Scientific
Research
Publishing

Perspective Projection Algorithm Enabling Mobile Device's Indoor Positioning

Seo Woo Han, Yun Jung Lee, Ji Hyeok Yun, Chang Yong Han, Dae Ho Lee, Doug Young Suh

Department of Electronics Engineering, Kyung Hee University, Yongin-si, Republic of Korea
Email: swhan0329@gmail.com, yunjung.erica@gmail.com, ychwh555@khu.ac.kr,
chang.yong.cyhan@gmail.com, nize@khu.ac.kr, suh@khu.ac.kr

Abstract

In order to improve the user's satisfaction with the augmented reality (AR) technology and the accuracy of the service, it is important to obtain the exact position of the user. Frequently used techniques for finding outdoors locations is the global positioning system (GPS), which is less accurate indoors. Therefore, an indoor position is measured by comparing the reception level about access point (AP) signal of wireless fidelity (Wi-Fi) or using bluetooth low energy (BLE) tags. However, Wi-Fi and Bluetooth require additional hardware installation. In this paper, the proposed method of estimating the user's position uses an indoor image and indoor coordinate map without additional hardware installation. The indoor image has several feature points extracted from fixed objects. By matching the feature points with the feature points of the user image, we can obtain the position of the user on the Indoor map by obtaining six or more pixel coordinates from the user image and solving the solution using the perspective projection formula. The experimental results show that the user position can be obtained more accurately in the indoor environment by using only the software without additional hardware installation.

Keywords

Camera Pose Estimation, Indoor Positioning, Perspective Projection, Homography

1. Introduction

Recently AR has made many technological advances, and the range that users can enjoy on smartphones and tablets has increased. AR is a technology that superimposes a virtual object in the real world seen by the user, and is used for

many purposes such as automobile navigation, education, tourism, advertisement, shopping, and games field [1]. AR is a location-based service that can provide a high-quality AR service only by knowing precisely the location of the user. GPS, which is most commonly used for locating a user, has a disadvantage in that an error in the range of 10 meters is generated outdoors [2]. Therefore, many researches on indoor position measurement have been made in various ways [3].

The current research trends of algorithms for obtaining the indoor position using a smartphone are as follows. Kalbandhe [4] suggested a method of detecting indoor BLE tags by scanning the surrounding area by turning on the bluetooth function of the smartphone, and measuring the received signal strength and transmission power from the BLE tags to find indoor location and Soewito [5] proposed a method to increase the accuracy of the user's position in the room by using the Pythagoras theory by receiving the information of 4 APs using the received signal strength indicator (RSSI) of the smartphone.

In calculating the position of the user in the room by the BLE tags method, the accuracy was calculated by dividing into three based on the BLE distance. Immediate at 0 - 0.5 m, Near at 0.5 - 3 m, Far at 3 m or more. The error of estimation of the indoor position is within 4 m maximum and the range of meter is different in terms of Immediate, Near, and Far. In addition, the error of the indoor location estimation of the method of obtaining the user position by the four pieces of AP information is 0.6 m on the average.

In this paper, we propose a method to extract indoor images of a fixed image of an indoor map, an indoor map, and a pixel image of a vertex of the object. We propose an algorithm to find the user's position. Chapter 2 explains perspective projection and homography for easy understanding of the concepts and equations used in Chapter 3. Chapter 3 describes how to obtain the pixel coordinates corresponding to the fixed object vertices in the user image and explains the formula for obtaining the user's position using the perspective projection algorithm. Chapter 4 verifies the proposed method through experiments and concludes in Chapter 5.

2. Background

This chapter will explain the perspective projection theory and homography, which are the main mathematical principle of the system that will be proposed in the next Chapter 3.

