

A Spot Navigation System for the Visually Impaired by Use of SIFT-Based Image Matching

Hotaka Takizawa¹(✉), Kazunori Orita¹, Mayumi Aoyagi²,
Nobuo Ezaki³, and Shinji Mizuno⁴

¹ University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan
`takizawa@cs.tsukuba.ac.jp`

² Aichi University of Education, Kariya, Japan

³ Toba National College of Maritime Technology, Toba, Japan

⁴ Aichi Institute of Technology, Toyota, Japan

Abstract. In this report, we propose a spot navigation system to assist visually impaired individuals in recalling memories related to spots that they often visit. This system registers scene images and voice memos that are recorded in advance by a visually impaired individual or his/her sighted supporter at various spots. When the individual visits one of the spots, the system determines the current spot from the results of image matching between the registered images and a query image taken by the individual at the spot, then plays a voice memo which corresponds to the spot. The system is applied to actual indoor and outdoor scenes, and experimental results are shown.

1 Introduction

In 2014, the World Health Organization reported that the number of visually impaired individuals was estimated to be approximately 285 million worldwide [1]. Many of them are trained by sighted supporters to move along paths, for example, from their houses to offices. During such training, they are often taught information related to important locations (called *spots*) on the paths. For example, a visually impaired individual is taught that there is a restroom just outside a ticket gate at a station. If the individual remembers the information about the restroom, he or she can use it later. However, otherwise, he or she cannot at all. The individual is strongly affected by whether he or she remembers such information. It is necessary to build an assistive system that helps visually impaired individuals remember information related to spots to visit.

There are several systems that help visually impaired individuals remember the information on everyday environments. The Digital Sign system [3] and the NAVI system [4] determine the current positions of visually impaired individuals by use of passive and AR markers, respectively. These systems need to deploy markers in large scale infrastructure for everyday use. A navigation system [2] determines the current position of a visually impaired individual by use of GPS, and then guides the individual along the predefined route. This GPS-based system cannot be used in, for example, reinforced concrete buildings. Sekai

Camera [5] is an AR application on mobile phones. Digital information, called Air Tag, can be virtually attached to the real world, and a user can know the local information from the Air Tag. The main targets of this system are sighted people, and thus it is difficult for visually impaired individuals to use this system. e.Typist Mobile [6] converts characters in environments to voices. Tap Tap See [7] and LookTel Recognizer [8] help visually impaired individuals to identify objects.

In this paper, we propose the concept of spot navigation to help visually impaired individuals remember information related to spots to visit. This concept is implemented as an application software on a mobile system, which is applied to actual indoor and outdoor scenes.

2 Concept of Spot Navigation

Spot navigation is a framework to assist visually impaired individuals in recalling the memories related to spots that they often visit. In this framework, first, a visually impaired individual visits several spots with a sighted supporter, and then records the position data and voice memos of the spots onto a mobile system. Later, when the visually impaired individual visits one of the recorded spots, the mobile system determines the spot position and then plays the voice memo that corresponds to the spot. The visually impaired individual can obtain the spot information by hearing the voice memo.

3 Implementation of a Spot Navigation System

3.1 Spot Navigation System

The spot navigation concept is implemented as an application software on an Android smartphone system (Google Nexus 4 [9]) and our Kinect cane system [10–12]. The application software has the following two modes:

1. **Registration mode:** A visually impaired individual visits each spot with his or her sighted supporter. The individual takes scene images with several perspectives by use of a camera on a mobile system, and records a voice memo about supplemental information related to the spot. The system registers the images and the voice memo to a dictionary in the system. The images are used as keys in the dictionary to determine the spot positions.
2. **Spot navigation mode:** When the visually impaired individual visits one of the recorded spots, he or she takes a scene image, and then inputs the image as a query into the system. The system determines the current spot from the results of image matching between the query image and the dictionary images. The system plays the voice memo corresponding to the matched dictionary image.

