

## DEVELOPMENT OF 3D CITY MODEL USING VIDEOGRAMMETRY TECHNIQUE

Hairuddin, A.<sup>1</sup>, Azri, S.<sup>1</sup>, Ujang, U.<sup>1</sup>, Cuétara, M.G.<sup>2</sup>, Retortillo, G.M.<sup>2</sup> and Mohd Salleh, S.<sup>1</sup>

<sup>1</sup>Dep. of Geoinformation, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia (UTM), Johor, Malaysia  
afina3@live.utm.my, suhaibah@utm.my, mduznir@utm.my, syahiirahsalleh78@gmail.com

<sup>2</sup>Planta 5ª Oficina 16, Calle Luis Álvarez Lencero, 3, 06011 Badajoz, Spain  
ceo@ecapture3d.com, guillermo@ecapture3d.com

**KEY WORDS:** Videogrammetry, 3D city model, point cloud, 3D building

### ABSTRACT:

3D city model is a representation of urban area in digital format that contains building and other information. The current approaches are using photogrammetry and laser scanning to develop 3D city model. However, these techniques are time consuming and quite costly. Besides that, laser scanning and photogrammetry need professional skills and expertise to handle hardware and tools. In this study, videogrammetry is proposed as a technique to develop 3D city model. This technique uses video frame sequences to generate point cloud. Videos are processed using EyesCloud3D by eCapture. EyesCloud3D allows user to upload raw data of video format to generate point clouds. There are five main phases in this study to generate 3D city model which are calibration, video recording, point cloud extraction, 3D modeling and 3D city model representation. In this study, 3D city model with Level of Detail 2 is produced. Simple query is performed from the database to retrieve the attributes of the 3D city model.

### 1. INTRODUCTION

Videogrammetry technique is one of the technique that can be used to get 3-Dimensional (3D) information of an object. A set of cameras were used to get spatial data for videogrammetry technique. Decision making time can be reduced by using videogrammetry and the complexity of data processing for infrastructure spatial data also can be minimized. This technique gives permission to user during data scanning to move the camera for time-dependent imaging. Due to its ability to capture millions of pixels, videogrammetry gained attention because of its high resolution of video sensors (Ioannis Brilakis et al., 2011). Compared to images, video sequence (videogrammetry) is able to give more valuable information that can help in processing. Post processing is done where small portion of video frames is selected to get gratification result (Abbas Rashidi et al., 2013).

Videogrammetry allows the capturing of thousands of points in a very short space of time like laser scanner and the objects geometry with texture (J Herráez et al., 2013). Videogrammetry is the process of measuring coordinates of object points from two or more video streams captured by camcorder. A videogrammetric approach requires little human intervention and can provide a high degree of automation (Habib Fathi et al., 2011).

3D city model is important visualisation of urban area and also the application of 3D city model is increasing rapidly these day (Gerhard Groger et al., 2012; Salleh et. al 2018; Mohd et. al. 2016). A 3D city model is useful in various applications like urban planning, training, as simulation of urban scenarios and virtual heritage conservation (Frueh et al., 2003). 3D city model is a representation of Earth's surface in digital format and any related objects that can be found on the surface. It is a computerized or city model that has graphic of buildings or other related objects in 2.5 or 3D (Jain

et al., 2013). Advantages of 3D city model when compared to 2D maps includes having 3D geometry of the features, illumination effect on the feature, shadow model on the features, texture information and change of perspective (Chen, 2011).

The aim of this study is to study videogrammetry technique in developing 3D city model. It is to see whether videogrammetry technique is compatible to produce 3D city model based on the point cloud that has been processed. The objectives of this study are to review existing technique for 3D city model, to explore videogrammetry technique in producing 3D city model and to generate 3D city model using videogrammetry technique.

### 2. STUDY BACKGROUND

3D city modeling is commonly done using photogrammetry technique and laser scanner which is known to be time-consuming to produce and an expensive instrument. 3D city model is widely used in applications in various fields (Azri et. al. 2019; Mohd et. al. 2017; Ujang et. al. 2018; Azri et. al. 2014). A lot of techniques can be used for 3d city modelling. An application that needs detailed building models for small area can use LiDAR. On the other hand, large areas which do not require detailed features are usually reconstructed from satellite or aerial images are more efficient (Kim et al., 2014).

