

Augmented Reality-Based Indoor Navigation using Google Glass as a Wearable Head-Mounted Display

Umair Rehman, Shi Cao
 Department of Systems Design Engineering.
 University of Waterloo
 Waterloo, Ontario, N2L 3G1 Canada
 (urehman, shi.cao)@uwaterloo.ca

Abstract—This research comprehensively illustrates the design, implementation and evaluation of a novel markerless environment tracking technology for an augmented reality based indoor navigation application, adapted to efficiently operate on a proprietary head-mounted display. Although the display device used, Google Glass, had certain pitfalls such as short battery life, slow processing speed, and lower quality visual display but the tracking technology was able to complement these limitations by rendering a very efficient, precise, and intuitive navigation experience. The performance assessments, conducted on the basis of efficiency and accuracy, substantiated the utility of the device for everyday navigation scenarios, whereas a later conducted subjective evaluation of handheld and wearable devices also corroborated the wearable as the preferred device for indoor navigation.

Index Terms—Wearable Head-Mounted Display, Augmented Reality, Indoor Localization and Navigation, Markerless Tracking, 3D Environment Scanning

I. INTRODUCTION

Indoor navigation systems can have many useful applications, and therefore they have been extensively studied with the aim to achieve accurate localization and better navigation. Indoor navigation systems however face a lot of technical challenges such as Non-Line-of-Sight (NLoS) conditions, high attenuation and signal scattering, greater concentration of physical impediments, transitory environment changes, high demand for accuracy etc. To address these challenges, different technologies have been introduced, which vary over a range of imperative parameters and user requirements like accuracy, cost, scalability etc. In order to find a suitable navigation technology for a particular application, the performance parameters need to be aligned with the requirements of the user [1].

Experimentation pertaining to indoor navigation has been mostly conducted on handheld mobile phones whereas proprietary wearable devices have not been widely explored. A wide range of wearable concepts, systems and applications are predicted to become very popular in the near future [2]. Amongst wearable devices, head-mounted displays (HMD) which employ head tracking can be highly facilitative in improving efficiency and performance of many human involved processes including navigation [3][4]. Proprietary head-mounted displays were not readily available in the past and therefore previous experimentation [3][4][5][6] was restricted

to just the lab environment [7] where internally engendered computing devices, comprising of head-mounted sensors and cameras, were normally used. Recently however Google and Microsoft released the prototype versions of their optical head-mounted displays [2]. These releases can now allow investigators to carry out more practical research and also develop easily deployable applications using their existential platforms.

An imperative aspect of an indoor navigation system is the user interface design. Head-mounted displays that can seamlessly portray navigational information on the real environment are considered to be very effective for user engagement [6]. Therefore augmented reality (AR) technology is considered ideal for wearable computing devices to achieve navigation tasks as this technology can improve situational awareness and human perceptibility [5][6][8]. AR based applications have been used as an impactful way of supplementing computer-generated information on the real environment, and therefore they have been comprehensively explored in various other fields as well including healthcare, defense, intelligence, transportation etc. [5]. AR interfaces for indoor navigation have been evaluated in previous studies [9][10], where it has been validated that AR does support accurate localization and a better user experience [11]; however the performance gets greatly dependent on the orientation and position of the device, and for hand-held devices the performance is bound to decline as users can get physically fatigued after holding the device for a long time therefore there was a need to explore the research on wearable devices.

Indoor navigation technologies like Ultra-wide Band (UWB), Wireless Local Area Networks (WLAN), and Radio Frequency Identification (RFID) etc. are potent technologies but require existing infrastructure to be deployed [1]. Most of these solutions also contain substantial localization errors; however these errors could be minimized when an inertial sensor based positioning approach gets supplemented using a probabilistic technique such as particle filtering [12]. Some technology solutions such as Bluetooth, Infrared etc. involve beacons to be attached in the building infrastructure which are not very expensive to manage but they render an unpleasant impression on the interior of the building and also experience high latency during detection phase [13]. While these sensor based technologies are popular positioning solutions, they

cannot approximate the user's orientation, which makes them inept for AR applications [14]. Computer vision techniques are therefore employed for AR based applications and previous studies have found computer vision technology to be more accurate when compared against Wi-Fi based fingerprinting which is one of the most investigated approach to achieve indoor localization [13].

