

IMPROVING SMARTPHONE POSITION AND ATTITUDE FOR GEOSPATIAL AUGMENTED REALITY APPLICATIONS

Jinwoo Chung¹, Hwiyoung Kim¹, Jangwoo Cheon¹, Impyeong Lee¹

¹ Dept. of Geoinformatics, University of Seoul, Seoul, Republic of Korea – (kandolii21, huddy, khai0614, iplee)@uos.ac.kr

Commission IV, WG IV/9

KEY WORDS: Augmented Reality, Reference Image, Single Photo Resection, Bundle Adjustment, Smartphone, Positioning

ABSTRACT:

Augmented reality is a technology that visualizes and synthesizes virtual objects and information in the real world. Augmented reality can provide useful information to the real world realistically. As the investment and demand for augmented reality increase, many researches have been conducted to visualize objects based on ground coordinate system such as cadastral information and underground facilities using augmented reality. In order to visualize objects based on the ground coordinate system on a smartphone using augmented reality, high accuracy of the position and attitude of the smartphone is essential. Accordingly, this study proposed position and attitude correction method for the first image using reference images and single photo resection. Then, the absolute position of the entire image was estimated using the first calibrated image and the tracking algorithm. When the reference image-based correction method was used, the accuracy was 0.74m, 0.94m. And when the single photo resection correction method was used, the accuracy was 3.13m, 1.24m. The method presented in this study showed higher accuracy when only the sensor was used, but it was confirmed that errors in position and attitude occur as the tracking error accumulates.

1. INTRODUCTION

Augmented reality is a technology that visualizes and synthesizes virtual objects and information in the real world. Augmented reality can be used to provide useful information to the real world more realistically. According to The Goldman Sachs' Virtual & Augmented Reality report, augmented reality technology is currently being used in game, entertainment, and education. In these fields of application, objects with absolute position constructed with ground coordinate systems are not visualized. However, as the usability of augmented reality technology is gradually expanding, research is being conducted to visualize objects having an absolute position using augmented reality. Objects such as underground facilities can be effective in inspection efficiency and stability when visualized using augmented reality technology. Håkansson (2019) conducted a study to visualize the cadastral boundary using augmented reality technology, and Stylianidis (2020) conducted a study to visualize underground facilities using augmented reality technology. Objects, such as cadastral boundaries and underground facilities, have an absolute position constructed with a ground coordinate system, so the location accuracy of objects must be secured for visualization. However, the GPS and geomagnetic sensors of smartphones have insufficient accuracy for position and attitude to visualize objects based on the ground coordinate system. Therefore, in order to visualize objects based on the ground coordinate system using augmented reality technology, it is necessary to accurately determine the location of the smartphone. Smartphones can be easily positioned using built-in GPS. However, a smartphone cannot obtain high location accuracy when determining a location using only GPS, and particularly includes more error factors in cities. Accordingly, many studies have been conducted to determine the location of the smartphone using other sensors of the smartphone. Among them, research on determining the position and attitude of a smartphone through a tracking algorithm using a camera attached to the smartphone is most actively conducted. The positioning method of the smartphone using the tracking algorithm has the advantage that it is not affected by indoor/outdoor, does not require separate

equipment and is independent of electronic devices compared to other positioning methods.

Davison (2003) estimated the location using MonoSLAM. MonoSLAM uses an extended Kalman filter to simultaneously estimate the camera position and the three-dimensional structure of the target area. Newcombe (2011) estimated the location using a direct method called DTAM(Direct Tracking And Mapping). DTAM consists of three steps: 1) Map initialization using stereo images, 2) camera position estimation using composite images generated from reconstructed maps, 3) depth information estimation considering spatial continuity. Newcombe (2011) estimated the location of the camera using Kinetic Fusion method using RGB-D camera. Using an RGB-D camera, you can directly obtain a three-dimensional structure of the environment that contains texture information. In this method, the target region is reconstructed in 3D by combining the depth map obtained in the voxel space, and the position of the camera is estimated by the ICP algorithm using the estimated 3D structure and the input depth map.

