Harnessing the Power of Technology to Improve Lives
P. Cudd and L. de Witte (Eds.)
IOS Press, 2017

doi:10.3233/978-1-61499-798-6-709

Image Based Location Estimation for Walking Out of Visual Impaired Person

Kazuho KAMASAKA[a], Itaru KITAHARA[b] and Yoshinari KAMEDA[b]
[a] *University of Tsukuba, Tsukuba, Ibaraki, Japan*
[b] *Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan*

Abstract. A new and intelligent walking navigation system could be helpful for visually impaired people so that they do not need helpers or guide dogs on going out. Conventional navigation systems using GPS are not available indoors or undergrounds. A new location estimation method that is available in such situations is necessary. We propose to estimate pedestrian's location with only a single camera attached to the pedestrian. Our proposed method uses only computer vision and no other sensors are needed. The location estimation is achieved by image retrieval. The retrieval database is built from a pre-recorded video taken along a planned walking path. Simple image retrieval using local image features does not work well when the images were taken in different conditions of time and weather. In this paper, we especially explore robustness for the change of sunlight condition due to its recording time. We propose a new method of robust location estimation based on our database built by combining more than one video that is taken under different conditions. Experiment results showed that the accuracy of the location estimation using the proposed database is better than the one from a conventional database built by using a single video.

Keywords. Image based location estimation, navigation, camera, computer vision, similar image retrieval.

1. Introduction

When visually impaired people go out, they usually need helpers or guide dogs. Intelligent navigation systems that can support them walking out without such helps are desired for visually impaired people. Important requirements of the navigation system are that it should work anywhere and anytime and without Internet access. As a location estimation method that can be used anywhere, Kameda et al [1] proposed an image based location system using only a camera. This approach assumes a video along the path to go has been taken in advance by a volunteer. Since visually impaired people usually follow a planned path to their destination such as a school or a workplace, it is not a problem that our location system can be used only on the path where the pre-recorded video is available. In navigation mode, a visually impaired person who is walking on the path takes query images at a fixed time interval by a worn camera. Then, the most similar video frame in the pre-recorded video to each query image is retrieved in the system so that the system can estimate the location in the path. That results in estimating the location of the person so as the path could be projected on a map. This approach can be applied anywhere if only a volunteer takes a video along the path beforehand. An advantage of our approach is that no investment on streets and social infrastructure are

required. However, location estimation accuracy is affected by the video quality – especially by the light conditions at the moment the pre-recorded videos are taken.

In this paper, a new database building method that uses two pre-recorded videos taken at different times is proposed. When a query image is given to the system, the most similar image at similar lighting condition in the two pre-recorded videos will be retrieved. The case of two videos is presented, but by using more videos, the robustness is further improved. Is also being considered to update the retrieval database by using query images taken by pedestrians. Since the two pre-recorded videos cannot be taken at the same walking speed, the frame-wise alignment from one video to the other is also presented in this paper.

Using multiple videos under different conditions is suitable for visually impaired people navigation because the path that visually impaired people need to follow alone is traced many times.

As for related works, many navigation systems for visually impaired people have already been proposed. Most of them are based on GPS [3, 4]. They could not work indoors or undergrounds. Therefore, new location estimation methods that do not rely on GPS have been proposed. For example, RFID (radio frequency identifier) tags [5, 6] may be installed along the path and they will be used to estimate the location using an RFID reader. These systems are only available in limited places where the tags are prepared. Other similar methods are promising in popular areas, such as NFC (Near Field Communication) [7], Beacons [8, 9], Wi-Fi [10], but they have the same problem. In less popular areas, alternative approaches are desired. As for an indoor localization method that does not require investment in infrastructures, Hu et al [11] used an RGB-D camera as a 3D sensor. But, their system needs a 3D map of the target building and an RGB-D camera is not yet a common device. The location estimation method [1] that uses only an ordinary monocular camera can be applicable on many situations. It is also good to integrate the camera-based method to conventional GPS and other sensor-based methods so that the integrated navigation system becomes more reliable.

This work is a part of a national funding project "Development of transportation support for visually impaired people by multi-generation co-creation." The project has been started in 2013 and it plans to conduct an integrated experiment on which our advanced technology will be evaluated by visually impaired people in actual street and building environment.