2.1. Perspective Projection

Projection is a flat representation of a three-dimensional (3D) object on one plane called the view plane. Let l be a line in the plane, and let V be a point not on the line. The perspective projection from V onto l is the transformation which maps any point P, distinct from V, onto the point P' which is the intersection of the lines $\overline{VP}$ and l, as illustrated in **Figure 1** [6]. Since the appearance
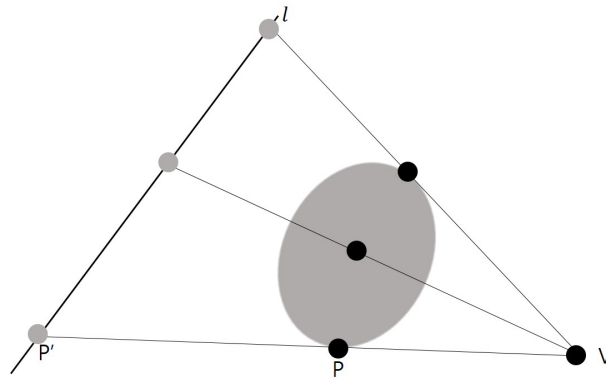
Figure 1. Concept of the perspective projection.

of the object varies depending on where you look at the object, it is possible to obtain the position of the user by using the perspective projection principle and formula.

2.2. Homography

When a plane is projected on another plane, a certain transformation relation is established between projected corresponding points, and this transformation relation is called homography. Homography is represented by a 3 × 3 matrix and is a transformation relation that establishes correspondence points on a homogeneous coordinate representation. That is, if points $(x_1, y_1), (x_2, y_2), \cdots, (x_i, y_i)$ on one plane are projected onto points $(x'_1, y'_1), (x'_2, y'_2), \cdots, (x'_i, y'_i)$ on different planes, there is always a 3 × 3 homography matrix H that satisfies Equation (1) between these corresponding points. In Equation (1), "s" is a scale factor [7]. When using the concept of homography in the image, it is assumed that the model of pinhole camera is used, and it is actively used in applications such as image rectification, image registration, and computation of camera motion.

$$s \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, i = 1, 2, 3, \ldots \qquad (1)$$

3. Proposal

3.1. Overview of the Proposed System

The database of the proposed system has an indoor map where the user wants to know the location. Also, among the fixed objects in the indoor map, the reference image taken at a position where the object is not distorted and the entire object can be contained in the image, for objects whose edges are sharp or contrasted with surrounding objects. The pixel coordinates information of the vertex is stored in the database. The coordinates of the indoor map are the 3D coordinates of the world coordinate of the object (frame, desk, door, window, computer, etc.) located in the indoor place. When a user takes an image, the reference image in the database and the image taken by the user are extracted, and

then the 2D pixel coordinate of the feature point and the 3D coordinate of the world coordinate of the feature point are used as input data. Input data is plugged in the perspective projection equation $q = Mp$ and the equation is rearranged to obtain the *M* vector by the least square method. Then, using the idea that the perspective projection is not the only user's position, the vector V is obtained to obtain the position of the user. The overall contents are shown in **Figure 2**.

3.2. The Method to Obtain Pixel Coordinates Using Homography

In this section, "obtain pixel coordinate at a user image using homography" in **Figure 2** will be described in detail.

Figure 3 and **Figure 4** are the plane and cross-sectional views of the Ingo Gallery at Kyunghee University, respectively, and the unit of each coordinate is meters. The symmetric points of the pixel coordinates of the fixed planar objects in the reference image can be found in the user image using the Speed Up Robust Features (SURF) algorithm [8]. The method consists of three steps as shown in **Figure 5**.