As in the above methods, the position estimation based on the camera image can more accurately estimate the relative position between the images by calculating the relative relationship in the continuous image. However, in order to visualize an object based on the ground coordinate system using augmented reality technology, an absolute position in each image is required, not only a relative position. If the exact absolute position and attitude of the first image are obtained, the absolute position of the overall image can be estimated through the relative relationship between the images.

In this study, a study was conducted to accurately estimate the absolute position and attitude of a smartphone in order to visualize objects based on the ground coordinate system through augmented reality. This study determined the absolute position and attitude of the first image using two methods: 1) position and attitude correction based on the reference image, 2) position and attitude correction using Single Photo Resection. Then, the absolute position and attitude of the entire image were determined using the tracking algorithm between camera images.

2. METHODOLOGY

The purpose of this paper is to determine the absolute position and attitude of the smartphone to visualize objects based on the ground coordinate system. Position and attitude between successive images may be relatively determined using a tracking algorithm. In order to change from a relative position and attitude to an absolute position and attitude, it is necessary to precisely determine the absolute position and attitude of the first image. This paper presents an algorithm for determining the position and attitude of a smartphone in two situations, with and without reference image data. The algorithm of this study can be configured as shown in Figure 1.

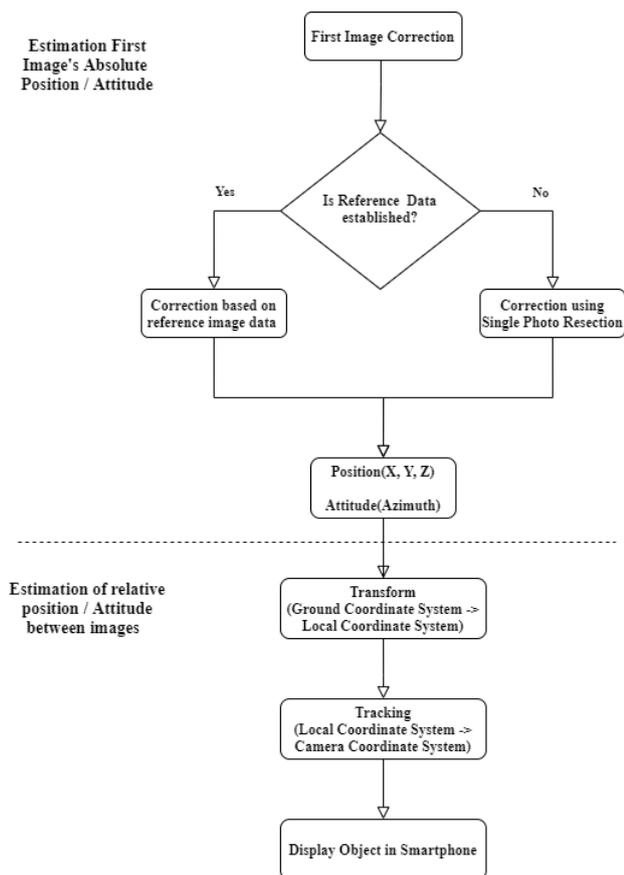


Figure 1. Process for visualizing objects based on ground coordinate system

2.1 Absolute position and attitude estimation for the first image.

In order to visualize objects based on the ground coordinate system on a smartphone, it is essential to define the relationship between the ground coordinate system and the local coordinate system based on the position and attitude of the smartphone. In this process, since the position and attitude of the smartphone has an error according to the smartphone sensor, it is essential to calibrate the position and attitude in order to visualize the object. The position and attitude of the smartphone can be corrected in two ways.

