AUGMENTED REALITY MARKERS AS SPATIAL INDICES FOR INDOOR MOBILE AECFM APPLICATIONS

Chen Feng & Vineet R. Kamat

University of Michigan, Ann Arbor, USA

ABSTRACT: *This paper presents a new methodology for utilizing Augmented Reality (AR) fiducial markers on mobile devices (smart phone/tablet) for indoor Architecture, Engineering, Construction, and Facilities Management (AECFM) applications such as navigation and inspection. On one hand, previous efforts on such applications focused mainly on exploring traditional non-visual-sensor-based methods to track user's position continuously, ignoring the advantage that in most of the built environment, human inspectors can easily navigate themselves to destinations given that they can make correct decisions at a set of discrete critical spatial locations (corner of hallway, stairs etc.). On the other hand, traditionally fiducial markers are extensively used to recover user's pose and thus serve as table-top AR display surface, benefiting from its cost-efficiency and high flexibility. A different view that combines these two observations is to attach these markers at critical locations whose global positions and orientations are known in advance, treating markers as spatial indices which help the automatic identification of key locations. Upon recognizing the marker, as well as estimating the relative pose between the user and marker, user's pose in the global coordinate frame can be calculated. Then decisions can be automatically made and users are guided to their destinations by 3D graphical instructions. An example application for complex indoor environment way-finding is built on Android platform and tested which demonstrates the efficiency of the proposed method.*

KEYWORDS: *Context-aware computing; Fiducial Marker; AprilTag; Indoor Navigation; Registration*

1. INTRODUCTION

Context-aware information delivery has been recognized as a critical component in many Architecture, Engineering and Construction, and Facilities Management (AECFM) applications (Anumba & Aziz, 2006; Khoury & Kamat, 2009; Andoh, et al., 2012). Identification of a user's location is one of the most fundamental and extensively studied problems in this area. Solutions that achieve good balance between cost and accuracy could lead to meaningful productivity improvement in applications such as facility management, construction inspection, and indoor way finding.

Previously, researchers' attention has focused more on traditional non-visual-sensor-based methods, such as Global Positioning System (GPS), Inertial Measurement Unit (IMU) (Akula, et al., 2011), Radio Frequency Identification (RFID) (Sanpechuda & Kovavisaruch, 2008; Andoh, et al., 2012), Wireless Local Area Network (WLAN) (Aziz, et al., 2005), and Ultra-Wide Band (UWB) (Teizer, et al., 2008). Recently computer vision and robotics communities have proposed methods such as Simultaneous Localization and Mapping (SLAM) (Thrun, 2008) and Visual Registration (Olson, 2011), utilizing visual sensors such as camera or lidar. Although most of these methods can track a user's 3D position continuously, they also have their own disadvantages respectively such as indoor unavailability, tracking drift, large infrastructure/special hardware requirement, cost inefficiency, and high computational power requirement.

However, in many of those methods' AEC domain applications, the fact that the end-user is human being makes human intelligence not only a non-negligible but rather important ingredient to achieve a good cost-accuracy balance (Akula, et al., 2011). This is different from robot navigation since robots need to know their own location at any moment for further decision making, while a human already possesses the ability of maintaining its own location within a certain range. For example, given a typical hallway with limited turnings, a human can easily navigate from one end point to the other; while a robot may need help of continuous SLAM. Only when the hallway contains lots of turnings and exit stairs, navigation becomes useful for humans, which is also true for outdoor road navigation. This observation in fact suggests that continuous tracking of user's position might not be necessary in many AEC applications; for human inspectors, to automatically extract discrete-spatial-distributed information could be sufficient to accomplish their jobs faster and better.

Naturally, marker based visual registration method for Augmented Reality (AR) becomes a good candidate for such discrete localization (Olson, 2011; Feng & Kamat, 2012). Registration problem means how to find the relative position and orientation between two coordinate systems in AR (Azuma, 1997). Thus, beside correspondences between different AR markers and discrete locations, these types of methods also provide estimation of not only position but also orientation. This additional dimension of information could help to obtain better visualization of extracted information and hence extend the application domain to where traditional sensor based methods such as RFID could not reach.