In the user image, the steps of obtaining the pixel coordinates of the object vertices are performed separately for each object. In the first step, the SURF algorithm is used to check whether an object composed of feature points existing on one plane in the reference image has the same object and corresponding feature points in the user image. At this time, candidates of corresponding minutiae points are varied, and only four points are extracted through the random sample consensus (RANSAC) algorithm [9]. Second, the homography matrix is obtained
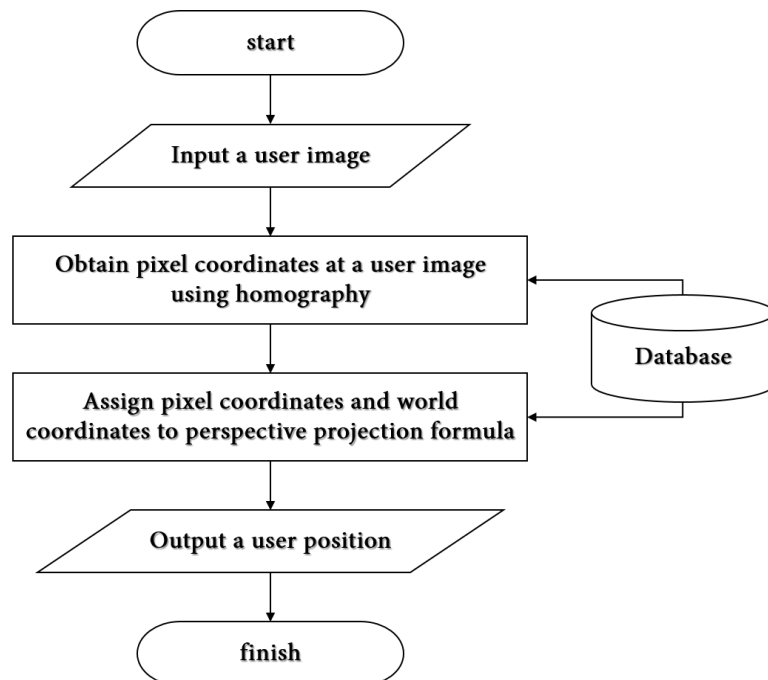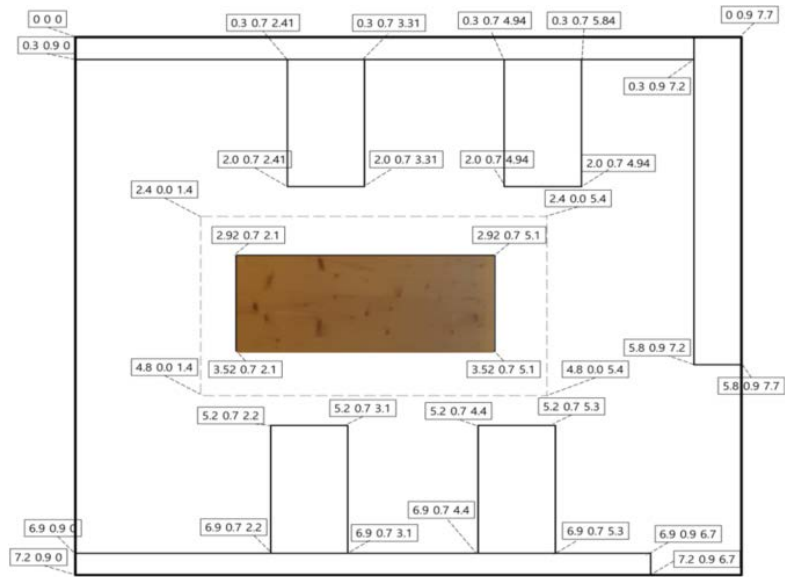

Figure 2. Flowchart of the proposed system.

Figure 3. Floor plan of innogallary in kyung hee university.
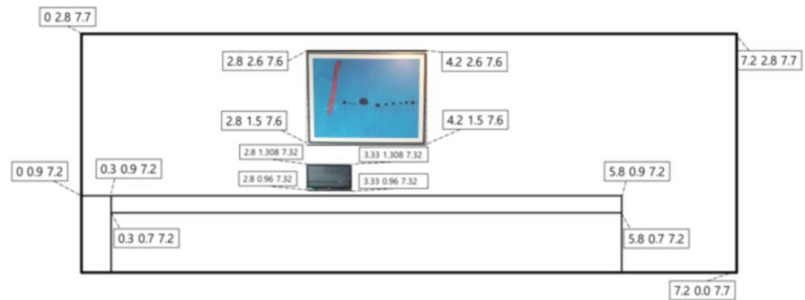

