Structure from Motion of an urban area on ground level perspective with an absent of control point

Tuntapat Sirikup[*1], Nipon Jongpitaksyl[1], Parnuhmesr Sirinaranun[1], Thumanoon Susumpow[1] and Yoshihisa Maruyama[2]

[1]Department of Civil Engineering, College of Engineering, Rangsit University, Pathum Thani, Thailand
[2]Department of Urban Environment Systems, Graduate School of Engineering, Chiba University, Nishi-chiba, Japan
[*]Corresponding author, E-mail: tuntapat.s57@rsu.ac.th

Abstract

With a remarkable efficiency of three-dimensional visual model of photo scanning, since the 1960s, the day that the computer vision was born, many pieces of research have been subjected to this study. Although a lot of methodology of 3D creation has been made so far such as 3D-model from the ultrasonic wave and 3D-model from point laser, however, it was found that creation from 2D-images could be done at rather lower cost. Even so, the time-consumption of the photo scanning process takes too long and the study of the way to fix this problem is yet to be done seriously. By using the tools that everyone already retained to fix the problem, the author assumes that this is going to be the strong point of this paper. The author uses a video camera instead of taking image shot by shot to capture the residential area. This will create the "continuity" of the images. Continuity of the images will help the software to be able to recognize the images without declaring any control point on images. The methodology can also be applied to the low-resolution images. Finally, the author creates his own error observation method, called "Triangular error comparison", to compare the error from a 3D visual model with 2D images. The error was found to be 1.3082 percent better than the shot by shot image shooting. The study concludes that by using the video camera, the model can be created with lesser effort, and with a better specification of the video camera, the quality of the model should be improved.

Keywords*: *Photoscanning, Structure from Motion (SfM), Agisoft Photoscan, ImageJ, CloudCompare, Computer Vision, three-dimensional visual model

1. Introduction

It is affirmed that surveying has occurred since around 2700 BC. Since then, surveyors had relied solely on chain and rope. Until the late 1950s, the surveyors sought for the surveying methodology that saves time and effort. In the 1950s, Geodimeter introduced Electronic Distance Measurement (EDM) equipment. The EDM uses a multi-frequency phase shift of light waves to find a distance (Mahun, 2014). In 1978, The US Air Force launched the first prototype satellites of the Global Positioning System (GPS). The GPS used a larger constellation of satellites and improved signal transmission to provide more accuracy (National Research Council (U.S.), 2013).

Due to the fact that the present technology is a lot cheaper, it allows people to create more effective products for surveying. One of them that has been pointed out a lot lately is three-dimensional scanning. People found it is efficient to work with because the human brain understands 3D objects better than 2D images. This is why a lot of 3D models have been applied to many fields of study, such as Geological surveying, disaster surveying, medication, etc.

A common approach to 3D scanning is Structure from Motion (SfM). It is a technique for creating a 3D visual model from 2D images (Prince, 2012). Although it has been used by geologists so far, however, for the highly detailed work such as archaeological site observation and urban surveying, it is still a too much time-consuming workflow. Therefore, in this paper, the methodology to decrease such a time-consumption is presented.

2. Objectives

The author's objective is to represent the method to collect data for generating a three-dimensional model without the declaration of any control points.

3. Materials and Methods

At the beginning, four data sets were given.

First Data set: Kumamoto, 2016
Second Data set: Chofu-shi, Tokyo, 2015
Third Data set: Hokkaidou & Kumamoto, 2018
Fourth Data set: Wako- shi, Saitama, 2018

Because creating a three-dimensional model is a time-consuming process, hence, a plan must be carefully made. Overall processes are "Observe all given data", "Conclude all the result and problems", and "Conduct self-experiment bases on the previous conclusion".

3.1 Observe all given data

In "Observe all given data" part, five things were observed,

- Speed of video snapshot (This will be declared only on video)
- Photo shooting interval
- Sky occupancy
- Amount of plugin image in software
- Motion blur effect (however, motion blur effect will be observed in self-experiment part)

All of the above was assumed to be the causes of the quality of the output 3D model. The author used VLC media player to take a snapshot from a video file (Julie, 2017) in the second, third, and fourth data. To generate 3D model, the author used Agisoft photoscan software (Agisoft, 2016). To observe sky occupancy-area on images, ImageJ software was used. However, the observation was done only on the first data set.