2. Method

2.1. Image retrieval and location estimation based on the retrieval

Similarity of two images is scored by the number of local small unique regions that are found in both images. To find and describe a small region in an image, SIFT [2] is used in computer vision. In SIFT approach, unique regions in an image are detected and they are called key points. A key point is represented by a key feature vector in 128 dimensional vector space, and if regions of the two key points look similar visually, their vectors are located close in the vector space. If a key point is given, the most similar key point can be found by searching the closest vector in the vector space. The authors call this a key pair.

In outdoor scenes, hundreds of key points are usually detected by SIFT an image of 320 by 180 pixels. If two images are taken at close time and space, many key pairs can

be found. In Figure 1 left, one green line denotes a key pair. Note that small regions around green dots at the end of each line look very similar. Key pairs could not be found well if the two images are taken at different time (Figure 1 right) because of lighting conditions and shadows in the scene.

The location estimation method based on image retrieval [1] is shown in Figure 2. The left side of Figure 2 shows pre-process, and the right shows the process of location estimation. In the pre-process, a pre-recorded video is decomposed into *reference frames*, SIFT key points are detected and their key features are calculated and stored into a database. Note that each key feature in the database is associated with the reference frame number in the pre-recorded video. In the process of location estimation, a set of SIFT key points are detected and their key features are calculated at a query image. For each key feature, the closest key feature is searched in the database. Since each result has a reference frame number, the frame number of the most similar one can be obtained simply by counting the occurrence of the reference frame numbers in the found key pairs.

Figure 1. SIFT key pair (left: small gap in time and space, right: large gap in time).

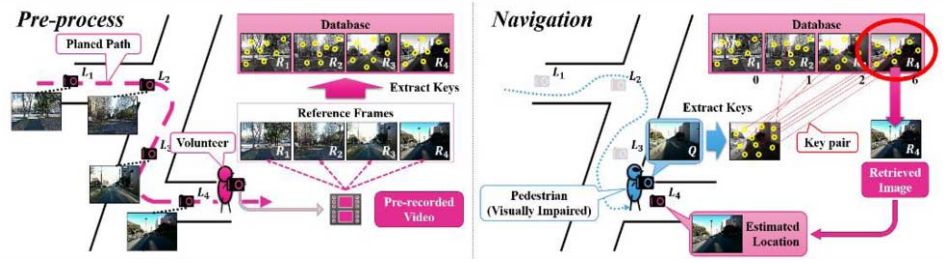

Figure 2. Location estimation based on image retrieval

2.2. Video Integration

Fortunately, the method described above can be directly applied for two pre-recorded videos. Figure 3 shows the extension of the method [1] to the two videos. The first pre-recorded video is named the main video and the second one the sub video. Note that the interval of reference frames might be different for the two videos. To integrate the two videos, it is necessary to detect reference image pairs that are captured at the same location in both videos. However, it is difficult to associate images since the appearances are different due to condition change. In the proposed method, the image retrieval approach is applied to find frame association.

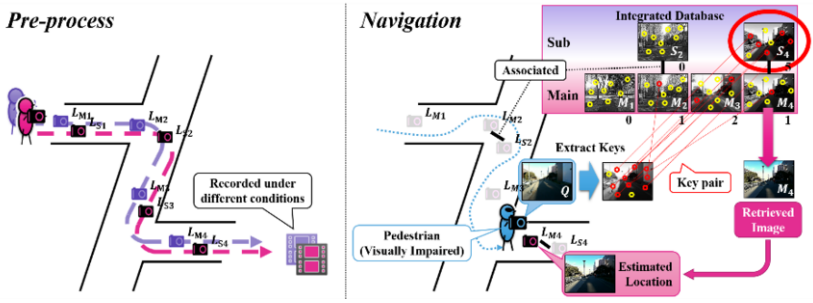

Figure 3. Proposed location estimation method.

2.3. Reference frame association

When location information is prepared for each reference frame for both videos, the procedure is exactly the same as shown above. However, the cost of adding location information to reference frames in a new video may be too high. Therefore, it is proposed to associate reference frames in the sub video to the ones in the main video. By assigning the reference frame in the main video, the location of the reference frame in the sub video is very close to the location of the associated reference frame in the main video.

A direct association is done basically just by treating each reference frame in the sub video as a query image to the main video. If the number of key pairs to the most similar reference frame is small, no frame association is made. In the case of Figure 4, only the frame No.12 and No.17 are directly associated to No.50 and No.53 in the main video respectively. Then, indirect association is done by interpolating the frame numbers between the direct associations. In this case, the reference frame No.13-16 in the sub video are indirectly associated to No.51 and No.52 in the main video.

