3D BUILDING RECONSTRUCTION BY MULTIVIEW IMAGES AND THE INTEGRATED APPLICATION WITH AUGMENTED REALITY

Jin-Tsong Hwang*, Ting-Chen Chu

Dept. of Real Estate and Built Environment, National Taipei University, Taiwan -
E-mail: jthwang@mail.ntpu.edu.tw; applechu0417@gmail.com

Commission I, WG I/Vb

KEY WORDS: Point cloud, Augment Reality, Unmanned aerial vehicle, Structure from Motion

ABSTRACT:

This study presents an approach wherein photographs with a high degree of overlap are clicked using a digital camera and used to generate three-dimensional (3D) point clouds via feature point extraction and matching. To reconstruct a building model, an unmanned aerial vehicle (UAV) is used to click photographs from vertical shooting angles above the building. Multiview images are taken from the ground to eliminate the shielding effect on UAV images caused by trees. Point clouds from the UAV and multiview images are generated via Pix4Dmapper. By merging two sets of point clouds via tie points, the complete building model is reconstructed. The 3D models are reconstructed using AutoCAD 2016 to generate vectors from the point clouds; SketchUp Make 2016 is used to rebuild a complete building model with textures. To apply 3D building models in urban planning and design, a modern approach is to rebuild the digital models; however, replacing the landscape design and building distribution in real time is difficult as the frequency of building replacement increases. One potential solution to these problems is augmented reality (AR). Using Unity3D and Vuforia to design and implement the smartphone application service, a markerless AR of the building model can be built. This study is aimed at providing technical and design skills related to urban planning, urban designing, and building information retrieval using AR.

1. INTRODUCTION

A conceptual model to organize the various tools for visualization in urban planning was designed by Al-Kodmany (2002). According to this model, non-computer-based tools can be classified as (1) a pen and paper, (2) paper maps, (3) photographs, and (4) three-dimensional (3D) physical models. The corresponding computer-based tools are (1) an electronic pen and paper; (2) geographical information systems (GIS), computer-aided mapping (CAM), and mapping information management systems (MIMS); (3) image editing programs, digital photographs, and digital video; and (4) 3D digital models, virtual reality, and urban simulators (Al-Kodmany, 2002). Therefore, augmented reality (AR) tools belong most strongly to the fourth category. Digital building models are necessary for urban design and planning, 3D city modeling, and several other applications. Aerial imagery has traditionally been a data source for digital building model generation, and various aerial imagery approaches have been investigated and developed. Aerial imagery shooting by rotorcraft-based unmanned aerial vehicles (UAVs) is suitable for reconstructing 3D building models because photographs can be clicked at fixed camera stations and multiview images can be obtained. In this study, a rotorcraft-based UAV was used for data acquisition. The flight plan entailed vertical shooting on the building to generate a 3D building model. Because vertical shooting suffers from shading by objects, multiview images

from the ground must be included to remove the shielding effect.

Classical photogrammetry is rarely adopted to obtain high-resolution and high-accuracy products through UAV images. New computer-vision algorithms and techniques enable superior output generation (Lee et al., 2014). Muneza et al. (2015) used UAV imagery to produce a high-quality orthophoto to update geospatial data. A DJI Phantom 2 Vision Plus quadcopter was used to collect 954 images at a flying height of 50 m. Using appropriate photogrammetric software, an orthophoto with a 3.3-cm ground sample distance (GSD) covering 0.095 km² was produced. With appropriate ground control points, an absolute positional accuracy of a 7.9 cm root-mean-square error (RMSE) was achieved.

Many software products are based on image-based modeling, such as Photosynth, Speeded Up Robust Features (SURF), VisualSFM, and Pix4Dmapper. These products reconstruct the spatial geometry of the shooting subjects by searching for feature points to match and stitch images, restore camera poses, and inquire about the coordinates of the shooting subject.