A commonly studied vision based indoor positioning approach involves image recognition of the real environment through live camera feed of the device in the form of ego-motion so that these images could be referenced against a pre-collected sequential database of orthographic images of the same environment. The pre-collected images are annotated with the position whereas the inertial sensors of the device can help deliver orientation [15]. This technique can therefore be used to deliver successful AR based directional instructions as well as user localization. The issue with this technique however is that it would entail extensive computational power since a large database map of images is being utilized which can cause subsequent delays during image referencing process [7].

Another computer vision based approach, widely studied before [4][5][7][8][16][17], uses augmented reality markers for optical tracking where directional information can be overlaid on these markers to conduct navigation. Various kinds of AR tracking is available for a wide variety of markers. Physical markers like ID Markers, Bar Codes and QR Codes, use fiducial tracking [18] for detection. These markers are recognizable due to their geometric shape and high contrast. Other physical markers like picture markers have adequate visual content to be distinctly recognizable. Physical markers are self-positioned strategically to cover the entire indoor environment while in some cases distinct features within the environment like furniture, signs etc. could also be used as picture markers. The issue with most physical markers is that they have to be physically placed in the environment in a way that they don't get obscured during navigation which impacts the interior of the building. As far as picture markers are concerned, not all indoor environments have enough distinct signs that could be utilized for navigation, and there is always a risk of not finding a trackable during navigation [19].

There is however a 3D markerless tracking approach available which is an advance form of optical tracking [19] not extensively explored for navigational purposes. 3D maps are created by environment scanning the potential area. Once adequate visual information of trackables (3D point clouds at different camera angles) is collected then they could be used for AR information overlay. This is computationally not very exhaustive for the device and the physical layout of the environment does not get disturbed. Finding distinct patches in an indoor area is not as challenging as finding a specific object to be represented as a picture marker and since entire areas are being taken into consideration as trackables; they are easily detectable from longer distances. Directional information can then be overlaid on these trackables using AR technology, which can produce a very accurate navigational experience. In

this research, we developed our application on the foundation of this novel tracking technology by implementing it on a head-mounted display for seamless navigation and localization. We tested the application and conducted comprehensive performance assessments on the basis of efficiency and accuracy of the technology while also illustrating a subjective analysis on user preference where we compared the wearable head-mounted display to handheld cellphones and maps.

II. PROPOSED SYSTEM

A. System Design

The major function of the system is to use environment tracking to assist people navigate in indoor areas using augmented reality based visual and audio instructions. The system design is developed to achieve optimum performance for a head-mounted display. The head-mounted display used is the Google Glass, which is ideal for augmented reality applications because different inertial sensors (gyroscope, accelerometer, magnetometer, light and proximity sensor) on the device ensures robust head-tracking capability that helps maintain the required orientation for the visual overlay which is then projected through a lens directly into the retina of the user. This delivers a very rich proprioceptive experience where the virtual content gets seamlessly integrated with the real environment. Developing applications on the Google Glass is straightforward as the Glass Development Kit (GDK) is an add-on to the Android SDK henceforth the android platform is utilized. The development of the Google Glass indoor navigation application requires a pre-deployment stage where the indoor environment has to be 3D scanned for information overlay. We developed our indoor navigation application using the Metaio SDK [20] which provides a multilayered environment to build AR applications on android platform.

B. System Overview

The system architecture shown in the Figure 1 depicts the process flow of the application. As described earlier, the pre-deployment data collected and appropriately configured in the first stage is stored on the application. Scanned environment consisting of visual features (3D point clouds) stored as trackables have location and navigation related information superimposed within them to be displayed during the tracking process. The camera and inertial sensors of the Google Glass are used to track the 3D point clouds. The discovery of these trackables from the real camera feed of the device is used to display the current location of the user as visual text. Once the location is configured then the route of the journey is ascertained. The potential route of journey, supplemented with directional instructions in a chronological order, is pre-stored in the app for experimental purposes. In the future however routing could be optimized to cover the entire environment but it was not considered for this experiment as the Google Glass suffered from repeated battery drainage due to excessive computational workload. Once the app configures the route then continuous AR based navigation instructions comprising of audio and visual aids leads the user to destination while

