Navilog: A Museum Guide and Location Logging System Based on Image Recognition

Soichiro Kawamura, Tomoko Ohtani, and Kiyoharu Aizawa

The University of Tokyo, 7-3-1 Hongo Bunkyo Tokyo 113-8656, Japan
(kawamura,fritz.tmk,aizawa)@hal.t.u-tokyo.ac.jp

Abstract. We developed a computer vision-based mobile museum guide system named "Navilog". It is a multimedia application for tablet devices. Using Navilog, visitors can take a picture of exhibits, and it identifies the exhibit and it shows additional descriptions and content related to it. It also enables them to log their locations within the museum. We made an experiment in the Railway Museum in Saitama, Japan.

1 Introduction

Recently, information technology for museum navigation is advancing: Wi-Fi based localization is utilized in some museums, and the visitors can easily choose the exhibits using their smartphones, and read the additional information. However, the museum needs to implement additional equipment that is many Wi-Fi stations, for the localization.

We made a system named "Navilog" which works on small tablet devices. The functions of system are identifying exhibits by image processing and showing additional information for users. Navilog performs those tasks by analyzing photos of exhibits taken by users.

2 A Museum Guide Prototype

In our system, the tablet device performs all image processing. The image database of museum exhibits is not large, so it can be stored in storage of the mobile device. This enables to make search executable without network connection. Such standalone implementation makes operation easy. We can use existing high performance descriptors without addition of data compression.

We made a prototype (Fig. 1) and asked museum visitors to evaluate its usability. Guide device runs along following steps.

1. Take query photograph of the museum exhibit (by the user)
2. Select region of interest (by the user)
3. Detect interest points in query
4. Extract descriptors
5. Match feature of query with those in database

S. Li et al. (Eds.): MMM 2013, Part II, LNCS 7733, pp. 505–507, 2013.

First, the user takes photo of exhibition by camera in a tablet. Photos in database we took when the museum is closed, so these photos contain only exhibit without visitors. However, the photos taken by users are likely to include other visitors. Then, the query photo might not match with those in the database. To resolve this problem, the user trims the exhibit region of the photo. After photo is taken, the device shows its picture on the display and the user interactively selects the region of interest. The user marks the exhibit using a touch operation (Fig. 2) and the system determines its bounding box of this region.

The system searches for the query image in the database. The pictures in the database have been taken from various positions and angles for each exhibit.

First, the system detects interest points in the query image. We use oFAST detector in ORB[1]. It is multi-scale fast detection algorithm.

Next, it extracts the feature descriptor around the interested points. We use SURF[2] which uses box filter and integral image, and is faster than SIFT. SURF describes image patch as 64 dimensional vectors. One query image has some hundred interested points. So it is computationally intensive to match descriptors of the query with these from database.

We use Bag-of-Features[3] method. We made "Visual word" vectors from images in the database. The pictures in database are converted to visual words histograms. These histograms are stored in the storage in the tablet device.

When the query image is shot, the system calculates its BoF histogram and compares those in database. Feature vectors common in most images are useless for image categorization so these should have less emphasis. *tf-idf* is a weighting method based on frequency.

When histograms are compared, these are weighted by *tf-idf* and then compared by L1 norm. System sorts these results in order of distance, allows the user to select proper one (Fig. 3).

When a user selects an exhibit from the list of results, Navilog shows the corresponding contents (e.g. its title, factory and guide movie).

Taken photos are showed in thumbnail view. The viewer displays photos and brief captions of exhibits. In order to read detailed explanation, the user taps button on the bottom of photo (Fig. 4).

Fig. 1. Museum guide prototype "Navilog"

Fig. 2. Selection of region of interest

Fig. 3. Result of image search

Fig. 4. User's photos and descriptions

Fig. 5. Camera icons show visited

The database also has coordinates of exhibition in exhibition room. The system shows the user's location history on the museum map (Fig. 5).

3 Conclusion

We developed a new museum guide using image recognition and conducted a user study to evaluate it.

In this research, the system determines foreground and background by touchscreen operation and reduces computational cost as all operations are done in mobile device. It requires neither networks nor servers.

We intend to improve this system for smaller devices such as smartphones.

References

1. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571 (2011)
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. 110(3), 346–359 (2008)
3. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, vol. 2, pp. 1470–1477 (2003)