OUTDOOR MOBILE AUGMENTED REALITY
FOR PAST AND FUTURE ON-SITE
ARCHITECTURAL VISUALIZATIONS

DANIEL HII JUN CHUNG, STEVEN ZHOU ZHIYING,
JAYASHREE KARLEKAR, MIRIAM SCHNEIDER, WEIQUAN LU
Interactive Multimedia Lab, National University of Singapore

ABSTRACT: Mobile devices have become widely used in outdoors on-the-move applications. Our research aims at developing a system which allows users to visualize the virtual buildings, streets, historic sites, landscapes or cityscapes (in the past or future) which are rendered seamlessly with the current actual scene captured by the camera (called Augmented Reality), and interactively exploring the virtual scene in real-time while the user is moving. We use the camera PDA, the GPS receiver and the gyroscope to track the mobile device’s position and orientation for virtual scene rendering. This article documents the issues regarding accurate alignment, tracking, errors and limitations with mobile devices and presents our solutions generated so far for the research in progress.

KEYWORDS: Mobile augmented reality / mixed reality, virtual reality, visualization, navigation

RÉSUMÉ : Les appareils mobiles sont devenus couramment employés à l’extérieur dans le cadre des applications en mouvement. Notre recherche vise à développer un système qui permet aux utilisateurs de visualiser les bâtiments virtuels, les rues, les sites historiques, les paysages ou les paysages urbains (dans le passé ou le futur) qui sont fondus avec la scène réelle courante capturée par l’appareil-photo (appelé « la réalité augmentée »), et permet d’explorer de manière interactive la scène virtuelle en temps réel tandis que l’utilisateur se déplace. Nous utilisons l’appareil-photo PDA, le récepteur de GPS et le gyroscope pour dépister la position et l’orientation de l’appareil mobile pour le rendu virtuel de scène. Cet article décrit les problèmes concernant l’alignement précis, dépistant les erreurs et les limitations avec les appareils mobiles et présente nos solutions produites jusqu’ici dans le cours de la recherche.

MOTS-CLÉS : La réalité augmentée mobile / éalité mélangée, réalité virtuelle, visualisation, navigation

T. Tidafi and T. Dorta (eds)
Joining Languages, Cultures and Visions: CAADFutures 2009
© PUM, 2009
1. INTRODUCTION

1.1. Research objective

Augmented reality visualization is still not well established in the architectural practice. Current practice of showcasing design proposals have always been using physical models, 2D still renderings, video animations and virtual reality simulations to show how it will look at the current site. The design is superimposed with photos taken on site or with 3D virtual surrounding buildings and site built entirely from scratch. This will require the architects or draftsmen to go to site to take images, buy or refer to site maps and surrounding building drawings as well as build the 3D site model if required.

Augmented reality visualization does not require all this because the proposed building can be placed directly on the real site in 3D so you can view it from any angle possible (Lonsing 2004). On top of that, when you are directly at the site, you can be more responsive to the site conditions, including all the other senses (sound and touch) apart from just vision. This will enhance the design process so it will be more sensitive to the site surroundings (Kieferle and Wössner 2003). An architectural design study conducted using such a system showed favourable response because the users think it is more intuitive and easy to understand a design on site than in the studio (Fukuda et al. 2006). Apart from used in design, it is also used to revitalize the existing built environments through the architectural planning process (Donath et al. 2001).

Instead of using personal computers, our research focuses on augmented reality visualization using a mobile device. Mobile devices provide the mobility to be at the site, which gives the architect flexibility to design anytime with the luxury of 3D CAD reference (Woodward et al. 2002). The biggest challenge is how to maintain real-time reliable tracking given the hardware constraints of camera resolution, memory, graphical card capability, and CPU power. In the architectural context, the research aims at using mobile devices to augment past and future buildings, streets, historic sites, landscapes or cityscapes on location, and for interactively exploration of the virtual scenes through time in the present setting. Exploring buildings, streets or entire cities from the past is currently only possible in a museum, and is often presented in the form of static displays (models, plans, drawings, photos). Our system allows the real-time displays of both the present physical scene and the virtual scene of the past. If the virtual scene is to be displayed at exactly the same location as the present scene, the user is able to toggle between the two views.