Photogrammetry and laser scanning requires expertise to handle both software and hardware. The need of expertise for both techniques is due to its complexity in data collection and data processing. Photogrammetry modeling is accurate but time-consuming. This is due to the registration accuracy which depends on wide-ranging network of control and tie points in the photographs that may need up to several days of preprocessing and strength or weakness of collected data is not readily apparent (Markley et al., 2008).

Facade texture created by using photogrammetric and laser scanning requires extensive laborious works and still the bottle neck in photorealistic building modelling process (Rasam et al., 2013). One of the issues of laser scanning is the large data volumes that necessitates increased computer processing and memory storage to support the data (Markley et al., 2008). Scanning using laser for large building required numbers of scans or low resolution of long range scanning (El-Hakim et al., 2003).

Level of Details (LODs) is an important concept of 3D modelling that can be defined as outlined degree or prior presentation of real world objects. This study attempted to generate 3D city models with Level of Detail (LoD) 2. LOD2 is more detailed model than LOD1, in which individual buildings are mandated and are usually modelled as simple structures containing standard and simplified roof structures (Biljecki, 2017). The building measured is only within a height of not more than 7 meters. The focus of this study is the building whereby any textures like trees, road and street furniture are not included as the texture is not part of LOD2 specification.

### 3. METHODOLOGY

#### 3.1 Tool



Figure 1. Huawei P10 Lite

Table 1 Camera specification of Huawei P10 Lite

No.	Parameters	Values
1	Mega Pixels (MP)	12
2	Aperture	f/2.2
3	Maximum aperture	1/ 2.8"
4	Size of pixels	1.25µm
5	Frame Per Second (fps)	30
6	Resolution	1080 (1920x1080)

#### 3.2 Selection of Software

Various types of software were used to complete this study. The software is used from data processing until data presentation. Figure 2 shows the collection of software used through this study.



Figure 2 The collection of software used in this study

#### 3.3 Flow Chart

For data acquisition, Huawei P10 Lite (Figure 1) is used. Figure 3 shows the flow chart for this research. It contains five (5) stages which are preliminary study, data acquisition, data processing, data manipulation and data presentation.

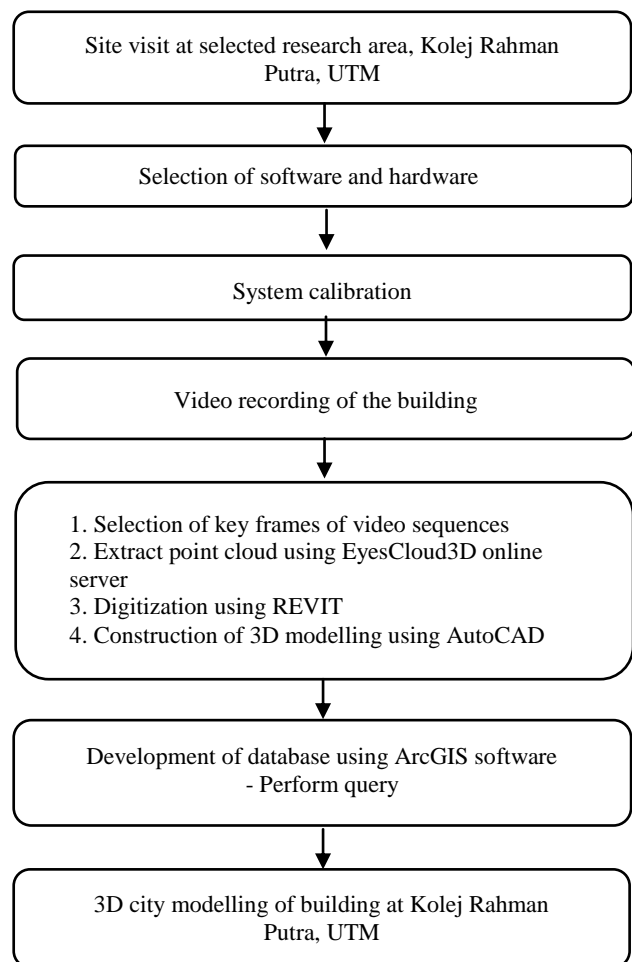


Figure 3 Flow chart of study

### 3.3.1 Preliminary study

Preliminary study is done before data collection started. It can help to identify the condition of the building, suitable distance between hardware and building, and to identify any obstacle that may interfere during video recording. The location of study area is located at Block G13, G14, G15, G16 and G17, Kolej Rahman Putra (KRP), UTM. Figure 4 shows the structure of the buildings in real world. The selection of software also is done in preliminary study to identify the software that may be suitable for data processing.