In the meantime, mobile devices are becoming more and more popular recently. Especially for smart phones and tablets, camera, CPU and even GPU are standard configuration, which unsurprisingly makes them ideal equipment for ubiquitous visual computing. Combining those observations, the authors were motivated to explore a new method utilizing AR markers as spatial indices to create links between physical locations and virtual information stored in databases, which runs on mobile devices. Our contribution in this paper is mainly a general computing framework as well as an indoor way-finding application based on such a framework. The fact that more and more people have smart devices makes it easier for people to access information using this method than all previous methods such as RFID and UWB.

The remainder of this paper is structured as follows. In section 2, previous work on both traditional and visual sensor based methods is reviewed; in section 3, the general computing framework of AR marker as spatial index is explained in detail; in section 4 an indoor way finding application based on that framework is developed; section 5 describes a set of experiments conducted to prove the new indoor way finding method's efficiency; and finally section 6 presents the conclusions of the paper.

2. PREVIOUS WORK

As mentioned before, researchers have extensively studied two types of localization techniques for context aware computing, i.e. traditional non-visual-sensor-based methods, as well as newly emerging visual-sensor-based methods.

Among the first type, GPS is mainly used for outdoor scenarios; IMU's tracking drift issue requires error correction either by human (Akula, et al., 2011) or by combining with other methods such as AR marker (Feng & Kamat, 2012); RFID-based methods usually depend on large infrastructure (i.e. enough RFID tags must be available) and also requires special tag reader which is not easily accessible by common people (Sanpechuda & Kovavisaruch, 2008; Andoh, et al., 2012); WLAN-based methods also require large number of footprints (Aziz, et al., 2005); UWB-based methods generally cost too much (Teizer, et al., 2008; Khoury & Kamat, 2009). Besides, none of these methods can easily provide instantaneous orientation information (even though there is angular sensors such as gyroscope, electrical compass or accelerometer, they themselves come with problems such as noise and sensitivity to the environment), which makes them not optimal for further 3D visualization purpose.

On the contrary, the second type of methods directly output orientation along with position information. By analyzing images captured from visual-sensors such as a simple webcam, the visual registration methods can recover that sensor's pose (position and orientation). Based on their different assumptions on the environment (i.e. the surrounding world where visual registration is going to be performed), these algorithms can be classified into two groups (Lepetit & Fua, 2005; Feng & Kamat, 2012): known environment vs. unknown environment (see Fig. 1, our proposed indoor way-finding application contains a module that can be classified as fiducial marker-based method).

The first group of visual registration algorithms is less computation-consuming and easier to design since the only unknown is the user's pose. Because they have been well studied, and many related powerful algorithms have been proposed in the last two decades, it is more realistic to apply them for solving real-world engineering issues. Within this class of methods, they can be further categorized into two groups: planar environment vs. non-planar environment. The first group is again easier to design because of the simple assumption made regarding the environment—a planar structure with known visual features. The second group is more often applied in a controlled environment with limited space, such as a small manufacturing workspace.

In the indoor way-finding applications, the authors chose to take advantage of plane-based methods since planar structures are abundant in buildings, construction sites, and other built environments where engineering operations are conducted, which makes this type of method very convenient to apply. In addition, a planar structure can simply be an image printed out on a piece of paper and attached to a wall/floor of a corridor, with negligible cost.

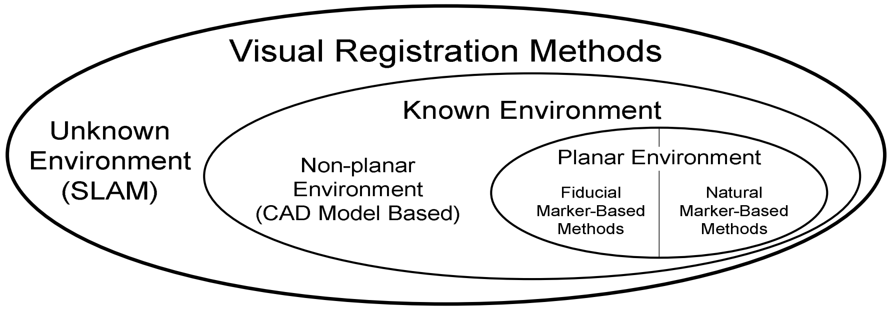

Fig. 1: A brief taxonomy of visual registration methods based on assumptions on environment.