First Data set: 2016 Kumamoto

Images were taken from six digital cameras, which installed on a vehicle in such a manner (Figure 1). However, only camera 1 is used in the analysis, because other three data have only a front-view image. Camera Arecont Vision 3110 with LM3NC1M lens was used in this case. Data came from Kumamoto in 2016. The data set is composed of 31 folders, each folder contains six sub-folders of images. Six images were taken once in each station every five meters. Image size is 2048 pixels by 1536 pixels. GPS information was attached to each image file. Figure 2 shows the first data set flowchart.

Firstly, 50 of images (In this data set analysis, only 11 residential routes are chosen, each route has 50 images) will be imported to the software. For the same route, the amount of image will be decreased by ten and then imported to the software, and so on until the amount has reached 20. Secondly, the author observed the percent of sky occupancy-area in each image using ImageJ software. Finally, the percent of sky occupancy will be analyzed together with the resultant model.

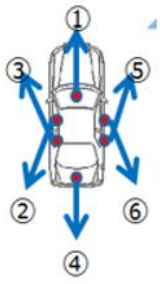

Figure 1 First data set, Cameras installation

Since the result is not in a good stage, the quality of the model will be determined by the appearance of the edges of the building rather than numerical quantity.

First data set result

Table 1 Percent Sky Occupancy

Nishihara Village				
Image Set	Sky occupancy percentage			
	50 Images	40 Images	30 Images	20 Images
A042201	61.4532	61.345	60.6277	61.3816
A042202	62.6121	61.9616	61.6866	62.8164
A042203	58.5562	60.0732	61.6563	62.3832
A042204	61.71	62.2482	62.6133	62.8063

Mishiki Town				
Image Set	Sky occupancy percentage			
	50 Images	40 Images	30 Images	20 Images
A042205	-	-	-	40.9667
A042301			52.871	54.3524
A042302	-	-	42.6041	41.5382
A042303	48.2595	47.5664	46.7344	46.8104
A042304	61.4594	61.3001	59.8411	57.1575
A042501	61.1851	60.0622	60.0767	60.5042

Sky occupancy started to fail the model after it reaches 60 percent (Table 1, shaded cells). This makes sense since the interested object should occupy the most area on images. However, the amount of image does not affect the quality of the model but the performance-time of the software.

Second Data set: 2015 Chofu-shi, Tokyo

Data were collected from one front-view of the vehicle by a digital video camera. The data came from Chofu-shi, Tokyo in 2015 database. Only one video file was received. It is 12:48 minutes long. Frame sizes are 1440 pixels by 1080 pixels with 29.97 frames per second. Since the result in the first data assure that the sky occupancy does affect the quality of the model for the occupancy of more than 60 percent, therefore, the percentage of sky occupancy will be no longer needed to be observed, and because the sky occupancy in the second data could be clearly seen that it is less than 50 percent. Also, since the amount of image does not affect the model, the author will not decrease the amount of image in any data set. The author uses VLC media player to take a snapshot in the film with an interval of 10 to 15 seconds (or 60 to 135 images) to minimize the size of the model. Recording ratio was set to one in VLC media player.

Second data set result

The result is good. As expected in the first data analysis, the percentage of sky occupancy of lesser than 60 percent (more interested object area occupancy) will give a better result. The frame interval was 1 meter per frame. This also proves that the presence of continuity affects the quality of the model. Also, the declaration of the CTP is no needed.

Third Data set: 2018 Hokkaido and Kumamoto

Data were collected by the VIRB Ultra 30 Garmin camera. The data came from Hokkaido and Kumamoto in 2018 database. The data folder of Hokkaido composed of 42 videos, while the data folder of Kumamoto composed of 30 videos. Only front-view data is used in this study.

In this data set, the camera was installed inside the vehicle, therefore, the instrument and objects inside the vehicle are reflected on the windshield (Figure 3).

Figure 3 Image from third data set camera

Hence, the reflection must be taken care first before the image alignment process. Although taking out the uninterested area seems to be a manual operation, however, the mask-shape can be the same for the entire image-set since there are no changes in the instrument position in the vehicle. Therefore, the masking process was conducted only once.

Even though the reflections were taken care, there is still distortion on images. In Figure 4, the reader can see such distortion by the comparison between the pole on street and the drown red line on the image.

Figure 4 Image distortion

Third data set result conclusion

The presence of the reflections and distortion affect the model, although the software has the auto-calibration using the imported images. According to the manual, it is recommended to pre-calibrate the camera before running the analysis. However, because the author did not participate in the data-collecting, the process could not be done.