Figure 4 Block G14, KRP, UTM

### 3.3.1 Data Collection

For data collection, system calibration of the hardware, Huawei P10 Lite was done. After calibration, video recording of the building can be started.

#### 3.3.1.1 System Calibration

Before data collection, system calibrations for hardware tools need to be done to ensure the accuracy of the hardware and also to avoid any error of measurement during data processing.

System calibration is done by recording a video of 3D calibration frame that has numbered markers. The frame (Figure 5) contains 146 markers which can be measured using Total Station to identify the actual distance between markers and the coordinate of the markers.



Figure 5 3D calibration frames

The coordinate of X, Y and distance, Z, is recorded to calculate the distance by using mathematical formula (Figure 6).

$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Figure 6 Pythagoras theorem

3D calibration frame application (figure 7) is developed to identify the distance between two markers. The application contains information like coordinate of X, Y, and Z, name of instrument, date of observation and also person in charge. User needs to insert the number of markers they observed in the application to get the predicted distance and compared the result between observed and predicted.

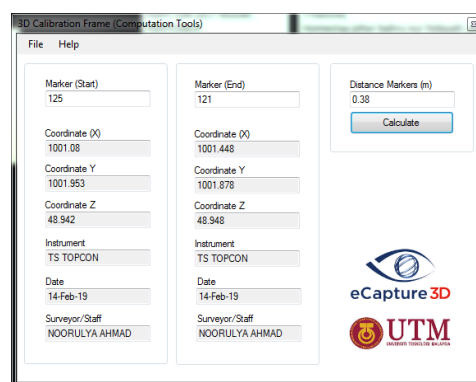


Figure 7 3D Calibration Frame application

Figure 8 shows the distance measured between two markers in X, Y and Z coordinate. The distance will be compared between observed and actual distance. Table 2 shows the result of calibration in X, Y and Z coordinates which were compared and the difference between actual and observed distance was computed.

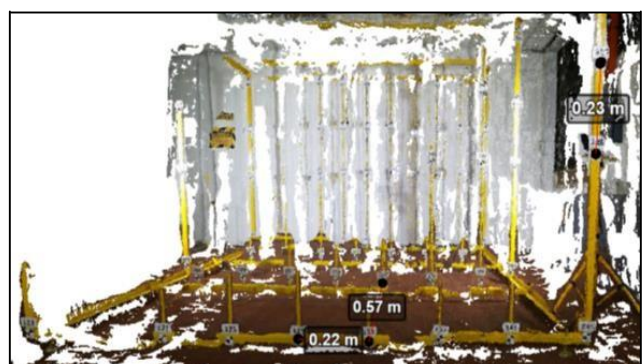


Figure 8 The distance between two markers in X, Y and Z coordinate

$$\text{Difference(m)} = \text{Actual Distance} - \text{Measured Distance}$$

Figure 9 Equation for calibration coordinate

Table 2 Calibration result for coordinate X, Y and Z

Coordinate	Marker	Actual distance (m)	Measured distance (m)	Difference (m)
X	129 – 133	0.37	0.22	0.15
Y	145 – 144	0.41	0.23	0.18
Z	133 – 100	1.01	0.57	0.44

### 3.3.1.2 Video Recording

Video recording of building is done by walking around the selected building. The video recording need to ensure that all building features is recorded by walking around the building at a slow pace. The distance between camera and building can vary. It should have 60% to 100% overlapped between two frames to get better result. Point cloud will be generated from the video by combining point cloud that have same properties with the other frames. The video recording should be not more than 60 seconds and the camera should walk along the building and not by open angle from one point. The building is being recorded part by part where one (1) part of building is one video, and the camera needs to be parallel with the building during video recording as shown in Figure 10.

