Handheld Augmented Reality

Distributed System Seminar FS 2013

Reto Lindegger
Student, ETH Zurich
lreto@student.ethz.ch

INTRODUCTION

The idea of computers helping people in their everyday life has been around for quite a while. One way a computer can be used to assist a user to accomplish a task is by providing important information relevant to this task. But to know which information is relevant, which information is needed at the moment and how to display this information in the best possible way is a difficult task to solve. Luckily, a computer can take advantage of a wide range of input devices like different sensors, cameras and user interfaces as well as sophisticated devices for displaying data like screens and projectors. The combination of processing input from the environment, selecting relevant data and displaying the data in a for the situation suited fashion is called augmented reality (AR).

Augmented Reality

A clear and proper definition of the term augmented reality is very difficult since augmented reality applications can have big differences in characteristics, purpose and usage of the input from the environment. When looking for a suitable definition, one is likely to find the work of Ronald T. Azuma about this subject. In his paper ”A survey of Augmented Reality” [1], he defines Augmented Reality as systems that have following characteristics:

- Combination of reality and virtual objects or information
- Interactive in real time
- Registered in 3D

He also explicitly states that rendered movies and 2D overlays on live video streams are not counted as augmented reality since a rendered movie does not comply with the second and a 2D overlay does not comply with the third requirement.

However, this definition is not the only one and as we will see later, there are some applications which can be seen as augmented reality application but violate one or more points from the definition above.

Another definition of augmented reality is the one on Wikipedia: ”Augmented reality (AR) is a live, direct or indirect, view of a physical, real-world environment whose elements are augmented by computer-generated sensory input such as sound, video, graphics or GPS data.” [2] This definition is less strict about the appearance of information or virtual objects in 3D, but simply says that the reality is augmented by sensor input. How the augmentation is done is left open.

In this definitions, nothing is said about the computer and the display used to generate and present the augmented reality. It can be a fixed installation in a room or it can be a integrated in a helmet. An augmented reality system can also be designed to fit in a user's hand. This type of AR systems is called handheld augmented reality.

Handheld Augmented Reality

As the name suggests, handheld augmented reality describes AR systems which can be hold by the user in his hand. So the main feature and also a great advantage is that such a system is very portable. The user can take the system to the place where it is needed.

With the ongoing improvement of smartphones, the development and deployment of AR systems gets easier and cheaper. The use of smartphones for AR applications has many advantages. Beside the already mentioned low price, they are also very widespread which makes it easy to distribute an application among the users. Almost everyone has a smartphone nowadays and downloading an app is as simple as it gets. Another advantage is, that they are by default equipped with a lot of useful sensors like accelerometer, gyroscope, camera and GPS. Smartphones are commodity hardware, which means nothing has to be produced only for the purpose of an AR application. All the required hardware already exists and can be used to design and produce a tool for helping the user in any situation.

However, smartphones or handheld AR systems in general have some limitations. First of all, they have limited computational power. Modern smartphones are well equipped and have incredible CPU speed and a lot of Memory, compared to desktop computers from some years ago. But still, there are tasks which are much easier to solve with a high end PC. Then of course battery can be an issue, as with all portable devices. Another drawback is that some tasks are very difficult or almost impossible without a supporting infrastructure. For precise localization, a infrastructure like Satellites (GPS) or some sensors or transmitters (indoor localization) are needed. So sometimes, the smartphone or portable AR device is just not enough. The last drawback I'd like to mention is linked to the user's convenience. When using a handheld augmented reality system, it has to be held in the hand all the time. So one hand is always busy holding the device while the other hand may be busy interacting with it.

Despite these limitations, handheld augmented reality devices may be of great help to the user in different situations. When carefully designed and tested, they can assist the user in his everyday life to accomplish tasks, which otherwise would be much more difficult and/or time consuming. In the rest of this document I shall present 4 different augmented reality applications. I will discuss their purpose, the problem they want to solve, their strengths and weaknesses and their usability.

INDOOR NAVIGATION

The first application I'd like to present addresses the problem of navigation inside a building. Since GPS and other satellite based navigation systems are not available inside a building, conventional navigation solutions are not feasible. Another, more suited method for indoor navigation has to be found.

