Crowd-Cloud Window to the Past: Constructing a Photo Database for On-Site AR Exhibitions by Crowdsourcing

Sohei Osawa[✉], Ryohei Tanaka, Takuji Narumi,
Tomohiro Tanikawa, and Michitaka Hirose

Graduate School of Information Science and Technology,
The University of Tokyo, 7-3-1 Hongo, Bunkyō, Tokyo 113-8656, Japan
{osawa,r_tanaka,narumi,tani,
hirose}@cyber.t.u-tokyo.ac.jp

Abstract. In this paper, we propose a crowdsourcing system that constructs a database for AR (Augmented Reality) contents by user generation for on-site AR exhibitions that runs on personal mobile devices. The AR exhibition systems superimpose images from the past onto the present scene in the photographed places. They require location information of past pictures and current pictures that were shot in same place at same angle. Previously, system designers had to prepare the contents by themselves because the estimation of photographed positions was an unautomated task. This is one reason why there were only a limited number of contents. We overcome this problem by implementing a system in which many users can participate in creating contents easily. Users can post necessary information for AR contents only by taking photos when the current scene matches the semi-transparent past image. Through experiments and workshops, we found that our system had many interesting and acceptable sides. For example, our application for looking for the photographed positions gives users entertainment akin to games. Although its user interface and user-motivating design require increased sophistication, our system could work well and gather valuable user-generated contents to provide a richer AR experience.

Keywords: User-generated contents · Digital museum · Mobile AR

1 Introduction

The number of AR applications has gradually increased with the rapid spread of mobile devices such as smartphones and tablets in recent years. The applications called on-site AR exhibition systems are among them. The AR exhibitions can superimpose past pictures onto the present scene in photographed places, which helps users perceive both the changed and unchanged parts of the scenery. This leads users to understand the history of the places more deeply [1, 2]. In our past research, we developed the AR exhibition system "Window to the Past," on which anyone can easily superimpose past pictures of certain places with their smartphones and tablets [3, 4]. This system works as follows. The present picture of the location is displayed semi-transparently in the

center of the screen as a reference. The video captured from the device camera is shown in the background. Users go to the location depicted in the semi-transparent picture and attempt to match the scenery between the semi-transparent reference picture and the live video while walking around. When the reference and captured images are sufficiently similar, the system overlays the past picture onto the current scene. As a result, users can see the past picture merged precisely onto the current scene. This system used only extracted feature points from both the reference image and that of the device camera, without the use of any other equipment or AR markers.

However, although this matching procedure is automated using computer vision technologies, it is difficult to specify the reference image automatically because the imaging conditions differ largely [4, 5]. In "Window to the Past," the reference was specified manually by pre-preparing both the past and present pictures from the same spot and using present pictures as a reference for image recognition. We discovered that it was time-consuming to identify the exact photographed place and camera angle from the past picture and to capture a photo in the present at the exact same location. Hence, the designers had to prepare pictures of all contents, and this is one of the bottlenecks to the enrichment of contents.

Thus, we add a new function to reduce the workload of the system administrators by crowdsourcing photo collection tasks among the users. In this function, users can create AR contents themselves and share their contents with many other users. We realize this process via easy operations. Users must identify the location of the past pictures and capture the current reference images. This will extend "Window to the Past" to a larger area with more data created by many users.

2   Related Works

Estimating the Photographed Places and Guiding Users There. The study of Kasada et al. [6] is an example of estimating photographed places only from past pictures. They constructed a method by which users looked for photographed locations and captured present images at the same position and angle as the past photos. First, users find objects that can be seen in both the old picture and the present web-camera capture and choose three correlating feature points. Users then are guided to the photographed position of the past picture estimated by the positional relations of these three points on the plane. When the live web-camera image changes according to the users' movements, the specified three points are tracked by using optical flow, and the direction of navigation and distance are calculated sequentially.