2.1.1 Case in which reference image data is constructed:

The reference image is an image that knows the exact exterior orientation. The reference image may be constructed through precise measurement or may be obtained by accurately estimating exterior orientation. If there is reference image data built around the area where the object is to be visualized, the location and attitude of the newly acquired image can be estimated using the reference image data. The exterior orientation of the new image can be estimated by finding the conjugate points between the reference image and the new image and using the relationship between the conjugate points.

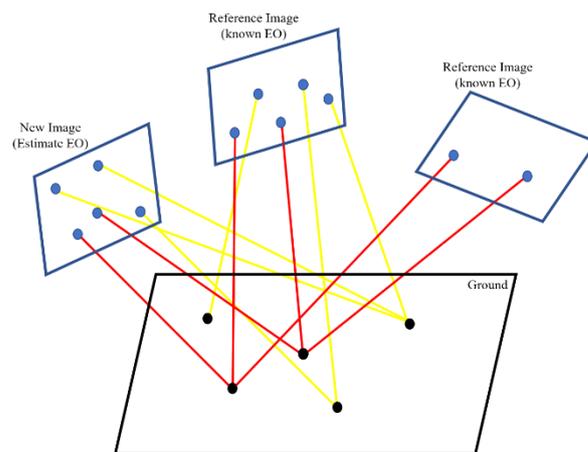


Figure 2. Concept of Estimating the EO of a new image using reference images

2.1.2 Case where reference image data is not constructed:

If there is reference image data, the position and attitude of smartphone can be estimated using a reference image-based correction method. However, it is not easy to construct reference image data for all regions that want to visualize objects. Therefore, a method for accurately estimating the position and attitude of the smartphone is needed even when there is no reference image data. In this study, if there is no reference image data, the position and attitude of the smartphone is estimated using the Single Photo Resection algorithm. Single Photo Resection is a method of correcting the position and attitude of a camera by using a known point in a single image. Basic Single Photo Resection uses the at least three known points on the image based on the collinear equation to determine the values of six elements of exterior orientation(position: X, Y, Z /Orientation: ω, ϕ, κ). However in this research, we modified the equation to estimate the four parameters(position: X, Y, Z /Orientation: azimuth).

2.1.3 Comparison of the two methods:

The correction of the image using the reference image and the correction of the image using Single Photo Resection have their respective advantages and disadvantages in terms of accuracy and usability. Table 1 summarizes the pros and cons of each.

	Strength	Weakness
reference image	More precision. Convenient for users.	Data construction must be preceded. The calculation time is relatively long.
Single Photo Resection	Just need ground control point data. Calculation time is short because only one image is used.	If there are many errors in the initial value, it cannot be estimated well. Inconvenient for users.

Table 1. Pros and cons of both methods

The position and attitude of the image based on the reference image is easy to use because the user does not operate directly on the application, and the estimation of the position and attitude is more accurate than the Single Photo Resection under the assumption that multiple reference images are constructed. However, this has disadvantages in that it is necessary to build reference image data for the region in use, and estimation accuracy may decrease when the seasons change and the environment changes.

The position and attitude correction of the image using Single Photo Resection has the convenience of knowing only two or more known points without having to construct other data, and since it uses only one image, it takes less computation time than using the reference image. However, Single Photo Resection requires an initial value, and if initial value is bad, it may not be easy to estimate. And the user is uncomfortable in selecting known points on the image.

2.2 Relative position and attitude estimation between images

Through the above steps, the absolute position and attitude of the first image can be determined. Accordingly, if the relative relationship between the first image and the remaining image is determined, position and attitude in all images can be determined. All of the images except the first image are determined using the rotation matrix obtained through tracking. The relative relationship between the first image and the rest of the image is obtained through two steps: 1) transform from the ground coordinate system to the local coordinate system, 2) local coordinate system to the camera coordinate system.