Problem Statement

Now lets look at the problem that this application tries to solve. Assume you are in a building you have never been before. It is a large and complex building and finding an office, a meeting room or the nearest bathroom on your own is quite tricky. Wouldn't it be nice, if your phone could give your directions? As mentioned above, navigation with GPS is not possible indoors. To overcome this limitation, another way of localizing the user has to be found. An operator of such a system on the other side is of course interested in keeping the effort needed to install a localization infrastructure and therefore the required cost low. As addition, as in all navigation systems, the presented solution should be as accurate as possible to increase usability.

Previous Work

Before I present a solution which meets the described requirements, let me show some attempts in providing an accurate and usable indoor navigation system.

Sensing Infrastructure

One way of localizing the user is by instrumentalizing the environment with a mesh of sensors. There have been several proposals for indoor navigation systems with a sensing infrastructure, of which I'd like to mention 3 interesting projects:

- Cyberguide by G.D. Abowd et al. [3]: Localization with infrared
- The BAT system by M. Addlesee et al. [4]: Localization with subsonic waves
- Evacuation system by L. Chittaro and D. Nadalutti [5]: Localization with RFID

The problem with these solutions is obviously that they are heavily infrastructure dependent. This is maybe applicable in a small area or when high accuracy is very important. However, to equip a large building with this dense infrastructure is not realistic and would also be very costly.

Sparse Infrastructure

Since a dense grid of sensors is complex and expensive, a more cost efficient solution is to have sparse infrastructure. The idea is to deploy checkpoints in selected spots all around the building. At these checkpoints the user is located, he is informed where he is and how he can continue to get to his destination. On the way between checkpoints however, the user is not assisted and has to find his way without help.

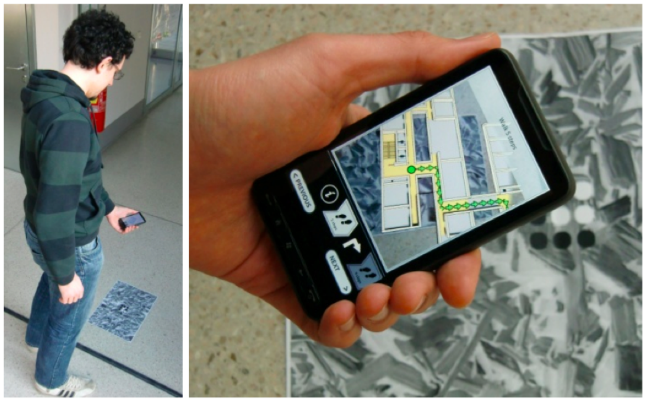

Figure 1: Indoor Navigation. User standing at info point, map with current location and path to destination.

Movement Measuring

This method of navigation does not rely on any infrastructure at all. The system measures the movement of the user with the help of different sensors like accelerometer and gyroscope. When the start point is known, the system can so guide the user through the building to the desired destination and provide a turn-by-turn navigation. The problems here is that the localization of the user becomes quite inaccurate over time when there are no checkpoints in between where the system could recalibrate the user's location. Also, it is hard to distinguish whether the user actually moved or if the user just moved the handheld device around.

Solution

When you have some halfway good solutions with different advantages and drawbacks, the best thing to do is combining them to get an accurate, reliable solution with high usability. This is what Alessandro Mulloni and others in their paper ”Handheld Augmented Reality Indoor Navigation with Activity-Based Instructions” did [6]. They combined exocentric and egocentric navigation to get an accurate indoor navigation system. To keep the costs low, they didn't use a dense infrastructure but rather just checkpoints all around the building. The checkpoints (also called info points) are in this case floor-mounted posters which can be detected by the back-facing camera of the smartphone (a seen in figure 1). Of course other technology like RFID could be used. When a user arrives at a checkpoint, the system knows exactly where he is, can recalibrate the navigation system and update the user's location. On the other side, the user also knows where he is, since the system can show him a map with detailed information about his location and the path that lies ahead. When the user is moving between the info points, the movement of the phone are measured to get an approximation of the user's location. This information is used to help the user find the next checkpoint by giving him directions in terms of so called activities. The activities are instructions like ”walk 4 steps”, ”turn right”, ”walk 6 steps”. The user interface is adapted to the current system state: at info points, a map with the exact location is shown. Between info points, no map is visible since the exact location is not known, but directions and navigation activities are shown. The two different user

