“Anywhere Access” with Annotated Environment Maps

Cha Lee, Jonathan Ventura, Chris Coffin, Sehwan Kim and Tobias Höllerer

1 Introduction

Currently, we are witnessing a potent confluence of trends in technology and society. First, hardware manufacturers are trending towards powerful, connected, yet ultra-portable mobile devices. These devices have advanced sensors for video, GPS, and orientation, and support a fast and available network connection. At the same time, there is a growing familiarity with and desire for social information sharing. People of all ages are becoming accustomed to broadcasting a multitude of personal data such as what they are doing and where, photos, and videos, as well as interactive input such as comments on other people's information streams, links, and recommendations.

Because of their ubiquity and capabilities, advanced mobile devices such as smart phones have great potential for improving and extending interactions on social networks. Already, users can upload text, photos, videos from wherever they are with their phone. However, this type of interaction only captures a small part of what could be achieved. In fact, researchers in the augmented reality (AR) and virtual reality (VR) fields have long studied the best ways to share experiences and communicate using high immersion devices.

We believe that environment maps (or panoramas) provide a great starting point for combining social interaction with mobile mixed reality. Panoramas already can be considered one of the first public successes of VR [1]. They are useful for remote collaboration and exploration, because they provide a high level of immersion (surround-view imagery), but can be stored compactly in any common image format.

2 Applications and Related Work

Existing applications such as Google Street View [3] already leverage a massive collection of geo-registered panoramas to provide remote exploration. However, Google Street View provides a passive VR experience, without spatial annotation other than street names. Related photographs and websites are displayed, but there is not a way to read or provide situated annotations in a true AR sense.

Layar [5] has developed software for mobile devices which does present spatialized annotations on live video, in the form of photographs and dots with text. However, data can only be added to Layar by requesting a developer key and submitting a full-fledged layer. We instead envision an open access model where all users can add comments and recommendations to the database. Also, Layar data can only be experienced from its actual location; instead, we see potential in also allowing remote experience of data.

On social networking sites such as Facebook and Twitter, millions of users stay in touch by sharing text, photos and videos. People use these systems as a way to connect with their friends and remotely share their lives. A globally connected AR system using mobile devices would greatly extend and enhance the social networking experience. For example, a mobile device could be used as a VR display which goes beyond the photo collections typically found on social networking sites. Currently, time is the most important organizing element for social data; consider for example the time-ordered stream of “status updates” on Facebook and Twitter. We see the spatial component of data as an equally important and interesting organizational tool. Augmented reality displays allow us to visualize the spatial component naturally, by situating content in the real world. For example, a party could be recreated out of the pictures taken and uploaded by its guests, with comments attached to elements of the reconstruction. Using the timestamps stored in the photos, we could play back the party in movie fashion, experienced virtually using a cell phone.

3 Our Previous Work

Live panorama acquisition The acquisition of surround-view panoramas using a single hand-held or head-worn camera relies on robust real-time camera orientation tracking. In absence of robust tracking recovery methods, the process has to be completely re-started when tracking fails.

While recovering to keyframes is not a new idea, we expanded on the ability of the system to recover by determining an ideal distribution of keyframes for recovery across the surface of a sphere [4]. This distribution was based on analysis acquired through performance tests in simulation and on live video using a pan tilt unit to achieve precise rates of speed.

Our system generates virtual keyframes from the captured environment map, allowing the generation of keyframes used for recovery which may not correspond to an image captured by the physical camera. We implemented our camera orientation relocalizer with the help of a GPU fragment shader for real-time application. Figure 2 compares the quality of panoramas made with and without robust tracking recovery.

A User-guided Robustness Metric Real-time computer vision based tracking is a central component of most mixed reality applications, and over the past decade there has been great progress in terms of the tracking accuracies that can be achieved. Despite the fact that several metrics have been established for measuring tracking errors, assessing the robustness of a particular tracking approach in relation to oth-