The traditional approach for applying a 3D building model in urban planning and design involves one person operating a computer to display the digital building on the planned location after designing and space planning. Thus, the location and the view angle are controlled unilaterally by the operator on the

---

* Corresponding author

scene of the display; other audiences can only watch the result passively. A modern approach is to rebuild the digital models; however, it is difficult to replace the landscape design and the building distribution in real time as the frequency of building replacement increases. AR is a potential solution to these problems.

The conceptual basis of AR is the integration of the real world with the virtual world. Using computer graphics techniques, an object is generated with necessary information and then added into the real environment. By using computer vision, AR can recognize markers, such as simple black-and-white 2D-markers or real environment surroundings (Ruan and Jeong, 2012; Lin et al., 2009). These tracking methods can be divided into two different classes: marker-based and markerless. In the marker-based approach, different information can be displayed on a variety of surfaces depending on the purpose of the system (Gao et al., 2014). Meanwhile, markerless tracking based on image processing uses natural features in the images to evaluate the camera pose. AR can improve the recognition tools for the real world and thus enable efficient interactions between humans and computers (Kim, Young-geun, and Won-jung Kim, 2014). Alexandro Simonetti Ibañez (2013) describes what is understood by AR and the different varieties of AR applications and then discusses the software development kit (SDK)'s features, architecture, and elements. This article describes the basis of the detection process realized by the Vuforia library

2. METHODOLOGY

2.1 Building Model Reconstruction

The 3D building reconstruction of urban areas using remote sensing and photogrammetry data has many applications, such as mobile navigation, environment simulation, urban planning, and tourism. The main techniques used for 3D model reconstruction are lidar scanning, close-range photogrammetry, and aerial photogrammetry. In the past decade, many studies have been conducted to detect and classify lidar data and aerial images. Moussa & EI-Shemy (2012) filtered point cloud to separate ground and non-ground points and then divided them accordingly into trees and buildings using the normalized difference vegetation index (NDVI). Lidar is a powerful system in photogrammetry and remote sensing and a useful source of high-accuracy 3D point cloud data. For example, Chih-An Chang and Liang-Chien Chen (2006) combined lidar point cloud data with topographic maps to reconstruct a housing model. However, the various approaches differ in their characteristics and purpose. Lidar is a more expensive means of obtaining more detailed information, and close-range photogrammetry can rebuild the 3D model through a calibrated digital camera and is processed by proper software. In terms of convenience of use, the close-range photogrammetry approach is the fastest approach for obtaining information, and it is easy to implement. When calibrated appropriately, a general-use digital camera can be used to efficiently construct a model. However, close-range photogrammetry is highly dependent on human intervention.

Traditional photogrammetry and modern automated structure from motion (SfM) and multiview stereo (MVS) algorithms use an object's images to reconstruct it in 3D space (Szeliski, 2010). The SfM and MVS algorithms are based on feature detection, extraction, description, and the matching of image features. The application of a 2D image has been changed from

measurement on the image to 3D visualization techniques (Zhou et al., 2009; Hwang et al., 2015). This type of procedure has developed during the last decade. For example, the multiview image modeling technique takes a series of photographs from different positions and angles and then extracts the feature points from each photograph. Multiview image modeling, which is based on image matching and epipolar geometry theory, uses these feature points to reconstruct the exterior orientation parameters of each camera station and 3D information about the shooting objects. Because of its convenience, the multiview image modeling approach was adopted in this study.

2.2 Augmented Reality

Virtual reality offers a digital recreation of a real-world setting, while AR delivers virtual objects as an overlay to the real world. AR refers to the addition of a computer-assisted contextual layer of information over the real world, creating a reality that is enhanced or augmented. The most difficult task in AR is the constant calculation and replacement of the visual angle to accurately project the virtual image on an object. Therefore, the main function of AR is to calculate the distance and orientation from the observation device to the predefined pattern in reality and to accurately project the setup model on the predefined pattern by using computer vision technology.