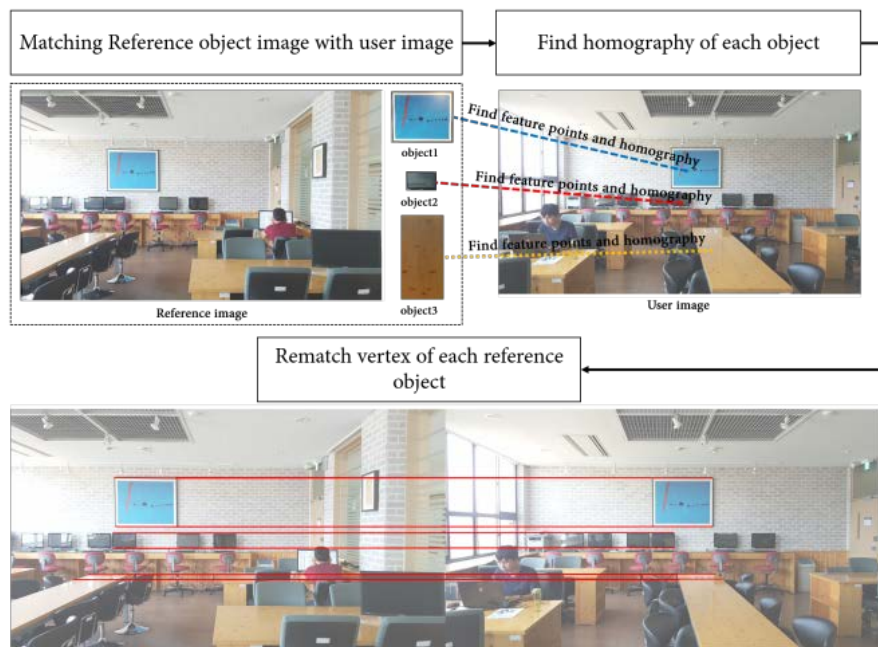**Figure 4.** Cross-section of innogallary in kyung hee university.

Figure 5. The steps for getting pixel coordinates.

by using the feature points used in the reference image and the four feature points finally selected in the user image. Finally, we obtain the vertex pixel coordinates of each object in the user image by matching the vertices of the objects in the reference image to the user image using the previously obtained homography matrix.

3.3. Derivation of the Equations for Perspective Projection

Let $p = (X_k, Y_k, Z_k, 1)^T$ be the homogeneous coordinates of a 3D point in the indoor map with $1 \le k \le n$, where n is the number of points. Let $q = (u_k, v_k, 1)^T$ be the homogeneous coordinates of the point p projected on the image. According to the theory of perspective projection [10], we may easily find a 3×4 matrix satisfying the following equation,

$$p = \begin{bmatrix} X_k \\ Y_k \\ Z_k \\ 1 \end{bmatrix}, q = \begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix}. \tag{2}$$

The equality should be understood in the sense of homogeneous coordinates. They are equal if one is a constant multiple of the other. Therefore, we have

$$u_k = \frac{m_{11}X_k + m_{12}Y_k + m_{13}Z_k + m_{14}}{m_{31}X_k + m_{32}Y_k + m_{33}Z_k + m_{34}}$$

and

$$v_k = \frac{m_{21}X_k + m_{22}Y_k + m_{23}Z_k + m_{24}}{m_{31}X_k + m_{32}Y_k + m_{33}Z_k + m_{34}}.$$

It follows that

$$u_k m_{31}X_k + u_k m_{32}Y_k + u_k m_{33}Z_k + u_k m_{34} = m_{11}X_k + m_{12}Y_k + m_{13}Z_k + m_{14} \tag{3}$$

and

$$v_k m_{31}X_k + v_k m_{32}Y_k + v_k m_{33}Z_k + v_k m_{34} = m_{21}X_k + m_{22}Y_k + m_{23}Z_k + m_{24} \tag{4}$$