Using the same techniques but different 3D scenes with proper registration and alignment, our system can also be used to visualize the future scene. As mentioned earlier, designs of new buildings are generally presented away from the actual building site with abstract plans or at best models being used. This stretches the viewers’ imagination, as it is often difficult for them to relate these
displays to the current location and looks, and to immerse themselves in the scenes. Our research aims to make it possible to visualize everything on site. This function provides the architects another factor to consider when they design, which is to take into account the response of the past and future buildings on site which will impact their design intention.

We propose a two step solution for outdoor mobile augmented reality tracking. The first step is to use GPS device connected to mobile device to obtain coarse location information of the mobile device (and the user). Unlike the usage of GPS to tag the location information as in previous architectural research like Sense-of-the-City (De Vries et al. 2008), landscape evaluation system (Kaga et al. 2006) and landscape architecture (Petschek 2005), we also combine the GPS information with the data from gyro device to generate a 6 degree of freedom (DOF) virtual scene. However, it should be noted that the consumer GPS and gyro devices have quite big errors that make visual augmentation misaligned. In the second step, we use the built-in camera of the mobile device to do vision-based tracking based on the physical video stream, hence to generate a refined rendering. The gyro is fast in getting orientation info but is not accurate enough for visual augmentation, while computer vision is slow but more accurate. By combining inputs from different sensors, our two steps solution can avoid the heavy computation which is generally required for the initial searching using vision-based tracking techniques, while compensating the errors carried by nowadays’ consumer mobile devices, compared to the previous work in mobile augmented reality (Fruend et al. 2001; Papagiannakis and Magnenat-Thalmann 2007; Lertlakkhanakul et al. 2005).

1.2. System overview

The main hardware used in the research consist of the single camera PDA (personal digital assistant) (HP iPAQ rw6828 Multimedia Messenger), the gyroscope (Vitec 3D Sensor TDS01V) and the GPS (global positioning system) device (HOLUX M-1000) in Figure 1. The iPAQ PDA is running on Windows Mobile 5 and generates a video resolution of 240 X 320 running at 15fps (which is real-time for mobile phones). The gyro is fast in getting orientation info but is not accurate enough for visual augmentation, while computer vision is slow but more accurate. By combining inputs from different sensors, our two steps solution can avoid the heavy computation which is generally required for the initial searching using vision-based tracking techniques, while compensating the errors carried by nowadays’ consumer mobile devices, compared to the previous work in mobile augmented reality (Fruend et al. 2001; Papagiannakis and Magnenat-Thalmann 2007; Lertlakkhanakul et al. 2005).

**FIGURE 1. THE GYROSCOPE, THE GPS DEVICE AND PDA.**
Figure 2 shows the overview of the whole system setup. Firstly, the GPS and gyroscope provide initial rough estimate of the camera location and orientation. Both these devices go through filtering to remove the jittering effects. Then, the initial camera pose is used to render (1st render) the graphical models of the present scene (not the models of the scene in the past or future yet) for further refinement of the tracking results. Next, the misalignment between rendered models and real data acquired from camera is estimated by silhouette tracking module. In the end, the final camera pose is estimated by combining initial pose and refinement obtained from silhouette tracking which is used to render (2nd render) the 3D graphical data (could be the scene of the past, present, or future) for final overlaying on real data. This is the final graphics displayed to the PDA user. The realignment process takes place for every frame of the video sequence.

**FIGURE 2. THE SYSTEM OVERVIEW.**

2. RELATED WORKS

Mobile devices are used a lot for augmented reality research these days, whether it is UMPC (Ultra Mobile Personal Computer), PDA or hand phones. The combination of using GPS, gyroscope and CAD on mobile devices, especially GPS and CAD has been done in a few augmented reality research projects (Lonsing 2007). There are those which use the client-server network to run the whole system (Pasman and Woodward 2003). They use various data compression methods to enable faster transfer, even if it means handling large 3D models (Pasman et al. 2004). Our approach is to avoid worrying performance issues through various communication networks, therefore not relying entirely on the client-server network performance. We also rely more on realistic textures than large 3D models.

