Camera Position Estimation with a Limited Number of Known Landmarks

Cheolwook Kim[1], Seonu Yi[2], Wansang Yoon[3], Sooahm Rhee[4]*.

[1] *Image Eng. Research Center, 3DLabs Co. Ltd., Seoul, South Korea. ileen3@3dlabs.co.kr*
[2] *Image Eng. Research Center, 3DLabs Co. Ltd., Seoul, South Korea. seon0620@3dlabs.co.kr*
[3] *Image Eng. Research Center, 3DLabs Co. Ltd., Seoul, South Korea. wsyoon@3dlabs.co.kr*
[4*] *Image Eng. Research Center, 3DLabs Co. Ltd., Seoul, South Korea. ahmkun@3dlabs.co.kr*

Abstract: In this study, we introduce a method for estimating the position of a mobile device using images acquired from its onboard camera. The experiment was conducted under a constraint in which the camera position is estimated using only a single known reference point, and the pre-mapped indoor spatial data. Since only the planar coordinates (X, Y) of the device are required, the formulation is simplified by estimating the viewing direction toward the landmark and the relative distance to it. The experimental results show that the estimated camera positions closely approximate the ground-truth values. These results suggest that the proposed approach can be effectively utilized in constrained scenarios, such as those involving mobile platforms and indoor environments.

Keywords: Indoor positioning; Camera position; Single reference point; Mobile device;

1. Introduction

In crowded and complex indoor environments such as shopping malls and subway stations, individuals often find it difficult to navigate or determine their current location. The problem becomes more severe in environments surrounded by tall buildings or located deep underground, where Global Positioning System (GPS) signals are weak or completely unavailable. With the recent advancement of the geospatial information industry, interest in indoor spatial data development and indoor positioning technologies has been steadily increasing. In general, much of the research on indoor positioning systems has focused on signal-based positioning techniques. Yang and Shao (2015) introduced a Wi-Fi-based indoor positioning technique that utilises received signal strength from pre-installed access points to estimate the user's position. However, its accuracy can be limited due to signal interference and multipath effect. In contrast, vision-based positioning methods typically estimate the camera's position by leveraging multiple known points within an image. A representative approach is the Perspective-n-Point (PnP) algorithm, such as the solvePnP function provided in the OpenCV library, which estimates the camera position based on correspondences between 2D images and their known 3D world coordinates (Lee et al., 2023). However, the estimation of camera position based on multiple known points is subject to significant constraints when applied to mobile devices. Unlike satellite or aerial imagery, image captured by mobile devices generally have a limited field of view (FOV), which makes a single frame unlikely to contain a sufficient number of identifiable landmarks. Consequently, this often results in either the inability to properly formulate the observation equations or the derivation of unstable estimates.

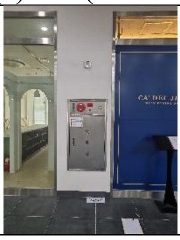
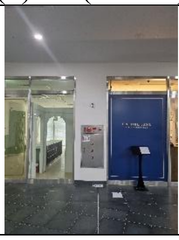
In particular, it is not an easy task to construct sufficiently detailed indoor spatial data that can distinguish landmarks in narrow indoor environments (Lee & Rhee, 2022). This study introduces a method for estimating camera position using single known landmark point in an image.

2. Methodology

2.1 Material

Two types of data are required for the experiment. The first type consists of landmark images and their coordinate information. In this study, we aimed to estimate the camera's position and orientation under the constraint of having only a single known landmark point. Subsequently, images were acquired to estimate the camera position relative to the known landmarks. In order to assess the accuracy of the estimated results against the ground truth, distance and angle measurements were carried out. To replicate conditions comparable to the expected experimental environment, we selected the first floor of the BT Center in Songdo, Incheon, as the study area. The image dataset was collected entirely with an Android-based mobile device, and all images were taken in portrait orientation. The 7 images were acquired, with data collected at distances of 3 m and 5 m from the landmark. In addition, to verify the accuracy of the estimated orientations, images were captured at observation angles of 20°, 40°, and 60° relative to the landmark. A fire hydrant was selected as the primary target. Table 1 presents the details of the images used in the experiment.

Table 1: Images used in the experiment.

