Simultaneous Localization and Mapping for Augmented Reality

Gerhard Reitmayr, Tobias Langlotz, Daniel Wagner, Alessandro Mulloni, Gerhard Schall, Dieter Schmalstieg and Qi Pan
Graz University of Technology
Graz, Austria
lastname@icg.tugraz.at

***Abstract*—Recently, the methods of Simultaneous Localization and Mapping (SLAM) have received great interest in the field of Augmented Reality. Accurate tracking in unknown and new environments promises to reduce the initial costs of building AR systems which often require extensive and accurate models of the environments, interaction objects and virtual annotations. However, it is still an open question how interesting and useful annotations can be created, attached and stored for unknown and arbitrary locations. In this paper, we discuss possible uses of SLAM in the different components of typical AR systems to provide meaningful applications and go beyond current limitations.**

***Keywords*-Augmented Reality, Simultaneous localization and mapping, Tracking, Interaction, Panorama**

I. Introduction

Augmented Reality systems deal with two fundamental technical challenges. In order to provide accurate registration of augmented visuals over the real world, two items of information must be known: The current view of the real world that needs to be augmented; and the virtual object geometry and its accurate registration with the real world. The former problem is usually referred to as the tracking problem [1] and can be approached in many different ways. Current work focuses on using cameras built into the AR systems to provide self-contained systems that provide fast and accurate pose estimation.

The later problem is often referred to as the authoring problem. Both tracking and advanced visualization in augmented reality require a good understanding of both the virtual and real parts of a scene. To avoid visual artifacts due to color, texture and saliency of the video background, good knowledge about the scene is necessary to help choosing a good visualization technique. If a sufficiently accurate 3D model of the real scene is available, correct occlusions, impostors for explosion views and visual effects for X-ray visualization can be rendered. Furthermore, interaction with the environment requires models for ray-picking or constrained interaction, such as snapping to object surfaces. Usually, models of the environment and the virtual annotations are prepared in advanced and only used passively at runtime.

Simultaneous localization and mapping (SLAM) has received much attention in the Augmented reality community in the last years. SLAM refers to a set of methods to solve the pose estimation and 3D reconstruction problem simultaneously while a system is moving through the environment. Initial work by Davison et al. [2], [3] demonstrated that a system using a single camera is able to build a 3D model of it's environment while also tracking the camera pose. Their system provided accurate and fast visual tracking of a handheld or wearable camera in an unknown environment. Rapid development culminated in the work of Klein and Murray [4] which demonstrated superior robustness and the ability to create models with thousands of 3D points.

However, these systems demonstrate an underlying problem with SLAM methods in AR. In an unknown environment, AR applications do not have the necessary information about what virtual objects and overlays to display. No framework of reference can be established and thus only toy applications are possible. Therefore, the possible integrations of SLAM systems into AR systems need to be carefully considered. In this paper we discuss some of the combinations we have investigated, and consider future research challenges.

With the prospect of user generated AR content in networked, social environments, the extension to unknown environments become even more important. A ubiquitous virtual reality (U-VR) should work in places that have not been mapped and modeled. Thus, SLAM can provide a way for end users to create the models required for both tracking and annotations on the fly. This can lead to continuous participation of users as described in CAMAR 2.0 [5] and provide a solution to the authoring problem.

II. Localization & Tracking

While SLAM provides an inherent tracking solution, it does not provide any reference to a known, global location. Therefore information that is referenced to such a real location, for example through a GPS position, cannot easily be rendered in a purely SLAM-based system. We have developed a panoramic mapping and tracking approach that is integrated with other sensors to provide global registration.

A. Panoramic Mapping

Our system is based on a simultaneous mapping and tracking approach, operating on cylindrical panoramic images. The algorithm is conceptually comparable to SLAM, however we do not create a 3D map of the environment, but