The VTT Technical Research Centre of Finland’s Augmented Reality Team started doing research in outdoor augmented reality since 2003. The use of
large size markers to position buildings caused users’ mobility to be limited (Woodward et al. 2007). Their upgraded version later is based on determining the building’s intended location by placing the 3D model in Google Earth, saving the information into a separate KML file (Honkamaa et al. 2007). The user’s location is determined by GPS, which is linked wirelessly to the handheld device. The use of GPS with gyroscope and computer vision techniques has been done (Azuma et al. 1999) but with assumptions like viewed objects are at a distance. Their tests are to track far distance structures only and assuming it is a static scene by annotating labels onto the tracked structures while ours are meant for real-time navigation from far to near distance tracking.

There are many others which integrate all or some of the gyroscope, the GPS device, computer vision techniques combinations to have augmented reality visualizations (Feiner et al. 1997; Hu and Uchimura 2006; Jiang et al. 2004; Reitmayr and Drummond 2006; You and Neumann 2001). Our research project has all those components integrated, without relying on a server and to be able to move freely in mobile to track 3D models in real-time.

3. 3D MODEL AND TEXTURE

The 3D model of the university campus built in Autodesk 3ds Max is used for the research project. The sources of reference are drawings of every single building’s plan, section and elevation as well as the terrain map with point height information. It consists of the terrain and building 3D surface models. The textures consist of photographs taken on site perpendicular to the elevation of the building surfaces. Every of these images are later stitched to combine into a single texture to be mapped over a building. Therefore, every single building’s texture consists of at least a combination of 10 images.

Since the rendering is done with Microsoft Direct3D Mobile, the 3D model has to be exported out in the Microsoft DirectX .x format. 3ds Max does not have such export plugin and Microsoft no longer produces such plugin for 3ds Max and Maya since its June 2007 SDK (Software Development Kit). There are two independent 3rd party plugin creators, namely Pandasoft and kW X-port. The texture format used in the research is the uncompressed BMP (bitmap). The Direct3D Mobile should also be read to read JPG, PNG, GIF and DDS texture formats but there is no success so far in loading them to run in the mobile device we used. We believe that it can be done and will surely improve the rendering speed since compressed formats like DDS, JPG, GIF and PNG are much smaller in file size and therefore lighter to run.

Both the Google Earth / Map and the terrain map are used to assist the accurate placement of these buildings on site. In addition to these sources, the individual drawings of the respective buildings on site help to make sure every building are created accurately like how they are in the physical world.
3.1. Google Earth X & Y axis alignment

As the first step to the research, the exploration to free access to satellite images with GPS data from Google Earth and Google Map are used as shown in Figure 3. It acts as a valuable reference point to the exact positioning of the 3D buildings. Satellite images are not 100% direct perpendicular to the non-flat earth surface so it will be less accurate. The images are also in low resolution with possible appearance of clouds to block the view. The altitude / height data is also not very accurate.

**FIGURE 3. ALIGNMENT WITH GOOGLE MAP.**

3.2. Terrain map X & Y axis alignment

The terrain map is in high resolution and extremely accurate with detailed terrain height / altitude point values as shown in Figure 4. The disadvantage of using it is the lack of updates so many alterations to the terrain as well as new buildings constructions are not added to the map. Still, the altitude / height information is crucial to accurately place the 3D virtual buildings on the terrain before augmentation.

**FIGURE 4. ALIGNMENT WITH TERRAIN MAP.**
3.3. Height alignment

Height alignment is used from stitching the video frames taken on site over the several buildings as shown in Figure 5. The tricky part of this alignment is the exact position of the camera on site as well as the orientation of the camera in terms of its yaw, pitch and roll values before doing an accurate alignment. From experience using this method, it should only be used as the last resort. The usage of individual building drawings with the terrain map point height information is still a better choice as shown in Figure 6.

**Figure 5. Multiple frames stitched from video taken on site.**

**Figure 6. Height alignment with terrain map.**

3.4. CAD and GPS coordinate systems geo-referencing

The final step before augmentation is to map the GPS geodetic coordinate system with the origin (0,0,0) of the 3ds Max Cartesian coordinate system for the entire virtual. We manually did this approximately through Google Earth map and 3ds Max 3d world reference. This will make sure whatever 3d model that is imported into the system will be converted to the geodetic coordinates. There is a research which does this automatically, converting the geodetic
coordinates to the Cartesian coordinates in the AutoCAD modeling system (Kouzeleas 2007).