Forth Data set: Wako-shi, Saitama, 2018

It is 40:58 minutes long video file. Data was collected by Garmin 360 VIRB camera. The video was presented as a two-dimensional image on each. However, the video was then stitched to three-dimensional executable video file, 360-degree video file. Data source came from Wako-shi, Saitama in 2018. Only one video file was received in this case. According to the third data conclusion, the result of the fourth data is expected to be the same as the third data set result due to the lack of calibration.

Forth Data set result conclusion

As expected, the presence of distortion does affect the quality of the model regardless of whether it has continuity or not.

3.2 Conclude all the result and problems

- Declaration of the CTP is no needed.
- The area of interested objects should be more than 60 percent**.**
- Amount of image affects only the software performance-time.
- The lesser the photo shooting interval or snapshot speed, the better the model**.**
- Pre-calibration is required to decrease the possibility of a faulty model.
- As in the manual, the stitched images are not recommended.
- Installation of the cameras should be done outside the vehicle.

3.3 Conduct self-experiment bases on the previous conclusion

Self-experiment will be conducted inside the Chiba university with awareness of the "Conclude all the result and problems" above. Olympus tough tg-5 camera will be used this time. Since the author does not possess any vehicle, therefore this time the experiment will be conducted without a vehicle. However, to mimic the effect of motion to the images, the author uses camera legs as a rotational axis for the camera. The camera will be rotated while taking the video on each station. To determine the effect of motion blur and the improvement of the model from the video camera, the author creates 2 scenarios.

Scenario 1: Compare the model from stationary photo shooting (called set A) with stationary images (Called set C).

Scenario 2: Compare the model from the video camera (Called set B) with stationary images (Called set C). By comparing two scenarios, the effect from motion blur to the model and the improvement of the quality of the model when using video camera can be known. The author uses error to compare two scenarios. The author creates his own method to determine the error from these two scenarios, called "Triangular error comparation". The author uses CloudCompare software to apply this method. The idea came from the fact that the ratio of the right triangle cancels out the units from the sides of the triangle. Why do we need to omit the units? Because the unit in each model was in different unit, e.g. meter and pixel. Also, the relative lengths are in different scale.

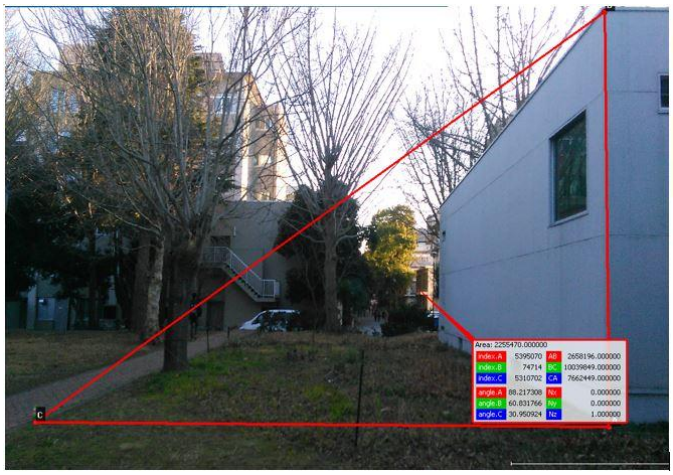

Figure 5 Triangle in CloudCompare

The author assigns the side parallel to the ground as X-side and the one perpendicular as Y-side (3.3.1, Figure 5). Although the triangle could not be created as a perfect right triangle since the side opposes to the right angle is not used, the author assumes that it is the right triangle.

$$\text{Average percent error of scenario 1} = \frac{\sum_{j=1}^{n}\left(\frac{\frac{Side\ X\ of\ setA}{Side\ Y\ of\ setA}-\frac{Side\ X\ of\ setC}{Side\ Y\ of\ setC}}{\frac{Side\ X\ of\ setC}{Side\ Y\ of\ setC}}\right)_j}{n}$$

$$\text{Average percent error of scenario 2} = \frac{\sum_{j=1}^{n}\left(\frac{\frac{Side\ X\ of\ setB}{Side\ Y\ of\ setB}-\frac{Side\ X\ of\ setC}{Side\ Y\ of\ setC}}{\frac{Side\ X\ of\ setC}{Side\ Y\ of\ setC}}\right)_j}{n} \qquad 3.3.1$$

Fifteen positions to draw triangle on are chosen (Figure 6). These positions are assumed to be the best since the edges of the structure can be clearly seen.