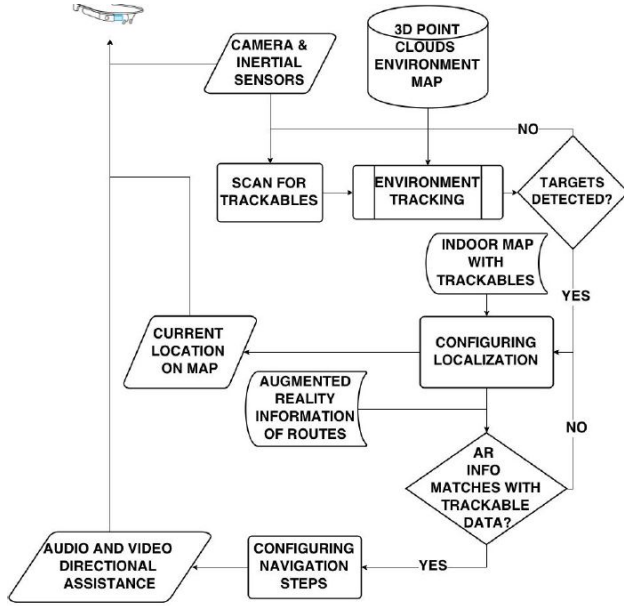


Fig. 1. System Architecture

location gets simultaneously updated and displayed on the screen. Inertial sensors, using gravity measurement for pose estimation ensure that AR based directional instructions get appropriately positioned during motion as any incongruity can create confusion between augmented and natural world scenarios.

C. 3D Environment Scanning

The location chosen for experimentation was the Games Institute at University of Waterloo. A potential route for navigation was determined and a possible task for experimentation, for example, navigating from one end of the route to the other using Google Glass was also finalized. 3D environment scanning was considered as the key process during the pre-deployment stage. Different areas of the route had to be scanned to develop an environment map. The environment scanning was conducted on Metaio Toolbox [20] using the camera of an android cell phone. Crucial areas within the trajectory were shortlisted for potential tracking as we did not intend to scan the entire environment since that would have been a lot of data to process during the tracking stage. We established that the minimum area to be scanned would be 2 meters in length so that larger areas could be detected during the tracking process. This would ensure that no discrepancy occurs when AR based positional information gets overlaid. Although all distinguishable surfaces within the environment were taken into consideration but highly textured surfaces were preferred to make certain that maximum number of visual features (3D point clouds) within an area were scanned. Environmental objects such as tables, chairs, bulletin boards, signs etc., which were part of areas we were scanning, were scanned

from all different camera angles and we also established that the minimum number of features to be scanned within an area would be 1500 so that the map gets adequately populated with trackables. Areas where a potential turn in navigation was expected was more comprehensively scanned as additional AR information was expected to be superimposed on those locations. In total nine different areas of varying length ($>2m$) from the route were scanned with trackables which were later gravity aligned using the inertial sensors of the cell phone. The process concluded once sufficient features on all the areas of the route of navigation were scanned.

D. Information Overlay and Tracking

After a route was fully scanned, the images were exported to Metaio SDK for AR information overlay. The 3D scans of all areas were placed in a sequential order to develop a movie like timeline progressing from the start of the destination and shifting towards the end. The information to be overlaid was twofold: navigational directions and localization information.

The first thing that was taken into account was user localization. We developed a static localization solution as user was not represented on the map actively however the current location of the user was displayed as visual text. The user could always read the location and figure out its position by referencing it on the annotated map which was part of the applications display interface. To accomplish this the overall map was strategically annotated with location numbers and alphabets whereas the route was specifically annotated into nine different zones which were specific parts within the scanned areas ($\leq 1m$) where each zone represented the distinct area that was scanned in the first stage. The trackables in the area were superimposed with the specific zone designations so whenever the user was present in that specific zone its designation was displayed as augmented information on the visual overlay of the Google Glass and hence the user could localize themselves in accordance with the map. We developed a non-dynamic localization solution mainly because we wanted to mitigate distraction as multiple objects moving on a small visual overlay could be concerning for certain individuals. We had also established the fact that the integral feature of the application was AR based navigation, not dynamic localization. The only reason to equip the application with a map based localization solution was to ensure users were contextually aware and thus a static localization solution made most sense.