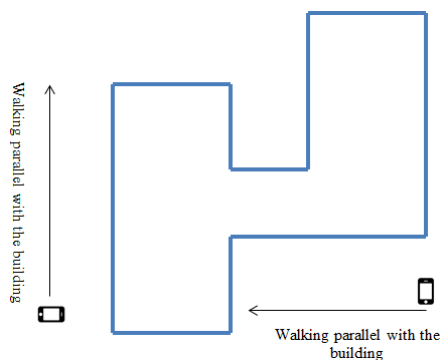


Figure 10 Illustration of the position of camera and how camera video records the building parts

### 3.3.2 Data Processing

After all parts of the building have been recorded, the video is processed to get point cloud of building before continue to generate 3D models. The data processing is started with selection of video key frames, extraction of point cloud, digitization and lastly, 3D reconstruction. Figure 11 shows the flow of data processing automatically.

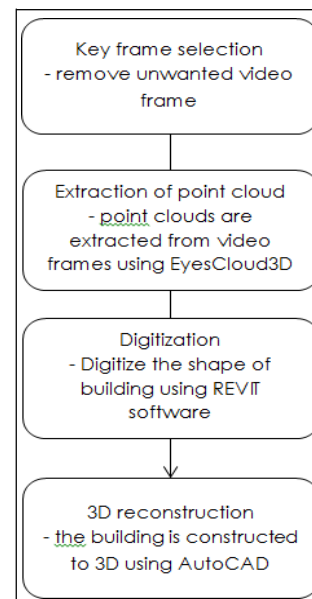


Figure 11 Flow chart of data processing

#### 3.3.2.1 Key Frame Selection

Key frame selection is done to minimize or reduce the video frames that been recorded. Any disturbances that are present in the video frames are deleted like trees, leaves or unclear frames. Although some video frames are removed, the video need to make sure it still covers all parts of the building. Adobe Premiere Pro is used to remove the video frames. It is a tool that specifically used for video editing and trimming. The selection of video frames can produce better results of point cloud, and the shape of building in point cloud can be seen clearly.

The extraction of different numbers of key frames is to find the optimum number of frames required for processing and also processed for different video streams captured from one specific scene in an order (Rashidi, 2014). The main purpose of key frame selection is to maintain as much view coverage of the target object as possible with the minimal necessary number of frames (Wang et al., 2014).

#### 3.3.2.2 Extraction of Point Cloud

The video frame is uploaded into Eyescloud3D for extraction of point clouds. The point cloud is produced and extracted from each video frames. Any point clouds that have same properties with other frames' point cloud will be combined to become one point cloud. To upload the video, the length of video should be not more than 60 seconds and internet connection is needed to upload the data into server. After point cloud has been generated, it will be downloaded from server into computer's local file.

EyesCloud3D is cloud software developed by eCapture3D. EyesCloud3D allows user to upload raw data either in picture format or video format to generate point clouds and mesh model. ECapture3D is a custom 3D technology services company that will develop a 3D technology model that universalizes access by all types of users easily and economically accessible. EyesCloud3D can be used and accessed from various platform like mobile phone, computer or tablet, and users also does not need to install it to any hardware because it used cloud platform. It also can generate 3D models



The extraction of point clouds is done by utilizing the relationship of epipolar and consistency of photo. It optimizes the original model by utilizing the match points achieved and the original space location. A method of fast extract algorithm is used to extract the feature points from the point visible field of neighbor images with an algorithm (Guojun Chen et al., 2010). To obtain accurate and dense 3D points, the feature points were extracted from every video frames that contain many features and can be found at all frames (J. Ishii et al., 2013). To get more accurate matching points, the selection of key frames is required to improve the efficiency (C. Zhang et al., 2017).

### 3.3.2.3 Digitization

Digitization is a process to get the shape of the building based on the point clouds that has been generated from video. For this process, AutoDesk REVIT is used to digitize the shape of building. The digitization is done to get the 2D wireframe of building which includes wall, roof and beam. A line is drawn along the sideline of building to get the 2D wireframe and the accurate shape of building parts. The 2D wireframe will be converted to 3D model.

