International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:8, No:9, 2014

Image Features Comparison-Based Position Estimation Method Using a Camera Sensor

Jinseon Song, Yongwan Park

***Abstract*—**In this paper, propose method that can user's position that based on database is built from single camera. Previous positioning calculate distance by arrival-time of signal like GPS (Global Positioning System), RF(Radio Frequency). However, these previous method have weakness because these have large error range according to signal interference. Method for solution estimate position by camera sensor. But, signal camera is difficult to obtain relative position data and stereo camera is difficult to provide real-time position data because of a lot of image data, too. First of all, in this research we build image database at space that able to provide positioning service with single camera. Next, we judge similarity through image matching of database image and transmission image from user. Finally, we decide position of user through position of most similar database image. For verification of propose method, we experiment at real-environment like indoor and outdoor. Propose method is wide positioning range and this method can verify not only position of user but also direction.

***Keywords*—**Positioning, Distance, Camera, Features, SURF (Speed-Up Robust Features), Database, Estimation.

I. Introduction

THESE days, due to the development of information technology and universal availability of smartphones, the use of position information has rapidly increased. Position estimation refers to estimating the user's position in indoor and outdoor environments and is essential not only to locate the user, i.e. its fundamental purpose, but also to improve the traceability of a system that requires driving such as unmanned robots or vehicles. With the increased need of position estimation, there have been various studies about it. The studies of position estimation are largely classified into triangulation, fingerprint, and image recognition. Triangulation estimates [3], [8] the position of a user by measuring the time transferring from the source of a signal to the receiving end, and the fingerprint technology [4] by saving the unique information of positions such as the earth's magnetic field and WLAN (Wireless Local Area Arrival) in the database and comparing the unique information with the information from the user. These techniques need an additional step of building their environments, but are the most popular subjects in the position estimation field because they provide quiet accurate information in terms of positions. Another technique of position estimation involves the use of images that are acquired through the camera sensor to estimate the user's position. Position estimation using image processing includes the marker-based technique for image recognition [5], the Stereo Vision technique using the difference of the angles of the stereo lenses, and the SLAM (Simultaneous Localization and Mapping) technique [1], all of which are currently studied in the position estimation field. Position estimation using a camera involves image processing, so requires rather a lot of processing and spends a lot of the processing capacity of the system, while providing users with visual as well as position information, for which this technique is a popular research subject [6], [9]-[11]. This study introduces a new technique that estimates positions using a camera sensor. Because this method does not require a marker that is used for position estimation, it is free of concerns over the marker being damaged and can contain directional information that the marker-based technique has been unable to provide. This technique overcomes a drawback of the stereo vision technique having a limit in terms of ranges for estimation, allowing a wide range of estimation thanks to its strong database. In addition, its use of a single camera sensor removes the process of combining the data from multiple sensors such as ultrasonic waves used in SLAM, LIDAR, and GPS.

Chapter II further explains about the three techniques of position estimation using image processing, and Chapter III explains the algorithm and the system structure of the suggested technique.

Jinseon Song and Yongwan Park with Yeungnam University, Gyeungsan, 712-749 Republic of Korea (corresponding author phone: +82-53-810-3942; fax: +82-53-812-8583; e-mails: ywpark@yu.ac.kr; siegsong@ynu.ac.kr).

II. Related Work

A. Position Estimation Technique Using Markers

Most of the position estimation techniques using image processing involve the use of markers. Markers are used to recognize a position through a camera sensor, requiring a pre-step in which the user installs recognizable markers on a user-designated position. The developer generates and attaches markers whose color, size, and shape should be standardized in order for the camera sensor to recognize them. And the markers installed on each position should be unique in their shape for a recognition purpose. A marker-based image recognition using a camera sensor requires a large volume of image processing. For the extraction of markers on an image, the threshold of the image that varies by environmental lights is determined, a binary image is generated, and then the markers are extracted by removing the background from the binary image. Out of the extracted markers, patterns, letters, and signs are recognized and stored in the database, and the position information contained in the stored markers is provided to the user. This is a brief explanation of position estimation using markers. The success of a marker-based position estimation technique is determined by how markers are recognized, and this technique is vulnerable to errors when the markers existing outside are damaged. Moreover, because the markers have to be structured differently by position, it requires attention to building and managing its infrastructure.