There are two primary types of AR implementations: marker-based and markerless. The field of marker tracking examines artificial structures placed in the natural environment. These artificial structures are called markers. When designing a marker, it is important to consider that markers must be easy to distinguish from their environment. In the marker-based approach, markers are usually square, allowing the AR software to recognize and track them in a video stream. That is, markers are the physical patterns that we create or print out and place in the environment. Markerless AR uses sensors in devices to accurately detect the real-world environment and allows users to place virtual objects in that environment. Markerless tracking based on image processing uses natural features in the images to evaluate a camera pose. This approach involves the extraction of feature points from the frames of a video stream. Next, the matching points between consecutive frames are identified. Based on these matching points, the camera's pose is determined. Finally, the AR system stores unmatched feature points in the previous frames and calculates their 3D coordinates.

In this study, 3D building models were generated in SketchUp format as data resources for a markerless AR model. The 3D building model was depicted using Unity, a mobile app that enables users to visualize 3D products in AR superimposed on the user's real environment, at real size and in real time

2.2.1 Qualcomm® Vuforia Augmented Reality SDK:
Vuforia is an AR SDK for mobile devices that enables the creation of AR applications in real-time video obtained from these devices. It uses vision-based computing technology to recognize and track image targets and 3D objects in real-time. Vuforia features a markerless tracking mode wherein a marker can be dynamically captured at runtime if the image contains enough detail. The target-tracking capability enables developers to position and orient 3D models in relation to real-world images when these images are viewed through the camera of a mobile device. Multi-target detection, target tracking, and virtual buttons are some of the features of Vuforia SDK. Vuforia supports the detection of several kinds of targets (e.g., objects, images, and English text).

The features of the SDK still include localized occlusion detection using virtual buttons. Vuforia provides application programming interfaces (APIs) in C++, Java, Objective C, and the .Net languages through an extension to the Unity game engine. Unity is an integrated authoring tool for creating 3D video games or other interactive content, such as real-time 3D animations. In this paper, version 5.5 of Vuforia SDK is used.

Vuforia is an AR SDK for mobile devices. A software development environment or integrated development environment (IDE) is necessary, such as Eclipse or Unity. In this case, we chose the Unity engine because it is the easiest to use and because of its compatibility with the Vuforia library and the Android and iOS systems. The app developed in this study is only for Android devices because Android makes it easy to check the feature function of the app.

2.2.2 Unity3D: Unity is a cross-platform game engine developed by Unity Technologies and used to develop video games for PC, consoles, mobile devices, and websites. Unlike traditional 3D tools, such as 3DS Max and Maya, which only generate images and video, Unity is a flexible and powerful development platform for creating multiplatform 3D and 2D games and interactive experiences. In Unity 3D, we can assemble art and assets into scenes and environments, adding special effects, physics and animation, lighting, etc. Meanwhile, the Vuforia extension for Unity enables detection and tracking functionality in the Unity IDE and allows the easy creation of AR applications and games (Antoni Serra Font, 2013).
Because Vuforia is only a library, it is necessary to install Unity first, and the version of Unity we used is 5.3.4. Second, the Vuforia Unity Extension must be imported into the Unity project. Although all Vuforia SDKs are available for both Android and iOS systems, the Android mobile OS used in this study was Android because of its high availability..