These may be written as a single matrix equation of the form

$$\begin{bmatrix} X_k & Y_k & Z_k & 1 & 0 & 0 & 0 & 0 & -u_k X_k & -u_k Y_k & -u_k Z_k & -u_k \\ 0 & 0 & 0 & 0 & X_k & Y_k & Z_k & 1 & -v_k X_k & -v_k Y_k & -v_k Z_k & -v_k \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{5}$$

Equation (5) can be repeated for n point and we obtain

$$
\begin{bmatrix}
X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\
0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\
 & \vdots & & & & \vdots & & & & \vdots & & \\
X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_1 & -u_nX_1 & -u_nX_1 & -u_n \\
0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_1 & -v_nX_1 & -v_nX_1 & -v_n
\end{bmatrix}
\begin{bmatrix}
m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34}
\end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \end{bmatrix}
\quad (6)
$$

Let

$$
R =
\begin{bmatrix}
X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\
0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\
 & \vdots & & & & \vdots & & & & \vdots & & \\
X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_1 & -u_nX_1 & -u_nX_1 & -u_n \\
0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_1 & -v_nX_1 & -v_nX_1 & -v_n
\end{bmatrix}
$$

and

$$
m = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} & m_{21} & m_{22} & m_{23} & m_{24} & m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}.
$$

If the measurement is all exact, Equation (6) has a nonzero solution for any number of data points, but this is not be the case in the real-world environment where the measurement of image and world coordinates is inexact (generally termed *noise*)—there will not be an exact solution to the overdetermined system $Rm = 0$ apart from the zero solution. Instead of demanding an exact solution, we attempts to find an approximate solution, namely a vector m that minimizes a suitable cost function. To avoid the solution $m = 0$, we require an additional constraint. Generally, a condition on the norm is used, such as $\|\vec{m}\| = 1$. Given that there is no exact solution to $Rm = 0$, it seems natural to attempt to minimize the norm $\|R\vec{m}\|$ instead, subject to the usual constraint, $\|\vec{m}\| = 1$. The solution is the (unit) eigenvector of $R^T R$ with minimum eigenvalue. To obtain the minimum value of $m^T A m$ where $A = R^T R$, the minimum eigenvalue among the eigenvalues of A must be obtained and the eigenvector corresponding to the minimum eigenvalue must be found.

Assume that n is the number of data used in matrix R where $n \in N$ and $n \geq 6$. Then matrix A is a 12×12 matrix. We may solve the 12th-order real-valued coefficient equation $\det(A - \lambda I) = 0$. The matrix A is a symmetric matrix with $m^T A m \geq 0$ and $\lambda \geq 0$. Then we can obtain the eigenvalues of A which are $\lambda_1, \lambda_2, \cdots, \lambda_{12}$ such that $\lambda_1 > \lambda_2 > \cdots > \lambda_{12}$.

The m value obtained in the above procedure is rearranged in the form of

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}.$$

Let $V = \begin{bmatrix} x & y & z & w \end{bmatrix}^T$ be a homogeneous coordinate of the position of the camera and it should satisfies $MV = 0$ since it is the only point that does not correspond to the plane of the image by the perspective projection transform M. Therefore, V is a nonzero vector in the null space of M. The matrix V divided into w, we finally obtain the user's position in 3D world coordinate as $V = (x / w, y / w, z / w)$.

4. Experiments

In this chapter, the following method was used to verify the proposed method at Inno Gallery, Kyunghee University. When the user takes an indoor image, obtain the pixel coordinate in the user image using the method described in section 3.2, and assign the 6 pairs of world coordinates corresponding to the vertices of the object captured as the feature points in the user image to the Equation (6) of chapter 3 to obtain the user's position. In order to analyze the error due to the increase in the number of data, the experiment is repeated in the same order as in the case of eight pairs of points.