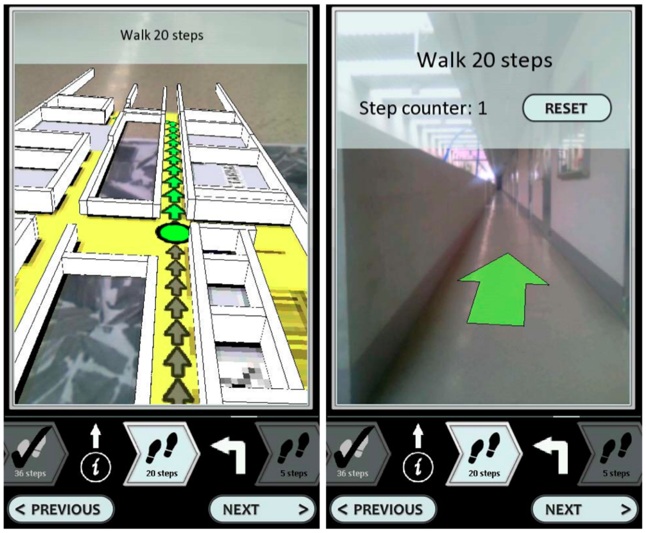

Figure 2: Indoor Navigation. Left, map with current location and path to destination. Right, view as seen between info points with directions and navigation activities.

interfaces can be seen in figure 2.

With this technique, the user is assisted in finding the next checkpoint and the system can recalibrate from time to time to minimize the errors made with the approximated localization that results from the sensor measurements and thus increasing the accuracy.

Evaluation

The system can combine the advantages of a sparse infrastructure and the movement measurements and on the other hand minimizing the disadvantages of these two systems like the cost and missing accuracy. The combined solution can be used to navigate indoors with little effort in the infrastructure and which good accuracy. A possible field of application could for example be a museum, where people should be guided through the museum or where visitors want to find specific exhibits or the nearest bathroom. Another place where this system could be used is an airport, where people are looking for the right gate or the correct baggage claim. In general, indoor navigation is most useful when people are spending a long time in a complex building (like museums) or when people are often in said building and try to find different destinations (like lecture rooms in an university).

SITE LENS

Another application which has also to do with walking around is Site Lens. But unlike in the previous presented application, here the user is outside and is not interested in finding his way but rather in getting information about his location. This application is designed to be used by architects, urban planners and urban designers.

Problem Statement

Architects, urban planners and urban designers like to visit a site before the design or planning activity. They do this to learn more about the location and the environment and to get a feeling for the place. They are interested in different values and measurements like carbon monoxide level, demographics, traffic flow, air quality and sunlight intensity. The problem is that this is much data and it all comes from different sources like statistics, tables, maps and perhaps photos. The idea behind the project I want to present is to provide the user with the desired information right at the spot where he needs it and to visualize the data right in the location where it was measured.

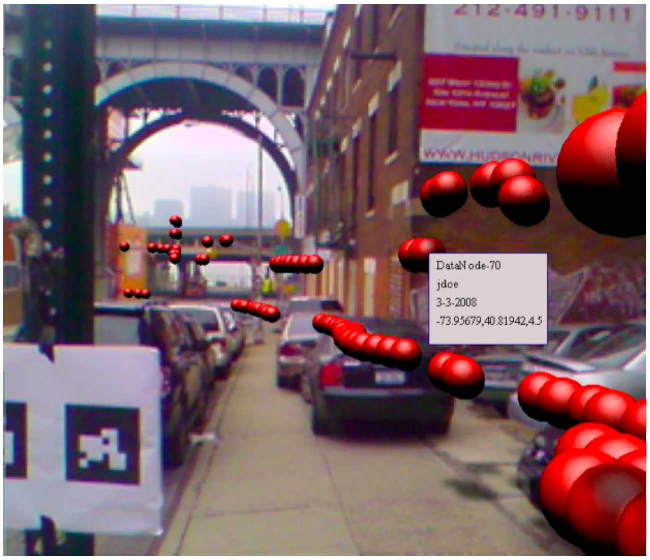

Figure 3: SiteLens. Live video feed with visualized measurements at the location where it was taken.