3. EXPERIMENT

3.1 Study Area

National Taipei University (NTPU), a public university in north Taiwan, founded in 1949, was used as the experimental area for aerial photogrammetry. The university's main campus is in Sanxia District, New Taipei City, Taiwan, as shown in Fig. 1. The campus is approximately 54.53 hectares in area. The College of Law building was chosen as the case study. This building has eight stories with a bilateral U-sharp symmetry, as shown in Fig. 2. The building's dimensions were approximately 85 m × 42 m × 35 m (length × width × height).
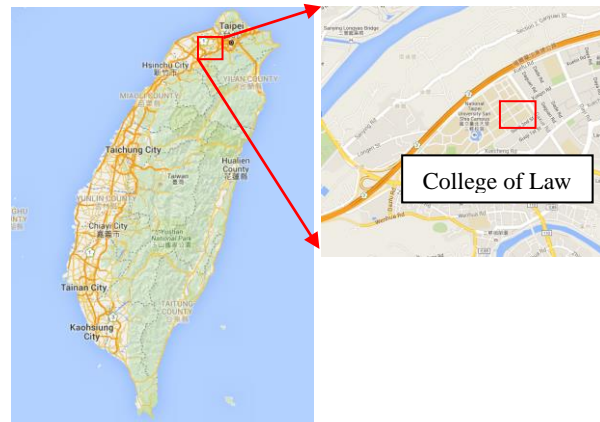

Figure 1. Location of the NTPU campus in Taiwan.

Figure 2. Case study building: College of Law.

Aerial photogrammetry was conducted using a rotorcraft-based UAV (YJ-1000-QC, AI-RIDER). The rotorcraft vehicle has the advantage of vertical takeoff and landing, which is useful for areas with many obstacles. The ground-monitoring system enables flight pre-planning, and the real-time images taken by a digital camera mounted on the UAV could be displayed on the flying controller. This type of UAV has a maximum load of 2.5 kg, a flight radius of 750 m, a GPS positioning accuracy of less than 2.5 m circular error probable, and a wind-resistant instantaneous wind speed of less than 18 m/s.

A rotorcraft-based UAV was used to conduct aerial photogrammetry on April 29, 2015. Two flying heights and shooting angles of the UAV were designed aslant to obtain photographs surrounding the building. Multiview images taken from the ground were also employed in the point cloud generation by Pix4Dmapper.

3.2 Multiview Images

The camera used in this experiment was a Canon EOS 5D Mark II with a focus length of 28 mm and a resolution of 0.0064 × 0.0064 mm per pixel. The frame size was 5616 × 3744 pixels per scene.

The building was photographed in two ways: vertical shooting using the UAV and multiview image acquisition from the ground. The flight altitude was approximately 200 m, and nine photographs were taken during the flight. The camera station distribution and the photographs from the flight are shown in Figs. 3 and 4, respectively. After the flights, 295 multiview images of the side of the building and the tree-shaded area were

taken in one round from the ground using the Canon EOS 5D Mark II.

Multiview images were obtained from distances of approximately 10–20 m from the building center. However, the locations of some of the camera stations were affected by the topography, causing slight variation in the shooting distance, and the photo overlap rate was approximately 80%. To build dense point clouds of the building while taking multiview images from the ground, each shot is designed to obtain as complete of a view of the building as possible. The distribution of the camera stations and the multiview images taken from the ground are shown in Figs. 5 and 6, respectively..
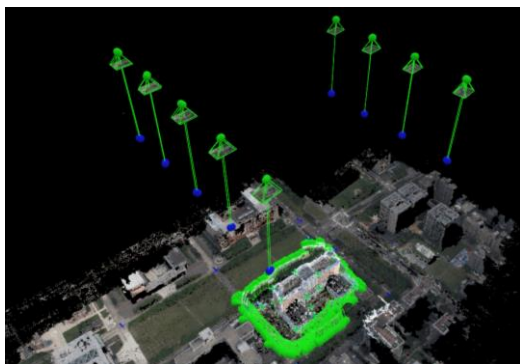

Figure 3. Camera station location for UAV and multiview images and the point cloud.

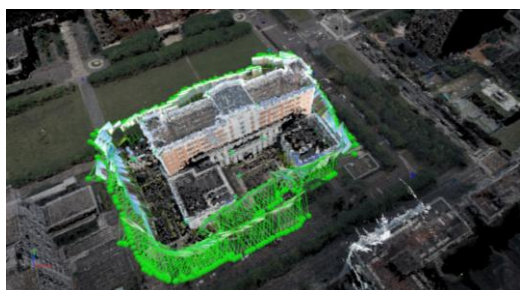Figure 4. Vertical shooting photographs of the building by UAV.