The second stage of the information overlay process was adding directional instructions as shown in Figure 2. Three different forms of assistive information were overlaid on the scanned areas. 3D arrows were the first information added to the scenario. These were giant, glossy, green colored arrows to make certain proper visibility on the small visual overlay of the Google Glass. Three different forms of audio instructions like “turn right, go straight, and turn left” were also added to the scenario on appropriate places. Lastly text based visual instructions of the audio was also superimposed on the trackables. The reason to add text and audio based instructions

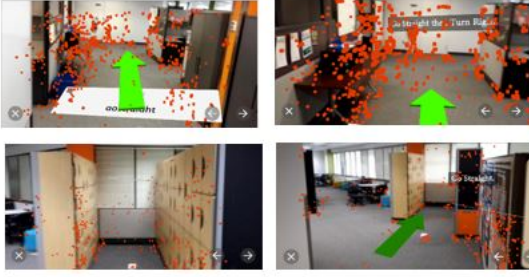


Fig. 2. Information being overlayed to the scanned 3D point clouds of the different areas within the environment map.

was to prototype a future hybrid solution for people with either hearing or vision based impairments. The trackables were properly translated, rotated and scaled to ensure that AR information was correctly positioned.

The tracking was analyzed when the application was tested in real environment. The interface described worked in exact accordance to the process flow. The trackables were quickly detected and application processing was swift. The audio augmentation was assistive and made the application more intuitive. Visual 3D arrows were impactful in showing direction and were successful in quickly guiding to destination location.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A performance analysis was conducted to evaluate the technology with respect to its accuracy and efficiency both giving us a detailed measure of the applications effectiveness in indoor environments. The efficiency was determined after carefully monitoring the individual times taken for successful feature detection, processing and information display whereas the accuracy was observed after analyzing the localization errors caused due to time delays. Since we were quantifying very small values of time, we had to write a separate program to look for specific triggers to note the time. The experiment was repeatedly conducted on nine random locations, each representing a distinct area of scan on the route. The locations were almost evenly apart to ensure that the experiment could be conducted at each position at variable pace.

Feature tracking was an extremely important aspect of the overall process as further navigation or localization information gets delivered after the analysis of a particular set of features in a given area. We expected a considerable number of features to be tracked in minimal amount of time so that AR information gets overlaid when required without any delay. We however did have certain concerns that pertained to specific scenarios. For example, if a user is expected to take a right turn at a point but inadequate features get tracked due to his extreme pace while walking at that location, then a delay in information delivery could be expected as it would take the application more time to process since system would continue to loop around searching for sufficient features. In rare situations, if a user passes a target location way too swiftly then it would be hard for the camera to capture the trackables

adequately and therefore no response from an application could also be a possibility. Analyzing such concerns, we ensured that the trackable area was large so the effects of fast pace of users gets substantially annulled. We also made certain that the feature extraction count on almost all occasions was uniform during the scanning process and the areas chosen had almost the same amount of visual content. To cater for these issues in detail however we did three redetection experiments on the navigational routes.

In the first experiment we set a maximum time limit of 0.7 seconds for features to be extracted and we walked by all imperative areas of the route. The results in Figure 3 depict that 50.6 % of features on average, for all nine locations, got successfully redetected. The pace was approximately maintained around 4 mph which is considered as fast paced for indoor environments but still the application was more or less able to detect half of the features. The second experiment, was conducted at a comparatively slower pace at approximately 2.7 mph where we calculated the maximum time it could take for 95 % of features to be redetected. The results in Figure 4 show that it can take on average 0.9 seconds for features to be redetected in a given area, which confirmed the fact that it takes a faster time for features to be spotted by the application after the initial features have been recognized. Our third experimentation was conducted at way faster pace where the goal was to deduce the minimum number of features needed at a position to initiate the processing for the AR information overlay. For this we wrote a separate internal script that gave us the needed feature count at the nine locations. We concluded that the minimum average feature count needed for the experiment was approximately 45% however it varied for individual locations as shown in Figure 5. While the speed was variable as we adjusted in accordance with the features being detected but we were able determine the maximum speed range which was between 4mph to 4.7mph. Although the results of these experiments were roughly construed, they were able to substantiate that the technology was practical enough for everyday indoor navigation scenarios. The maximum speed range from experiment 3 was extremely fast and was highly unlikely to be observed in ordinary circumstances. We however did realize that on a few locations, feature detection took significantly longer mainly due to poor lighting. During the experiment we ensured that head orientation of the subject did not drift drastically at any point and the route was mostly insight for features to be adequately detected.