4. CURRENT MATCHING TECHNIQUES

We need accurate camera position and orientation for successful augmentation. However, the current GPS and gyroscope we use are not having extremely high accuracy. Therefore, we need computer vision based approach to estimate the correct camera parameters. There are many tracking methods which uses wireframe or surface model.

Wireframe modeling is a method to export out only important splines or edges to represent the specific building to be tracked. It is used because it can be done pretty fast. It has been done to track bridges (Comport et al. 2007; Wuest and Stricker 2007; Simon and Berger 1998) and industrial objects. However, not all 3D file formats support spline rendering. Those that do not support it will require some additional input to translate it out for possible tracking. IGES format is one example which does support spline rendering. The DirectX .x file format only support meshes export.

4.1. Feature point tracking

Feature tracking is not suitable for model based approaches as the available features can be difficult to be tracked. These include illumination changes, inaccuracies arising from texture acquisition and mapping process etc.

Edge tracking is robust as feature involved are edges which are detected under varying illumination conditions and fast as matching is performed in one-direction only, i.e. the direction perpendicular to edge orientation (motion along the edge is not perceived due to aperture problem). We propose silhouette tracking approach as edges corresponding to model and real image gets cluttered due to large viewing distances which renders the edge tracking approach impractical for AR applications. Camera based tracking is employed as a fallback mechanism whenever silhouette tracking fails due to unavailability of clear outline, too small viewing distances.

4.2. Silhouette tracking

Silhouette tracking is performed on grayscale images and overall approach is presented in tabular format. The approach assumes the presence of well-structured sky-line within the field of view of the camera.

Edge extraction is performed using Canny edge detection on the grayscale version of these images obtained from camera feed and rendered data. As mentioned earlier, due to large viewing distances edges get cluttered making them unsuitable for tracking without further processing which is obvious from
these images. Extraction and tracking of outline is difficult in urban scenarios due to high density of objects, presence of other objects/subjects such as trees, traffic, pedestrians etc. We resolved these issues by using silhouette tracking. As mentioned, silhouettes are free of cluttered edges, easy to extract and track as opposed to previous approaches. From these correspondences, camera parameters are obtained by assuming orthographic camera projection. All the steps are shown in Figure 7.

**Figure 7. Silhouette Tracking for Correct Estimation of Camera Parameters.**

5. SYSTEM TESTING

Our approach in the research is to make the whole system work robustly on the laptop. Later, the implementation is done to the mobile PDA version. In the initial field test, an open field was chosen beside the campus football field. This choice is to avoid any surrounding structures or plants to interfere with
the GPS data reading. We used a laptop attached with the webcam, gyroscope and GPS device for our test as shown in Figure 8.

There are twelve 3D models being viewed in the test. The triangles count ranges from 44 to 777 while the vertices count ranges from 40 to 550. The .x DirectX file size ranges from 16 to 245 KB. The BMP texture resolutions used are 512 X 512 pixels and 1024 X 1024 pixels. The texture file size ranges from 769 KB to 1026 KB. The idea is to use as minimum surfaces as possible to represent a building and be dependent of the textures to represent them. This approach is chosen to lessen the burden of the limited processing power of the mobile devices. The PDA can render the 3D model at around 30 frames per second to augment it on the video in real-time as shown in Figure 9. Currently, under the open sky condition, we can track all other surrounding buildings well. The result may be different in conditions where the sky illumination contrast is not huge enough to differentiate the building from the sky as well as obstructions by surrounding buildings at the background with higher height, bus stops, trees and lamp posts obstruction the edges. As mentioned in Section 4.1 above, feature point tracking will be used as a fallback mechanism.

**FIGURE 8. THE WEBCAM, THE GYROSCOPE AND GPS DEVICE ATTACHED TO THE TRIPOD.**
During our research in this project, we encountered some problems:

- **Access to the back camera:** Currently, we successfully augment the 3D models on the older PDA version which only has the back cameras. The new generation PDA and mobile phones have two cameras (front and back). The manufacturers are not willing to provide the camera access interfaces required to select the use of either camera.

- **Flickering display:** For this flickering version of our program, each frame of the display is presented twice. The 3D models are rendered to the device; the live video stream is rendered to the DirectDrawSurface. The lock texture is another approach where the background would be rendered to a dynamic texture on a flat model at the back of the scene. The problem here is that since this would be fine on the desktop, the DirectX Mobile does not allow
us to lock the texture so we can’t draw to it. We may switch to OpenGL ES to solve this problem.