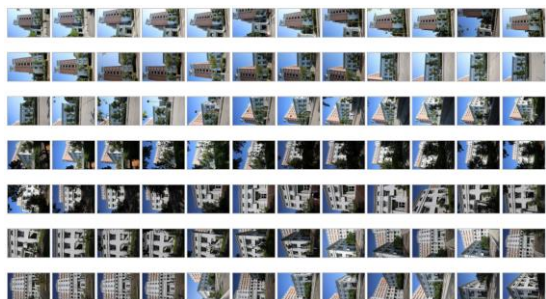Figure 5. Camera station distribution and multiview image photographs.

Figure 6. Part of multiview images taken from the ground of the area surrounding the building.

3.3 Augment Reality

AR refers to the addition of a computer-assisted contextual layer of information over the real world, creating a reality that is enhanced or augmented.

There are two primary types of AR implementation: marker-based and markerless. The markers in the marker-based approach are squares that are recognized and tracked by the AR software in a video stream. That is, the markers are physical patterns that we have created or printed out.

The basic equipment for AR is as follows:
- Video capture devices for video input, such as digital video cameras or mobile phones.
- A monitor for displaying images. After the augmented information is combined with the real video image, the superimposed image is shown on the screen.
- Computer and software. Using Unity as an example, a programmer can use the SDK provided by Vuforia and Unity to compile and build an application..

3.3.1 Marker AR: In the commonly used AR system, the orientation pattern of a system usually consists of a square border containing asymmetrical patterns. The advantage of AR is that it can provide meaningful content for a small region and easily detect the spatial location. However, the disadvantage of AR is that identification is only possible after viewing the entire card, meaning that the card cannot be partially obscured. AR can track markers; therefore, the model is superimposed on the real video stream. The resulting combination of the virtual model and the real environment is output to the monitor.

3.3.2 Markerless AR: Markerless AR uses sensors in devices to accurately detect the real-world environment and allows users to place virtual objects in that environment. Markerless tracking based on image processing uses natural features in the images to evaluate a camera pose. First, feature points are extracted from the frames of a video stream. Next, the matching points between consecutive frames are located. Based on these matching points, the camera's pose is determined. The AR system stores unmatched feature points in the previous frames and calculates their 3D coordinates. The procedure for markerless tracking systems is as follows:
- the AR system extracts natural features in the current frame;
- the system compares the features and identifies matches;
- the approximated camera poses are evaluated using the matching points whose 3D positions are already known for these calculations;
the system calculates the 3D positions of the matching points whose 3D positions are unknown (Ziegler, E., 2010).

4. RESULTS

4.1 Building Model Reconstruction

In this study, the models built from photographs taken by a UAV and multiview images were merged to minimize the effects of shielding, as images taken by vertical photogrammetry are usually shielded by trees.

Nine photographs were taken by the UAV, and 295 multiview images were taken by the Canon EOS 5D Mark II. The buildings were reconstructed as 3D models from the images.

The images were processed into point clouds using Pix4Dmapper, as shown in Fig. 7.

The 3D models were reconstructed using AutoCAD 2016 to generate vectors from the point clouds and using SketchUp Make 2016 to rebuild a complete building model with textures. The 3D building model is shown in Fig. 8.

Figure 7. Point cloud of the building generated by Pix4dmapper.

