Panoramic Image-Based Navigation for Smart-Phone in Indoor Environment

Van Vinh Nguyen, Jin Guk Kim, and Jong Weon Lee

Sejong University, Seoul, Korea
vinhnv2@yahoo.com, jingoogi@hotmail.com, jwlee@sejong.ac.kr

Abstract. In this paper, we propose a vision-based indoor navigation system for a smart-phone. The proposed system is designed to help a user traveling around an indoor environment to determine his current position and to give him the direction toward a chosen destination. For sensing user's position and orientation, the system utilizes panoramic images, which are pre-captured the environment and then processed to create a database. For matching images captured from user's smart-phone with the database, we use SURF[1], a robust detector and descriptor. Besides, to minimize responding time, the system employs client-server architecture in which a server module is mainly in charge of time consuming processes. Also, a tracking mechanism is applied to reduce matching time on the server. The experimental results show that the system can work well on a smart-phone in interactive time.

Keywords: Indoor navigation, panorama tracking, augmented reality.

1 Introduction

Generally, navigation in unknown environment considers Global Positioning System (GPS) to estimate positions. Many GPS-based navigation systems running on robot or vehicles are available and work effectively. Recently, with the growing of new generation of smart-phones that are equipped with GPS sensor, navigation on a smart-phone becomes realistic. However, a GPS-based navigation method only works well in outdoor environment. Since GPS signals are not available for indoor environment and other sensors such as WiFi-based estimation [2] and RFID-based tracking [3] are not robust yet, in this paper we focus on vision-based navigation method that is normally simpler, cheaper and easier to deploy broadly than other solutions.

Vision-based localization methods are normally based on the analysis of captured images to get information of current position. In our proposed system, indoor localization is based on the comparison of runtime captured images with the database of reference images, which are pre-captured from the targeting indoor environment. Each image can be described as a set of natural features, thus the comparison of two images becomes feature matching problem. To deal with this problem, we apply a SURF feature [1] which has been found to be highly distinctive and repeatable. However, SURF-based matching shows low speed when running on a smart-phone due to limited resources of the device, thus we employ client-server architecture to

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 370–376, 2011.

reduce system's responding time. Because reference images in the database should cover entirely the targeting environment for efficient localizing, we rely on pre-captured panoramic images to reduce the number of required reference images.

The remainder of this paper is structured as follows. Section 2 surveys related works in vision-based mobile phone localization and natural feature descriptors. Section 3 describes our approach in detail. Experiment results and discussions are presented in Section 4. We conclude with a discussion of limitations and future work in Section 5.

2 Related Works

The AR-PDA project [5] develops an image-based object recognition and tracking method for AR applications that requires the 3D-model of an object to be matched. A fully indoor navigation system targeting on camera phone has been proposed in [6], which is based on floor corners and SIFT features. This work relies on client-server architecture to reduce response time, but the system still suffers from late response that is about 10 seconds. S. Saito et al. takes a different approach. They present their indoor positioning system [7] based on fiducial markers that are pre-designed for containing position information. Using markers for determining user's position is easy way to deploy, but it locks a user to fixed positions for sensing current position. In our proposed approach, we try to make a user freely move around environment while getting navigation guides at anywhere he is staying. To enable this feature, we apply natural feature-based matching method for our system.

For comparing two different images, many approaches have been proposed. Point-based approach uses interest point detectors and local descriptors to localize and describe image features. SIFT (scale invariant feature transform) method [8] is well-known for its robustness. The SIFT feature descriptor is invariant to scale, orientation, affine distortion and partially invariant to illumination changes. A research mentioned in [9] indicates that the SIFT-based descriptors outperform other local descriptors on both textured and structured scenes. The SURF descriptor [1] is partly inspired by the SIFT descriptor and has similar performance to SIFT, while it is much faster than SIFT. Considering the system performance, we choose SURF as our main detector and descriptor.

3 Proposed System

3.1 System Overview

Considering the overall performance, the proposed system is designed in client-server architecture. It consists of a client module running on smart-phone(s) connected to a server via a wireless network. The system's architecture is demonstrated in Fig. 1.