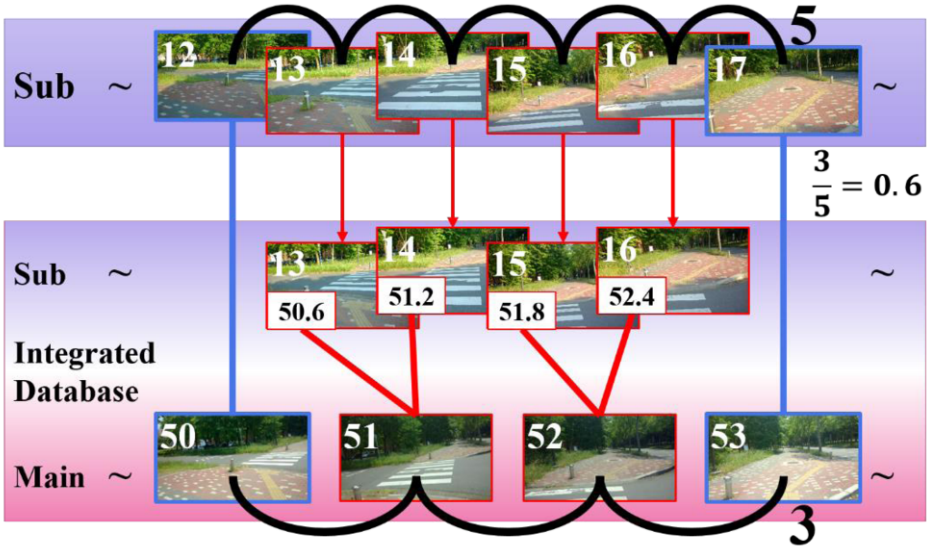

Figure 4. Reference frame association and database integration.

3. Experimental Results with Discussion

Three videos were taken on a real walking path for about 200 seconds that includes both outside and inside scenes. Walking speed was almost the same for the three videos, but not exactly so. The videos were taken at around 9, 11, and 13 o'clock on a sunny day. Figure 5 shows some frames in the videos. The videos were taken by the rear camera of Microsoft Surface Pro 3 and they are decomposed at 3 fps and resized to 360x180 pixels. A database composed by only the 9 o'clock video is denoted as DB_{09} and by the 13 o'clock video DB_{13}. The proposed integrated database DB_{09+13} was built from the main video of 9 o'clock and the sub video of 13 o'clock. 613 frames of the 11 o'clock video are used as query images taken at 3 fps. Comparisons were made with the results of location estimation using DB_{09} (conventional single database) and DB_{09+13} (proposed integrated database).

Figure 5. Video frames (top: captured at 9, bottom: captured at 13).

Figure 6 shows some results of query images and the retrieved reference frames. Six queries along the path were shown from the top row to the sixth row. 1st and 3rd columns are the same query images in the 11 o'clock video, and right next are the retrieved reference frames of the integrated database DB_{09+13} (2nd column) and single database DB_{09} (4th column). Green points and lines show the key pairs. If the number of key pairs is less than 3, which are marked with a red line, the estimation result is rejected due to a low reliability. The retrieved reference frames indicate that at locations of the 1st, 5th, and 6th, the queries were successfully done in the both methods. The 3rd queries also show the same location, but the integrated database returns more similar reference frame than that of DB_{09}. At locations of the 2nd and 4th, only the proposed method successfully returns the most similar reference frames.

Figure 7 shows the results of location estimation. The left graph shows the result of the proposed method (DB_{09+13}) and the right shows the conventional method (DB_{09}). The horizontal axis indicates the number of frames in the query video of 11 o'clock. The red dots of the left vertical axis indicate the estimated location on the path and the green crosses of the right vertical axis indicate the number of the key pairs. When the number of key pairs is less than 3, the retrieved result is rejected because of the low reliability of location estimation. Since all the videos were taken at very similar walking speed, the red dots are expected to form a linear correlation. In Figure 7 right, the location

estimation of the conventional method was not successful for some sections such as 1500-3000 and 3500-5000 frames. In contrast, as shown in Figure 7 left, the queries based on the proposed method were successful in such sections too. 124 query images were rejected by the conventional method, whereas only 55 query images are rejected by the proposed method.