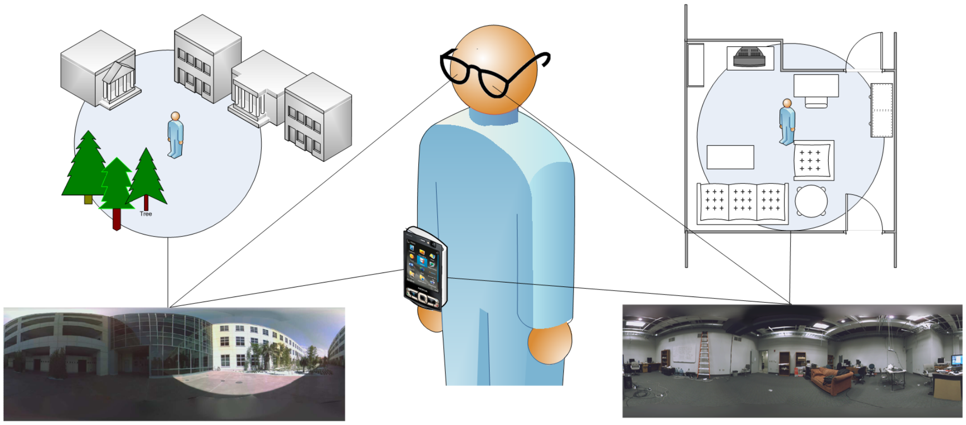

Figure 1: Indoor and outdoor panorama acquisition and annotation on a cell phone.

ers remains a challenge. Our work provided a formula by which the robustness of multiple systems can be compared assuming that there is a set of representative video data with associated ground truth information [2]. Our focus was on orientation tracking with the goal that the metric should be generally applicable.

We established our metric based on the analysis of three variations of an orientation tracking system. We then tested the resulting metric by implementing a fourth method and comparing the output of our metric's estimation of the robustness to the results of expert evaluations of the robustness.

In order to ensure that the robustness metric was accurately made for augmented reality applications, we recorded 40 sets of one minute long head movement data as user went about search and casual exploration tasks, then played this orientation information back using a pan tilt unit in both indoor and outdoor environments in order to obtain ground truth orientation data for each frame of video. To establish our metric we then compared this to rankings of expert users on the robustness of the systems based on separate evaluations of the environment maps generated by the orientation tracking, and hands on use of the systems themselves.

Adding the third dimension Although panoramas already contain a wealth of information, they lack the third dimension: depth. Even a simplified three-dimensional model of an environment can improve us to display and organize a scene. In previous work, we augmented a camera with a single point laser rangefinder, which captures a depth sample for each frame recorded [9]. Using plane detection and fast image-guided interpolation, we can extrapolate this point depth across the entire panorama to achieve a full panoramic depth map alongside the color imagery. See Figure 3 for an example. This depth map enables several important augmented reality technologies. The depth samples give us an extra cue to help segment out objects for annotation. Also, using the depth map, we can render virtual objects with correct occlusion.

The AR simulator To develop good interfaces for mobile AR, we need to study the effects of varying factors on the user experience. However, it is extremely challenging to run controlled studies comparing multiple Augmented Reality (AR) systems. We have begun experimenting with an AR simulation approach, in which a Virtual Reality (VR) system is used to simulate multiple AR systems.

But are the results of experiments using AR simulation even valid for real-world AR systems? There are multiple steps required to validate AR simulation. We must analytically compare the level of immersion of our final simulator to real world AR systems so that these values make sense and are reasonable. Then we need to replicate a small set of experiments from the literature and show that the results from simulation are comparable to the established results. Finally, we need to do direct comparisons between studies run on our simulator and studies with real, practical systems.

In prior work we have begun to address the validity of AR simulation. We first present our results on replicating a well known AR experiment, and in a follow-up study, we investigate the effects of simulator latency on a 3D tracing experiment performed in AR simulation [6].

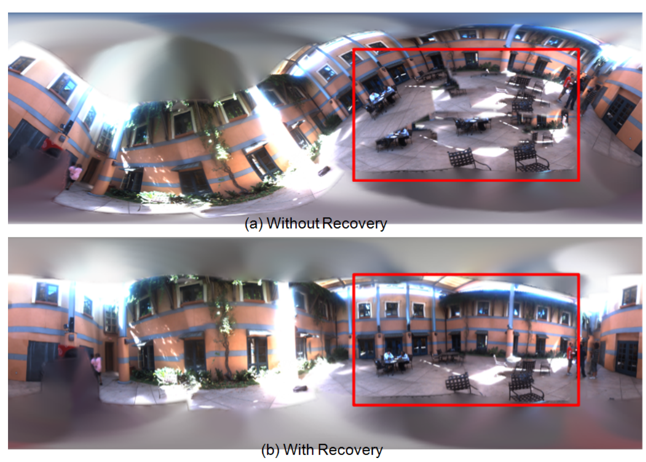

Figure 2: Robust recovery from tracking failure greatly improves the quality of acquired panoramas.