B. Position Estimation Technique Using Stereo Vision

Stereo vision was created in a perspective in which human have a pair of eyes to see objects and is a vision sensor good for seeing objects and their backgrounds with a three-dimensional vision. Depending on the technique, stereo vision is classified into an active technique and a passive technique. The active technique projects laser at the point from which the user wants to acquire depth information and receives two images, which are used to find a corresponding point at which lasers from the two images meet. This technique eliminates the process of matching the images. Unlike the active technique, the passive technique uses only two cameras without using laser, matches the images acquired from the cameras, and extracts depth information. Most of the position estimation studies about stereo vision are about this passive technique. Two images acquired by stereo vision are taken at different times, so have to be matched if no laser has been used to find a corresponding point. Matching is a process of discovering the corresponding relationship between two images to find the point at which both images project it in a three-dimensional space. It is classified into area-based, feature-based, and energy-based techniques. The process of matching produces not only thee-dimensional images from the acquired images, but also enables calculating distance by acquiring the depth between the user and the object. However, stereo vision-based position estimation calculates the depth using the difference of the angles of the two images acquired, so reliability substantially decreases when the depth information was earned from a long distance of 5M or greater. Due to this drawback, the stereo vision technique is not popular for calculating distances and estimating positions and mainly used to generate a three-dimensional image from images in the left and the right.

C. SLAM (Simultaneous Localization and Mapping)

SLAM is, when the position of a robot is unknown from an indoor environment, to sense its surroundings, create a map, and estimate the position of the robot in the map [2]. The core of SLAM is, as the name suggests, localization and mapping. Localization recognizes the moving direction and route of a robot against its surroundings, so needs an accurate map for an accurate estimation of its position. Mapping is a process of analyzing information that the robot's sensor collects and drawing an answer, i.e. drawing a map. Vision-based SLAM using a camera sensor extracts features – an essential process for image processing, after which the features on each image frame that are shown as a single point in a real-life space, are matched. This process uses the position of the images captured, i.e. the position information of the robot, as its base and needs sensors, e.g. accelerated sensor and a gyro sensor, which to find position information of the robot. Keys to a successful SLAM are the calculation speed or capacity of the computer, simplification of the algorithm, and optimization for better accuracy, allowing real-time processing and increasing efficiency. However, this technique should perform localization and mapping at the same time and requires calibration of the data acquired from the vision sensor and the sensors for the purpose of finding position information, causing a substantial overhead to the system.

III. Proposed Algorithm

For estimating the position of a user using a single camera sensor, we captured images at a regular interval and stored them and their position information in the database. The images stored in the database were used to compare them against the images inputted by the user and to provide position information. To determine similarity between the database images and the user images, a SURF-based [7] image-matching algorithm was used, and after the user's images were compared against each database image, the position information of the image in the database most similar to the user's image was provided to the user.

A. Formation of Database

The database is comprised of images, which also contain the position information of where they were captured. The capturing interval of images for the database directly affects the accuracy of the images' position estimation, so the shorter the interval is for the database, the more accurate estimation becomes. Throughout the paper, our database has been continually updated in order to provide more information and more accurate estimation to users.

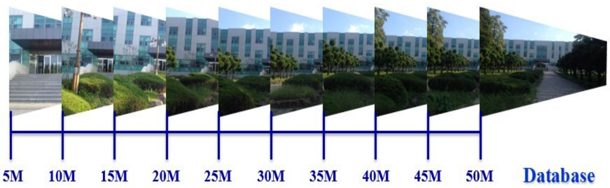

Fig. 1 Position estimation of a single-dimensional, Single-directional outdoor space

First of all, to determine workability of our system, for an outdoor experiment, only single-directional images corresponding to the direction of the user were captured. In this experiment, the images were captured at an interval of 5M up to the target point, and a total of ten images in a 50M-range were saved in the database and compared with the user image.

As the first experiment proved its workability, for a better database, the single-dimensional straight-lined range for our first experiment was upgraded to an indoor two-dimensional coordinated space, followed by an appropriate update of the database to update the user with appropriate position estimation. The images stored in the database were designated with a pair of coordinates (x, y) and contained the position information of where they were captured. In addition, each pair of coordinates in the database contained images taken in the four – east, west, south, and north - directions for which to determine the direction that the user faces.

B. Determination of Similarity between Images

The SURF-based algorithm was used to overcome the problems occurring due to the changing environment and rotation and angles of the acquired images and to increase the detection of similarity between the images.

TABLE I
POSITION OF THE DATABASE MATCHED IN THE OUTDOOR, SINGLE-DIMENSIONAL STRAIGHT-LINED RANGE

Input Data	Matched Result
19M	20M
23M	23M
31M	30M

The SURF-based algorithm is advantageous because the features extracted remain the same regardless of the angle and size of an image, so it is a representative image-matching algorithm in the image recognition field. Each feature extracted through the SURF algorithm holds a feature value of the pixel composition, which is used to calculate the Euclidian distance between the database images and the user images, determining the same features between the two images.