Previous Work

A similar project but with a slightly different objective is the Vidente Project [7]. Its purpose is the visualization of subsurface features like pipelines or power cables. This information is integrated in the video feed of the back-facing camera of a smartphone. So when the user looks at the display and points the camera to the ground/street, he is able to see the subsurface infrastructure right where it is located like in a X-Ray image.

Solution

For displaying data and measurements on the smartphone Sean White and Steven Feiner have created the application SiteLens [8]. The idea is that most measurements and other data has a location where it is recored or where it belongs to. For example a CO measurement can be coupled with the coordinate where it was taken, traffic and critical spots for traffic congestions can be linked to specific roads or junctions.

In combination with the smartphone's location (GPS) and its orientation (gyroscope, compass) the data can then be visualized in the video feed of the camera to represent it exactly at the spot where it belongs. When the user then looks at the display, he can clearly see the data integrated in the environment and can so get a better feeling for the data and the location. How the user sees the data and its visualization in the video feed can be seen in figure 3.

Evaluation

The system is great to make the process of visiting a site and comparing the data with the location easier. The user can so see the data in relation to its relevant location. What is not yet integrated but would be a good improvement is the acquiring of data. So far, the system only can display measured data. The possibility to collect and share data while walking around would be a good idea and increase the amount of available data and thus be a great advantage to all users.

LOOKING AT IMAGES

The next application I'm going to present tries to solve another task. Here it is not data that is visualized in a video feed. Actually, unlike in the other two examples, the video feed of the back-facing camera is not used at all and neither is the location. However, this time it's the front-facing camera that is used. But more about this later, first I'm going to show the problem that is to be solved.

Problem Statement

As digital camera improve, the produced pictures are getting bigger and bigger. But it's not only the resolution that improves over time. Today's pictures are not always flat 2D photos anymore. There are panoramic 360 images, multi-perspective and multi-view images and much more. Of course there are also the normal 2D stills, of course in high resolution to get the best possible result.

On the other side, the displays where these pictures are being viewed are not getting much bigger. If anything, they are getting smaller since pictures are often viewed on mobile devices like phones or tables nowadays. The question is now how to make the viewing of large imagery as convenient as possible on these mobile devices.

Previous Work

Of course, there has been a lot of people thinking about this problem and there are also some quite interesting techniques and solutions to solve it. A well-known solution is Google's StreetView. When using StreetView on a mobile phone, one can spin around 360 and move the phone up and down to navigate in the panorama and to see all possible spots in the picture. This hand free interaction with the phone is accomplished by using the input from the available sensors (accelerometer and gyroscope). Another similar product is TourWrist [9], where users can create 360 panoramas, share them with other users and explore the pictures my spinning around and moving the phone.

The Glasses-free 3D display [10] is also an application for smartphones for viewing images but with another goal, namely to see 3D pictures. Imagine an object which was photographed from different angles. Of course on the phone you can only watch one picture (one perspective) at a time. But by tracking the face of the user, smartphone application can decide which picture from what perspective should be shown to create the impression of a 3D object where the user can see different sides of the object by moving his head around.

Solution

Instead of sticking to one type of input, Neel Joshi, and others [11] combined different input sources to get a convenient and usable way of interacting with the device for viewing large imagery without having to use the hands. Controlling the phone with the hands while viewing images has some drawbacks. First of all, the hand that is used to control the application obscures a big part of the display and thus hides the picture that is displayed. Another disadvantage is that the application can't always distinguish between a dragging or an interaction with the picture like drawing.

Also, they wanted to create an input interface that doesn't require spinning in place for 360 for watching a 360 image. For this reason, they combined the input from the gyroscope with face tracking using the front-facing camera. So like in the real world, were we analyze an object by rotating the object or by moving our gaze relative to the object, the navigation is done by either rotating the device or by moving our head and therefore looking in a different angle at the screen. The application moves the picture according to device's orientation and the angle in which the viewer looks at the screen. With this technique, a smooth navigation in the picture is possible as well as the already described viewing of 3D objects by moving the head relatively to the screen.

Evaluation

Using face tracking as input for viewing images is a great idea and very interesting for creating the impression of a 3D screen with 2D images. Hand free smartphone interaction is a good way to improve usability and convenience when using smartphones and aims to make the interaction easier and feel more natural.