3.2 Image Matching Based on the SIFT

The Scale Invariant Feature Transform (SIFT) [13] can extract pixels that have distinct features, which are described by 128-dimensional vectors. The feature vectors are invariant against the changes of scale, rotation and illumination. Such pixels are called *key points*.

Let k^q and k^d denote key points in a query image q and a dictionary image d , respectively, and v_i^q and v_i^d denote the i -th feature value of k^q and k^d , respectively. The system searches for the key point pair which minimizes the following distance:

$$\delta(v^d, v^q) = \sqrt{\sum_{i=1}^{128} (v_i^d - v_i^q)^2}. \quad (1)$$

Fig. 1 shows a matching result where lines represent the key point pairs.

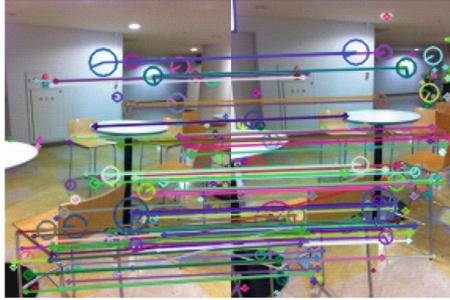


Fig. 1. Example of key point pairs in a query image and a dictionary image at an indoor spot.

The system evaluates the following six criteria based on geometrical relations between key point pairs:

1. too few pairs,
2. size consistency,
3. direction consistency,
4. 2D affine constraint,
5. area size and,
6. axis inversion,

which are proposed by Kameda et al. [14, 15]. If all the criteria are satisfied, the query image is determined to correspond to the dictionary image.

4 Experimental Results

4.1 Image Matching Test 1

Conditions: 22 indoor spots and 22 outdoor spots were selected for the experiment, and eight images were taken at each spot. The resolutions of the images were 144×192 pixels. A two-fold cross-validation test were employed with the following parameters: $K_0 = 10$, $t_{size} = 0.35$, $t_{dir} = 45$, $t_{affine} = 12$, $t_{area} = 15$, which are used in the image matching.

Results: Table 1 lists the accuracy of the image matching. *Correct detection* represents a situation where the system can find a dictionary image that corresponds to a query image. *False detection* represents a situation where the system mistakenly selects a dictionary image that does not correspond to a query image. *No correspondence* represents a situation where the system determines that the dictionary does not include any images that correspond to a query image. In this test, no correspondence is the failure of the image matching, because the image dictionary includes at least one image corresponding to a query image.

In this result, the 317 query images were matched successfully, but the 35 query images cannot be matched. Figure 2 (a), (b) and (c) show the matching results where the right and left images are dictionary and query images, respectively. Figure 2 (a) shows the images which are taken at the same indoor spot. 25 key point pairs are obtained correctly from the same objects. All the criteria are satisfied (i.e. $K = 25$, $E_{size} = 0.05$, $E_{dir} = 2.4$, $E_{affine} = 5.5$ and $E_{area} = 38.8$), and thus the system successfully determines the spot. Figure 2 (b) shows the images which are taken at different spots, and no key point pairs are obtained from the images. The number of the key point pairs is smaller than the threshold K_0 , and thus the system successfully determines that they are the different spots. Figure 2 (c) shows the images which are taken at different spots. 19 key point pairs are obtained, but the size consistency, the direction consistency and the axis inversion are out of permissible ranges. Therefore, they are successfully determined to be different spots.

4.2 Image Matching Test 2

Conditions: We verified whether the system correctly returns no correspondence in cases where the dictionary does not include any images corresponding

Table 1. Matching accuracy in Test 1.

