Using Aerial Photographs for Improved Mobile AR Annotation

Jason Wither* Stephen DiVerdi† Tobias Höllerer‡

Four Eyes Laboratory§
University of California
Santa Barbara, California, USA

Abstract

We present a mobile augmented reality system for outdoor annotation of the real world. To reduce user burden, we use aerial photographs in addition to the wearable system's usual data sources (position, orientation, camera and user input). This allows the user to accurately annotate 3D features with only a few simple interactions from a single position by aligning features in both their first-person viewpoint and in the aerial view. We examine three types of aerial photograph features – corners, edges, and regions – that are suitable for a wide variety of useful mobile augmented reality applications, and are easily visible on aerial photographs. By using aerial photographs in combination with wearable augmented reality, we are able to achieve much higher accuracy 3D annotation positions than was previously possible from a single user location.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality H.5.2 [Information Interfaces]: User Interfaces—Interaction Styles

Keywords: wearable system, outdoor augmented reality, annotation, modeling, anywhere augmentation

1 Introduction

Annotation of outdoor scenes is an important part of mobile augmented reality research. Generally, the situated content displayed by a wearable system is carefully constructed offline using many different technologies, including modeling programs, GIS data, and aerial photographs. In this project, our focus is on annotating an outdoor scene from within the wearable system, providing an appropriate interface to allow accurate markup in a mobile context. To reduce the amount of manual work that must be done by the user, we have modified our system to use aerial photographs of the region in conjunction with the wearable's acquired data. This allows the user to accurately place 3D annotations from a single position by providing a means of accurately gauging depth.

With orientation tracking, from a static position a user can easily cast a ray to select a visible feature in the scene, but setting the depth of that feature is more difficult. Previous work in this area requires the user annotate the same feature from multiple viewpoints to triangulate a position [10], or estimate depth from a static viewpoint using artificial depth cues [15]. However, commonly available aerial photographs [6, 16] can be used to allow accurate 3D position input from a single location. After casting a ray by looking at a feature and pushing a button, our system presents an aerial view of the scene and the cast ray and allows the user to adjust the ray and set a distance. The result is a significant improvement in the accuracy of 3D positions over previous AR distance estimation work [15], as well as the ability to annotate features that may not be directly visible from the user's location, such as the opposite side of a building.

*e-mail: jwither@cs.ucsb.edu
†e-mail: sdiverdi@cs.ucsb.edu
‡e-mail: holl@cs.ucsb.edu
§web: http://ilab.cs.ucsb.edu/

Figure 1: The aerial photograph view from within the system. In the lower left corner, an insert of the video feed from the head-worn camera can be seen. The user's position and orientation are represented on the photograph with a small cone avatar. A small set of features have already been annotated – two corners (the green points), one edge (the green line), and one region (the transparent green rectangle).

We examine three different types of features, corners, edges, and regions, a user may want to place in the outdoor scene, based on how they appear in the aerial photograph. Corners can correspond to the vertices of building silhouettes and are useful for modeling geometry [2]. Vertical walls appear as edges that can be used to properly orient and position world-aligned billboard annotations [7]. Uniform regions in aerial photographs can denote buildings, fields, etc. and can be annotated with a label and a bounding box for wearable navigation purposes [5]. Our interface is thus geared towards finding these types of features in our aerial photographs. We use these annotations as a representative set of possible information a user may want to input, but our system is not geared towards any particular application and only minor modifications are needed to tailor the approach to other task scenarios.

This paper is the first step towards our goal of *anywhere augmentation*, where the usual AR initial costs of manual modeling, calibration and registration are alleviated to make augmented reality readily available in any unprepared environment. Our contribution is to significantly reduce the work necessary to create physically-situated annotations in an unprepared, large-scale outdoor scene.

1-4244-0651-X/06/$20.00 ©2006 IEEE

2 Previous Work

Rekimoto et al. [12] introduced the idea of Augment-able Reality with a system that allows users to annotate the environment with contextual information at specific locations that have been prepared ahead of time with active or passive markers. They envisioned extending the system to allow annotations for any position of known GPS coordinates. Our system expands on this concept by allowing annotations of unprepared environments at arbitrary locations.