The main problem with this new application is that most users already are familiar to the traditional interaction techniques like zooming and panning by finger movements. When introducing a new technique, it first feels unnatural and takes some time to get used to it.

KINECT FUSION

So far we have seen an indoor navigation system, a sophisticated solution for displaying measurements in the environment where it was taken and an application for viewing large images without having to touch the screen. In all these systems, the goal is mainly to display data or information for the user. Now I'm going to present an other application of augmented reality which does not only display data, but is also capable of acquiring enough data do represent the environment as a 3D model. In this AR application, the handheld device is not a display but actually a camera that is used to reconstruct the environment. So let's first look at the task that this system is trying to solve.

Problem Statement

As stated above, the goal is to get an accurate digital reconstruction of the environment. For this purpose, the user should be able to walk around with a camera in his hand which captures the necessary information that is used to build a replica of the real world. The idea is to do this in real time, so that the model is build at the time the camera is moved around. The real time constraint is also necessary to notice changes in the scene and to react according to these changes. We will see later how the system can react to sudden changes in the scenery. The 3D reconstruction should also be accu-

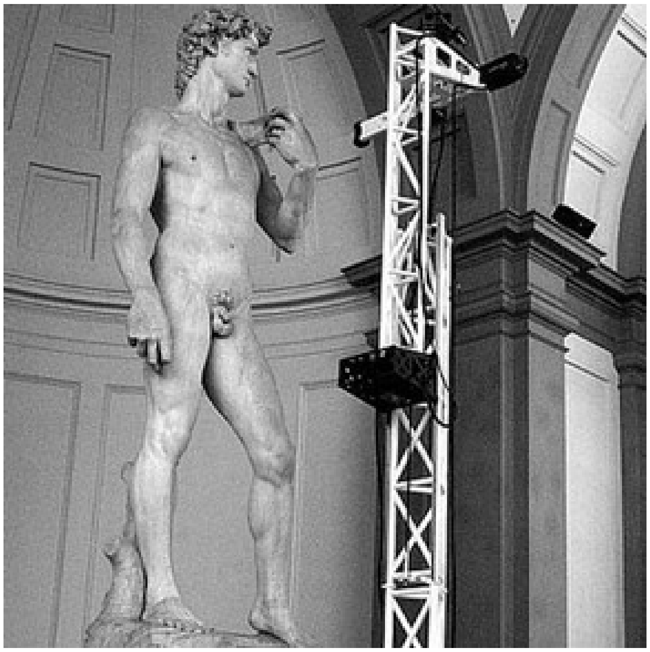

Figure 4: Michelangelo Project. Large infrastructure is used to scan the statue.

rate enough to enable physically correct interaction with the model. Furthermore, the whole system should be infrastructure less, which means that the system should only consist of the camera and a connected computer to process the image data.

Previous Work

The idea to capture the reality on images (video or still images) and generate an 3D model out of it is not new and there are already some interesting solution around. The first on I'd like to mention is the Digital Michelangelo Project [12]. The system uses laser rangefinders and cameras to scan a statue and create a virtual replica of the statue. The authors choose this name because they tested the system on Michelangelo's David statue. While the system produces a quite detailed high quality 3D model, it has some drawbacks. First of all, it is large and heavy as you can see in figure 4. The second drawback is that the scanned object or scenery can't move during the scan. While this is no problem when scanning a statue, it can't be used to scan a room with an interacting person in it. So this project does not meet the previous listed goals.

Another project which also aimed at generating a 3D model by capturing images was developed by P. Merrell et al. and described in the paper *Real-time visibility-based fusion of depth maps* [13]. First, they generate several depth maps from pictures captures by a moving camera. Then they combine these depth maps to build a 3D model of the captured scene. Again, the problem is that the model is not generated in realtime and the system does not allow a moving or changing scene. However, it is well suited for modeling large objects, like a building for example.