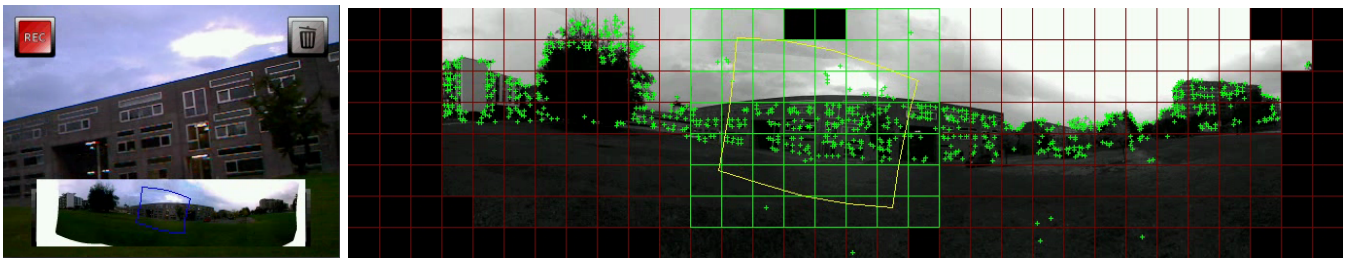

Figure 1. The panorama mapping in operation. (Left) a video frame and the location in the created panorama. (Right) Keypoints points used in the tracking algorithm.

limit the map to a 2D panorama. This simplification works well for users who are standing still while turning the phone. Our method creates a cylindrical map of the environment on the fly and simultaneously uses this map for tracking the camera orientation. Our approach requires 15ms per frame on a smartphone and allows for applications running at interactive frame rates (30Hz). A detailed description of the approach is given in [6].

The panoramic mapping method assumes that the camera undergoes only rotational motion. Under this constraint, there are no parallax effects and the environment can be mapped onto a closed 2D surface. The individual video frames are mapped to a cylindrical surface to create the panorama map. Interest points are detected in the map and used in active search in the following video frames to estimate the camera rotation (see Figure 1). Although a perfect rotation-only motion is unlikely for a handheld camera, our method can tolerate enough error for casual operation. Especially outdoors, where distances are usually large compared to the translational movements of the mobile phone, mapping errors tend to be negligible.

B. Sensor fusion

Using just the panorama tracker in a handheld device provides only relative orientation. For many outdoor applications, geo-referenced data is available and an absolute position and orientation is necessary to accurately register the annotations. A magnetic compass can provide the direction to magnetic north and is therefore used in most outdoor AR setups. However, magnetometers suffer from noise, jittering and temporal magnetic influences, often leading to deviations of 10s of degrees in the orientation measurement. Through combination with the relative, but accurate orientation tracking from panoramas, a global and accurate orientation can be estimated [7].

The integration of the magnetic compass with the visual tracker uses a state machine. Initially the panorama tracker is started with the orientation from the compass. If the compass starts to drift due to an external magnetic field, the relative rotation between the two sensors changes and only the vision tracker is used. Conversely if the vision tracker fails - such as under fast motion and blurry images - only the magnetic compass is trusted (see Figure 2). Together both components complement each other and provide improved overall performance.

C. Place recognition

The panoramas created in the mapping and tracking approach can also be used to recognize locations of annotations. Two distinct methods are possible. Firstly, whole panoramas can be stored together with their GPS location on a server. If a user is close to a location with a referenced panorama, that panorama is downloaded and tracking continues directly from the stored panorama. In this way, the panoramas form an image database used for locating the device. Annotations, such as landmarks and tourist sights, are stored within the panorama image, and can be rendered in the video view. However, this mode requires the user to stand very close to the same location as the original panorama.

To overcome this limitation, we developed a different approach [8]. Instead of storing a complete panorama with several information items, all annotations are stored separately. Together with an annotation, a small visual template of the area surrounding the location in the panorama is stored as well. Whenever a user browses annotations, a new panorama is created on the fly. Any annotations stored

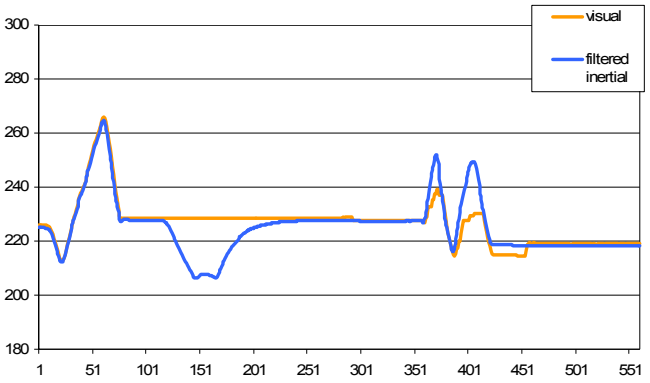

Figure 2. Angle to north estimated through sensor fusion. The visual orientation tracker is not influenced by magnetic disturbances at frame #161, while the magnetic compass deals with fast motions at frames #360 - #460.