More recent work has been done in using wearable systems to acquire accurate positions of arbitrary locations, towards the goal of modeling outdoor scenes from within a wearable system. Baillot et al.'s [2] wearable modeler is based on first creating 3D vertices by casting intersecting rays from multiple viewpoints, and then creating primitives fit to those vertices for the final model. Our annotations are a more easily acquired version of their construction points, and could be used for the same sort of modeling application as they describe. The paper also shows an example using an indoor scene's architectural floor plan as a guide for creating vertices. This served as one of the inspirations for our use of aerial photographs, but we extend the concept in many new directions – we use commonly available aerial photographs that are automatically registered to the user's location, and we use them not only for creating points, but for several other types of annotations as well.

Piekarski and Thomas' outdoor wearable modeling system [9, 10] implements a wide variety of techniques based around the concept of working planes that are created by sighting down a wall or feature to be modeled from one viewpoint, and then moving to another location to create points on that plane at the correct depth. Our solution replaces working planes with the overhead view, removing the need for the user to move around large buildings to distant locations for accurate modeling.

To avoid requiring multiple viewpoints for determining 3D positions, Reitmayr and Schmalstieg's system [11] has a complete model of the environment (obtained offline) that users can annotate by casting rays that intersect with the model's surfaces. In our system, we avoid the cost of acquiring such a model and use aerial photographs for the same purpose.

Maps and aerial photographs have also been used in many mobile systems for passive localization purposes. The Cyberguide system [1] uses a hand-held device to display a rough estimate of the user's position and orientation on an abstract map with point-of-interest annotations. ARVino [8] uses aerial photographs in a virtual reality view of GIS data, to aid the user in mentally mapping abstract information onto the physical environment the data annotates. We extend the functionality of both Cyberguide and ARVino by using aerial photographs for passive localization of the user, as well as active annotation of the environment by the user.

3 System

Our wearable system can be see in Figure 2. At its core is an Alienware Area-51 m5500 laptop, which is worn on the user's back. The display is an SVGA Sony Glasstron PLM-S700E hanging from the front of a helmet, used in video see-through mode. Mounted directly above the display are a Point Grey Firefly firewire camera and an InterSense InertiaCube2 orientation tracker, and on top of the helmet is a Garmin GPS 18 position tracker. User input is through a hand-held ErgoTouch RocketMouse. All of these devices are relatively inexpensive, off-the-shelf components.

Many services now exist on the internet to provide access to aerial photographs and high-resolution satellite imagery of the world, including Yahoo Maps (at 1.0m resolution for the entire United States) [16] and Google Maps (with variable resolution) [6]. In our current system, we acquire 0.5m resolution aerial photographs offline from Google Maps and stitch them together into a single large view of the University of California, Santa Barbara campus. However, it is possible with a wireless internet connection to download map data on the fly based on the wearable's reported GPS coordinates. This automatic map acquisition would allow the system to work in new environments covered by map services without the initial setup cost.

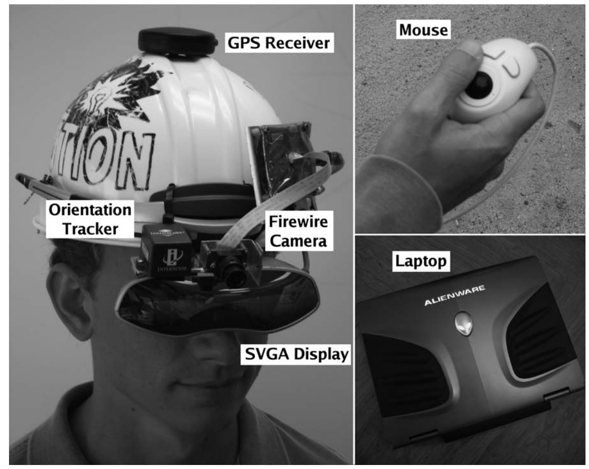

Figure 2: Our wearable system hardware. An Alienware Area-51 m5500 laptop (worn in a backpack), an SVGA Sony Glasstron PLM-S700E display, a Point Grey Firefly firewire camera, An InterSense InertiaCube2 orientation tracker, a Garmin GPS 18 receiver, and an ErgoTouch RocketMouse.