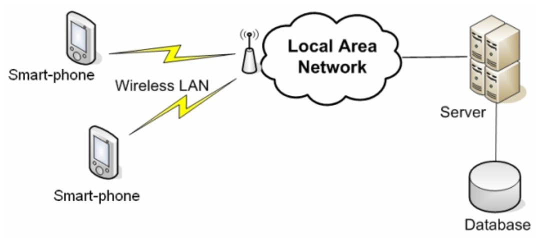

Fig. 1. System architecture diagram

The client module uses smart-phone's camera for capturing images of the real world and sends those images to the server. When the server receives an image, some filters are applied to the image to remove noises. After that, the received image is compared with ones stored in the database. The comparison of images is based on SURF features [1] that are detected in the image. If any match is found, the server estimates current user's position and orientation and they are sent back to the smart-phone. Based on the estimated position and orientation, the client module will display a direction to a chosen destination. Fig.2 shows the system's processes and their relations.

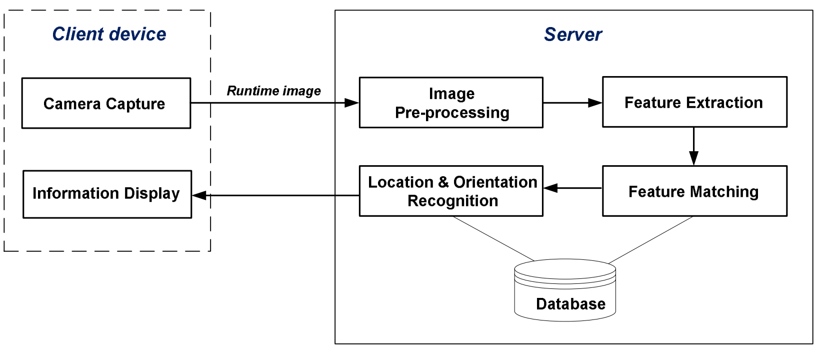

Fig. 2. System's workflow diagram

3.2 Database Creation

Because our position recognition approach is based on image matching, we need to create a database of features of reference images. Features on the DB are compared with features of smart-phone's images during the runtime phase. A large set of reference images should be taken so that they cover the targeting environment. For minimizing the number of required images, we capture panoramic images that are taken in the environment. Specifically, creating the database has following steps:

- *Step 1*. We setup a 2D reference coordinate system in the indoor environment and then the coordinates of featured places (i.e. rooms in a building) are roughly estimated.

- *Step 2*. For each featured place, we create several 360° panoramic images. Panoramic images are easily created by using some available applications such as panorama tools using a stitching method. In our work, we use wide-angle camera lens for capturing environment and then use a stitching tool to create panoramas. Besides, the number of panoramic images in each featured place has to be minimized to reduce the matching time while maintaining the required resolution of position estimation. Practically, we divide each place into 4 m x 4 m cells and take one panoramic image in the center of each cell. The position and orientation are recorded with each panoramic image. The orientation of panoramic image means the angle between the view's center and Y-axis of the reference coordinate system.
- *Step 3*. We divide horizontally each panoramic image into four sub-images. That means one sub-image covers 90° of the view and it is easy to calculate the orientation of each sub-image based on the orientation of the original panoramic image. Fig. 3 shows one set of sub-images of a panoramic image.
- *Step 4*. We run SURF detector and descriptor on each sub-image and then store the descriptors with corresponding positions and orientation information into the database.