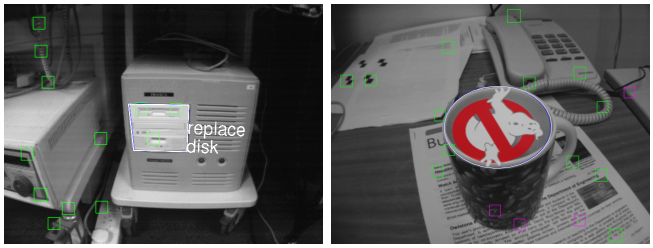

Figure 3. Two views of annotations created in unknown environments. (Left) instructions for a remote worker, (right) message left by a friend.

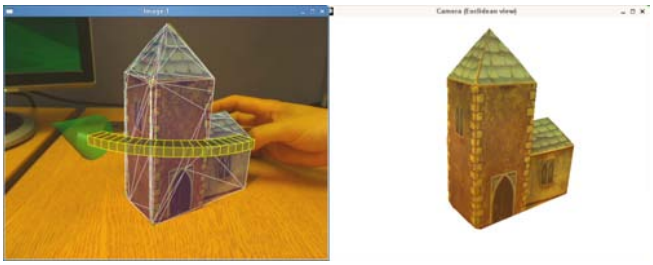

Figure 4. Reconstructing a model on the fly with ProForma. The system shows the current partial reconstruction and provides hints to the user to complete the model.

close to the user's location are then downloaded together with their visual templates. These templates are search for in the new panorama and are located with NCC template matching. Once such a template is located, the corresponding annotation is displayed in the view.

III. Models

Models have many different uses in U-VR. Firstly, visual tracking usually refers to a known model, either built online with SLAM methods or created offline. Furthermore, placement of annotations and virtual objects happens with reference to known models of the environment. Within unknown environments, different methods are required to enable these functions.

A. Annotations

The first issue is to place and register virtual objects within the unknown environment. Usually, placing 3D annotations in a video stream of a moving camera is not easy for a combination of reasons. To place a virtual object in 3D space the user has to specify 6 parameters with respect to a given world frame. This can be accomplished either through direct manipulation interfaces such as seen in CAD software, or through specifying the location in multiple views to triangulate the true pose [9], [10]. Both approaches are severely hindered, if the user cannot control the viewpoint of the camera to select appropriate views.

To simplify the creation of annotations in unknown environments, we extended a state-of-the-art SLAM system to support geometrical landmarks that represent more closely the real world objects that are of interest to the user [11]. Instead of relying on fragile methods to select these geometrical landmarks automatically, the user indicates interest by placing an annotation on a feature which then is measured and added to the model maintained by the system. Thus, the presented system combines the complementary strengths of a human operator, who understands a scene and can make informed decisions about it, and a computer system which performs accurate measurements beyond any human capability. The result is a simple interface coupled with accurate operation.

Figure 3 shows some example views from the application. The highlight box and round texture geometry are applied with a single click during the online operation of the application. The SLAM system estimates the 3D location of the underlying surfaces and provides an accurate location for the annotation geometry.

B. Content creation

Similar methods can be employed to create virtual models of real objects for the use in augmented reality applications. In the ProForma system [12], we developed an online model reconstruction system that uses AR overlays to guide the user during the reconstruction process (see Figure 4). Online reconstruction and space carving methods are used to track corners and points on the object. The resulting model is directly used to track the real object and also to provide a preview of the reconstruction. In a next step, the virtual model can be uploaded to a U-VR service and become part of the user's content library.

IV. Challenges

The described work demonstrates the usefulness of online reconstruction in various augmented reality and networked, social VR applications. The ability to create environmental models for localization and virtual models for AR content is a necessary feature of any future networked U-VR system. However, there are still open research challenges to be addressed.

Sensor integration: Outdoor augmented reality typically relies on GPS, magnetic compass and inertial sensors for estimating position and orientation of the user's point of view. These sensors can exhibit large errors depending on the environment and transient disturbances. By combining these robust and absolute, but noisy, sensors with an online reconstruction system we can merge the relative, but accurate measurements of the SLAM system with the noisy global estimates. Efficient implementations of visual SLAM for mobile and handheld devices and appropriate sensor fusion models are investigated to achieve a more robust outdoor localization

Enhanced visualization in unknown environments: Advanced visualization in augmented reality requires a good understanding of both the virtual and real parts of a scene.