As another method to estimate the photographed position more exactly, the "On-site virtual time machine" by Nakano et al. [4, 7] generates a point cloud of the current target landscape in advance with Bundler [8]. Users manually associate more than eight points that are not changed between objects in the current point cloud and those in the past picture. The system then calculates the exact photographed position using the eight associations and guides users to the position.

However, these guiding systems that estimate photographed positions require users to perform more numerous and complicated operations. In some cases, users can find

the photographed position more quickly without using the systems. Users' operations in our system should be as simple as possible because we must encourage many users to use it proactively. In addition, as seen from the example in which many people enjoy "film-induced tourism [9]," "pop-culture tourism [10]," and so on, it is fun for them to find and visit the photographed positions themselves. In these trips, tourists often capture photos with the same composition as that of works. This method is called "rephotography" and has a long history [11]. It will motivate users to use our system spontaneously and continuously by giving the fun to the users. We wanted users' voluntary search for photographed positions to be available in our user-generated estimation system.

Solutions to Difficult Tasks for Computers by Crowdsourcing. For unautomated tasks, many studies have found that great success can be achieved crowdsourcing. The study of Luis et al. [12] succeeded in labeling images using computer games. Labeling images is an unautomated task for computers, but they solved this by crowdsourcing using a computer game. They created an online game called ESP Game in which two players provided words associated with each image and scored points if their words corresponded. Players' answers were accumulated in a large database and used to improve accuracy of labeling. In addition, Wikipedia is a classic example of user-generated contents composing an encyclopedia that features enormous articles by crowdsourcing. Many studies explain why people are motivated to contribute to the Wikipedia project. Kuznetsov analyzed the motivation of Wikipedia editors using the Value Sensitive Design (VSD) approach [13] and showed that they enjoy a sense of accomplishment, collectivism, and benevolence [14]. Nov analyzed motivational factors based on eight general motivations used extensively in research on open source software development and volunteers [15]. They revealed that fun and ideology enhance users' motivation. These studies show that user-generated contents would be successfully obtained only by substantial feelings, such as fun and a sense of accomplishment.

3 Design of the Proposed System

3.1 Overview

In this chapter, we introduce our implementation of the crowdsourcing system that constructs a database for AR contents by user generation for on-site AR exhibitions and runs on personal mobile devices such as smartphones or tablets. To be useful, the database for on-site AR exhibitions must contain present reference pictures of the past photographs, GPS information of the locations, and annotation of the pictures. The main required functions of our proposed system are as follows:

1. Posting Past Pictures (PPP)
2. Identifying Photographed Positions and Angles by crowdsourcing (IPPA)
3. Appreciating an On-site AR Exhibition (AOAE)

3.2   Procedure of Posting Past Pictures

The main function of PPP is to collect various and valuable images from many users. We thus create Web pages on which users can contribute past photos and view them freely (Fig. 1). They can also add comments to the photo annotations. On the right side of the page is the submission form through which users can contribute past photos under pseudonyms and comment on the photos. On the left side, thumbnails of the contributed past photos are listed. When the thumbnail is selected, the larger image is displayed along with the thumbnails of uploaded current photos that were taken from the same position and angle as the past one. Users can evaluate these uploaded photos on this page (Sect. 3.3 explains this in detail.).