### 3.3.2.4 3D Reconstruction

3D reconstruction is a step where 2D wireframe is converted into 3D surface. The wireframe is extruded with specific value of depth to give a view of 3D. All parts of building like walls, beam and roofs is converted to 3D and all parts were combined as one to obtain a representation of the building in 3D which is as close to reality as possible. For this study, the 3D building generated is only until Level of Detail 2 that does not include doors and windows

### 3.3.4 Data Manipulation

A database (Figure 12) is created that contains attribute or information of the building. The database stores attribute data to facilitate simple queries and make any data selection easier. The database contained information such as block name, total floor, total single and double room, and other related information. The database is created using ArcScene software as the 3D city model is being viewed using ArcScene.

FID	Shape*	Name	No. Floor	DoubleRoom	SingleRoom	TotalRoom	RoadName
0	MultPatch M	G13	5	15	14	29	Jalan Meranti 2

Figure 12 The database of Block G13, KRP, UTM

### 3.3.5 Data Presentation

The model is represented by using ArcGIS software, ArcScene as shown in figure 13. The position of the building matches the real world position of the buildings to have better understanding and visualisation of the buildings.





Figure 13 The visualize of 3D city model

## 4. RESULTS AND DISCUSSION

Based on the resulting point clouds, the number of points are varied because it depends on the area of building being measured and also disturbances such as trees and leaves. The color of building can also affect the number of points such as bright colors will result in high reflection. Table 3 shows the number of point clouds according to each part of the building.

Table 3 The number of point clouds

Point clouds	No. of point cloud
	773, 509
	2, 919, 398
	674, 137

	1, 305, 664
	1, 305, 664
	1, 235, 656
	1, 249, 913
	1, 100, 742
	1, 100, 742
	1, 299, 961
	685, 678



1, 148, 833

From the point clouds, the building frame is digitized based on the shape of point cloud. The digitization will produce 2D wireframe which will be used to generate 3D model. Each part of the building is digitized to be 2D wireframe. Figure 15 shows the result of 2D wireframe of a building part.

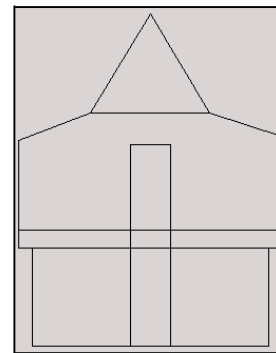


Figure 15 2D wireframe of building part

From 2D wireframe, 3D building is converted and generate the shape of the building (figure 16). Part by part of the building is generated into 3D model. The building parts are then combined to get one complete building that hold all parts as shown in Figure 17.

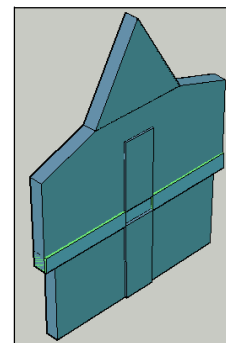


Figure 16 3D model of building part

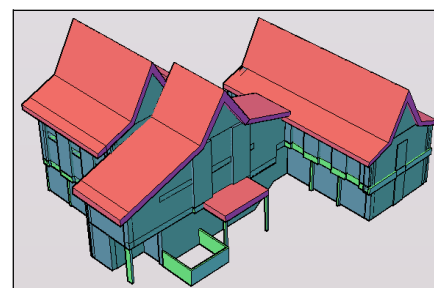


Figure 17 The complete 3D building

Once the assembly of the building is finished, the building is place at its position to be visualised as a 3D city model as shown in Figure 18.

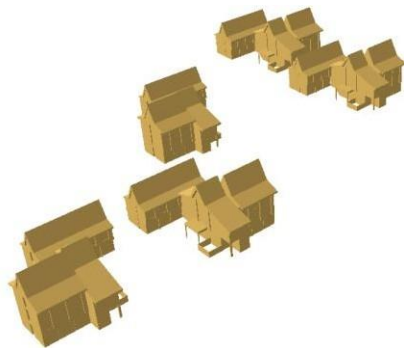


Figure 18 the presentation of 3d city model

Figure 19 shows the syntax to perform a query based on the database developed where a simple query is performed to retrieve information based on the condition. Figure 20 shows the query result.