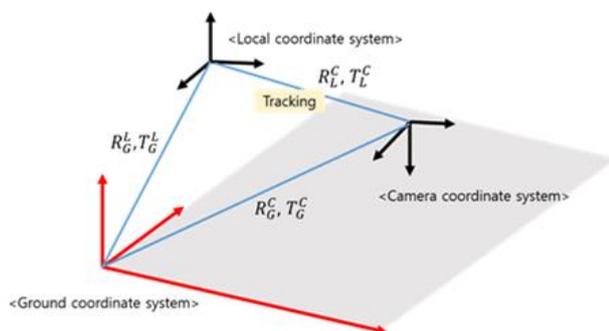


Figure 3. Relationship between coordinate systems

2.2.1 Transform(Ground Coordinate to Local Coordinate):

Since the relative relationship from the first image through tracking is presented based on the local coordinate system of the first image, the camera position and attitude based on the ground coordinate system must be changed to the local coordinate system. Transformation to the local coordinate system is performed using the corrected azimuth angle. Since the Y-axis of the smartphone local coordinate system always indicates the opposite direction of gravity, it is possible to convert from the ground coordinate system to the local coordinate system using the azimuth of the smartphone. At this time, the transformation formula is as in equation (1).

$$R_{GL} = R_z(0^\circ)R_y(\text{azimuth})R_x(90^\circ) \quad (1)$$

Where

R_z = rotation matrix to z axis

R_y = rotation matrix to y axis

R_x = rotation matrix to x axis

R_{GL} = rotation matrix Ground to Local Coordinate

2.2.2 Tracking(Local Coordinate to Camera Coordinate):

The local coordinate system converted through the azimuth can be converted into a camera coordinate system for each image using a tracking direction vector through VIO(visual inertial odometry). VIO is an algorithm that estimates the relative relationship by matching feature points between images through vision sensor. The result can be expressed as a tracking direction vector. In this study, tracking direction vectors were used to convert each image into a camera coordinate system. The coordinate transformation is as shown in equation (2).

$$R_{LC} = [U_x \ U_y \ U_z]^T, U_x = [U_{xx} \ U_{xy} \ U_{xz}]^T \quad (2)$$

Where U_{xx}, U_{xy}, U_{xz} = tracking direction vector

R_{LC} = rotation matrix Local to Camera Coordinate

The absolute position and attitude of all images based on the ground coordinate system can be determined using equations (1) and (2). Through the above process, objects based on the ground coordinate system can be accurately visualized on the smartphone screen.

3. EXPERIMENTAL RESULTS

An experiment was conducted to compare the coordinates of the points calculated with the location of the image determined by the proposed method and the coordinates of the measure points. Since the accuracy of the coordinates of the object is affected by the accuracy of the location of the smartphone, it is possible to evaluate the location of the smartphone images. Because the boundary is clear, the parking area was selected as the test area, and 25 points were measured using GPS-RTK.



Figure 4. Experiment area & distribution of GCP

The experiment was conducted using a Samsung Galaxy Note9 smartphone and interior orientation of the smartphone was accurately estimated. A total of 43 images were acquired using a smartphone, each divided into 32 reference images and 11 test images. Test images were acquired by dividing them into two sets of 5 sheets and 6 sheets at distance of about 10m. The accuracy was tested by calculating the coordinate values by projecting the ground points visible on the smartphone screen to the ground and comparing the calculated values with the measured values. The following four types of smartphone positions and attitudes were used. 1) Position and attitude acquired using a smartphone sensor, 2) Position and attitude accurately estimated through bundle adjustment, 3) Calibrated position and attitude using reference images, 4) Calibrated position and attitude using Single Photo Resection. We verified the research method by comparing the position of the ground point calculated through the four types of smartphone positions and attitudes.

3.1 position / attitude using smartphone sensor information

In order to check the accuracy of the smartphone sensor, the coordinate values of the ground point were calculated based on the position and attitude obtained through the smartphone sensor, and the difference from the actual measured value was compared.