Analyzing the timings for AR information augmentation was also crucial to determine the efficiency of the technology. The device after recognizing the features would have processed it using the data present in the internal memory. While there maybe individual factors that might influence the feature detection process, we still came up with a test bed that could give us an approximate result representing the time system entails for processing each kind of augmentation shown in Figure 6. We then calculated the average time it took for the Google Glass to output audio augmentation, which was 0.18 seconds whereas the augmentation of localization

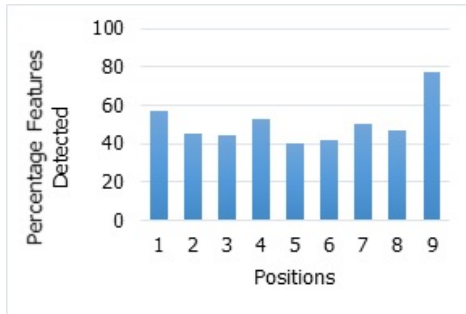


Fig. 3. Maximum (%) of features detected at 0.7 seconds.

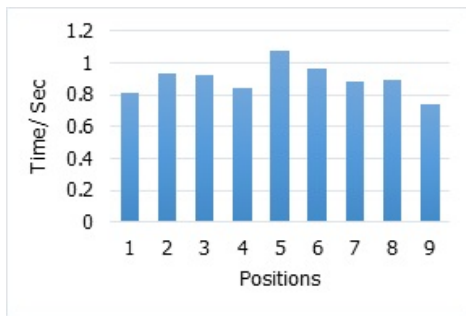


Fig. 4. Time taken to detect 95 % of features

information was the quickest in about 0.04 seconds and the time required to supplement visual arrows and text into the environment was about 0.14 seconds on average for a location. The localization information was the fastest since it was the first in the loop to be displayed on the screen followed closely by audio information. Visual information and text required inertial sensors to continuously ensure that the information gets supplemented on the right area on the camera generated feed therefore it was comparatively time consuming however overall these results portrayed extremely quick information processing and display.

Lastly, we analyzed the localization error. Since we developed a static localization system, we were expecting a delay in the location update which could cause a slight error. The AR

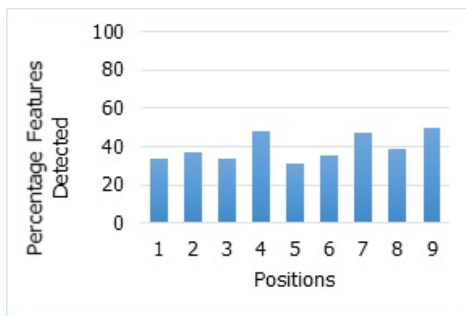


Fig. 5. Minimum (%) of features needed at the given position to initiate processing.

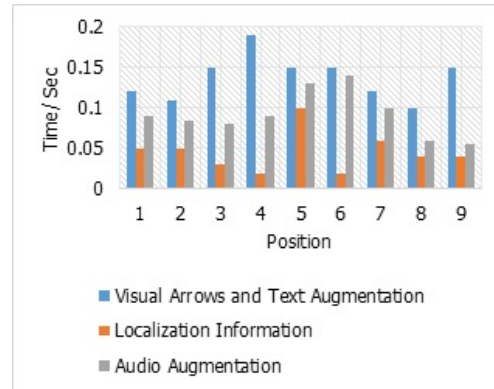


Fig. 6. Time delays to due to information processing.

