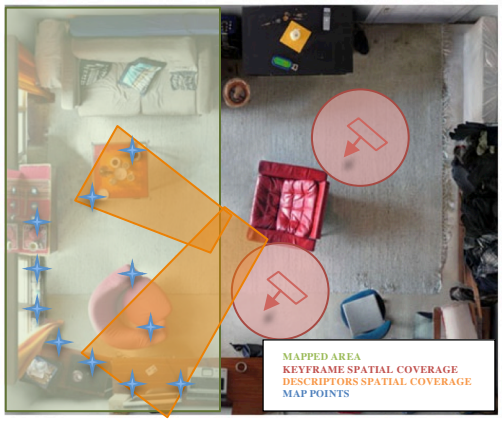

Figure 2: An illustration of a 2d spatial re-localization coverage where the user can move in the whole room.

We first consider an image-based re-localization where we use the position of a keyframe, as the starting pose of the camera for the tracker. It is done by a ZMSSD on a subsample of the real image, with their orientation aligned along the Z axis. This method allows for real-time re-localization but has a pretty small spatial coverage around the reference keyframe. We can clearly see that although we have the capability to manually add keyframes, it is not a viable solution for a large 3D scene.

Secondly, to increase the coverage, we decided to compute FREAK descriptors [4] on each map point and in each keyframe during the mapping step. They are then matched with descriptors on the current frame computed at FAST corners [5], which have a Shi-Tomasi score beyond a threshold on the two lowest levels of the image pyramid. A RANSAC-based POSIT then determines the new pose of the tracker.

The first method is fast but unreliable most of the time, while the second is more robust but significantly slower. We decided to use both at the same time, running them concurrently in two threads. We chose a strategy consisting of using the image-based re-localization while the descriptor-based is still running. If we are still lost when the second one finishes, we use its result instead.

4 Results

The system has been tested on both Android and iOS. The mobile phones were Samsung Galaxy S2, Samsung Galaxy S3 and iPhone 4S. On both the Galaxy S3 and the iPhone 4S (which are not current high-end mobile phone) the tracking and image-based re-localization are done in real-time for an image of sizes, respectively 320x240 and 480x360. A descriptor based re-localization lasts between 800ms and 1500ms when it's a success. The framerate on the Samsung GS2 is near real-time, and can only really be used with the image-based re-localization. The descriptor-based re-localization increases the spatial coverage by a factor three.

Figure 3: System in action in various environments (see also on http://youtu.be/hnmqMc3hZgM)

5 Conclusion and Future Work

We have established various benefits of decoupling localization and mapping for augmented reality. It is meaningful for performance and optimization, and is mandatory when we want to augment contextually the environment.

For future work, we plan to do further optimizations on the mapping step to allow for a better re-localization and a quicker tracking, such as making statistics on the value of the information contained in each version of a map point descriptor.

References

[1] G. Klein and D. Murray. Parallel Tracking and Mapping for Small AR Workspaces. *Proc. IEEE/ACM ISMAR'07*, November 2007.

[2] C.-C. Wang, C. Thorpe and S. Thrun. Online Simultaneous Localization And Mapping with Detection And Tracking of Moving Objects: Theory and Results from a Ground Vehicle in Crowded Urban Areas. In *IEEE ICRA*, September 2003.

[3] G. Klein and D. Murray. Parallel Tracking and Mapping on a camera phone. In IEEE *ISMAR*, October 2009.

[4] A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. In IEEE CVPR, june 2012.

[5] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In *IEEE ICCV,* October 2005.