Plane-based methods can be further classified based on different visual features they adopt: fiducial marker vs. natural marker. A fiducial marker is composed of a set of visual features that are “easy to extract” and “provide reliable, easy to exploit measurements for the pose estimation” (Lepetit & Fua, 2005). Usually those features are a set of black and white patterns forming simple geometry by circles, straight lines, or sharp corners and edges. Well known fiducial markers include ARToolKit (Kato & Billinghurst, 1999) and the newly proposed AprilTag (Olson, 2011).

Distinct from a fiducial marker, a natural marker does not require special predefined visual features. Instead, it treats any visual features in the same way. In this sense, almost any common image, ranging from a natural view to a company logo, can immediately be used as a natural marker. Even though natural marker methods have advantages of more accurate and stable as well as larger tolerance to partial occlusion, its relatively higher computational requirement than fiducial marker methods limits its sphere of application to high-end smart mobile devices. Taking all these factors into consideration, the authors chose to apply fiducial markers in the indoor way finding application so as to lower the requirement of its targeted devices.

3. METHODOLOGY

In this section, a general computing framework is described, representing the authors’ proposed idea to use AR marker as spatial index which links between physical location and virtual information related to that location. Since the framework is generic, the descriptions are not fixed to any specific algorithms. As long as an algorithm can meet the requirements of each module to be described below, it can fit into the framework. This makes it possible to adapt new algorithms for existing system with little difficulties.

The system overview is shown in Fig. 2. The system operates based on the following procedure. Firstly, an image potentially containing an AR marker is captured by the camera on the mobile device. This image is then sent to the marker recognition module. If the image contains an AR marker and it is recognized by the marker recognition module, then the ID of the AR marker is sent to the database as a key value to search for attached information. Simultaneously, the estimated pose of the mobile device relative to the marker is sent to the 3D visualization module. This visualization module then takes the estimated pose to render the information sent back from the database on the screen of the mobile device for further decision making by its end-user. In order to achieve such functionalities, each module needs to conform to certain requirements, which are explained in more detail below.

3.1 Physical Space

Generally speaking, the physical space in this framework, as the targeted environment of such a system, could be any indoor scene. For example, a complex shopping mall, an international airport, a subway station, a big warehouse or a building construction site. In fact, in some application scenarios, even outdoor scene such as a public park could be equipped with such a system.

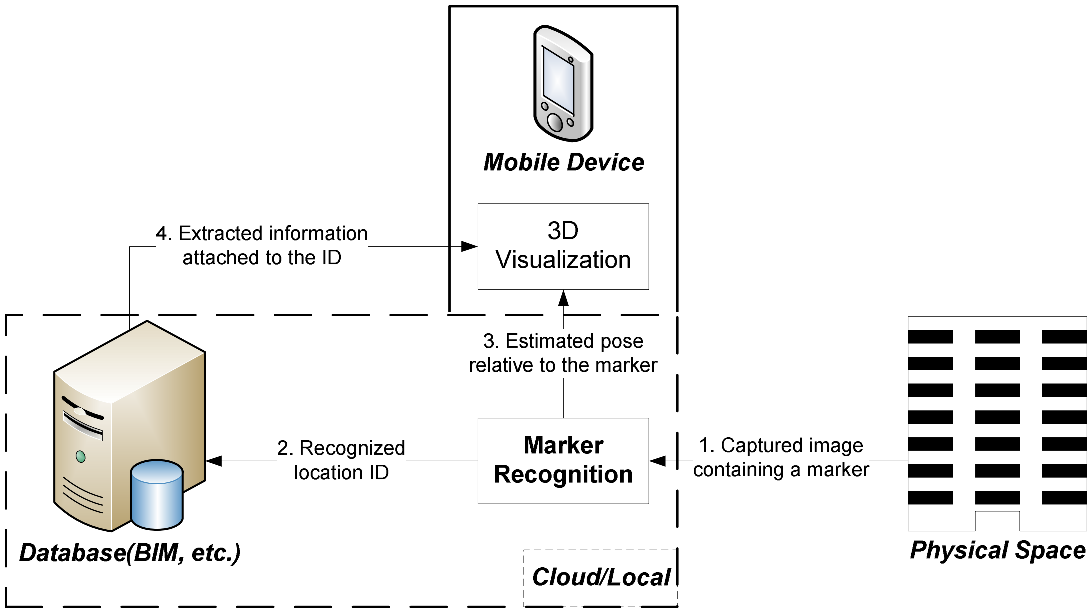

Fig. 2: System overview of AR marker as spatial index.