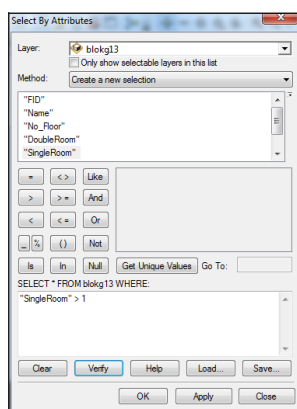


Figure 19 Simple syntax for query

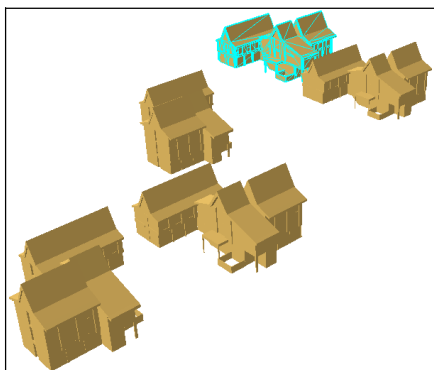


Figure 20 result of query

For this study, level 2 of level of detail (LoD) is used to visualise the building. LoD 2 is a model with a simplified roof shape, and where the object's part can be modelled in multiple semantic classes. Level of detail is an omnipresent concept in geographic information. It is used for the communication of how thoroughly real-world features have been acquired and modelled. LoDs are not always ordinal. The textures can be added to LoD though it is not part of the LoD specification. The LoD can also include other thematic classes like roads and vegetations which have their LoD description, but they are less prominent and the acquisition is not prescribed for each LoD (Biljecki, 2017). The accuracy of the 3D building is in centimetre level of 15-20cm.

## 5. CONCLUSION

Based on the study, videogrammetry can be used as one of the technique to produce 3D city model. Videogrammetry technique is not awell known technique like existing technique of photogrammetry and laser scanning which are usually used to produce 3D city model. Due to disadvantages of the existing techniques, videogrammetry technique can be used to overcome those disadvantages. This technique can help in reducing the time taken for data acquisition and processing, reducing cost for hardware as user can use their own camera, and also videogrammetry does not require high expertise to handle both hardware and software for data processing. Videogrammetry technique also can give permission to its user to move from their position to record a video of the building or object.

## ACKNOWLEDGEMENT

This research was partially funded by UTM Research University Grant, Vot Q.J130000.3552.05G34 and Vot Q.J130000.3552.06G41. This research is also part of the collaboration between 3D GIS Research Lab, UTM and eCapture3D (Ecapture Research and Development SL). eCapture has provided eyescloud3d for this research via their platform [www.eyescloud3d.com](http://www.eyescloud3d.com).

## REFERENCES

- Abbas Rashidi, Fei Dai, Ioannis Brilakis, & Velad, P. (2013). Optimized selection of key frames for monocular videogrammetric surveying of civil infrastructure. *Advanced Engineering Informatics, Volume 27*(Issue 2), Pages 270-282. doi:<https://doi.org/10.1016/j.aei.2013.01.002>
- Azri, S., Ujang, U., Abdul Rahman, A., 2019. 3D Geo-clustering for Wireless Sensor Network in Smart City. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-4/W12*, 11-16.
- Azri, S., Ujang, U., François, A., Darka, M., Alias, A.R., 2014. Spatial Access Method for Urban Geospatial Database Management: An Efficient Approach of 3D Vector Data Clustering Technique, 9th International Conference on Digital Information Management (ICDIM). IEEE, Bangkok, Thailand.
- Biljecki, F. (2017). Level of Detail in 3D City Models. 353. doi:10.4233/uuid:f12931b7-5113-47ef-bfd4-688aae3be248
- C. Zhang, H. Wang, & Liu, J. (2017). A Fast Key Frame Extraction Algorithm And An Accurate Fetaure Matching Method for 3D Reconstruction From Aerial Video. *2017 29th Chinese Control And Decision Conference (CCDC)*, 6744-6749. doi:10.1109/CCDC.2017.7978392
- Chen, R. (2011). The Development of 3D City Model and Its Applications in Urban Planning. *2011 19th International Conference on Geoinformatics*, 1-5. doi:10.1109/GeoInformatics.2011.5981007
- El-Hakim, S.F., Beraldin, J.A., Picard, M., & Vettore, A. (2003). *Effective 3D modeling of heritage sites*. Paper presented at the Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.