3.1 Calibration

While our position and orientation sensors report absolute coordinates in a global reference frame, both have too much error to be used without further calibration. We found that our GPS tracker can be off by as much as 10 meters, but that it also drifts slowly. To compensate for the initial offset users are asked to specify their exact location on the aerial photograph, which greatly reduces the position error for a single run of the application.

Our orientation sensor often has a large amount of error as well, particularly in the presence of nearby ferromagnetic materials. We found that this can distort our tracking results by as much as twenty degrees. To overcome this error we did two things. First, we add a second step to the calibration process. In this step the user centers the view at a distinct feature and then clicks a button to bring up the aerial photograph. The user's position and orientation are overlaid on the photograph, and the user is able to modify the displayed orientation to directly coincide with the user's chosen feature. This calibration procedure is similar to the single point calibration technique formalized by Baillot et al. [3], except that in our system the calibration location does not have to be predefined because the user chooses the necessary points on the aerial photograph during the process. However, because the user input is limited to adjusting the orientation in the overhead view, only error in yaw is accounted for. Roll and pitch can be roughly corrected by assuming the user's head is level and oriented vertically during the calibration. The second way we compensated for orientation tracker error was to integrate a modified version of our hybrid tracking system [14] which is based in part on previous work by Satoh et al. [13] and You et al. [17].

4 Annotation Interface

While aerial photographs provide many opportunities for annotations by themselves, they are especially useful in combination with

Figure 3: Example annotations as seen by the user in first person view mode. *Left to right:* (a) Corner annotations on the corners of two buildings are rendered as cubes. (b) An edge is annotated with a texture mapped onto the plane of the wall it denotes. (c) A region annotation is rendered as a wireframe bounding box. These renderings are not geared towards a particular application; rather, they are for illustrative purposes. Applications using these annotations would have visual representations tailored to their needs.

the first person views of the physical world a wearable system affords. With only an aerial photograph, it is possible to annotate many types of features, but only in 2D – modeling the accurate height of a building would be very difficult. Our wearable system provides an advantage by allowing the specification of a 3D position from the combination of the aerial photograph with the first person viewpoint. The usefulness of aerial photographs is also greatly increased when the user can be situated in the environment they are annotating. For example, it may be difficult to distinguish features in an aerial photograph alone, but when a user can stand in the scene and look at the buildings from a ground-level viewpoint, these ambiguities can be more easily resolved. Having both the aerial photograph and the first person view is analogous to having a perspective view and a top-down view in a CAD modeler – while many things can be done in either view independently, having both views is often faster and more powerful.

Our annotation system utilizes these two views of the scene to make specifying annotations very easy. First, the user casts a ray in the direction of the feature to annotate by centering it in the field of view in first person mode. Once a ray has been cast, the view switches to an aerial photograph mode, with the user's position and the ray overlaid. Then only a few simple interactions are necessary for the user to create any type of annotation at the correct location in the overhead view.

The interaction techniques we chose to use for annotating are focused on the ray that is cast from the first person view. To place an annotation, the user casts a ray in the appropriate direction, and then sets the distance of the annotation along that ray. We chose to break this interaction into two one-dimensional tasks to reduce user burden and increase the resulting precision in the less-precise wearable environment [4], and to keep the ray cast by the user central to the annotation interaction.

The three types of annotations we examine are corners, edges and regions. See Figure 1 for examples of each of these features in the aerial photograph view. The specifics of the interface for each type of annotation, as well as example applications for each type of annotation are described in the three sections below.

4.1 Corners

The most general type of feature, corners, can be used for many different applications. They are not limited to corners of buildings but can represent any feature that has a distinct, visible location in the aerial photograph. This could include objects like light poles along a street, doors of buildings, trees, or even features on the ground like street or sidewalk corners.