- **Gyroscope and GPS device reading accuracy**: The gyroscope can only give 10 data readings per second. This could be because of the translation of data from serial to USB output. The GPS device we are using has a resolution of 10 to 15 meters.

- **The z- / DEPTH-buffer**: The access to the z-buffer, which manages the image depth coordinates in 3D graphics, is still not possible in the mobile version. This is because the Direct3D Mobile does not support the depth buffer formats required for us to access and read the buffer surface information. The accessibility is required to enable the retrieval of depth coordinates for the tracking to work to estimate the buildings’ exact positions in 3D.

### 6. CONCLUSION AND FUTURE WORK

In this paper, we propose a two step solution for outdoor mobile augmented reality tracking. The first step is to use GPS and gyro device connected to a mobile device to obtain coarse location and orientation information of the mobile device to generate a 6 degree of freedom virtual scene. The second step is to use the built-in camera of the mobile device to do vision-based tracking based on the physical video stream, hence to generate a refined rendering. By combining inputs from different sensors, our two steps solution can avoid the heavy computation which is generally required for the initial searching using vision-based tracking techniques, while compensating the errors carried by nowadays’ consumer mobile devices.

Based on this technique, we also build application which allows user to see the real-time displays of both the present physical scene and the virtual scene of the past (or future). If the virtual scene is to be displayed at exactly the same location as the present scene, the user is able to toggle between the two views. This function provides the architects with another factor to consider when they design, which is to take into account the response of the past and future buildings on site that will impact their design intention. They can take snapshots of particular views on site with the past or future developments together with their proposed design sketches or CAD model superimposed on the scene which can be used later for presentation.

There are several possible ways to further improve the proposed system. In terms of the accuracy of the positioning and orientation of the user at the site, a better quality gyroscope and GPS device can help improve our work. A gyroscope with higher amount of readings per second will help minimize errors. We are currently testing the OS5000-S gyroscope, which is reading at 40fps and more accurate. A DGPS (Differential GPS) device can also improve the accuracy of reading the position of the user on site. The DGPS uses a fixed ground station as reference to help improve the data resolution to minimize errors.
We currently have broadcast services available in the country by the Maritime and Port Authority of X and X Land Authority. It is quite accurate that it has been used for site measurement for preliminary architectural design process (Tedrumpun and Nakapan 2004). Another choice is the Real-Time Kinematic (RTK) GPS system which is quite identical with how DGPS works (Knight et al. 2006). This will improve the current inaccuracy issue of our GPS device which can be caused by atmospheric effects, satellite geometry, measurement noise, ephemeris data, satellite clock drift and selective availability (degraded by the USA Department of Defence).

Apart from that, the possibility to work with the company who owns the proprietary software to access any of the cameras on the mobile devices will help significantly. The access granted will enable the research of higher end latest mobile devices which can run our application more efficiently. Otherwise, another option is to explore the possibilities of using the UMPC or perhaps another mobile operating system like Symbian. This is to overcome the z-buffer accessibility since the non-mobile operating systems do allow access to the z-buffer fully as well as solve the flickering problems.

We believe mobile devices will exponentially become more powerful like their desktop counterparts. We will definitely upgrade our 2006 iPAQ to a more updated version as the research progresses. The usage of compressed textures (DDS, JPG, PNG, GIF) on 3D models will also improve the rendering performance as they consume less texture memory.

ACKNOWLEDGEMENTS

The project is funded by Singapore A*Star Thematic Strategic Research Programme (TSRP) on Mobile Media, Project Title, I-Explore Interactive Exploration of Cityscapes through Space and Time, WBS No. R-263-000-458-305. We will like to thank Dr. Christian Gilles Boucharenc for helping us to translate the abstract and keywords to French.

REFERENCES


Donath, D., Beetz, J., Grether, K., Petzold, F. and Seichter, H. 2001, Augmented Reality Techniques for Design and Revitalisation in Existing Built Environments, Reinventing the


Lonsing, W., 2007, Combining GPS and CAD, Predicting the Future [25th eCAADe Conference Proceedings, Frankfurt am Main, pp. 655-661.