	Mean(m)		
	X	Y	Z
Set 1	10.73144	-2.29421	-0.0036
Set 2	4.363726	-3.29007	0.053842

	RMSE(m)			
	X	Y	Z	Total
Set 1	11.24532	4.637826	0.077072	12.16439
Set 2	6.2571	4.667287	0.082137	7.806511

Table 2. Mean & RMSE using only smartphone sensor

The X, Y position and azimuth direction of the smartphone used values obtained from the sensor. However, depending on the GPS characteristics, the Z(altitude) value contains too many errors, so this was determined using the average altitude value of the experimental area. It can be confirmed that the deviation of the sensor value is severe, and it can be confirmed that the error in the X direction is more severe than in other directions. When checking the RMSE value, it can be confirmed that the error of Set 1 is greater than Set 2. In addition, each RMSE appears to be about 12m and 8m, so it can be judged that it is impossible to visualize objects based on the ground coordinate system using augmented reality using only sensors.

3.2 Position / attitude calibrated using bundle adjustment

In this study, it was judged as the maximum level to secure the position and attitude of the smartphone determined through bundle adjustment. Table 4 shows the results when using the calibrated smartphone position and attitude through the bundle adjustment.

	Mean(m)		
	X	Y	Z
Set 1	0.130926	0.023439	-0.000094
Set 2	0.135702	-0.04115	-0.0028

	RMSE(m)			
	X	Y	Z	Total
Set 1	0.155693	0.089778	0.023444	0.181246
Set 2	0.164407	0.131395	0.022269	0.211637

Table 3. Mean & RMSE using Bundle Adjustment

The accuracy of the ground point was confirmed based on the position and attitude of the smartphone accurately estimated using the bundle adjustment. when using the location of the smartphone calibrated through the bundle adjustment, it was confirmed that there was almost no deviation. In addition, RMSE was calculated as 0.2m for both sets, and it was judged that it is possible to more accurately visualize objects based on the ground coordinate system/

3.3 Position / attitude calibrated through reference image

The accuracy of the ground point was confirmed based on the position and attitude of the smartphone adjusted through the reference image. 32 reference images were used, and exterior orientation was accurately estimated through bundle adjustment.

	Mean(m)		
	X	Y	Z
Set 1	0.163423	0.441353	-0.0007
Set 2	0.343693	0.251372	-0.00827

	RMSE(m)			
	X	Y	Z	Total
Set 1	0.242426	0.701226	0.02511	0.742374
Set 2	0.497948	0.797117	0.026767	0.940247

Table 4. Mean & RMSE using reference image

RMSE and average values were much smaller when using the position and attitude of the smartphone calibrated using the reference image than when using the sensor values. However, it showed a relatively large error than when using the position and attitude of the smartphone calibrated through the bundle adjustment.

3.4 Position / attitude calibrated using Single Photo Resection

The accuracy of the ground point was confirmed based on the position and attitude of the smartphone adjusted using Single Photo Resection. When performing Single Photo Resection, previously acquired ground control points were used.

	Mean(m)		
	X	Y	Z
Set 1	2.866719	0.150848	-0.0041
Set 2	-0.22548	0.688522	-0.01589

	RMSE(m)			Total
	X	Y	Z	
Set 1	3.034681	0.765788	0.020401	3.129878
Set 2	0.381645	1.175065	0.033544	1.235943

Table 5. Mean & RMSE using single Photo Resection

When the position and attitude of the smartphone calibrated using Single Photo Resection was used, it showed a much lower error than when only the smartphone sensor was used. However, it showed a greater degree of error than the bundle adjustment and reference image-based methods that use multiple images together. In addition, different levels of error can be checked according to the set, unlike the method based on the bundle adjustment and the reference image. Since the Single Photo Resection method uses the initial sensor value, the level of adjustment is determined according to the sensor value. Even when the sensor value was used, Set 1 shows a larger error than Set 2, and Single Photo Resection was also affected by this, so it was confirmed that Set 1 had a large error.