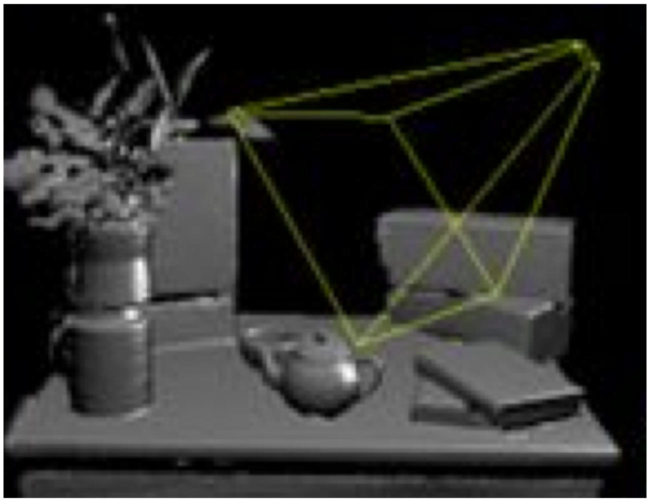

Figure 5: KinectFusion. Reconstructed scene and camera position.

Solution

We have seen two good solutions for modeling immobile scenes and objects. Now I present a new approach which generates a 3D model in real time and also allows interaction with the virtual replica. The project was developed by a group at Microsoft Research and is called KinectFusion [14]. As the name already reveals, the system uses a Kinect sensor to generate the 3D model. The advantage of this is that the Kinect is commodity hardware, which means it is easy to get and cheap. It is also quite mobile and can be moved around. Furthermore, it already provides a depth map as well as an RGB video stream.

To generate the model, only the depth map is used. However, the RGB images can be used to generate texture for the model. The idea of the project is, that the model evolves over time as the camera is moved around. Moving around the camera leads to new perspectives: the scanned objects are seen from a different angle and more detailed are revealed to the camera. So basically every camera movement adds more details to the reconstruction. In figure 5 you see a reconstructed scene with a table and some objects on in as well as the current position and angle of the Kinect sensor.

The interesting part is now the interaction with the objects in the scanned scene. A user can move an object in a scanned scene and this object is then separated from the rest. So basically the user can get a 3D model of a small object by moving it round in the scene and separating it so from the background, which remains static. But not only can the user interact with already existing objects. Virtual objects can also be placed inside the 3D model and interact with the replica of the reality. The interaction can be physical, for example when virtual particles bounce from the reconstructed 3D model (shown in figure 6), or it can be visual, for example when a virtual object obscures a real world object and casts a shadow on the scene or when the real world object are reflected on a virtual object. The last type of interaction I'd like to mention is the touch interaction. The system can distinguish between static background and moving foreground

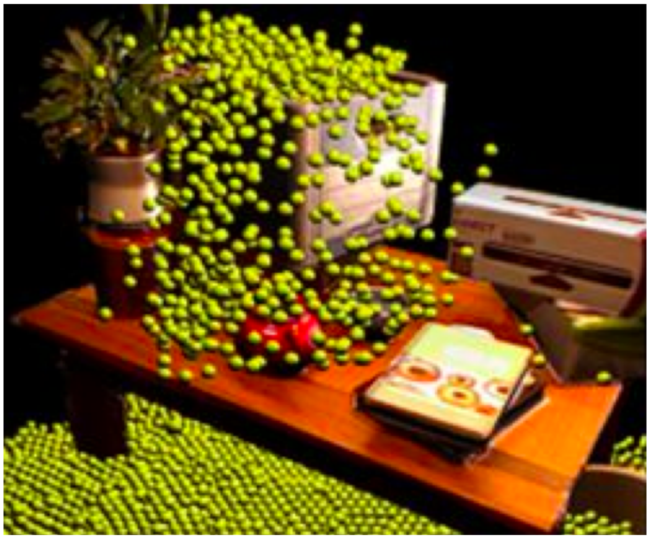

Figure 6: KinectFusion. Interaction of virtual particles with scene reconstruction.

(a user for example) and can then calculate intersection between the two. These intersections can then be recognized as touch interaction which allows multi touch recognition on every surface.

Evaluation

This very interesting project generates a 3D model from a scanned scene in real time and also allows to interact with this model. It meets all the previous listed requirements and is truly a fascinating technique. Possible applications of such a system are numerous and very promising. For example it could be used in robotics, where a robot could scan the environment in realtime and use the generated model for navigation. It could also be used for planning or designing, for example by a interior architect. Of course, an application in entertainment is also highly possible.