To avoid visual artifacts due to color, texture and saliency of the video background, good knowledge about the scene is necessary to help choosing a good visualization technique. If a sufficiently accurate 3D model of the real scene is available, correct occlusions, impostors for explosion views and visual effects for X-ray visualization can be rendered. However, in unknown environments most of the direct approaches fall down, because no model is available. Here on-line reconstruction of the unknown environment can provide the required models on the fly. We propose different approaches, ranging from simple 2D segmentation of the scene, to sparse and dense 3D reconstruction of the environment. In the 2D case, salient regions and different appearance properties of the scene can be analyzed and used in X-ray visualizations that depend on occluders and texture. With 3D reconstruction, occlusions, shadows and lighting effects between real and virtual objects can be accurately simulated.

Interaction in unknown environments: On-line reconstruction can also support user interfaces in unknown environments by providing the necessary models for interaction methods such as ray-picking or constrained interaction on the fly. Furthermore, by identifying objects in the scene, based on simple shapes, object detectors or through human interaction, applications can provide specific interactions with specific objects. For example, labels can be aligned to distinct features such as windows and doors on a facade, instead of just being placed parallel to the building wall. Individual objects that are meaningful to the application can be highlighted and made available for further interaction.

Semantic models: Future applications may require more complex scene understanding than current SLAM systems can provide. For example, only surface geometry alone allows only for a static description of the scene. Recognizing objects and independent parts of the scene enables the application and the user to attach meaning to these items. The user could interact with the application through modifying the scene, instead of direct input through a user interface.

V. Conclusion

The addition of simultaneous localization and mapping to the toolbox of the AR systems engineer opens new possibilities for ubiquitous and social AR applications. Through combination with other sensors and tracking systems, clever user interaction and systems design, online reconstruction allows us to extend the scope of AR to any environment. In this paper, we explored some of these directions and identified future research challenges. This is of particular interest for networked and social applications as these will be used in a casual way by many users in different environments. Therefore, it will be a requirement for these applications to be able to deal with unknown surroundings and to provide users with tools to create annotations on the fly.

Acknowledgment

This work was funded by the Austrian Research Promotion Agency (FFG) under contract no. FIT-IT 820922, through the Christian Doppler Laboratory for Handheld Augmented Reality.

References

[1] G. Welch and E. Foxlin, "Motion tracking: No silver bullet, but a respectable arsenal," *IEEE Comp. Graph. Appl.*, vol. 22, no. 6, pp. 24–38, Nov/Dec 2002.

[2] A. J. Davison, W. W. Mayol, and D. W. Murray, "Real-time localisation and mapping with wearable active vision," in *Proc. ISMAR 2003*. Tokyo, Japan: IEEE, October 7–10 2003, pp. 18–27.

[3] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proc. ICCV 2003*, Nice, Italy, October 13–16 2003, pp. 1403–1410.

[4] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *Proc. ISMAR 2007*, Nara, Japan, Nov. 13–16 2007.

[5] C. Shin, W. Lee, Y. Suh, H. Yoon, Y. Lee, and W. Woo, "CAMAR 2.0: Future direction of context-aware mobile augmented reality," in *Proc. ISUVR'09*, July 8–11 2009, pp. 21–24.

[6] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg, "Real-time panoramic mapping and tracking on mobile phones," in *Proc. VR 2010*, March 20–26 2010.

[7] G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg, and B. Hofmann-Wellenhof, "Global pose estimation using multi-sensor fusion for outdoor augmented reality," in *Proc. ISMAR 2009*, Oct. 19–22 2009, pp. 153–162.

[8] T. Langlotz, D. Wagner, A. Mulloni, and D. Schmalstieg, "Online creation of panoramic augmented reality annotations on mobile phones," *submitted for publication*, 2010.

[9] Y. Baillot, D. Brown, and S. Julier, "Authoring of physical models using mobile computers," in *Proc. ISWC'01*. Zurich, Switzerland: IEEE, October 8–9 2001, pp. 39–46.

[10] W. Piekarski and B. H. Thomas, "Augmented reality working planes: A foundation for action and construction at a distance," in *Proc. ISMAR 2004*. Arlington, VA, USA: IEEE, November 2–5 2004, pp. 162–171.

[11] G. Reitmayr, E. Eade, and T. W. Drummond, "Semi-automatic annotations in unknown environments," in *Proc. ISMAR 2007*, Nara, Japan, Nov. 13–16 2007, pp. 67–70.

[12] Q. Pan, G. Reitmayr, and T. W. Drummond, "ProFORMA: Probabilistic feature-based on-line rapid model acquisition," in *Proc. BMVC 2009*, London, UK, Sept 7–10 2009.