	# of images	Ratio
Correct detection	317	90%
False detection	0	0%
No correspondence	35	10%

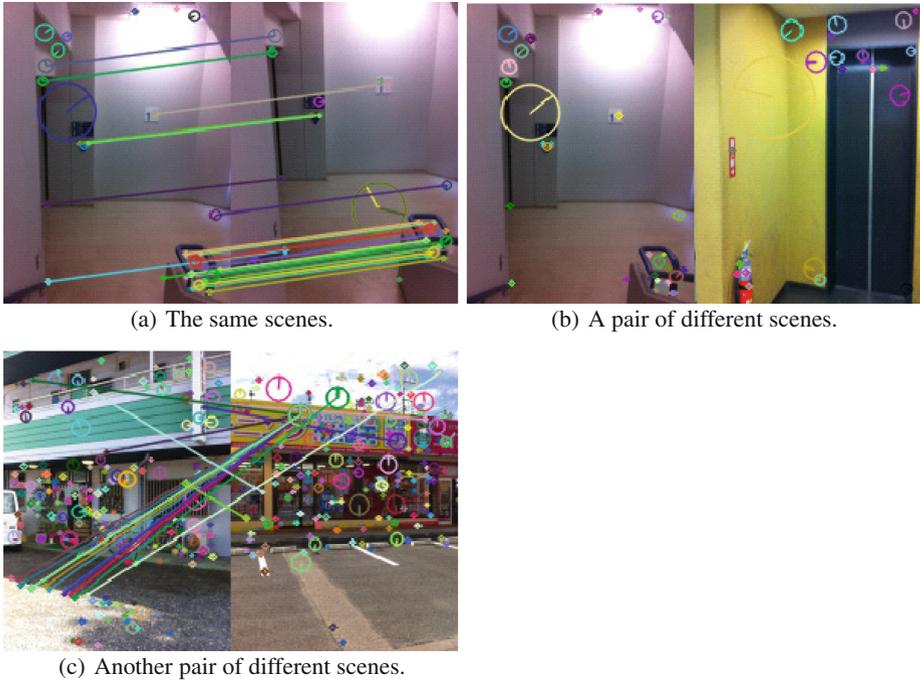


Fig. 2. Matching results in Test 1.

Table 2. Matching accuracy in Test 2

	# of images	Ratio
False detection	0	0%
No correspondence	28	100%

to a query image. 28 images were taken at other spots (14 indoor spots and 14 outdoor spots), and the image matching were performed by use of the 28 images and the dictionary images.

Results: Table 2 lists the matching results. In this test, no correspondence represents that the system successfully indicates that there is no dictionary images corresponding to a query image. All the query images are correctly determined to be no correspondence.

Figure 3 (a) and (b) show matching results. Figure 3 (a) shows the images which are taken at different outdoor spots. The system successfully determines that they are different spots, because the size consistency and the direction consistency are out of permissible ranges. Figure 3 (b) shows the images which are taken at indoor and outdoor spots. These images provide only two key point



(a) A pair of different scenes.

(b) Another pair of different scenes.

Fig. 3. Matching results in Test 2.



(a) Registration 1.



(b) Registration 2.



(c) Registration 3.



(d) Spot Navigation.

Fig. 4. Registration and spot navigation modes in user study.

pairs. The number of the pairs are smaller than the threshold K_0 . Therefore these spots are successfully determined to be different spots.

4.3 User Study

We conducted a user study where a blindfolded subject used an Android smartphone system in which the spot navigation method was implemented. In Fig. 4(a), a person having a white cane played a role of a visually impaired individual, and a person in a white T-shirt played a role of a supporter. They were in an entrance of a building. The supporter set a smartphone, took an image of the scene, and input a voice memo, “*Here is an entrance. There is a direction board.*” In Fig. 4(b), they were in front of a multi-purpose room. The supporter input a voice memo, “*Here is a multi-purpose room. There is a kitchen inside.*” In Fig. 4(c), they were in front of our laboratory. In this case, the visually impaired individual input a voice memo, “*Here is our laboratory*”, under the advice of the supporter. Figure 4(d) shows a situation where the visually impaired individual comes to the multi-purpose room by himself. He took an query image, and then the smartphone system played the correct voice memo. By hearing the voice memo, he could remember the kitchen in the multi-purpose room.