4 Future Work

We plan to implement the basic framework of a Virtual Geo-Caching system. Users with capable phones will be able to log onto our system in conjunction with their Facebook account. The phone's camera and display are then used to create a "magic lens" view, in which the display is treated as a window to the real and augmented world. Once logged on, a user will be able to generate location-tagged videos, panoramas, and single images which will be stored on our central server. Users are then able to view, edit, and annotate this data via a mobile device or desktop machine. By using their Facebook friend's list, users will also be able to view and annotate their friends' content, which should allow for novel and rich social interaction [7]. For example, users will be able to take videos of events and geo-cache it at the location where it occurred. Friends will be able to view these videos online, or at those locations at a later time. Multiple viewpoints can be saved by different friends, allowing for a multi-view show by non-resident friends remotely, or at the same location at a later time. When panoramas and images are created, they are registered and co-located with the real world.

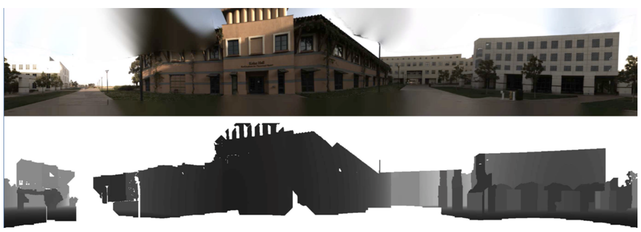

Figure 3: Using a a laser range finder, we can capture a depth map alongside the panoramic image.

Improving panorama acquisition We are currently working on an extension of our virtual keyframe recovery for 6 DOF tracking. Our goal is to allow for recovery using projective texturing on captured model information. We can then capture virtual keyframes which we can use to recover orientation tracking.

Current and future work will include work on combining information from environment maps and other commonly available information such as aerial photographs and simple geometric models possibly generated from Google SketchUp.

One possible project includes using aerial photograph segmentation to extract lines of buildings. By then extracting vertical planes from the lines we extracted, we can texture map the vertical planes with information from environment maps. Using the textured planes we can then extract the tops of buildings and use the rough outlines for virtual augmented reality applications.

Dense 3D modeling We would like to extend our modeling approaches beyond a single viewpoint to full, dense 3D modeling of a scene. A rich, textured geometric environment model enables significant advances for AR interfaces. We have already found promise in simplified modeling approaches such as plane finding [8]. A crucial factor in the success of dense modeling, however, is the scalability of the process. We are currently researching algorithms which can handle very large image sets efficiently and also can handle streams of new imagery as they are arrive. With this system in place, users of mobile AR across the globe could be constantly participating in a global modeling effort, just by using their phones for normal everyday tasks.

User studies on interaction & immersion An important part of developing the foundations of mobile AR lies in controlled evaluation of the interfaces and techniques proposed. We are currently investigating the best method for viewing environment maps on small-screen, tracked mobile devices like the iPhone. Fundamentally we would like to investigate whether more immersive interfaces, such as a magic lens, actually improve performance over more traditional interaction techniques such as using a stylus or finger input.

References

[1] S. E. Chen and L. Williams. View interpolation for image synthesis. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 279–288, New York, NY, USA, 1993. ACM.

[2] C. Coffin, S. Kim, and T. Höllerer. A metric for tracking robustness in real-time panorama acquisition and beyond. 2010. Currently under review.

[3] Google. Maps. http://maps.google.com/, October 2009.

[4] S. Kim, C. Coffin, and T. Höllerer. Relocalization using virtual keyframes for online environment map construction. In *The 16th ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2009.

[5] Layar. Layar. http://www.layar.com/, October 2009.

[6] C. Lee, S. Bonebrake, T. Höllerer, and D. A. Bowman. A replication study testing the validity of ar simulation in vr for controlled experiments. In *Mixed and Augmented Reality. IEEE/ACM International Symposium on*, New York, NY, USA, Oct. 2009. ACM.

[7] R. Simon, P. Rohlich, and T. Grechenig. Geopointing: evaluating the performance of orientation-aware location-based interaction under real-world conditions. *J. Locat. Based Serv.*, 2(1):24–40, March 2008.

[8] J. Ventura and T. Höllerer. Online environment model estimation for augmented reality. In *Mixed and Augmented Reality. IEEE/ACM International Symposium on*, New York, NY, USA, Oct. 2009. ACM.

[9] J. Wither, C. Coffin, J. Ventura, and T. Höllerer. Fast annotation and modeling with a single-point laser range finder. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 65–68, New York, NY, USA, Sept. 2008. ACM.