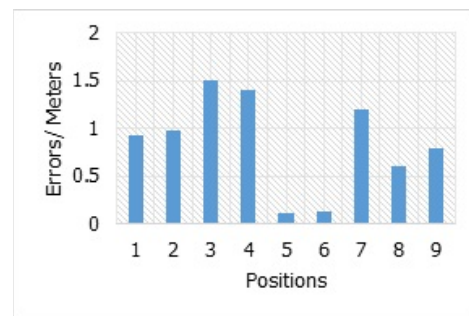


Fig. 7. Localization error due to time delay.

information shows a zone where the user is located as visual text on the display and the user can then look at the given 2D map also labelled with all the zones to figure out their current location with respect to the environment. The problem that potentially occurs here is that when a location is updated, it takes on average 0.04 s extra for the device to show the new location of the user. The distance travelled within this time is denoted as the approximation of the localization error due to time delay. To further delve down into this issue, we again randomly selected nine points and configured position error with respect to time on these location shown in Figure 7. On average a 0.3m localization error is expected due to time delay however for further experimentation, where a dynamic localization module is developed, a more comprehensive position-error analysis could be conducted for finding the localization error due to feature recognition which would set foundations for this technique to be compared against Wi-Fi and other positioning technologies.

After conducting objective performance evaluation of just the Google Glass, we wanted to construe a comparative performance analysis with respect to the android cell phone. It was easy for us to develop an application and implement it on the android cell phone since we were essentially using the same framework. The performance of the android cell phone was very similar to that of the glass since the duration for our experimentation were very small. On hand held devices, user

performance gets affected in larger run studies as they tend to get fatigued which impacts the orientation and positioning of the device. In the end we tried to subjectively enquire user preference of 27 participants (N=27) who underwent same navigation tasks at our test location (Games Institute) using our developed application on Google Glass and on an Android Cell Phone. We also added paper map as the third assistive device since it is often utilized for indoor environments as well. The tasks were completed on three different trajectories for the three devices however we ensured that all trajectories get uniformly tested by all devices. The details of this usability comparison study have been reported in another paper [21]. The results showed that people mostly preferred the Google Glass, mainly due to the low cognitive workload required and a better AR navigation experience as the device was hands-free. The preference for android cell phone was almost the same as people preferred the cell phone mainly due to the better display and the fact that they found phones to be intrinsic and easily manipulative in nature as the participants had been using them for a long time. The paper map was only preferred by 3 participants who were expert map readers and had been using the paper map for a very long time for other applications.

IV. CONCLUSION AND FUTURE WORKS

Computer vision technologies provide an effective alternative to other sensor based localization and navigation methods especially for indoor navigation purposes. While many of them had been extensively investigated before, 3D point cloud based environment tracking has not been thoroughly studied. We therefore detailed an implementation of this tracking technology on Google Glass since these head-mounted wearable devices were anticipated to greatly improve navigation experience when complemented with AR technology. The framework presented in this paper explicated the entire development and testing process that involved 3D environment scanning, rendering AR information and testing the technology under varying situations for comprehensive performance evaluation.

Various aspects of this study could be potentially pursued for future research. Firstly, a dynamic localization module within the application should be developed in order to accomplish a more comprehensive position-error analysis so that this technology could be quantitatively compared to other positioning technologies like Wi-Fi, Bluetooth etc. In this experiment, we investigated 3D point cloud tracking for AR information overlay however other markerless techniques, like CAD based edge model tracking, should also be explored. The impact of environmental features, such as lighting, visual content of objects etc., should be examined in order to determine their bearing on the tracking capacity of the system. Future applications of this technology in different industrial settings such as healthcare, gaming, manufacturing etc. should be individually studied and other sensor equipped wearable devices should also be considered for such experiments.

ACKNOWLEDGMENT

The authors would like to extend their sincere appreciation to the members of the Games Institute for their repeated support and appreciation throughout this research.