The locations were successfully estimated by retrieved reference images of the sub database in the sections 1500-3000 and 3500-5000 frames, where the estimation was difficult only with the main database. Figure 8 shows the contributed video of each query in Figure 7 left. Some sections filled by the green circles were successfully done because of the high similarity between the query image and the sub video.

If the timing of walking out is close to that of the pre-recorded video, the conventional single video method works fine and less memory is needed. However, it is sometimes impractical to going out at a designated time. Our proposed method can be applicable to multiple sub videos, and it can also have videos in different weather conditions. This is very useful when the lighting conditions unexpectedly change when he/she goes out because our proposed method can prepare an integrated database that can cope with different times and weather conditions without requiring to know how much they are different.

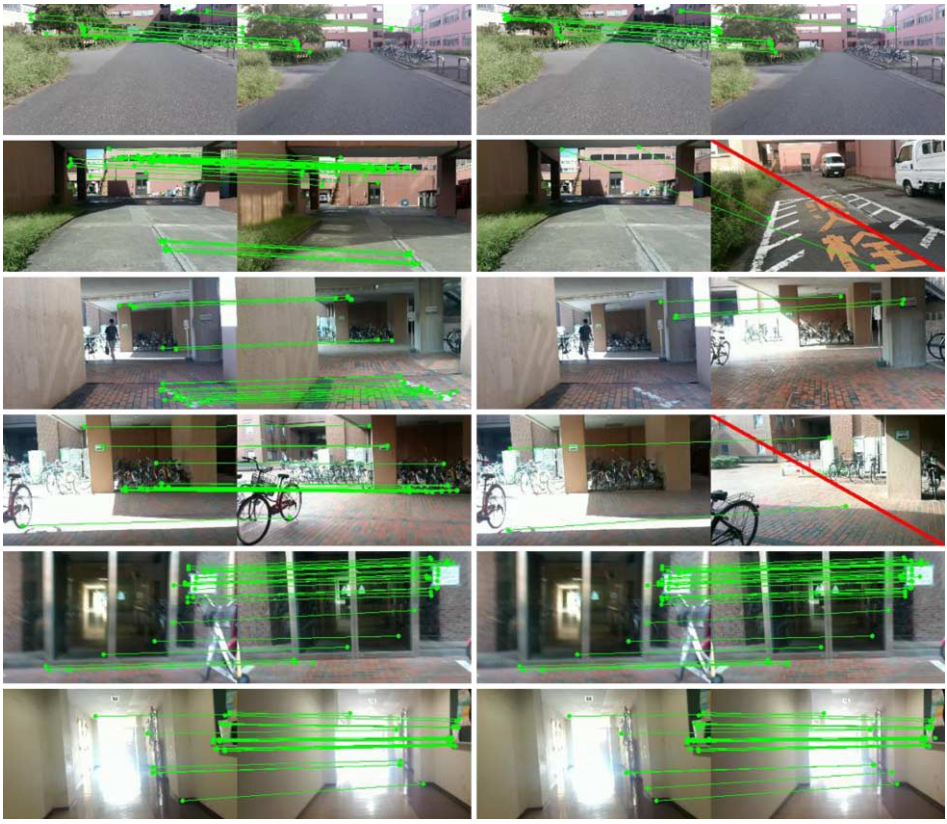

Figure 6. Six queries and their results along the path.

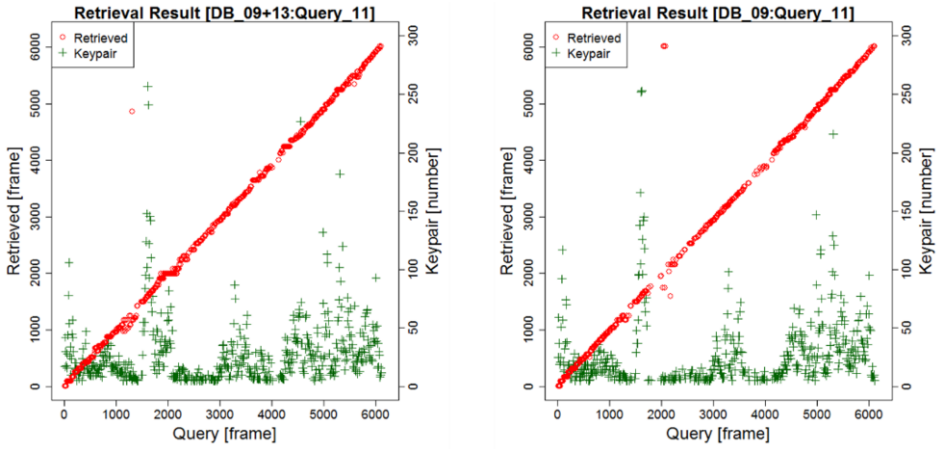

Figure 7. Location estimation results (left: proposed, right: conventional).