(a) 3m (frontal)	(b) 5m(frontal)	(c) 3m(20°)	(d) 5m(20°)
			
(e) 3m(40°)	(f) 5m(40°)	(g) 3m(60°)	(h) landmark
			

2.2 Method

the approximate location of the landmark was acquired using a mobile device. In addition, both landmark images and images captured with the mobile device were acquired. Based on this information, the camera orientation is estimated from the line-of-sight vector to the target in the mobile image, while the approximate distance is computed by comparing the width ratio of the object in the two images. Equation (1) presents the model for camera position estimation.

$$X_{mobile\ device} = P_{landmark} + D * \vec{d}_{enu} \tag{1}$$

where, $X_{mobile\ device}$ = The position of the mobile device

$P_{landmark}$ = The position of the landmark

D = The distance between the mobile device and the landmark

$\vec{d}_{enu}$ = The line-of-sight vector in ENU coordinates

The process of camera position estimation was structured into three steps. First, the orientation of the camera was determined using the line-of-sight vector representing the viewing direction, which was derived from the centroid of the four corner points of the landmark. The pixel coordinates were collected manually and converted from the image coordinate system to the ENU coordinate system through the sensor model. Subsequently, the distance between the landmark and the mobile device was estimated using a straightforward method based on the width ratio of the object in the images. The ideal scenario is when the landmark is captured from a frontal view; however, in practice, images are more often acquired from oblique views of the landmark. To address this, a homography transformation was computed using the four corner points of the landmark image and their corresponding points in the captured image, allowing the object to be rectified into a frontal view for width ratio computation. In the final step, the camera position was estimated relative to the landmark reference, based on the line-of-sight vector and the distance derived from Equation (1).

3. Results

The calculated positions were evaluated through relative azimuth comparison. The azimuth between the landmark and the camera position obtained from the frontal image (Table 1(a)) was set as 0°, and relative azimuths were derived from the images (Table 1(b)-(g)). The relative azimuths were then compared with the ground-truth measurements. Table 2 shows the results of the azimuth and distance comparisons.

Table 2: The results of comparing azimuths and distances.

Name of data	Estimated azimuth(°)	Ground-truth azimuth(°)	Difference azimuth(°)	Estimated distance(m)	Ground-truth distance(m)	Difference distance(m)
3m (frontal)	0.0	0.0	0.0	2.969	3.0	0.030
5m (frontal)	-18.254	0.0	-18.254	4.998	5.0	0.002
3m (20°)	16.422	20.0	-3.578	3.094	3.0	-0.094
5m (20°)	-10.473	20.0	-30.473	5.239	5.0	-0.239
3m(40°)	34.151	40.0	-5.849	3.808	3.0	-0.808
5m(40°)	18.213	40.0	-21.787	6.490	5.0	-1.490
3m(60°)	0.166	60.0	-59.834	5.333	3.0	-2.333

First, we confirmed that the distance estimation between the landmark and the mobile device exhibited a gradual increase in error with increasing angle and distance. For the data captured at observation angles below 40°, the distance estimation errors were less than 1 m. When the observation angle was increased to 60°, however, the estimated distances showed errors greater than 2 m compared with the ground-truth values. In contrast, the azimuth estimation results exhibited greater variability. At a distance of 3 m, the estimated azimuths deviated from the ground-truth values by no more than 5°, except in the case of the 60° observation data. In comparison, when the landmark was observed from a distance of 5 m, most of the estimated positions deviated from the ground-truth values by more than 20°.

4. Conclusion

For distance estimation, the error tended to increase as the observation angle deviated from the frontal view. This phenomenon can be attributed to errors introduced in the homography-based rectification process, during which the captured image is transformed to a frontal view. Accordingly, for practical implementation, observations acquired at angles exceeding 60° should be excluded, following the same approach adopted in the experimental dataset. Subsequently, from the azimuth comparison, maintaining a sufficient distance from the landmark was identified as an important factor for reliable indoor positioning. In this experiment, accuracy declined when the landmark was observed from beyond 5 m. The degradation may be attributed to errors in the line-of-sight vector used for position estimation. For practical implementation, it would be preferable to apply the method in cases where the landmark is observed from around 3 m away and within an angle of 40°. In future research, we aim to integrate object detection techniques to automatically extract the landmark pixel coordinates—previously obtained manually in this experiment—and to apply the positioning method in an automated manner.

Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement grant funded by the Ministry of Land, Infrastructure and Transport (RS-2022-00141819, Development of Advanced Technology for Absolute, Relative, and Continuous Complex Positioning to Acquire Ultra-precise Digital Land Information)

References

Lee, Y. J., and Rhee, S. A., 2022. Study of Feature Based Algorithm Performance Comparison for Image Matching between Virtual Texture Image and Real Image, Korea Journal of Remote Sensing, 38(6-1), pp. 1057-1068.

Lee, Y. J., Yoon, W. S., and Rhee, S. A., 2023. Mobile Camera-Based Positioning Method by Applying Landmark Corner Extraction, Korea Journal of Remote Sensing, 39(6-1), pp. 1309-1320.

Yang, C. and Shao, H. -r., 2015. WiFi-based indoor positioning, *IEEE Communications Magazine*, 53(3), pp. 150-157.