Towards Collaboratively Mapped Multi-View Mobile Augmented Reality

Klen Čopič Pucihar
School of Computing and Communications
Infolab21
Lancaster University
Lancaster, LA1 4WA, UK
klen.copicpucihar@lancaster.ac.uk

Paul Coulton
School of Computing and Communications
Infolab21
Lancaster University
Lancaster, LA1 4WA, UK
p.coulton@lancaster.ac.uk

MobileHCI 2011, Aug 30–Sept 2, 2011, Stockholm, Sweden.
ACM 978-1-4503-0541-9/11/08-09.

Abstract
Whilst Augmented Reality (AR) offers the possibility of exciting new interaction modalities for mobile users, the prospect of providing a collaborative multi-view, in which multiple viewers may view the same shared AR workspace from multiple perspectives, offers another equally exciting prospect. Currently, this has only been researched using marker-based AR approaches that present significant scalability issues, as they require the wide scale physical augmentation of our environment with markers. However, recent advances in marker-less systems using natural feature tracking, coupled with the use of wireless, ad-hoc communication could be used to create a collaboratively mapped multi-view AR space.

Keywords
augmented reality, collaborative, multi-view, marker-less, natural feature tracking.

ACM Classification Keywords
H.5.3 Group and Organization Interfaces.

General Terms
Design, Human Factors, Theory

Introduction

Two architects are inspecting the site of their theme park development and decide that the natural landscape suggests alternate layouts. Rather than waiting to return to office to test these ideas on their scale model, the first architect points his phone at the ground where a 3D model of development appears, and begins manipulating the model using gestures. The second architect then points his phone at the same area, and the same model appears, and he starts making his own alterations. Importantly, the model is shared and correctly orientated to each architect's point of view so that they both combine to produce collaborative changes as if it was on physical model. At the end of their deliberations, the architects switch the model to full scale and inspect it within the actual landscape.

It is clear from this scenario that the field of Mobile Augmented Reality (MAR) presents exciting possibilities for new mobile users experiences. However, to realize such Collaborative Multi-View Mobile Augmented Reality (CMV-MAR) scenarios, a number of fundamental research questions must be addressed. In the following paragraphs we discuss the challenges ahead that need to be solved in order to realize this vision.

Background

MAR systems are normally classified through the method used to obtain a 3D map, from which the camera pose can be determined, thus allowing the virtual objects to be rendered onto the scene. These methods can be subdivided into sensor and vision based approaches.

Sensor-based approaches take advantage of increasing number of sensors, such as GPS, accelerometers, magnetometers (digital compass), and recently gyroscopes, on commercial mobile phones. Combining readings obtained from such sensors allow the position and orientation of the phone camera to be estimated [5]. Such systems are relatively easy to implement and are the basis of many of the AR browsers, such as Layar[1] and Wikitude[2]. Nevertheless, the main problems of such systems are the relatively crude accuracy, jerky augmentation, relatively long response times to camera pose changes, limitation to outdoors environments, and problems with ensuring the augmentation with correct context.

The simplest and most widespread of the vision techniques involves the use of 2D fiducial markers to provide a pose estimate relative to the environment [4] [10]. The problem with this methodology is that augmentation of the environment with fiducial markers is required, which limits the possibility for widespread use. The alternate approach is natural feature tracking, which remedies this problem by using features within the environment to produce the 3D map. Nevertheless, this comes at the expense of high processing cost, which is always an issue when implementing mobile AR [11]. Whilst there have been a number of systems developed using natural feature tracking on mobile devices [9], the majority of them require apriori knowledge of the environment, which limits the wide scale deployment of such systems.

[1] www.layar.com

[2] www.wikitude.org

The main challenge of estimating camera pose in a completely unknown environment is the system initialization. To initialize the system, the 3D map of the environment is required. For planar surfaces, this can be done using a model-based approach, where 3D model of a planar scene can be created by acquiring a view with no perspective distortions. Such a view can be captured by ensuring that the scene plane and the camera image plane are parallel, as in the case of MAR system Nestor [3], or by un-distorting the initial image by utilizing sensing capabilities of a mobile phone, as shown by mobile AR systems SID-MAR [1] and mobile specific implementation of Gepard [8]. In the later, a single textured patch is used for camera pose registration and relocalization. The system supports patch information sharing through a Bluetooth connection and presents the first attempt of CMV-MAR. However, the system only tackles the map-sharing problem in the simplest form of single file synchronization. Further, there is no evidence of the augmented scene information sharing as well as the usability and the success of the proposed collaboration.

