Visual Tracking for Augmented Reality

Manfred Klopschitz, Gerhard Schall, Dieter Schmalstieg, Gerhard Reitmayr
Institute for Computer Graphics and Vision, Graz University of Technology, Austria
{klopschitz, schall, schmalstieg, reitmayr}@icg.tugraz.at

1 Summary

Localization of mobile devices is an essential task in Augmented Reality and has therefore been an active research topic for many years. Typically, indoor tracking approaches, such as methods based on infrared or ultra-wide-band, require preparations of the environment and special hardware sensors. Conversely, image feature tracking approaches can provide orientation estimates without special tracking hardware installations. With the advent of mobile devices equipped with sensors such as digital cameras, image based localization gains importance in Augmented Reality. Typically, fiducial marker tracking was considered as a standard image based localization method. We propose the use of natural image feature based tracking methods, which are a generalization of the same principals but do not require the presence of fiducial tracking targets.

2 Vision based Localization for Augmented Reality

Augmented Reality (AR) is a powerful user interface for mobile computing. AR superimposes registered 3D graphics on the user's view of the real world, allowing the user to perceive overlaid information that is spatially registered to the environment (see Figure 1). In this context, tracking provides indoor localization to correctly register digital information, e.g. navigational hints, to the user. Computer Vision-based localization techniques offer great advantages over other localization methods based on infrared, WLAN/Wi-Fi or ultra-wide-band. Image-based measurements allow very high precision in the pose estimation and self-contained operation without complex and expensive infrastructure. In computer vision, the problem of location recognition has been addressed in the past by a variety of approaches. The most successful methods rely on wide baseline matching techniques based on sparse features such as scale invariant interest points and local image descriptors. The basic idea behind these methods is to compute the position of a query image with respect to a database of registered reference images, planar surfaces or 3D models. Assuming a static scene, geometric verification can be used to determine the actual pose of the camera with respect to the exemplary database. Different viewpoints or illumination changes are largely handled by robust feature descriptors, such as SIFT [5], that provide invariant descriptions of local image patches. Vocabulary tree based image retrieval and inverted file scoring [2] allows for fast search of large SIFT descriptor databases.

3 Localization for Mobile Devices

In recent years mobile computing devices and state of the art mobile phones in particular, have seen immense progress in miniaturization and performance. These devices are well-established and offer a convincing hardware package containing all components necessary for vision based localization and information visualization. The first approaches for mobile phone localization were client-server based where tracking is outsourced to a PC connected via wireless link. But, these approaches suffer from restricted bandwidth, the imposed infrastructure dependency limiting the scalability in the number of client devices and high response times. As a consequence server-based approaches are not suitable for AR.

However, recent approaches have shown that natural feature tracking with 6 Degrees Of Freedom (DOF) can be realized in real-time using mobile phones [1]. Point based visibility of data base features are used in this work for reduced computational- and memory requirements. Moreover, Takacs et al. [3] present an outdoor localization system directly performing keypoint detection and matching. Features are clustered in a 2D grid and pre-fetched by proximity. Each grid element contains a set of clustered meta-features representing the most repeatable features. Geometrically consistent meta-features are created. However, no 3D model is used and thus true geometric consistency is not enforced and no full 6DOF pose is computed.

Figure 1: User is equipped with a mobile device that accurately renders 3D structures and information directly on top of the orthographic photo [4].

4 Tracking Data Acquisition and Outlook

For 6DOF pose estimation, a full 3D localization of feature points is necessary. If a simple known planar texture is used as tracking target, these positions are given by design. For larger work spaces, this 3D database acquisition step becomes challenging. We propose to use triangulated natural image features, obtained with a structure from motion (SfM) reconstruction system. These 3D points are registered into a global coordinate system and partitioned into a representation suitable for efficient localization on mobile phones (see Figure 2). The need for efficient and reliable image based tracking increases as the development of mobile devices featuring video cameras as potential positioning sensor continues.

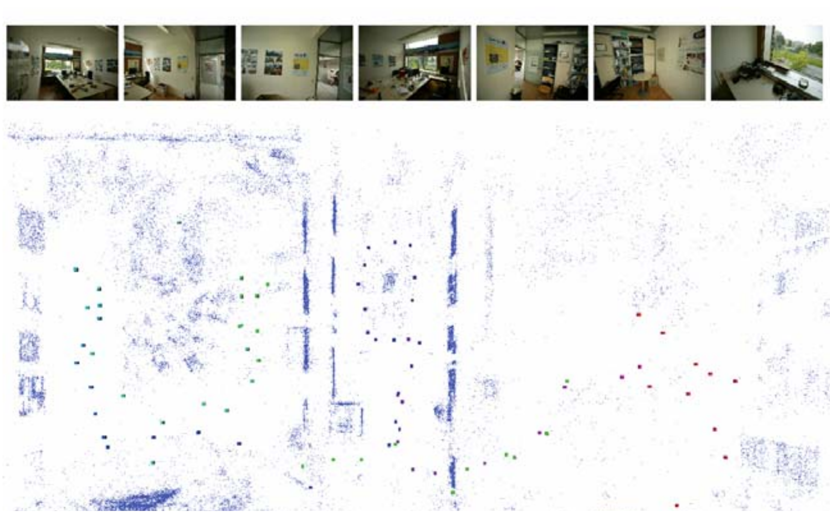

Figure 2: Sample source images used to create the 3D reconstruction (top). 3D point cloud database resulting from the reconstruction. The point cloud shows two rooms from above. Small blue dots represent triangulated feature points and larger dots represent the computed 3D locations of the mobile phone. Colours indicate the timeline and show the user's motion. (bottom).

References

[1] C. Arth, D. Wagner, M. Klopschitz, A. Irschara, D. Schmalstieg. Wide Area Localization on Mobile Phones. Proceedings of Int. Symposium on Mixed and Augmented Reality 2009 (ISMAR´09), IEEE, 2009-October

[2] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In Proc. of CVPR´06, pages 2161–2168, 2006.

[3] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W.-C. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, and B. Girod. Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In Multi.Med. Inform. Retrieval´08.

[4] G. Schall, Grabner H., Grabner M., Wohlhart P., Schmalstieg D., Bischof H., „3D Tracking Using On-line Keypoint Learning for Mobile Augmented Reality", Workshop on Visual Localization for Mobile Platforms in conj. with IEEE CVPR 2008, USA, 2008

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