In order to compare the performance of the proposed method, we will compare it with the result of user location using homography. The method of using homography should know the camera position of the reference image in advance. Using the SURF algorithm, we find feature points that are invariant to environmental changes such as scale, illumination, and viewpoint from two images from the reference and user images, and match reference and user images and obtain homography between the two images. If you de-compose the homography, you can see how much the user's image has been rotated and translated from the reference image. The results can be used to determine where the user took the image.

4.1. Experimental Conditions

Since the perspective projection was designed for the pinhole camera, the camera parameters were obtained with the image resolution of 3264×1836 using the camera calibration tool to eliminate the distortion of the camera. The results are shown in **Table 1**.

World coordinates are calculated in meter units, and the unit of pixel coordinates is pixel. The image taken by the user is shown in **Figure 9**. This image was taken with a camera corresponding to **Table 1**.

Table 1. Camera parameters used in the experiment.

Device	Camera intrinsic parameter			
	Focal length x	Focal length y	Center of the image x	Center of the image y
SM-N920S	2518	2513	1632	918

Figure 6 and **Figure 7** are same as **Figure 8**, and the images indicate the case where the feature points are held at 6 or 8, respectively. The yellow dot in the image below is the feature point.

Table 2 and **Table 3** show the 2D pixel coordinates of the feature points obtained from the user image. The 3D world coordinate of the object matching the feature point used as the pixel coordinates was taken from the data base and used as the input data.
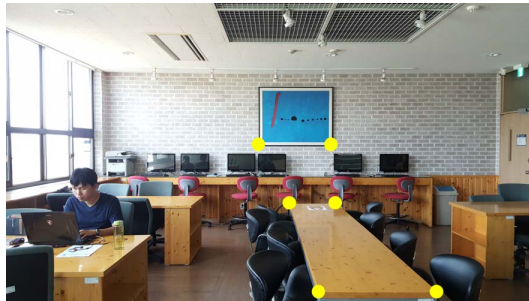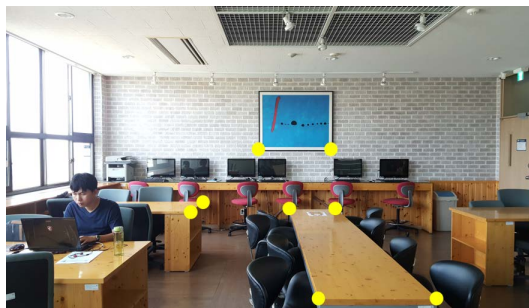

Figure 6. 6 features of the image.

Figure 7. 8 features of the image.

Figure 8. A user image of innogallary.

Table 2. The input data used in **Figure 6**.

Points	1			2			3			4			5			6		
3D world coordinate	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
	2.8	1.5	7.6	4.2	1.5	7.6	2.92	0.7	5.1	3.52	0.7	5.1	2.92	0.7	2.1	3.52	0.7	2.1
2D pixel coordinate	u	v	u	v	u	v	u	v	u	v	u	v						
	1555	880	2007	881	1743	1236	2030	1235	1924	1781	2652	1780						

Table 3. The input data used in **Figure 7**.

Points	1			2			3			4			5			6			7			8		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
3D world coordinate	2.0	0.7	4.94	2.0	0.7	4.94	2.8	1.5	7.6	4.2	1.5	7.6	2.92	0.7	5.1	3.52	0.7	5.1	2.92	0.7	2.1	3.52	0.7	2.1
	u	v		u	v		u	v		u	v		u	v		u	v		u	v		u	v	
2D pixel coordinate	1138	1247		1214	1189		1555	880		2007	881		1743	1236		2030	1235		1924	1781		2652	1780	

4.2. Experiment Result

In order to compare the accuracy of the proposed experimental results, we extracted the homography matrix of the reference and user images and found the position of the user. The result of extracting feature points using SURF algorithm in both images is shown in **Figure 9**. The position of the user obtained using this method is listed in **Table 4**. **Figure 10** is an image that warps the reference image, the user image and decomposes the homography estimated from the two images to apply the rotation, translation values obtained from the decomposition and make the feature points of the reference image and the user image the same. In **Figure 10**, it can be confirmed that warping is performed based on the blue frame.