Nevertheless, there are also some disadvantages in this system. First of all, while the camera is small and very mobile, the connected computer might not be. Even though a laptop could be used, it's still not as mobile as for example a smartphone and thus it doesn't quite qualify as handheld augmented reality. The second disadvantage is that the visualization of the reconstruction is on a screen which might not be in the user's field of view. So when a user interacts with the scene or even with a virtual object that is placed in the reconstruction, he might not actually see what he is doing. The combination with other techniques could be helpful in this situation, for example the combination with a projection based system.

CONCLUSION

We have seen 4 different applications of handheld augmented reality. One application for indoor navigation, a system for visualizing measurements on the spot where they were taken, an application for viewing large images and an application for capturing the reality and generate a 3D model. Now let us look at each application and compare it to the criteria for handheld augmented reality.

Indoor navigation:
This application meets all requirement of the definition from above [1]. It mixes reality with virtual objects in the navigation view and the visualized walking direction which is place in the image in 3D, it is in real time and interactive and reacts to the users movements. As a smartphone app it is also very portable.

SiteLens:
This application also meets all the requirements. It is portable, it mixes reality with virtual objects by displaying measurements in the video feed in 3D and it is interactive.

Touch-less phone interaction:
This application does not meet the definition of Azuma [1], but since it is interactive, in real time and reacts to sensor input, it counts as AR according to the Wikipedia definition.

KinectFusion:
Here, the definition of Azuma applies. All the listed requirements are fulfilled. However, the system is not truly handheld, since an additional computer is used which does not fit into a hand. Also the user might not be able to see his actions in the generated 3D model when the screen for visualizing the AR is not in his field of view.

As we see, a clear definition of handheld augmented reality is very difficult and its application can have many form. While these presented projects are all quite different, they all have in common that they try to assist the user in his everyday life or in his work.

REFERENCES

1. Ronald T. Azuma. A Survey of Augmented Reality. In *Presence: Teleoperators and Virtual Environments 6* (August 1997), 355-385

2. Wikipedia, article about augmented reality, *http://en.wikipedia.org/wiki/Augmented_reality* (15.4.2013)

3. Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R., and Pinkerton, M. Cyberguide: a mobile context-aware tour guide. In *Wireless Networks 3* (1997), 421-433.

4. Addlesee, M., Curwen, R., Hodges, S., et al. Implementing a Sentient Computing System. *Computer 34* (2001), 50-56.

5. Chittaro, L. and Nadalutti, D. Presenting evacuation instructions on mobile devices by means of location- aware 3D virtual environments. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (2008) 395-398.

6. Alessandro Mulloni, Hartmut Seichter, Dieter Schmalstieg. Handheld Augmented Reality Indoor Navigation with Activity-Based Instructions. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services* (2011) 211-220.

7. Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B., and Schmalstieg, D., Handheld Aug-

mented Re- ality for Underground Infrastructure Visualization. In *Journal of Personal and Ubiquitous Computing* (2008)

8. Sean White, Steven Feiner. SiteLens: Situated Visualization Techniques for Urban Site Visits. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2009) 1117-1120.

9. TourWrist, *http://tourwrist.com*

10. 3D displays on mobile devices: HCP. Engineering Human-Computer Interaction Research Group. *http://iihm.imag.fr/en/demo/hcpmobile*

11. Neel Joshi, Abhishek Kar, Michael Cohen. Looking at You: fused gyro and face tracking for viewing large imagery on mobile devices. In *Proceedings of the ACM SIGCHI conference on human factors in computing systems* (2012) 2211-2220.

12. M. Levoy et al. The digital Michelangelo project: 3D scanning of large statues. In *ACM Trans. Graph* (2000)

13. P. Merrell et al. Real-time visibility-based fusion of depth maps. In *n Proc. of the Int. Conf. on Computer Vision (ICCV)* (2007)

14. Shahram Izadi, Richard A. Newcombe, David Kim, Otmar Hilliges, David Molyneaux, Steve Hodges, Pushmeet Kohli, Jamie Shotton, Andrew J. Davison, Andrew Fitzgibbon. KinectFusion: real-time dynamic 3D surface reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (2011) 559-568