Fig. 2 Features in the extracted image

Our system was designed to compare the user image against the database images, to count the number of the matching features, and to find the image holding the highest matching count of features in the database. The features contained in a user image acquired at a random position are extracted; an image is brought out of the database; its features are found; and then the matching features are counted. This process is repeated until the database image with the highest number of matching features is found, and the position information of that image is provided to the user for position estimation.

IV. EXPERIMENTS AND RESULTS

A. Position Estimation in an Indoor Single-Dimensional Straight-Lined Range

We captured an image 1500mm high above the ground and converted it to a size of 640x480 prior to use. For a fast extraction of features and their matching, the pre-processed image was converted to a gray image.

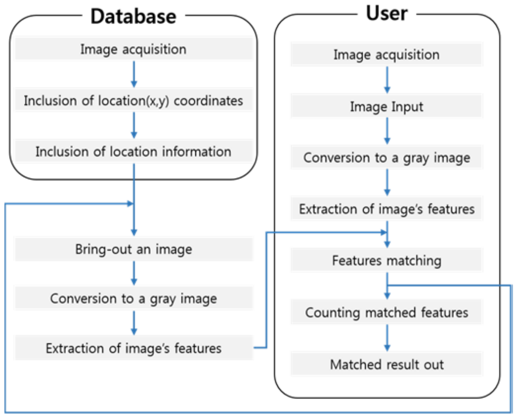

Fig. 3 Position estimation algorithm based on comparison of image features

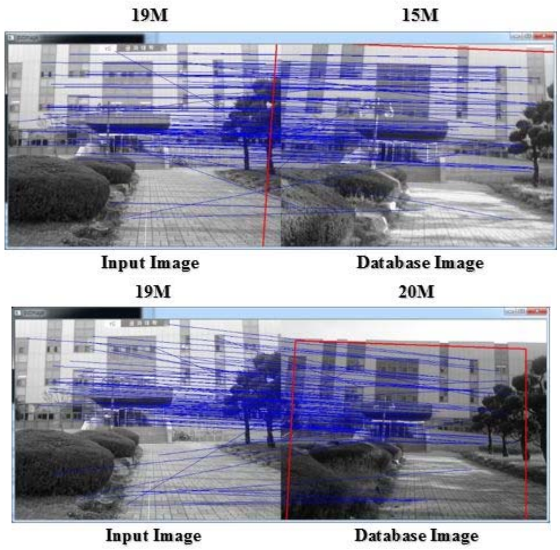

Fig. 4 Features matching in an outdoor, single-dimensional straight-lined space

For our first experiment with an outdoor environment, a total of ten images were captured from a 50-metered single-dimensional straight-lined range at an interval of 5M, which consisted of the database, and user images were captured at random ranges, 19M, 23M, and 31M. The images captured from the random ranges were compared with the database images to determine similarities, which were examined how they are matched with the most similar database image.

As a result, the images captured from the three random points are deemed the highest similarities with the database image in their nearest point, and therefore were matched with the position value in the database. Table I

B. Position Estimation in an Indoor Two-Dimensional Coordinated Space

For the experiment performed with an indoor two-dimensional coordinated space, as shown in Fig. 5, we designated a total of 38 coordinated points each with a pair of the x and y coordinates (x,y) at every three meters in order to estimate the position. The x coordinate was set for the distance from the west to the east, and the y coordinate from the north to the south, all of which fell between the coordinate points (3,3) and (75,12). In addition, unlike the experiment performed with the indoor, single-dimensional straight-lined range, this experiment captured images in the four directions and stored them in the database in order to provide directional information to the user.

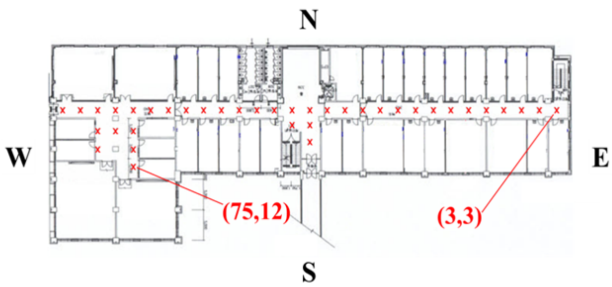

Fig. 5 Coordinates in the indoor, two-dimensional space

To examine the result of this experiment, we captured images in the four directions - east, west, south, and north at five random points in the indoor two-dimensional space and estimated the position. As a corridor-based range in the indoor environment contains only two directions, i.e. the front and back, the database was formed only with images of the west and east, and the database of a wider space such as the lobby was formed with the four directions to provide directional information to the user.