Fig. 1. Web page screenshot for Posting Past Pictures

3.3   Procedure of Identifying Photographed Positions and Angles by Crowdsourcing

The main function of IPPA is to identify the photographed positions of contributed photos by crowdsourcing. We realized this function by creating an application and Web pages that help users identify photographed positions fairly easily (Figs. 2 and 3). The application enables users to capture pictures with compositions similar to the past pictures and
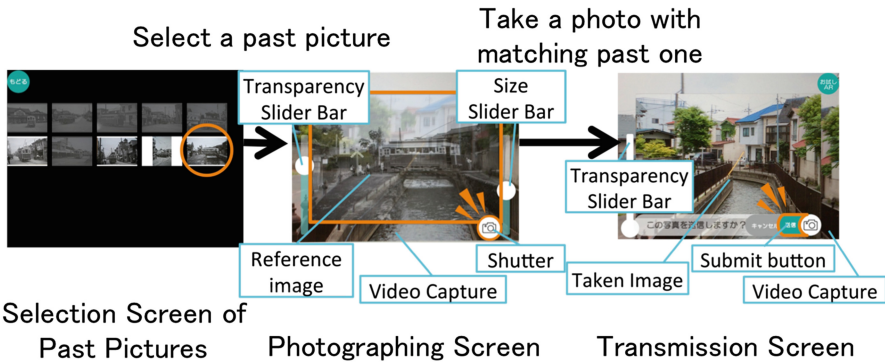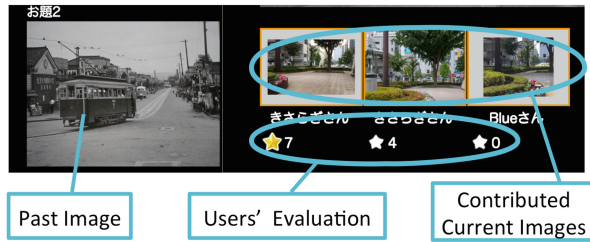

Fig. 2. Application to capture and collect current pictures at estimated photographed positions by users.

Fig. 3. Web pages to evaluate pictures captured by other users

post necessary data via simple operations. The Web pages enable users to evaluate how similar the pictures captured by other users are to the corresponding past photos.

Application to Capture and Collect Current Pictures at Estimated Photographed Positions by Users (Collection IPPA). This proposed system is realized as a mobile device application (Fig. 2). In this application, past photos uploaded by PPP are first displayed as a list. Users select a favorite from among them, and the selected picture is displayed semi-transparently in the center of the screen as a reference. The video captured from the device camera is shown in the background. Users explore the spots where old photos were captured. They need only to capture photos at the correct position and angle where the semi-transparent past image matches the background, and then they can easily add a series of necessary information for the AR exhibitions to our database. The slider bar on the right can change the size of the reference image to capture photos matching any of contributed photos under the settings of a device camera. The left slider bar can change the transparency of the reference image for easy comparison between the reference image and the background. It can be hoped that the users become familiar with the area and help revitalize the area by strolling around.

Web Pages to Evaluate Pictures Captured by Other Users (Evaluation IPPA). The present pictures collected via Collection IPPA have a risk of being associated with wrong locations, because the accuracy of the pictures are guaranteed only by the photographers themselves. It is necessary to evaluate the correctness of the posted present pictures, but automatic evaluation is difficult using feature points if current scenery has significantly changed from the past picture. We solve this problem also by users' power. We create the Web pages where users vote on pictures (Fig. 3). On these Web pages, the present pictures are listed with the corresponding past one, and each user can vote on which picture best matches the past picture. The present picture that receives the most votes is registered as a reference image. The image is used as a marker for AR exhibition, and its GPS data are also used for the corresponding past picture.

3.4 Procedure of Appreciating an On-Site AR Exhibition

The main function of AOAE is to appreciate past pictures superimposed onto current scenery at the photographed position. Thus, we extended "Window to the Past," on