If the error is calculated by the general error rate calculation method, as the position of the camera is located closer to the origin of the world coordinate, it can be confirmed that even if the difference between the calculated user position and the actual user position is small, the error rate is large. Therefore, in this paper, the error rate as Equation (7) is newly defined to calculate a reasonable error.

$$\text{error rate} = \frac{\text{user location} - \text{calculated user location}}{\text{Diagonal length of indoor space}} \times 100 \qquad (7)$$

Since the Inno gallery of Kyung hee University has a width of 7.7 m, a length of 7.2 m and a height of 2.8 m, the value to be deducted from the denominator of the error rate is $\text{diagonal length of innogallary} = \sqrt[3]{7.7^2 + 7.2^2 + 2.8^2} = 4.9182 \text{m}$.

Table 4 summarizes the proposed method and the results obtained by estimating the homography of the two images and the error rate.

5. Conclusions

The method proposed in this paper can find the user's position with high accuracy by knowing more than 6 pairs of 3D coordinates of Indoor map and 2D image pixel coordinates. The proposed method provides users with a wide range of AR services because it can find the position of the user with high accuracy without the cost of constructing additional infrastructure than the method of using the hardware infrastructure studied by many papers to obtain the indoor user position. It is expected to give high satisfaction. In the proposed method, it is shown that the error rate is 4.705% p lower than the x, y and z coordinates when the user position is obtained by using 6 points. In addition, the user's posi-
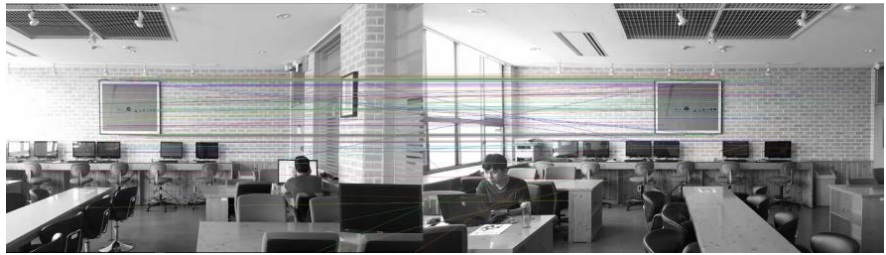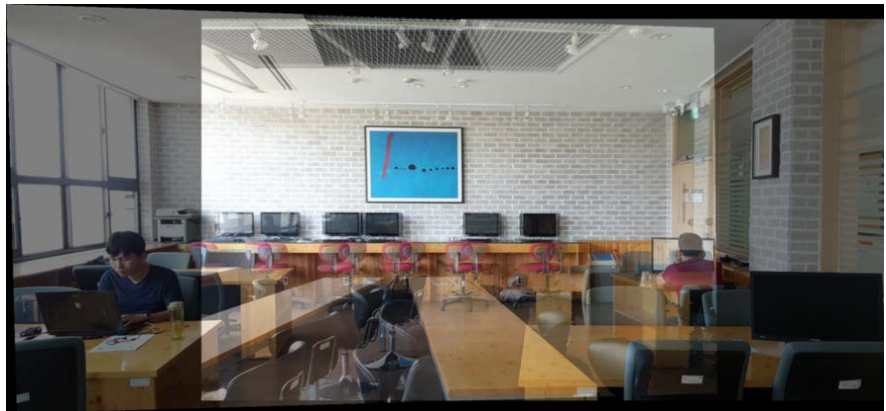
Figure 9. A feature-matching image.