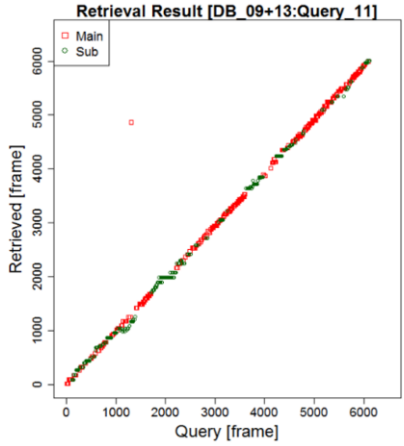

Figure 8. Contribution of the integrated database of Figure 7 left.

4. Conclusion

Our location estimation method using a single camera has an advantage that it can be used anywhere. It is easy for volunteers to help because they are just asked to take a video along a planned path. In the proposed method, robust location estimation resistant to condition changes is achieved by integrating multiple pre-recorded videos.

In our approach, it is necessary to take some pre-recorded videos by volunteers at first. The videos taken by ordinary people who walk along the path are also acceptable to build our new database for image retrieval.

Performance evaluation of the proposed method for integrating more sub databases should be done as a future work. Weather and other condition changes should be examined in more practical situations. Also, a new user interface that informs visually impaired people the location by sound should be invented.

The authors belong to a research group aiming to support the transportation of visually impaired people by developing a new navigation technology. By combining

other technologies such as GPS based and/or other sensor based location estimation methods and obstacle detection ahead of walking visually impaired, the authors aim to realize a system that supports all visually impaired people going out freely, by themselves.

Acknowledgement

This research is supported by a fund from Research Institute of Science and Technology for Society (RISTEX) and JSPS KAKENHI Grant Number 17H01773.

References

[1] Y. Kameda, and Y. Ohta, Image Retrieval of First Person Vision for Pedestrian Navigation in Urban Area, *Proceedings of IAPR 20th International Conference on Pattern Recognition* (2010), 364-367.
[2] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision* **60** (2004), 91-110.
[3] E. Carrasco, E. Loyo, O. Otaegui, C. Fosleitner, J. Spiller, D. Patti, R. Olmedo, and M. Dubielzig, Autonomous Navigation Based on Binaural Guidance for People with Visual Impairment, *Assistive Technology: From Research to Practice: AAATE 2013* **33** (2013), 690-694.
[4] R. Ivanov, Mobile GPS Navigation Application, Adapted to Visually Impaired People, *Journal of Information Technologies and Control* **1** (2009), 20-24.
[5] H. Fernandes, V.M. Filipe, P. Costa, and J. Barroso, Location Based Services for the Blind Supported by RFID Technology, *5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion* **27** (2014), 2-8.
[6] S. Chumkamon, P. Tuvaphanthaphiphat, and P. Keeratiwintakorn, A Blind Navigation System Using RFID for Indoor Environments, *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* **2** (2008), 765-768.
[7] A. Ganz, J.M. Schafer, Y. Tao, L. Haile, C. Sanderson, C. Wilson, and M. Robertson, PERCEPT Based Interactive Wayfinding for Visually Impaired Users in Subways, *Journal on Technology & Persons with Disabilities* **3** (2015), 33-44.
[8] D. Jain, Path-Guided Indoor Navigation for the Visually Impaired Using Minimal Building Retrofitting. *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility* (2014), 225-232.
[9] S. Bohonos, A. Lee, A. Malik, and L.R. Manduchi, Universal Real-Time Navigational Assistance (URNA): An Urban Bluetooth Beacon for the Blind, *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments* (2007), 83-88.
[10] T. Moder, P. Hafner, and M. Wieser, Indoor Positioning for Visually Impaired People Based on Smartphones, *14th International Conference on Computers Helping People with Special Needs* **2** (2014), 441-444.
[11] F. Hu, N. Tsering, H. Tang, and Z. Zhu, Indoor Localization for the Visually Impaired Using a 3D Sensor, *Journal on Technology & Persons with Disabilities* **4** (2016), 192-203