which users can easily superimpose past pictures of certain places with smartphones and tablets (Fig. 4). This system works as follows. First, the places already registered by IPPA are indicated on a map. When users select one location, the present picture of the place is displayed semi-transparently in the center of the screen as a reference. The video captured from the device camera is shown in the background. Using slider bars, users can transform the size and transparency of reference image freely. Users visiting the place can try to match the camera image with the semi-transparent reference picture. When the reference and captured images are sufficiently similar, the system overlays the past picture onto the current scene. The similarity of the images is recognized by matching feature points based on ORB [16].
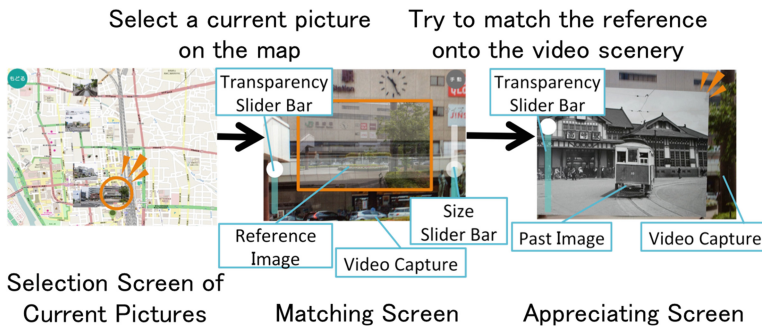

Fig. 4. Application for appreciating an on-site AR exhibition

4   Experiments

4.1   Experimental Purposes

This experiment aimed to evaluate IPPA. There are two processes in IPPA: Collection IPPA and Evaluation IPPA. In the first experiment, we judged the accuracy of the photos collected in Collection IPPA by the distance from the photographed location to the correct location and by visual inspection. The participants actually used the Collection IPPA and the posted present pictures, which were evaluated in terms of how well they matched with correct photographed positions and angles of the past pictures.

In the second experiment, we verified whether the users are able to select accurate photographed position photos from the various collected photos in Evaluation IPPA. The participants responded with the quantitative matching degree of their photographed points on a range of 0 to 1. This experimental result would suggest how well users can evaluate the photographed position of the images by comparing the current image and the past image.

4.2   Experiment 1: Evaluation of Collection IPPA

Detailed Procedures. We prepared 15 past images taken by various-angle cameras at the University of Tokyo Hongo Campus. Their accurate photographed positions were already known. The participants used Collection IPPA to post present pictures that match past images for 1 h. They then answered free description type questionnaires on the usability of Collection IPPA. Twelve participants in their twenties participated in the study. They were students at the university and were slightly acquainted with Hongo Campus. The posted present pictures were evaluated in terms of matching degree with the accurate photographed positions and angles of the original past pictures.

Result and Discussion. Figure 5 indicates an example in which we plotted some photographed points of uploaded images and past pictures on the map. The colors indicate the types of target past pictures used in the experiment. As the figure shows, the participants were able to capture images near the accurate positions. In 10 out of 15 pictures, there were some uploaded images in which the distance from the photographed location to the correct location was within 2 m. In the range of 2 m, there was almost no difference in the appearance of current scenery because most of the main objects in the past pictures used in this experiment were distant. All photos were captured at approximately the same camera angle, which could be sufficiently used as reference images for the AR exhibitions. In two out of 15 pictures, the uploaded images were within 5 m. One reason for this is that the participants had very little clue how to find the correct photographed positions because most of the main objects in the past pictures were hidden behind obstacles. There were only two cases that were not within 10 m. (red circle points in Fig. 5.) The two past pictures of these points were imaged by wide-angle cameras. Thus, the appearance of these uploaded pictures was more similar than those from accurate points (Fig. 6). The participants in this study were slightly acquainted with Hongo Campus as noted above. To sum up, users who knew a bit about the photographed place of the past picture were able to identify the locations and captured photos there. Moreover, with regard to the other place, nobody uploaded current pictures there. This occurred by accident because the participants used our application for their favorite pictures within the limited time in this experiment.