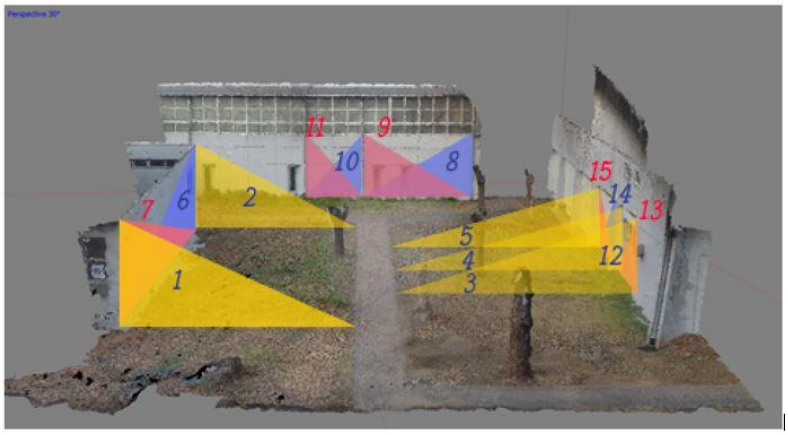

Figure 6 All Triangle position in CloudCompare

4. Results and Discussion

Table 2 Triangular error comparation

	Triangle Number							
Case	1	2	3	4	5	6	7	8
model from Stationary photo shooting	3.3975	4.0614	11.672	12.247	12.177	11.429	16.1	3.4435
model from Video	3.0943	3.9402	12.208	11.69	12.372	10.646	13.751	3.0499
Ordinary image	2.8826	4.5171	10.482	11.371	10.897	9.3286	14.14	3.2476
Scenario 1 %Error	17.864	10.087	11.353	7.7029	11.742	22.51	13.863	6.0337
Scenario 2 %Error	7.3436	12.77	16.47	2.8046	13.537	14.127	2.7482	6.0873

	Triangle Number						
Case	9	10	11	12	13	14	15
model from Stationary photo shooting	3.3487	0.8045	0.8654	1.5514	1.4811	1.5363	1.5433
model from Video	3.5007	0.8099	0.8099	1.636	1.4369	1.5369	1.3852
Ordinary image	3.4246	0.8276	0.8182	1.4796	1.485	1.5378	1.482
Scenario 1 %Error	2.2174	2.7993	5.7694	4.8516	0.2599	0.0936	4.1417
Scenario 2 %Error	2.2214	2.1455	1.0208	10.567	3.2371	0.0542	6.5322

Scenario 1 Average %Error	8.0859
Scenario 2 Average %Error	6.7777

Scenario 1 and 2 gave the average-error of 8.0859 and 6.777, respectively. Therefore, the model that was created by the video camera was improved by 1.3 percent. The motion blur only affects some local shape of the model.

5. Conclusion

- By using the video camera, the declaration of CTP is not needed.
- Model's quality is improved when the video camera is used.
- The area of interested objects should be more than 60 percent**.**
- Amount of image affects only the software performance-time.
- The lesser the photo shooting interval or snapshot speed, the better the model**.**
- Pre-calibration is required to decrease the possibility of a faulty model.
- As in the manual, the stitched images are not recommended.
- Installation of the cameras should be done outside the vehicle.

6. Acknowledgements

We would like to thank to the president of Rangsit University and of Chiba University for made a MOU. This did us a big favor. It gave us an opportunity to do a research between our universities. We also would like to thank any supports until the very end of the research.

7. References

Prince, S. J. D. (2012). *Computer vision: models, learning and inference*. New York: Cambridge University Press.

Agisoft. (2016). *Agisoft PhotoScan User Manual: Professional Edition, Version 1.2*. Retrieved from https://www.agisoft.com/pdf/photoscan-pro_1_2_en.pdf

Mahun, J. (2014). *Electronic Distance Measurement*. Retrieved from http://www.jerrymahun.com/

National Research Council (U.S.). Committee on the Future of the Global Positioning System; National Academy of Public Administration (1995). *The global positioning system: a shared national asset: recommendations for technical improvements and enhancements*. National Academies Press.

Julie. (2017). *How to Take Batch Screenshots or Screencaps in VLC Media Player*. Retrieved from https://turbofuture.com/computers/How-to-take-batch-screenshots-or-screencaps-in-VLC-Media-Player.