Figure 8. Building model generated from the point cloud by AutoCAD 2016 and SketchUp Make 2016.
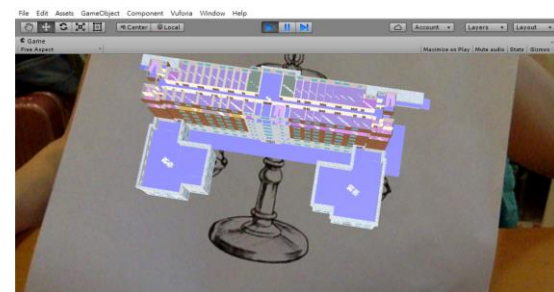
4.2 Markerless AR

A mobile phone with the Unity app can be used to visualize this 3D building model in real environments by using the predefined markers of real buildings or real graphs. We can easily visualize the 3D model by using AR, as shown in Figs. 9(a) and (b).

(a)
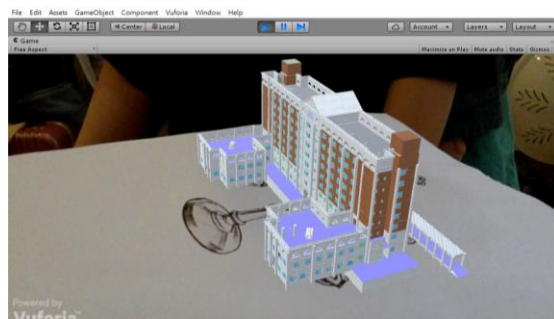
(b)

(c)

(d)

Figure 9. 3D building model created for a real environment by using the predefined marker in AR: (a) marker of the real graph; (b) the 3D building model displayed on a mobile phone by using the Unity app; (c) interior spatial arrangement of the selected floor; (d) the main designs of the interaction functions, which include object manipulation and information retrieval.

4.3 Human-Object Interaction

In this study, the interaction functions are mainly object manipulation and information retrieval. The object manipulation function includes select, hide, show, resize, translation, and swap. Meanwhile, the information retrieval function includes text, image, and hyperlink. The object manipulation method is a feature that allows the user to select any object and perform a function to manipulate the objects. The manipulation method enables the selected objects to be hidden or shown and can isolate particular objects from others.

In this case, the building model consists of eight floors, each of which was set up as a user-selectable object. In this way, the user can select a desired floor by clicking on the outside surface of the building model. When the user clicks on one floor, the pop-up canvas and a hyperlink button connected to the floor appear; when clicked, this button simultaneously presents the user with detailed information about this building. The canvas is a game object with a canvas component on it, and all user

interface (UI) elements should be contained inside. Additionally, those floors above the floor that was clicked on will be automatically hidden, allowing the floor plan to be observed by looking down on the building model. The hyperlink button was designed to link to the College of Law website, making it convenient for users to obtain information about the department, as shown in Figs. 9(c) and (d).

In the application fields of design study and urban design study, combining real and virtual environments with this kind of interaction allows the user to flexibly examine and recognize objects and their contents, with the object swap, rotate, resize, and translation manipulation functions being particularly important.

5. CONCLUSIONS

In this study, to obtain the building model, photographs taken around the building by the UAV were combined with multiview images taken from the ground to remove plant shielding. Next, the 3D models were reconstructed using AutoCAD 2016 to generate vectors from point clouds and using SketchUp Make 2016 to rebuild the complete building model with textures. The building model was used as a resource for the markerless AR in Unity.

The rotorcraft-based UAV was a suitable platform for obtaining images for the building model reconstruction, especially when configured for shooting around the building and different flight altitudes with a high view angle tilts. This vehicle has the advantages of vertical takeoff and landing, which are useful for the areas with many obstacles. Thus, for building model reconstruction, a rotorcraft-based UAV is a more flexible choice than a fixed-wing UAV.

In urban planning and design, a modern approach is to build digital models instead of entities; however, it is still difficult to replace the landscape design and relocate the buildings in real time. One potential means of solving these problems is the use of AR. The building model was expressed in SketchUp format and linked with the marker AR programmed using Unity. Upon the retrieval of the building information, the concepts presented by an urban planner or a developer can be executed. Additionally, users can access detailed information for each floor without visiting the real building by interacting with it in AR. This application can be used as a tool for conducting urban-scale visual representations with highly flexible manipulation of objects in real time.