The most straightforward application for corner annotations is modeling. These corners could be used like Baillot et al.'s construction points [2], or as a sparse point cloud in any other modeling application. Corners could also simply be used as 3D points of interest to bind contextual information to.

Placing a corner in 3-space is a three step process in our system. First, the user finds the feature they want to annotate and centers that feature in their field of view. A mouse click changes to the aerial photograph view, with the user's position and the cast ray drawn on top. If tracking error has caused the ray not to intersect the selected feature, the user can change its direction by rolling the trackball in the direction the user wants the ray to move. When satisfied, another mouse click creates an annotation on the ray at the user's position. Rolling the trackball away from the user moves the annotation further away along the ray – once the annotation is at the same distance as the feature, a final mouse click completes the annotation. The view returns to first person mode and the user can see their corner annotation as a small cube (see Figure 3a). An important note here is that the annotation will appear at the correct height in the first person view, because the user originally cast a 3D ray in the first step.

4.2 Edges

Edges in aerial photographs are useful for many different kinds of annotations. One common use for annotated edges is placing world registered planar labels. For instance Güven and Feiner [7] use world-stabilized images in their authoring environment to localize the information they are presenting, by displaying the annotation as a billboard on an existing structure. Sets of multiple edges could also be used to model anything with sharp image boundaries, such as building perimeters, fields, pools, sidewalks and roads.

Creating an edge annotation in our system follows the same basic procedure as creating a corner annotation. The user centers the edge to be annotated in first person view, adjusts the cast ray and sets the distance along the ray. Instead of a point, the edge annotation is drawn as a line segment perpendicular to the cast ray. Once the annotation is positioned correctly, a final step is needed to adjust its orientation to align with the feature being annotated. This is done in the same way the cast ray is adjusted, by moving the trackball in the desired direction of rotation. After the annotation is fully specified, the display returns to the first person view where the user can see the new annotation (see Figure 3b).

4.3 Regions

Regions are the third type of feature we have chosen to annotate. Feiner et al. [5] demonstrate the use of region orientations by using screen oriented, world stabilized annotations to label buildings. These annotations can be particularly useful if tracking is not robust enough to support more tightly registered annotations such as edge or corner annotations. We also give our region annotations a width and depth (the x- and y-axis on the aerial photograph, respectively), so regions can very quickly be used as a rough axis-aligned bounding box. Generally, any large aerial photograph feature could be usefully annotated by regions, such as fields and parking lots, or even more visually complex semantic regions like a park full of trees or a group of buildings.

Specifying a region annotation follows the same basic steps as the corner annotation. This time, the user casts a ray through the center of the area to annotate and sets the distance along the ray so the final position is at the center. To finish the region, the user then drags out a corner of the bounding box to fit the area on the aerial photograph, and that action is mirrored for the other three corners of the bounding box. The result is the region annotation bounds the area in the aerial photograph, and its height is set to the height of the original ray at the distance to the annotation's center. Afterwards, the display is returned to the first person view, where the region annotation is drawn as a wireframe box (see Figure 3c).

5 Results and Discussion

Informal testing shows that the use of aerial photographs allows users to annotate scene features in 3D from a static viewpoint with much greater precision than was previously possible [15]. The longitude and latitude accuracy of an annotation position is limited only by the accuracy of the map and the ability of the user to manipulate the position accurately. Google Maps provides data at 0.5m per pixel resolution for Santa Barbara [6], and user input is generally accurate within a few pixels, so our final annotation precision is ± 1.5m. Since height information is computed from the ray cast by the user, its accuracy is dependent on the quality of the orientation tracking.

Once annotations are placed, the accuracy of their appearance in the first-person augmented reality view is determined by our system's tracking accuracy. Standard PC GPS units make cheap, wide-area position tracking possible, but at low accuracy. Our system regularly experiences drift of up to a few meters over short periods of time even in clear conditions. Differential GPS or a hybrid GPS and vision or inertial position tracker would improve this result. These small position and orientation errors result in apparent mismatch between annotations and image features in first-person view, even with good annotation position accuracy.