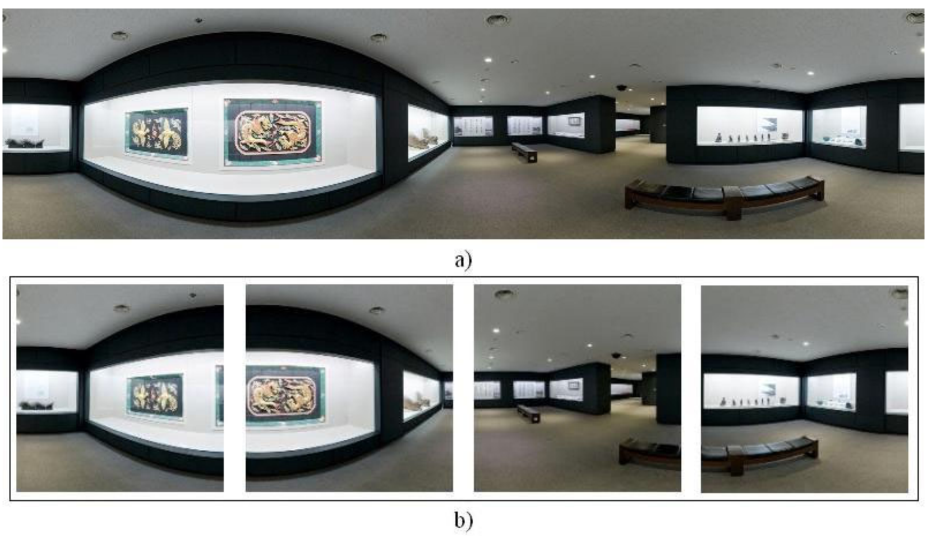

Fig. 3. a) 360° panoramic image b) Four sub-images created from the panorama

3.3 Position and Orientation Sensing

An image sent from a smart-phone to a server is processed using SURF. Generated SURF descriptors are compared to ones in the database. If any match is found, the corresponding position and orientation of the matched sub-image in the database are considered as the position and orientation of the smart-phone. However, the captured image from the smart-phone is usually smaller than reference sub-images and these images are not often aligned. Therefore, we apply homography to revise the real orientation of the smart-phone: we use RANSAC[4] to estimate the central point and four corners of the matched reference sub-image in the captured image's frame. Having that information, we can calculate the error of the orientation using the equation 1. After estimating the current coordinates and the real orientation of the smart-phone, it is straightforward to calculate the direction from the current position to the chosen destination since any chosen destination's coordinates is known.

$$\Delta\beta = \frac{\Delta x * 90}{Ws} \tag{1}$$

Ws: Width of reference image in captured image's frame.
Δx: Difference of 2 central points' x-coordinate in captured image's frame.

To reduce matching time, not all of the reference images in the database are used to compare with smart-phone's captured image. Each captured image is firstly compared to reference sub-images that are closely located to previous position of the smart-phone. Assuming slow user's motion, this mechanism can avoids comparing a captured image to all reference images, which is significant time consuming process.

4 Experimental Results

We have implemented a test version of the proposed navigation system for performance evaluation. The client module is deployed on an iPhone-3GS whereas the server module is running on a PC-2.4GHz. The tested indoor environment is a museum and a database of features is created from 11 panoramic images. The system is then tested in the real museum with a following scenario:

Fig. 4. An example of server-side operation. Left image is one received from iPhone. Right one shows the direction to the target location.

"A user travels around the museum with an iPhone. When he needs to move to another place, he uses the iPhone to capture images (or objects) near his current position. Based on the captured image, the navigation system running on the iPhone will show a virtual arrow indicating the direction to the destination. Moreover, an estimated distance to the destination is also displayed".

The testing results show that the system can guide a user successfully to a destination in an interactive time by displaying a correct direction and an approximate distance. For each image received from an iPhone client, the server consumes around 0.7 second to determine position and orientation of the iPhone. Considering time for transferring information from the server to the phone, the average response time of the proposed system is around 1 second.

5 Conclusion

We have proposed a navigation system for mobile phone in indoor environment. The proposed system exploits the characteristics of panoramic images, which cover 360^O of surrounding scenes, to create a database of reference images. Smart-phone's positions and orientations are estimated based on positions and orientations of matched panoramas. Besides, to improve the performance of the system, the system employs a client-server architecture that outsources all time-consuming tasks to a server. Also, a tracking mechanism is applied to reduce matching time on the server. Therefore, the system achieves reasonable responding time as proven in the experimental evaluation. The limitation of the system is the difference between estimated smart-phone's position and real position. Currently, the estimated position is same as the nearest (or the most similar) panorama's position. For any applications which require more precise position of a smart-phone, we need to improve the position estimation method. This problem is addressed as our future work.

Acknowledgments. This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA), under the Culture Technology(CT) Research & Development Program 2010.

References

1. Bay, H., Tuyelaars, T., Van Gool, L.: Speed-up Robust Features. In: ECCV, pp. 404–417 (2006)
2. Gezici, S.: A Survey on Wireless Position Estimation. Wireless Personal Communications: An International Journal 44(3), 263–282 (2008)
3. Bouet, M., dos Santos, A.L.: RFID tags: Positioning principles and localization techniques. In: Proceedings of IFIP Wireless Days - 2nd International Home Networking Conference, IHN 2008 (November 2008)
4. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM 24(6), 381–395 (1981)

5. Gausemeier, J., Frund, J., Matysczok, C., Bruederlin, B., Beier, D.: Development of a real time image based object recognition method for mobile AR-devices. In: Afrigraph, pp. 133–139. ACM, New York (2003)
6. Hile, H., Borriello, G.: Information overlay for camera phones in indoor environments. In: Hightower, J., Schiele, B., Strang, T. (eds.) LoCA 2007. LNCS, vol. 4718, pp. 68–84. Springer, Heidelberg (2007)
7. Saito, S., Hiyama, A., Tanikawa, T., Hirose, M.: Indoor Marker-based Localization Using Coded Seamless Pattern for Interior Decoration. In: Virtual Reality Conference (2007)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2), 91–110 (2004)
9. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence (2005)