3.5 Tracking error

When using the smartphone position and attitude using the reference image and Single Photo Resection correction, it showed a higher error than the bundle adjustment. Both the reference image and the Single Photo Resection correction method calibrate the position and attitude of the first image in each set and determine the position and attitude for the entire image through tracking. Therefore, the standard of correction becomes the first image, so when the tracking error exists, the error is inevitably accumulated as the image progresses. Table shows the results.

Number of Set	Number of Image	RMSE
Set 1	1	0.199989
	2	0.948781
	3	0.926347
	4	0.879754
	5	0.496759
Set 2	1	0.172641
	2	0.380487
	3	0.710102
	4	0.893703
	5	1.649872
	6	1.282918

Table 6. Tracking Error

The first image in each set is the result of correction based on the reference image. The first image had an error similar to the result of the bundle adjustment. However, the remaining images accumulate errors due to tracking errors. Therefore, in order to estimate the position and attitude of the smartphone more accurately with respect to the entire image, correction for the tracking error is essential.

To visually confirm the result, the parking lot boundary line was visualized on the image. The OpenCV library was used to visualize the parking lot boundary on the image. The visualized screen is shown in Figure 5 and Figure 6. Figure 5 visualizes the parking lot area on the first image that was directly calibrated, and Figure 6 visualizes the parking lot area on the last image estimated through tracking. As shown in Figure 5, the parking lot area of the result of the correction based on the reference image and the result of the correction using Single Photo Resection was visualized similar to the bundle adjustment. However, Figure 6 was visualized to deviate from the parking area when compared to the bundle adjustment.

4. CONCLUSIONS

In this study, the accurate position and attitude of the smartphone was estimated to visualize objects based on the ground coordinate system using augmented reality technology. In this study, to determine the exact position and attitude of a smartphone, smartphone position and attitude correction methods using reference image and Single Photo Resection were presented. In order to verify the proposed method, an experiment was conducted to compare the calculated ground point coordinates with actual coordinates. As a result of the experiment, the accuracy was 0.74m, 0.94m when the reference image-based method was used, and the accuracy was 3.13m and 1.24m when the Single Photo Resection correction method was used. Both methods showed higher accuracy than using only the sensor, but lower accuracy than the bundle adjustment. In addition, the Single Photo Resection using only one image showed lower accuracy than the reference image-based correction method using multiple images to correct.

As a result of analyzing the cause of the two methods that showed lower accuracy than the bundle adjustment, the accumulation of errors due to the tracking error was confirmed. Since the two methods presented are corrected for the first image of each set, the error may increase as the tracking error accumulates as the image progresses. Therefore, in order to determine a more accurate position and attitude for the entire image, a technique for correcting the tracking error is required. If such a tracking error can be corrected, objects based on the ground coordinate system can be sufficiently visualized using augmented reality technology.