The alternative approaches that also allow initialization of non planar scenes are Simultaneous Localization and Mapping (SLAM) techniques, originally developed for robot navigation, and successfully adopted to hand-held AR system such as MonoSLAM [2]. The work of MonoSLAM was extended by Klein's and Murray's creation of PTAM [6], where they demonstrate how localization and mapping processes can be separated and executed in parallel, providing significant increases in both the accuracy and robustness of handheld AR. A heavily modified version of PTAM has later been adopted for the mobile phone [7] and is, up until now, the only 6-DOF SLAM system running on a mobile phone.

Requirements for CMV-MAR

To enable CMV-MAR, a number of challenges remain and differ according to the type of AR system. In case of sensor based systems, the provided position and orientation of the users' camera is absolute, thus the camera point of view of collaborating users does not need to be aligned. To enable collaboration, users need to only share information regarding changes to the augmented scene (i.e. object's location and shape). Whilst sensor mobile AR systems are already creating collaborative environments by overlaying landmarks with geo-tagged information placed by other users, the problems previously defined remain.

One of the challenges for collaboration in vision-based systems is the nature of camera pose estimation, which is defined relative to the 3D map used for camera tracking purposes. To enable collaborative AR, 3D maps of landmarks need to be shared or aligned. In marker-based or extensible tracking systems, map sharing is not a problem, as the map is shared or aligned through the system initialization, where the same marker is used.

In case of markerless AR systems, each system initialization produces a 3D map with different coordinate system, thus map information needs to be shared amongst the collaborative users for map alignment purposes. The goal of alignment is to move the coordinate system of maps into the same point, scale maps to the same unit, and to ensure that all users use the same base plane for augmentation. This can be accomplished by sharing a minimum of three reference points. If the assumption is made that all systems select the same base plane, map alignment requires only two reference points. In markerless AR systems where feature descriptors are used for camera tracking, reference points could be shared by sharing feature descriptor amongst the collaborating users.

In addition to aligning the map, the sharing of the map information is distributing the computational cost of initializing and maintaining 3D maps, which is in the case of long term mapping systems, a resource hungry operation. Nevertheless, the feasibility and performance impact of maintaining an ad-hock communication channel open while executing a processor intensive camera tracking still needs to be studied.

To conclude, CMV-MAR can be achieved through the use of wireless ad-hoc communication, where changes to the augmented scene and 3D map information can be shared between the mobile clients. This will enable the creation of a distributed version of the AR space amongst the users viewing the scene. However, it raises a number of research questions namely: performance impact of maintaining an ad-hock communication channel open while executing other crucial resource hungry operations; organisation of the network; management of the distributed map; system optimisation by distributing computational loads. An extension of this research could be to investigate how collaborative virtual maps could be retained within the physical space for future use by other groups of users; how these maps may be communicated to such users; and how they should be developed to evolve with the inevitable changes to the environment.

References

[1] Čopič Pucihar, K., Coulton, P. and Hutchinson, D. Utilizing Sensor Fusion in Markerless Mobile Augmented Reality *MobileHCI 2011*, Stockholm, Sweden, 2011.

[2] Davison, A.J., Reid, I.D., Molton, N.D. and Stasse, O., MonoSLAM: Real-Time Single Camera SLAM. in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2007), 1052-1067.

[3] Hagbi, N., Bergig, O., El-Sana, J. and Billinghurst, M., Shape recognition and pose estimation for mobile augmented reality. in *Proceedings of the 8th IEEE ISMAR*, (2009), IEEE Computer Society, 65-71.

[4] Henrysson, A., Billinghurst, M. and Ollila, M. Face to face collaborative AR on mobile phones. *Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on*. 80-89.

[5] Kähäri, M. MARA – Sensor Based Augmented Reality System for Mobile Imaging Device. *ISMAR*. 180-180.

[6] Klein, G. and Murray, D., Parallel Tracking and Mapping for Small AR Workspaces. in *Proceedings of the 6th ISMAR*, (Nara, Japan, 2007).

[7] Klein, G. and Murray, D., Parallel Tracking and Mapping on a Camera Phone. in *Proceedings of the 8th IEEE ISMAR*, (Orlando, 2009).

[8] Lee, W., Park, Y., Lepetit, V. and Woo, W., Point-and-shoot for ubiquitous tagging on mobile phones. in *Proceedings of the 9th IEEE ISMAR* (2010), 57-64.

[9] Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T. and Schmalstieg, D., Pose tracking from natural features on mobile phones. in *Proceedings of the 7th IEEE ISMAR*, (Washington, DC, USA, 2008), IEEE Computer Society, 125-134.

[10] Wagner, D. and Schmalstieg, D., Artoolkitplus for pose tracking on mobile devices. in *Proceedings of 12th Computer Vision Winter Workshop (CVWW'07)*, (2007), Citeseer, 139-146.

[11] Wagner, D., Schmalstieg, D. Making Augmented Reality Practical on Mobile Phones, Part 1. *IEEE Computer Graphics and Applications*.