Fig. 5. Map plotting some photographed points of uploaded images and past pictures (Color figure online)
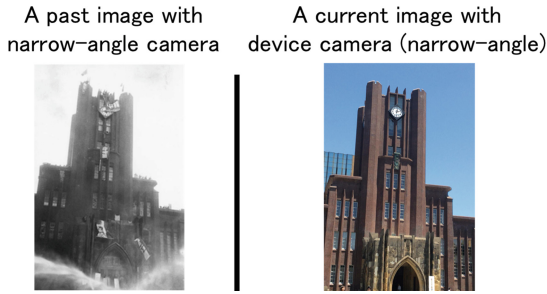
A past image with   A current image with
narrow-angle camera device camera (narrow-angle)

Fig. 6. Visual difference using narrow-angle cameras and device cameras

A past image with   Current images with
wide-angle camera   device camera (narrow-angle)

Taken image from a   Taken image from
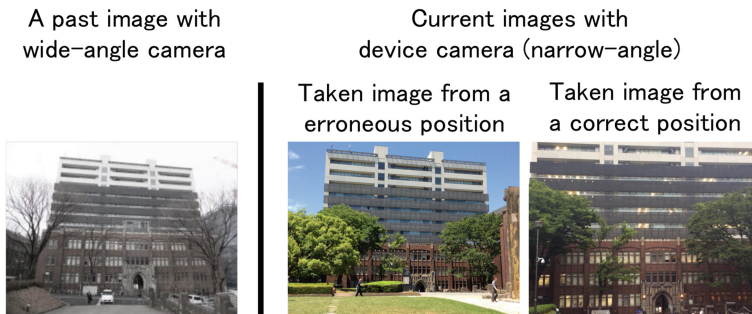erroneous position   a correct position

Fig. 7. Visual difference using wide-angle cameras and device ones

In addition, seven of the 12 participants mentioned that they did not have a sufficient angle of view (AOV). Past pictures were taken with wide-angle cameras, which cannot be imaged within the AOV of the mobile device cameras. (The AOV of mobile devices is approximately 42–48°.) Five of the 12 participants answered that they wanted to enlarge the reference images sizes. In this Collection IPPA, we limited the reference image size to the device's screen size. Users must be able to magnify reference sizes larger than the screen size, and the application must be usable for wide-angle pictures.

4.3   Experiment 2: Evaluation of Evaluation IPPA

Detailed Procedures. The participants provided quantitative matching degrees of photographed points between past and current images within a range of 0 to 1. We used images collected through Experiment 1 for this experiment. Nine participants in their twenties participated in the study. We obtained an average score of all participants' matching degree of the photographed position for each uploaded picture. We judged an evaluation as correct when the average value was over the 75 % difference limen. The result of users' evaluation is classified according to the three patterns above.

Result and Discussion. There are three patterns in the collections by Collection IPPA according to the photographed position and appearance. They are as follows:

- Pattern 1: The appearance is correct, and the photographed position is also correct.
- Pattern 2: The appearance is greatly different, but the photographed position is correct; this is caused by new, obstructing buildings or demolition or by a great difference in AOV.
- Pattern 3: The appearance is similar, but the photographed position is different owing to a great difference in AOV.

In the case of Pattern 1, the participants had high opinions of the images and achieved correct evaluation (Fig. 7). There are 20 uploaded pictures classified under Pattern 1. Eighteen of 20 were correctly evaluated. In the case of Pattern 2, the participants had low opinions of them and evaluated that they were not correctly taken (Fig. 8). There are 7 uploaded pictures classified under Pattern 2, all of which were evaluated incorrectly. In the case of Pattern 3, the participants had low opinions of them, and the pictures were correctly evaluated (Fig. 9). There are 5 uploaded pictures classified under Pattern 3, all of which were correctly evaluated (Fig. 10).
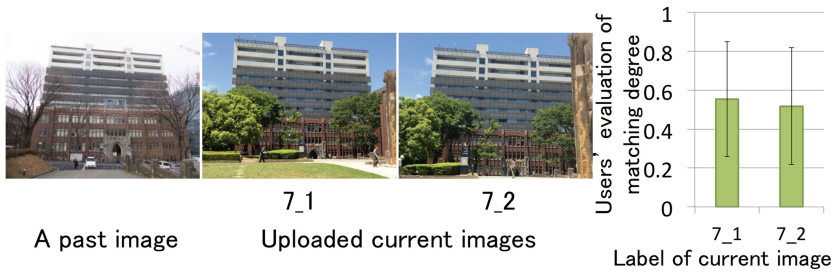