REFERENCES

- [1] Mautz, Rainer, "Indoor Positioning Technologies," ETH Zurich, Department of Civil, Environmental and Geomatic Engineering, Institute of Geodesy and Photogrammetry, 2012.
- [2] H. Moustafa, H. Kenn, K. Sayrafian, W. Scanlon, and Y. Zhang, "Mobile wearable communications [Guest Editorial]," *Wirel. Commun. IEEE*, vol. 22, no. 1, pp. 10–11, 2015.
- [3] S. L. Joseph, X. Zhang, I. Dryanovski, J. Xiao, C. Yi, and Y. Tian, "Semantic Indoor Navigation with a Blind-User Oriented Augmented Reality," 2013, pp. 3585–3591.
- [4] M. Kalkusch, T. Lidy, M. Knapp, G. Reitmayr, H. Kaufmann, and D. Schmalstieg, "Structured visual markers for indoor pathfinding," in *Augmented Reality Toolkit, The First IEEE International Workshop*, 2002, p. 8–pp.
- [5] J. Kim and H. Jun, "Vision-based location positioning using augmented reality for indoor navigation," *Consum. Electron. IEEE Trans. On*, vol. 54, no. 3, pp. 954–962, 2008.
- [6] J. B. Kim, "A personal identity annotation overlay system using a wearable computer for augmented reality," *Consum. Electron. IEEE Trans. On*, vol. 49, no. 4, pp. 1457–1467, 2003.
- [7] S. Kasprzak, A. Komninos, and P. Barrie, "Feature-based indoor navigation using augmented reality," in *Intelligent Environments (IE)*, 2013 9th International Conference on, ACM, 2013, pp. 100–107.
- [8] L. C. Huey, P. Sebastian, and M. Drieberg, "Augmented reality based indoor positioning navigation tool," in *Open Systems (ICOS)*, 2011 IEEE Conference on, 2011, pp. 256–260.
- [9] A. Möller, M. Kranz, S. Diewald, L. Roalter, R. Huitl, T. Stockinger, M. Koelle, and P. A. Lindemann, "Experimental evaluation of user interfaces for visual indoor navigation," 2014, pp. 3607–3616.
- [10] A. Mulloni, H. Seichter, and D. 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, pp. 211–220.
- [11] N. Bhanage and A. Zhong, "Improving User Experiences in Indoor Navigation Using Augmented Reality," EECS Department, University of California, Berkley, Thesis UCB/EECS-2014-73, May 2014.
- [12] Plamen Levchev, Michael N Krishnan, Chaoran Yu, Joseph Menke, and Avidesh Zakhor, "Simultaneous Fingerprinting and Mapping for Multimodal Image and WiFi Indoor Positioning," presented at the *Indoor Positioning and Indoor Navigation*, 2014.
- [13] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based localization in indoor environments," in *Computing for Geospatial Research and Application (COM. Geo)*, 2013 Fourth International Conference on, 2013, pp. 70–75.
- [14] JZ Liang, E Turner, A Zakhor, and N Corso, "Image-Based Positioning of Mobile Devices in Indoor Environments," in *Multimodal Location Estimation of Videos and Images*, Springer International Publishing, 2015, pp. 85–89.
- [15] M. A. Lakhani, "Indoor Navigation based on Fiducial Markers of Opportunity," 2013.
- [16] S. S. Chawathe, "Marker-based localizing for indoor navigation," in *Intelligent Transportation Systems Conference*, 2007. ITSC 2007. IEEE, 2007, pp. 885–890.
- [17] B. A. Delail, L. Weruaga, and M. J. Zemerly, "CAViAR: Context Aware Visual Indoor Augmented Reality for a University Campus," 2012, pp. 286–290.
- [18] D. Amin and S. Govilkar, "Comparative Study of Augmented Reality Sdk's," *Int. J. Comput. Sci. Appl.*, vol. 5, no. 1, pp. 11–26, Feb. 2015.
- [19] C. Koch, M. Neges, M. König, and M. Abramovici, "Natural markers for augmented reality-based indoor navigation and facility maintenance," *Autom. Constr.*, vol. 48, pp. 18–30, Dec. 2014.
- [20] "Metaio SDK," May-2013. [Online]. Available: <http://www.metaio.com/sdk/>.
- [21] U. Rehman, and S. Cao, "Experimental evaluation of indoor navigation devices," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59. Accepted.