In order to make the original physical space compliant with the system, AR markers must be attached to a discrete set of locations which are critical to the application. The AR markers should have their own IDs which serve as the spatial indices of that set of locations. The marker recognition algorithm corresponding to those markers should contain two functions: 1) the ability to extract the ID of a marker presents in an image; 2) the ability to recover the pose of that image, i.e. the mobile device, relative to the marker. Some possible choices of marker could be AprilTag (Olson, 2011) or KEG Tag (Feng & Kamat, 2012).

Even though the requirement of one-to-one mapping between AR markers and physical locations makes our approach slightly infrastructure dependent, the facts that AR marker is flexible and easy to use/install, cost efficient, and more importantly capable of providing orientation information make it more ideal than RFID tag.

3.2 Mobile Device

A mobile device is the core of such a system. Its camera serves as the main input where information enters into the system. While cameras are part of the standard configuration of common mobile devices such as mobile phone or tablet, a rear facing camera setup is more desirable than a front facing one, since the direction of sight of the camera could be aligned with user's viewing direction more naturally so that user could see the 3D visualization more conveniently.

Also, the 3D visualization module naturally locates within the mobile devices. It visualizes the information sent from the database in the means of augmented reality by pose estimated from the marker recognition module. Here an implementation choice must be made for any application conforming to this framework: whether the marker recognition module should locate locally within mobile device or remotely in the server/cloud. This usually depends on the invoking frequency of this computation framework. For some applications, the whole procedure needs to be performed for every frame of the live video stream captured by mobile device's camera. In this situation, a local marker recognition module is more reasonable, otherwise the captured image needs to be transferred to the server/cloud side very frequently, which will increase the data volume transferred through the mobile network resulting in higher cost and longer latency. For some other applications, the invoking frequency is relatively low, say around 1 time per minute. Then turning the marker registration module into an online service could be a good choice.

3.3 Database

A database is the foundation of such a system. With physical locations mapped into IDs through AR markers, the database module is very flexible and can be implemented from almost all kinds of databases available, such as MySQL, SQLite or even a self-defined text file. Similarly, the storage location of this database is also an

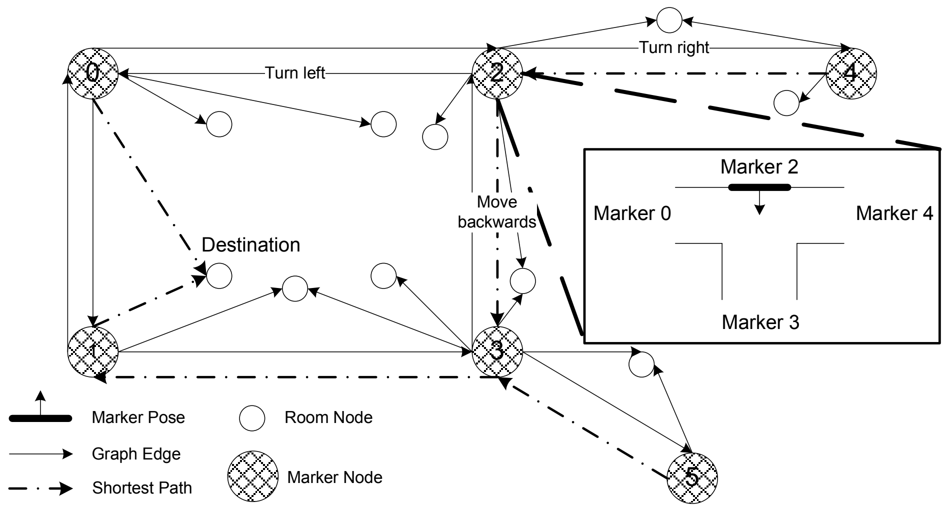

Fig. 3: An example of oriented graph for indoor way finding application.

implementation choice, depends on the size of the whole database as well as the size of information to be sent back to the 3D visualization module. For relatively smaller-size database, say less than 1 Mb, downloading the whole database from cloud/server side to mobile device once and for all is a reasonable choice. Otherwise, server-side storage of the whole database could be more efficient, as long as the size of the information to be sent back in step 4 of Fig. 2 is relatively small.