Figure 10. A warped image between reference image and user image.

Table 4. At **Figure 6** and **Figure 7**, user location calculation and homography user location and error rate.

	6 points			8 points		
	X	Y	Z	X	Y	Z
Real View point	3.02	1.47	0.1	3.02	1.47	0.1
Proposed method	3.0163	1.4939	0.2735	3.0314	1.2935	0.4946
Proposed method error rate	0.075%	0.56%	3.52%	0.23%	3.58%	8.02%
Opencv Decompose	3.776	1.369	0.031	3.776	1.369	0.031
Opencv Decompose error rate	15.37%	1.5%	1.4%	15.37%	1.5%	1.4%

tion should be more accurate as the number of input data is larger when there is no error in the input data value when the user position is obtained by using 8 points. In this experiment, since there is an error in the input data, it has grown. However, its average is 2.14% p lower than when calculating the position of the user by estimating the homography.

In the future, the application of the AR technology to the smart phone will increase, and if the technology provided by the application needs the accurate location information of the user, the proposed method is expected to be needed. In this paper, further research is needed on how to reduce errors that occur when input data is not accurate.

Acknowledgements

through the Ministry of science, ICT and Future Planning/Gyenggi Province (20171218).

References

[1] Grillo, G., Lamberti, F., Manuri, F., Sanna, A., Paravati, G., Pezzolla, P. and Montuschi, P. (2014) Challenges, Opportunities, and Future Trends of Emerging Techniques for Augmented Reality-Based Maintenance. *Transactions on Emerging Topics in Computing*, **2**, 411-421. http://ieeexplore.ieee.org/document/7004838/

[2] Mahiddin, N.A., Madi, E.N., Dhalila, S., Hasan, E.F., Safie, S. and Safie, N. (2013) User Position Detection In an Indoor Environment. *International Journal of Multimedia and Ubiquitous Engineering*, **8**, 303-312.
https://doi.org/10.14257/ijmue.2013.8.5.30

[3] Dinh-Van, N., Nashashibi, F., Thanh-Huong, N. and Castelli, E. (2017) Indoor Intelligent Vehicle localization Using WiFi Received Signal Strength Indicator. 2017 *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility* (*ICMIM*), Nagoya, 33-36. https://doi.org/10.1109/ICMIM.2017.7918849

[4] Kalbandhe, A.A. and Patil, S.C. (2016) Indoor Positioning System Using Bluetooth Low Energy. 2016 *International Conference on Computing, Analytics and Security Trends* (*CAST*), Pune, 451-455. https://doi.org/10.1109/CAST.2016.7915011

[5] Soewito, B., Faahakhododo, I. and Gunawan, F.E. (2016) Increasing the Accuracy of Distance Measurement between Access Point and Smartphone. 2016 11*th International Conference on Knowledge*, *Information and Creativity Support Systems* (*KICSS*), Yogyakarta, 1-6. https://doi.org/10.1109/KICSS.2016.7951423

[6] Marsh, D. (2005) Applied Geometry for Computer Graphics and CAD. 2nd Edition.

[7] Jeon, S.J., Kim, J., Ji, M.G., Park, J.H. and Choi, Y.W. (2017) Position Error Correction Using Homography in Discretized Positioning Circuit for Gamma-Ray Imaging Detection System. *IEEE Transactions on Nuclear Science*, **64**, 816-819.
https://doi.org/10.1109/TNS.2016.2639532

[8] Bay, H., Tuytelaars, T. and Van Gool, L. (2006) Surf: Speeded up Robust Features. *Computer Vision-ECCV* 2006, 404-417. https://doi.org/10.1007/11744023_32

[9] Fischler, M.A. and Bolles, R.C. (1981) Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, **24**, 381-395. https://doi.org/10.1145/358669.358692

[10] Hartley, R. and Zisserman, A. (2000) Multiple View Geometry in Computer Vision. 2nd Edition, Cambridge University Press, Cambridge.