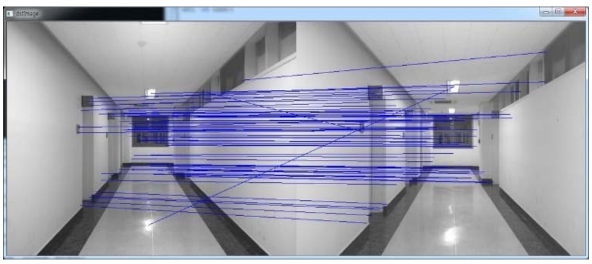

Fig. 6 Images matching with indoor, two-directional coordinates

For the indoor position estimation experiment with a coordinated space, unlike the straight-lined space in the single-dimensional experiment, its range was added with another dimension, and the direction that the user faced was provided as well.

As a result, as shown in Table II, images facing different directions from five random points were captured and used for estimating the position and were matched with the database images nearest from the user from which we were able to acquire its coordinates and direction.

V. Conclusion

The proposed position estimation technique using image processing estimates the position, based on the comparison between the images stored in the database that was built beforehand and the user images that were captured from random positions. The keys for this position estimation technique matching features are a precise database and the detection of the image that is most similar to the input image via comparison between the features of the images.

TABLE II
Position of the Database Matched in an Indoor, Two-Dimensional Coordinated Space

Random position coordinates	(10,3) W	(46,5) N	(47,7) S	(76,7) N	(80,10) E
Nearest position coordinates	(9,3) W	(46,6) N	(48,6) S	(75,6) N	(81,9) E
Matched position coordinates	(9,3) W	(46,6) N	(48,6) S	(75,6) N	(81,9) E

In this paper, to detect where the user is located in a two-dimensional coordinated space, for the database, the images were set as (x, y) coordinates and captured in the four directions. The suggested technique does not require managing its infrastructure like the marker-based estimation technique and also overcomes the problem that the stereo vision technique faces in terms of estimated ranges. In addition, unlike the SLAM-based technique that needs various sensors, our suggested technique uses a single camera sensor without the process of combining data.

Our study estimated positions only using still images and built a database containing more images for accuracy and more position information. As a result, estimating a single position even consumed quite a lot of computing time, disabling a real-time processing.

Our future studies should focus on building a simpler and better organized database containing a large volume of images and on filtering images only with essential elements in the process of matching the extracted features, then enabling a real-time processing of the imaged inputted by the user.

Acknowledgment

This work was supported by the BK21 Plus Program (Creative Human Resource Development Team for ICT-based smart devices, 214C000066) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF).

References

[1] N. Karlsson, E.D. Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, M.E. Munich, "The vSLAM Algorithm for Robust Localization and Mapping," *Proc. IEEE Int'l Conference Robotics and Automation*, 2005.
[2] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. *In Proceedings of the AAAI National Conference on Artificial Intelligence, Edmonton, Canada*, 2002.
[3] Yanying Gu, Anthony Lo, Ignas Niemegeers "A Survey of indoor positioning systems for wireless personal networks" *Proc. IEEE communication surveys & tutorials 2009.*
[4] K. Kaemarungsi, "Efficient design of indoor positioning systems based on position fingerprint" *International conference on wireless networks, communication and mobile computing*, 13-16 June 2005 vol. 1, pp. 181-186.

[5] J. B. Kim, H. S. Jun, "Vision-Based Position Positioning using Augmented Reality for Indoor Navigation", *IEEE Transactions on Comsumer Electronics*, vol. 54, No. 3, August 2008.
[6] A. Baumberg. “Reliable feature matching across widely separated views” *In CVPR*, 2000, pp. 774 – 781.
[7] M. Brown, D. Lowe, "Invariant features from interest point groups" *In Proceedings of the 13th British Machine Vision Conference(BMVC)*, 2002, pp. 253-262.
[8] Hui Liu, Houshang Darabi, Pat Banerjee, Jing Liu, "Survey of wireless indoor positioning techniques and systems", *IEEE Transactions on System, Man, and Cybernetics—Part C: Applications and reviews*, vol. 37, NO. 6, November 2007.
[9] H. Bay. T. Tuytelaars, L. Van Gool, “SURF: Speeded-Up Robust Features", *in ECCV*, 2006, pp. 404-417.
[10] D. Bradley, G. Roth, “Adaptive Thresholding Using the Integral Image”, *Jonrnal of Graphics, GPU, and Game Tools* 2007, vol. 12, Issue 2, pp. 13-21.
[11] J. Wolf, W. Burgard, and H. Burkhardt. “Robust vision-based localization for mobile robots using an image retrievel system based on invariant features” *In Proc. of Int. Conf. on Robotics and Automation (ICRA)*, 2002.