4. WAYFINDING APPLICATION

As mentioned previously, the above methodology is best suitable for the decision making process which involves a set of discrete spatial locations. Indoor way finding is one good example application for which the above methodology could be very helpful.

Consider the following first-person scenario: you are a new student on the first day of class. You were given a room number which is the first class you are going to take. But the building is large and unfamiliar, and even though you entered it for a few times previously, you still find it difficult to find the correct direction. You try to look for an indoor floor plan view of the building but it takes you more than 5 minutes and you fail to understand the map. Since it is already time for class, very few people are available for inquiring the directions. When you finally find someone for help, the instructions you get are “go along this hallway and turn left at room 1318, then walk to the end of the hallway and turn right. Then pass the exit door and go upstairs to second floor and turn right again. After another two exit doors and a right turn is your destination.” After you find the room probably half of the class has passed. Similar situations happen frequently inside many complex buildings for a person who is unfamiliar with the environment or doesn’t have a good sense of direction, and even for someone who works/studies inside the building every day since s/he has never been to the room before.

In this case one can see that the most useful piece of information, i.e. the instruction which greatly influences the decision making process, contains such a set of discrete locations (e.g. room 1318, exit door, corner, stair, etc.) as described before. It is natural to consider that an AR marker should be attached to each of such critical spatial locations to map these locations into spatial indices. In the database module, the position and orientation of each marker under the building’s global coordinate system is stored, as well as positions of all the rooms of the building. Thus an oriented graph is generated from this information, with nodes as rooms and markers, edges as instructions of how to move from starting node to ending node, edge weights as lengths of physical paths (see Fig. 3, shaded large circle means marker node, small circle means room node). There are two situations for adding edge to the graph:

1. When there is a physical path between a room node A and a marker node B without passing other markers, an edge from B to A can be added to the graph; No edge from room node to marker node should be added;
2. When there is a physical path between a marker node C and a marker node D without passing other markers, an edge from C to D, as well as an edge from D to C, should be added to the graph.

Fig. 4: Two screenshots showing graphical instruction of how to move to the destination (Best viewed in color).

The instruction stored within each edge, say edge starting from node A to node B, is the method of how to move from A to B assuming user is currently facing towards the marker. For example, as the orientation of marker 2 shown as an arrow in the detailed floor plan view of marker 2 in Fig. 3, instruction stored in edge from marker 2 to marker 0 could be "turn left", since when the user gets this instruction s/he should be looking at this marker 2.

After the graph is constructed, the user is asked to select a node as destination. Then the Dijkstra algorithm (Dijkstra, 1959) is performed on the graph to compute the shortest path from all other marker nodes to this node. Thus whenever user comes to a marker, the marker recognition module will extract the corresponding spatial index and then find the marker node in the graph. Then the shortest-path edge starting from this node is retrieved and the moving instruction is displayed graphically in the 3D visualization module (see Fig. 4). When user is standing at the marker node which has a direct edge pointing to the destination, special logo is shown (the animated eye shown at the top-right corner of the left image in Fig. 4) to remind user to slow down and look carefully around the surroundings since s/he is very close to the destination room.

The authors implemented the described indoor way finding application on Android platform. Here the two design choices are made as follows. Marker recognition module is located in the mobile device, since real-time performance is desired; database is stored in server side as a simple self-defined text file describing the whole graph of a building (termed as a map file). This application assumes that user is aware of which building s/he is in. When user selects the building, the corresponding map file will be downloaded to the mobile device. The marker recognition module in this application adopts the AprilTag (Olson, 2011). As mentioned in section 2, the reason to adopt this specific algorithm is that AprilTag is proved to be superior than previous proposed fiducial markers in the sense of speed, accuracy and tolerance to critical view conditions (long distance, large viewing angle, partial occlusion etc.).

The authors have made this application, named as Mobile AR Navigator (MARvigator), publicly available online at http://www.umich.edu/~cforrest/upload/marvigator/. Although similar methods or systems have been proposed before (Kalkusch, et al., 2002; Wagner & Schmalstieg, 2003; Augmented Reality & Assistive Technology Laboratory of NUS, 2011), MARvigator is superior since it takes advantage of the state-of-the-art fiducial marker system and is implemented conforming to the general framework described previously, resulting in faster real-time performance (about 15 frame-per-second as shown in Fig. 4), high ease of use and flexibility (could be easily setup at any complicated indoor environment or some outdoor environment).