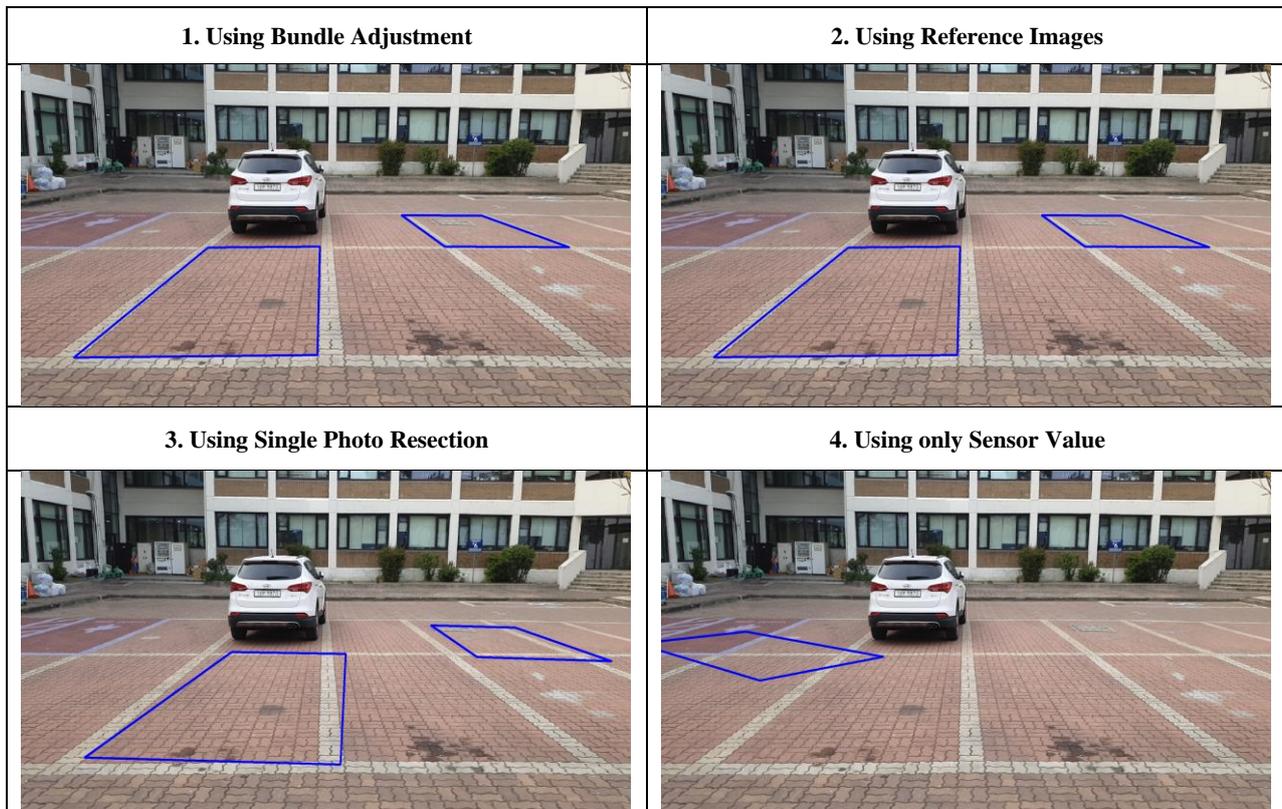


Figure 5. Visualization of the first image

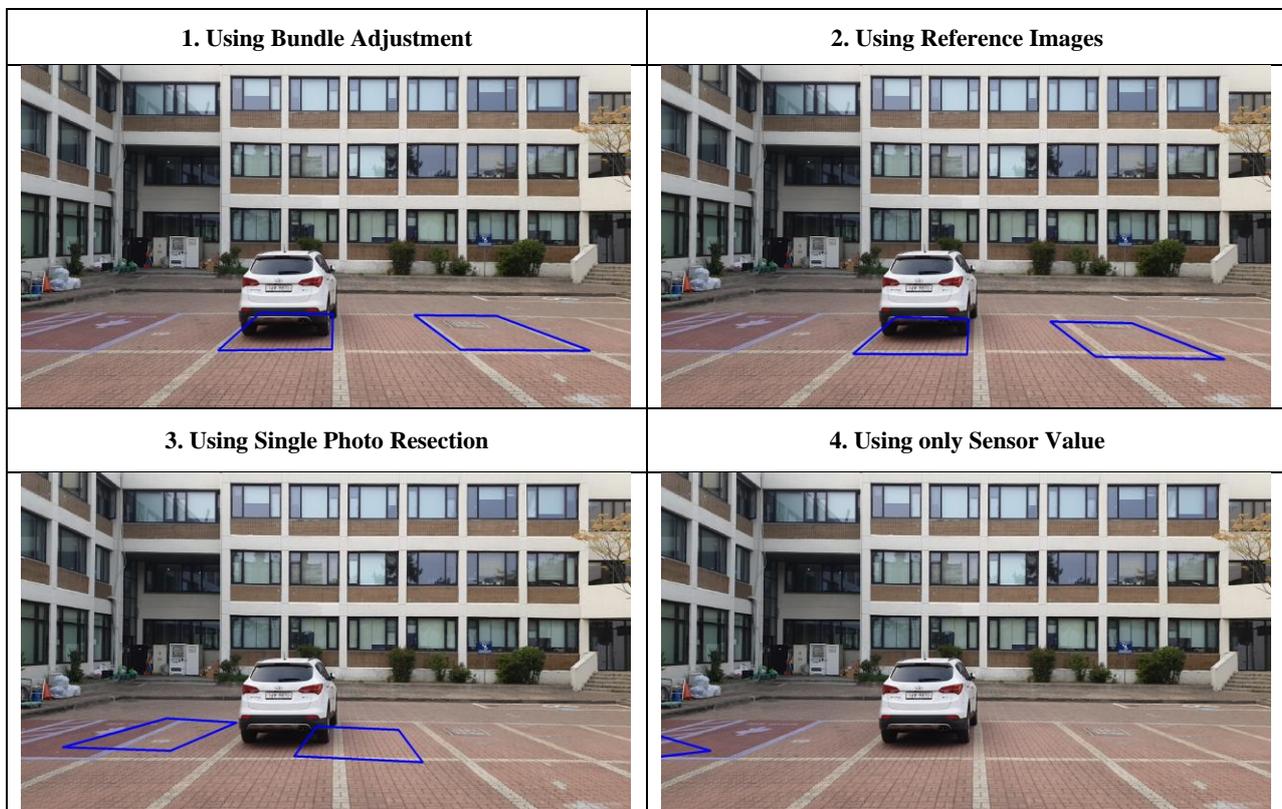


Figure 6. Visualization of the last image

ACKNOWLEDGEMENTS

This research was supported by the Institute of Spatial Information Research & Korea Land Information Corporation (A-GNSS location accuracy and augmented reality based cadastral information display research through development of image-based geocoding technology).

REFERENCES

International Data Corporation. 2019. Worldwide Augmented and Virtual Reality Spending Guide. <https://www.idc.com/>

Goldman Sachs. 2016. Virtual & Augmented Reality(Understanding the race for the next computing platform). <https://www.goldmansachs.com/>

Håkansson, L. 2019. Visualizing cadastral parcels for surveyors using handheld Augmented Reality.

Stylianidis, E., Valari, E., Pagani, A., Carrillo, I., Kounoudes, A., Michail, K., & Smagas, K. 2020. Augmented Reality Geovisualisation for Underground Utilities. *PFG-JOURNAL OF PHOTOGRAMMETRY REMOTE SENSING AND GEOINFORMATION SCIENCE*.

Kim, K., Billingham, M., Bruder, G., Duh, H. B. L., & Welch, G. F. 2018. Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008–2017). *IEEE transactions on visualization and computer graphics*, 24(11), 2947-2962.

Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. 2007. MonoSLAM: Real-time single camera SLAM. *IEEE transactions on pattern analysis and machine intelligence*, 29(6), 1052-1067.

Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., ... & Fitzgibbon, A. 2011, October. KinectFusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality* (pp. 127-136). IEEE.

Knauth, S., & Koukofikis, A. 2016. October. Smartphone positioning in large environments by sensor data fusion, particle filter and FCWC. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (pp. 1-5). IEEE.

Dongdong, Y. F. Z. 2012. WiFi location based on Android platform [J]. *Electronic Measurement Technology*, 9.

Kim, H. 2018. Generation of individual ortho-images based on reference images for facility inspection. (Masters dissertation). University Of Seoul, Seoul, South Korea.

Kim, H. 2019. Multi-sensor fusion for car localization using reference images. (Ph.D. dissertation). University Of Seoul, Seoul, South Korea.