Aerial photographs bring with them a number of limitations. First of all, while we use nearly top-down orthographic images, there is still a slight off-axis view angle that causes the roofs of tall buildings to shift a small, non-uniform distance from their ground perimeter. Currently, we do not account for this effect. Actual top-down orthographic aerial photographs would alleviate this problem.

6 Conclusion

We present a novel mobile augmented reality system for the annotation of features in large, outdoor scenes. Our primary contribution is the integration of a new data source, aerial photographs, to significantly reduce user burden while increasing annotation accuracy. The result is an improved user experience for the traditional augmented reality task of outdoor annotation.

Opportunities for future work include improving tracking accuracy, supporting additional annotation types, and integrating additional data sources such a ground images, GIS data, and road maps. This work would improve the robustness and general applicability of our system. Most important however, is to remain true to the goal of anywhere augmentation. The work must continue to emphasize low startup cost and quick, easy integration to new scenes. This way, the traditional barriers to high quality augmented reality can be overcome, significantly increasing the general appeal of augmented reality solutions.

Acknowledgments

Special thanks to Ingrid Skei and John Roberts for their work on the initial prototype. This research was in part funded by a grant from NSF IGERT in Interactive Digital Multimedia #DGE-0221713, and an equipment donation from Microsoft.

References

[1] G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide: A mobile context-aware tour guide. In *Proceedings of Wireless Networks*, pages 421–433, 1997.

[2] Y. Baillot, D. Brown, and S. Julier. Authoring of physical models using mobile computers. In *Proceedings of the International Symposium on Wearable Computers*, pages 39–46, 2001.

[3] Y. Baillot, S. Julier, D. Brown, and M. Livingston. A tracker alignment framework for augmented reality. In *Proceedings of the the International Symposium on Mixed and Augmented Reality*, pages 142–150, 2003.

[4] D. Bowman. *Interaction Techniques for Common Tasks in Immersive Virtual Environments*. PhD thesis, Georgia Institute of Technology, Atlanta, GA, 1999.

[5] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In *Proceedings of the International Symposium on Wearable Computers*, pages 74–81, 1997.

[6] Google maps, 2006. http://maps.google.com/.

[7] S. Güven and S. Feiner. A hypermedia authoring tool for augmented and virtual reality. In *Proceedings of the International Symposium on Wearable Computers*, pages 89–97, 2003.

[8] G. King, W. Piekarski, and B. Thomas. ARVino – outdoor augmented reality visualisation of viticulture GIS data. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 52–55, 2005.

[9] W. Piekarski and B. Thomas. Interactive augmented reality techniques for construction at a distance of 3D geometry. In *Proceedings of the Workshop on Virtual Environments*, pages 19–28, 2003.

[10] W. Piekarski and B. Thomas. Augmented reality working planes: A foundation for action and construction at a distance. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 162–171, 2004.

[11] G. Reitmayr and D. Schmalstieg. Collaborative augmented reality for outdoor navigation and information browsing. In *Proceedings of the Symposium on Location Based Services and TeleCartography*, pages 31–41, 2004.

[12] J. Rekimoto, Y. Ayatsuka, and K. Hayashi. Augment-able reality: Situated communication through physical and digital spaces. In *Proceedings of the International Symposium on Wearable Computers*, pages 68–75, 1998.

[13] K. Satoh, M. Anabuki, H. Yamamoto, and H. Tamura. A hybrid registration method for outdoor augmented reality. In *Proceedings of the International Symposium on Augmented Reality*, pages 67–76, 2001.

[14] J. Wither. Interaction and annotation at a distance in outdoor augmented reality. Master's thesis, University of California, Santa Barbara, Santa Barbara, CA, 2005.

[15] J. Wither and T. Höllerer. Pictorial depth cues for outdoor augmented reality. In *Proceedings of the International Symposium on Wearable Computers*, pages 92–99, 2005.

[16] Yahoo maps, 2006. http://maps.yahoo.com/.

[17] S. You, U. Neumann, and R. Azuma. Hybrid inertial and vision tracking for augmented reality registration. In *Proceedings of Virtual Reality*, pages 260–268, 1999.