5. EXPERIMENTAL RESULTS

In order to show MARvigator's efficiency in helping people find destinations, a set of experiments is conducted on the first floor of a building on the author's university, which is known on the authors' campus for its complex network of corridors and hallways. Seventeen apriltags are placed on critical locations among the region whose positions and orientations are shown in Fig. 5. Six target positions are selected (room 1351, 1318, 1069, 1188, 1040 and 1504) which need to be sequentially found by each of the 10 experiment volunteers, among which 6 volunteers work/study within a part of this region and the other 4 have never been to this region or are not familiar with it.

Each volunteer starts from nearby marker 0. They are instructed on the usage of MARvigator for one minute before the experiment. The time for each volunteer's reaching target room is recorded with resolution of half a minute. In order to reduce to the minimum the influence of other factors such as variations in walking speed and sense of direction, volunteers are asked to switch between using and not using MARvigator during the way

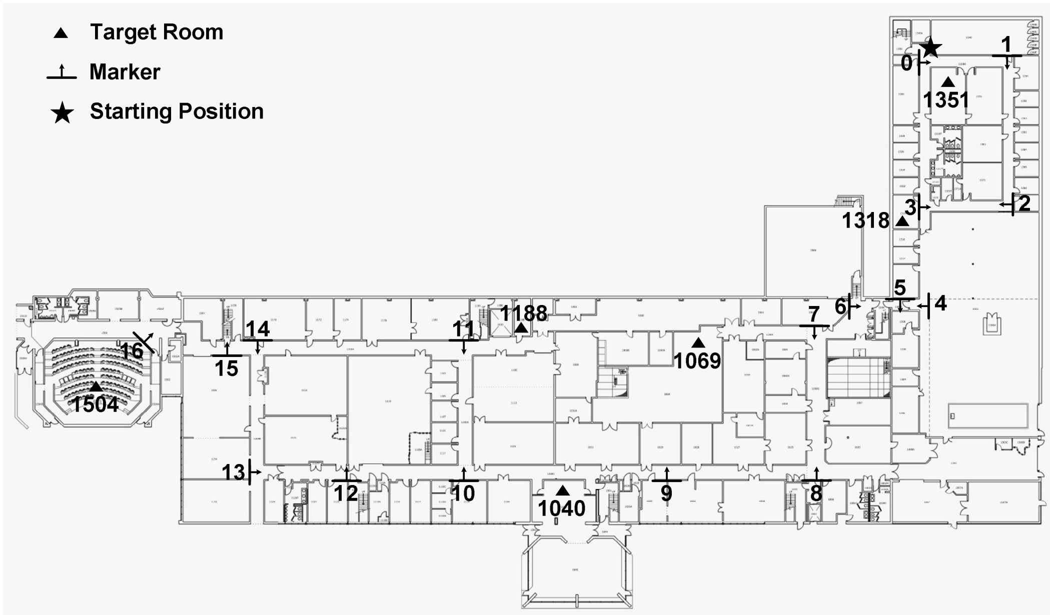

Fig. 5: Indoor way finding experiment setup.

finding. For example, if a volunteer start to find the first target room 1351 with the help of MARvigator, then s/he should not use MARvigator but any other common means of way finding (looking at the map on some part of the building or asking other people for help) to discover the second target room 1318, and vice versa. In the experiment, 5 people start with MARvigator and the other 5 people start without.

The experiment results are shown in Table 1. The shaded grid shows the time to find the target room assisted with MARvigator. From these results it is clear that MARvigator does help people to find destinations faster. Notice that for the first two targets, using MARvigator took slightly longer time. This could be explained by the fact that the user needs time to learn how to use the application. Also the first two targets are very close to the start position and can be found easily. However, by comparing the average time of finding later targets, MARvigator's efficiency is better highlighted.

Table 1: Experimental results for the 10 volunteers (shaded cells shows the time with MARvigator assistance).