- Frueh, C., & Zakhor, A. (2003). *Constructing 3D city models by merging ground-based and airborne views*. Paper presented at the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.
- Gerhard Groger, & Plumer, L. (2012). CityGML - Interoperable Semantic 3D City Models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 12-33.
- Guojun Chen, Jie Jiang, & Su, H. (2010). An Optimized Model of Polyhedral Visual Hull Based on Feature Points Matching. 2010 International Conference on Intelligent Computing and Cognitive Informatics. doi:10.1109/ICICCI.2010.123
- Habib Fathi, I.B., & Brilakis, I. (2011). Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features. *Advanced Engineering Informatics*, 25(4), 760-770. doi:https://doi.org/10.1016/j.aei.2011.06.001
- Ioannis Brilakis, Habib Fathi b, & Rashidi, A. (2011). Progressive 3D reconstruction of infrastructure with videogrammetry. *Automation in Construction*, Volume 20 (Issue 7), Pages 884-895. doi:https://doi.org/10.1016/j.autcon.2011.03.005
- J Herráez, J Martínez-Llario, E Coll, Rodríguez, J., & Martin, M.T. (2013). Design and calibration of a 3D modeling system by videogrammetry. *Measurement Science and Technology*, 24. doi:https://doi.org/10.1088/0957-0233/24/3/035001
- J. Ishii, S. Sakai, K. Ito, T. Aoki, T. Yanagi, & Ando, T. (2013). 3D Reconstruction of Urban Environments Using In-Vehicle Fisheye Camera. 2013 IEEE International Conference on Image Processing, 2145-2148. doi:10.1109/ICIP.2013.6738442
- Jain, Kamal & Gurjar, Surendra Pal & Mandla, & Venkata. (2013). Virtual 3D City modeling: Techniques and Applications. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. doi:10.5194/isprsarchives-XL-2-W2-73-2013
- Kim, H., & Han, S. (2014). A framework for the automatic 3D city modeling using the panoramic image from mobile mapping system and digital maps. Paper presented at the 2014 IEEE Virtual Reality (VR).
- Markley, J.D., Stutzman, J.R., & Harris, E.N. (2008). Hybridization of Photogrammetry and Laser Scanning Technology for As-Built 3D CAD Models. Paper presented at the 2008 IEEE Aerospace Conference.
- Mohd, Z.H., Ujang, U., Choon, T.L., 2017. Heritage House Maintenance using 3D City Model Application Domain Extension Approach. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Mohd, Z.H., Ujang, U., 2016. Integrating Multiple Criteria Evaluation and GIS In Ecotourism: A Review. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-4/W1*, 351-354.
- Rasam, A.R.A., Hamid, N.A., Hamid, M.Z.A., Maarof, I., & Samad, A.M. (2013). 3D photorealistic modeling of university building using Digital Close Range Photogrammetry. Paper presented at the 2013 IEEE 4th Control and System Graduate Research Colloquium.
- Rashidi, A. (2014). Improved Monocular Videogrammetry for Generating 3D Dense Point Clouds of Built Infrastructure. Georgia Institute of Technology,
- Syahiirah, S., Ujang, U., 2018. Topological information extraction from buildings in CityGML. *IOP Conference Series: Earth and Environmental Science* 169, 012088.
- Ujang, U., Azri, S., Zahir, M., Abdul Rahman, A., Choon, T.L., 2018. Urban Heat Island Micro-Mapping via 3D City Model. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-4/W10*, 201-207.
- Wang, G., Lu, Y., Zhang, L., Alfarrarjeh, A., Zimmermann, R., Kim, S.H., & Shahabi, C. (2014). Active key frame selection for 3D model reconstruction from crowdsourced geo-tagged videos. Paper presented at the 2014 IEEE International Conference on Multimedia and Expo (ICME).

*Revised August 2019*