5 Conclusion

In this report, we propose a spot navigation system to assist visually impaired individuals in recalling the memories related to spots that they often visit. The system can identify spots by use of the image matching technique based on the SIFT, and give a visually impaired individual the supplemental information related to spots by use of voice memos. The experimental results indicate that the proposed system is promising to help visually impaired individuals.

One of our future works is improve the accuracy of the image matching by use of, PCA-SIFT [16], BSIFT [17], and CSIFT [18].

Acknowledgment. This work was supported in part by the JSPS KAKENHI Grant Number 25560278.

References

1. World Health Organization, Media Centre, Visual impairment and blindness, Fact Sheet No 282. <http://www.who.int/mediacentre/factsheets/fs282/en/>
2. Gaude, M., Candolkar, V.: GPS Navigator for Visually Impaired. Int. J. Electron. Signals Syst. (IJESS), vol-2 (2012), ISSN: 2231-5969, ISS-2,3,4
3. Legge, G.E., Beckmann, P.J., Tjan, B.S., Havey, G., Kramer, K., Rolkosky, D., Gage, R., Chen, M., Puchakayala, S., Rangarajan, A.: Indoor navigation by people with visual impairment using a digital sign system. PLoS One **8**(10), e76783 (2013)

4. Zöllner, M., Huber, S., Jetter, H.-C., Reiterer, H.: NAVI – a proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 584–587. Springer, Heidelberg (2011)
5. Sekai Camera Support Center BEYOND REALITY. <http://support.sekaicamera.com/ja/service>
6. e.Typist Mobile. MEDIA DRIVE CORPORATION. <http://mediadrive.jp/products/etmi>
7. TapTapSee - Blind and Visually Impaired Camera. TapTapSee. <http://www.taptapseeapp.com>
8. LookTel Recognizer Documentation. LookTel. <http://www.looktel.com/recognizer-documentation>
9. Nexus 4 - Google. <http://www.google.co.jp/nexus/4/>
10. Takizawa, H., Yamaguchi, S., Aoyagi, M., Ezaki, N., Mizuno, S.: Kinect cane : an assistive system for the visually impaired based on three-dimensional object recognition. In: The Proceedings of the 2012 IEEE/SICE International Symposium on System Integration, vol. 1, No. 1, pp. 740–745 (2012)
11. Takizawa, H., Yamaguchi, S., Aoyagi, M., Ezaki, N., Mizuno, S.: Kinect cane : object recognition aids for the visually impaired. In: The proceedings of the 6th IEEE International Conference on Human System Interaction (HSI 2013), 6 p. (CDROM proceedings) (2013)
12. Orita, K., Takizawa, H., Aoyagi, M., Ezaki, N., Mizuno, S.: Obstacle detection by the kinect cane system for the visually impaired. In: 2013 IEEE/SICE International Symposium on System Integration (SII 2013), pp. 115–118 (CDROM proceedings) (2013)
13. Hironobu, F.: Gradient-based feature extraction : SIFT and HOG. In: PRMU, CVIM 160, pp. 211–224 (2007)
14. Kameda, Y., ohta, Y.: Image retrieval of first-person vision for pedestrian navigation in urban area. In: ICPR, pp. 364–367 (2010)
15. Kurata, T., Kouroggi, M., Ishikawa, T., Kameda, Y., Aoki, K., Ishikawa, J.: Indoor-outdoor navigation system for visually-impaired pedestrians: preliminary evaluation of position measurement and obstacle display. In: Proceedings of ISWC 2011, pp. 123–124 (2011)
16. Ke, Y., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511–517 (2004)
17. Stein, A., Herbert, M.: Incorporating background invariance into feature-based object recognition. In: Proceedings of IEEE Workshop on Applications of Computer Vision (WACV), pp. 37–44, January 2005
18. Abdel-Hakim, A.E., Farag, A.A.: CSIFT: a SIFT descriptor with color invariant characteristics. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1978–1983 (2006)