Time(min)	Volunteer who starts with MARvigator					Volunteer who starts w/o MARvigator						
Target	A	B	C	D	E	F	G	H	I	J	Average	
1351	0.5	1	1	1	0.5	0.5	0.5	0.5	0.5	0.5	0.80	0.50
1318	0.5	0.5	0.5	1	0.5	0.5	1	1	1	0.5	0.60	0.80
1069	0.5	1	1	1	1	0.5	2	5	2	2	**0.90**	2.30
1188	2	2	7	4	2	1.5	2	4	3	2	3.40	**2.50**
1040	1	2	1	1	2	2	0.5	1	4	4	**1.40**	2.30
1504	3	3	2	2	2	2	1	2	3	2	2.40	**2.00**

6. CONCLUSION

In conclusion, the proposed general computing framework of using AR marker as spatial index to link physical locations with virtual information related to that location offers a new perspective of utilizing AR markers. The indoor way finding application MARvigator conforming to this framework is proved by experiments to be very efficient and convenient. Future research directions could be adopting more advanced computer vision and machine learning algorithms to replace the marker recognition module with landmark recognition module which further improves the infrastructure dependency of the framework.

7. REFERENCES

Akula, M. et al., 2011. Integration of Infrastructure Based Positioning Systems and Inertial Navigation for Ubiquitous Context-Aware Engineering Applications. *Advanced Engineering Informatics, 25(4), ICCCBE 2010 Special Issue on Computing in Civil and Building Engineering*, pp. 640-655.

Andoh, A. R., Su, X. & Cai, H., 2012. *A Boundary Condition-based Algorithm for Locating Construction Site Objects Using RFID and GPS.* Proceedings of 2012 Construction Research Congress, West Lafayette, IN.

Anumba, C. & Aziz, Z., 2006. Case Studies of Intelligent Context-Aware Services Delivery in AEC/FM. *Lecture Notes in Computer Science*, Vol 4200, pp. 23-31.

Augmented Reality & Assistive Technology Laboratory of NUS, 2011. *AR Navi NUS.* [Online]
Available at: http://serve.me.nus.edu.sg/ongsk/ARLab/arnavi.htm [Accessed 12 06 2012].

Aziz, Z. et al., 2005. *Context Aware Information Delivery for On-Site Construction Operations.* Proceedings of the 22nd CIB-W78 Conference on Information Technology in Construction, Dresden, Germany, p. 321–327.

Azuma, R., 1997. A survey of augmented reality. *Presence:Teleoperators and Virtual Environments*, pp. 355-385.

Dijkstra, E., 1959. A note on two problems in connexion with graphs. *Numerische mathematik*, Vol 1, pp. 269-271.

Feng, C. & Kamat, V. R., 2012. *A plane tracker for AEC-automation applications.* Proceedings of 2012 International Symposium on Robotics and Automation in Construction, Eindhoven, NL.

Kalkusch, M. et al., 2002. *Structured visual markers for indoor pathfinding.* Proceedings of the First IEEE International Workshop On Augmented Reality Toolkit, Darmstadt, Germany.

Kato, H. & Billinghurst, M., 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. *Proceedings of the 2nd International Workshop on Augmented Reality*, pp. 85-94.

Khoury, H. M. & Kamat, V. R., 2009. Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, Volume 18, pp. 444-457.

Lepetit, V. & Fua, P., 2005. Monocular Model-Based 3D Tracking of Rigid Objects. *Foundations and Trends in Computer Graphics and Vision*, pp. 1-89.

Olson, E., 2011. *AprilTag: A robust and flexible visual fiducial system.* Proceedings of IEEE International Conference on Robotics and Automation, pp. 3400-3407.

Sanpechuda, T. & Kovavisaruch, L., 2008. *A review of RFID localization: Applications and techniques.* Proceedings of 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, pp. 769-772.

Teizer, J., Venugopal, M. & Walia, A., 2008. Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and material positioning and tracking. *Transportation Research Record: Journal of the Transportation Research Board*, Vol 2081, pp. 56-64.

Thrun, S., 2008. Simultaneous localization and mapping. *Robotics and cognitive approaches to spatial mapping*, pp. 13-41.

Wagner, D. & Schmalstieg, D., 2003. *First Steps Towards Handheld Augmented Reality.* Proceedings of the 7th International Symposium on Wearable Computers (ISWC'2003), White Plains, New York, USA..