ACKNOWLEDGEMENTS

REFERENCES

Al-Kodmany, K., 2002. Vizualization Tools and Methods in Community Planning: From Freehand Sketches to Virtual Reality. *Journal of Planning Literature* 17, 189-211.

Alexandro Simonetti Ibañez, 2013. Vuforia v1.5 SDK: Analysis and evaluation of capabilities, Master thesis, Master in Science in Telecommunication Engineering & Management, Universitat politecnica de catalunya, 119 pages.

Antoni Serra Font, 2013. Location based augmented reality application o ted reality application on Unity 3D, Master thesis, Master in Science in Telecommunication Engineering & Management, Universitat politecnica de catalunya, 55 pages.

Chang, Chih-An and Chen, Liang-Chien, 2006. Building Shaping from LIDAR Data, Journal of Photogrammetry and Remote Sensing, 11(2), June, pp. 175-189.

Gao, X., Tian, J., Liang, X., and Wang, G. (2014). ARPP: An Augmented Reality 3d ping-pong game system on Android mobile platform. Pages 1–6. IEEE.

Hwang, Jin-Tsong, Ting-Chen Chu, and Chien-Hong Chen, 2015. Accuracy Assessment of 3D Building Construction by Multiview Images and the Integrated Application with Augmented Reality, The 36th Asian Conference On Remote Sension, Metro Manila Philippines.

Kato, H. Billinghurst, M. Poupyrev, I. Imamoto, K. Tachibana, K., 2000. Virtual object manipulation on a table-top AR environment, IEEE and ACM International Symposium on Augmented Reality, pp.111-119.

Lee, J. N., K. C. Kwak, 2014. A Trends Analysis of Image processing in Unmanned Aerial Vehicle, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 8(2), pp. 2-5.

Lin, L., Wang, Y., Liu, Y., Xiong, C., and Zeng, K. (2009). Marker-less registration based on template tracking for augmented reality. Multimedia Tools and Applications, 41(2):235–252.

Lowe, D. G., 2004. Distinctive Image Features. Internation Journal of Computer Vision, 60(2), pp. 91-110.

Mikolajczyk, K. and C. Schmid, 2005. A Performance Evaluation of Local Descriptors. IEEE Trans. Pattern Analysis and Machine Intelligence, 27(10), pp. 1615-1630.

Muneza J. Maurice1 , Mila N. Koeva2 , Markus Gerke 2 , Francesco Nex 2 ,Caroline Gevaert, 2015. A photogrammetric approach for map updating using UAV in Rwanda, International Conference on Geospatial Technologies for Sustainable Urban and Rural Development, November 18 - 20, 2015 , Kigali – Rwanda.

Ruan, K. and Jeong, H. (2012). An Augmented Reality System Using Qr Code as Marker in Android Smartphone. Pages 1–3. IEEE.

Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media.

Young-geun Kim and Won-jung Kim, 2014. Implementation of Augmented Reality System for Smartphone Advertisements, International Journal of Multimedia and Ubiquitous Engineering Vol.9, No.2 (2014), pp.385-392.

Ziegler, E., 2101. Real-time markerless tracking of objects on mobile devices, thesis of Bachelor of Science, im Studiengang Computervisualistik, University KOBLENZ, LANDAU, 83 pages.

Zhou, K., 2010. Structure & Motion. Structure in Pattern Recognition, Vienna University of Technology, Faculty of

Informatics, Institute of Computer Graphics and Algorithms, Pattern Recognition and Image Processing Group.

Zhou, K., M. Zillich, and M. Vincze, 2009. Reconstruction of three dimensional spatial clusters using monocular camera, Information Sciences and Systems, CISS 2009, 43rd Annual Conference, pp. 689-694.