Fig. 8. The values of users' matching degree in the case of Pattern 1
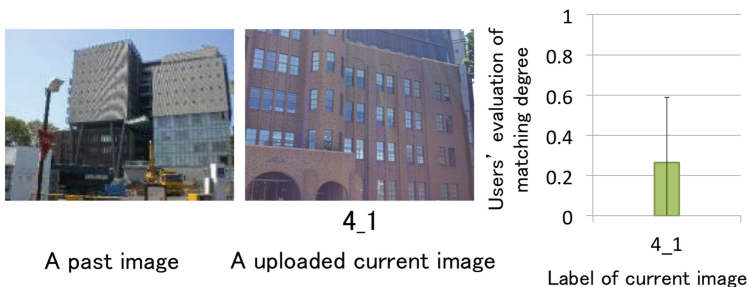
Fig. 9. The value of users' matching degree in the case of Pattern 2

If the appearance of uploaded images is significantly different from those of the corresponding past pictures such as in Pattern 2, the participants were unable to judge

Fig. 10. The values of users' matching degree in the case of Pattern 3

them accurately because very little information was available for making a decision. Therefore, if similar present images are uploaded in spite of greatly different appearances, we must increase the evaluation value of the images.

4.4   User Study in Workshops

We integrated "Window to the Past" and Collection IPPA into one application named "Crowd-Cloud Window to the Past." We then distributed this application, "Crowd-Cloud Window to the Past" through the Internet (http://nozokimado.org/). We held workshops using this application in cooperation with many local communities and collected questionnaires and hearing information to evaluate the interest and feasibility of this application. The participants enjoyed using our application in these workshops. We obtained some positive opinion such as follows:

- Searching for photographed positions and taking pictures were interesting, like games.
- I thought it was troublesome to find the photographed positions but surprisingly fun to try.
- I enjoyed appreciating past scenery superimposed onto current scenery.
- I want AR contents of my town.

The result suggests that our application is substantially acceptable. However, there were also some negative opinions as follows:

- I want to be motivated to go far.
- I have no motivation to look for places in a familiar area.
- The behavior of the application was a little heavy.
- Sometimes, I properly superimposed the reference image onto the scenery, but the AR contents did not start.

To maintain users' motivation, some techniques should be introduced such as gamification methods that reward users for the number of their uploaded positions. We also should sophisticate its user interface and system behavior to improve usability. Finally, we found that our application was slightly complex and time-consuming, but it had many interesting and acceptable aspects.

5   Conclusion

In this paper, we presented a crowdsourcing system that constructs a database for AR contents by user generation and applied it in "Window to the Past," which runs on personal mobile devices. In "Window to the Past," the system designers had to prepare the contents by themselves, so there were only a limited number of contents. We overcame this problem by implementing a system in which many users can participate in creating contents by crowdsourcing. This will extend "Window to the Past" to a larger area with more data created by many users.

We proposed a crowdsourcing system in which users identify the location of the past pictures and capture the current reference pictures. This system helps users generate contents for "Window to the Past" with enjoyment. The database contains current reference pictures of the past pictures, GPS information of the locations, and annotation of the pictures. Our proposed system is realized as a smartphone and a tablet device application. Users explore the spots where old photos were captured on behalf of the designer. They only need to take photos at the correct position and angle where the semi-transparent past image matches the background, and then they can easily add a series of necessary information for "Window to the Past" to our database. We integrate "Window to the Past" and this crowdsourcing database construction system into one application named "Crowd-Cloud Window to the Past."

Through our experiment, we evaluated this proposed crowdsourcing system. The results showed that users who knew a bit about the photographed place in the past picture were able to identify the locations and take photos there. In most uploaded pictures, the distance from the photographed place to the correct location was within 10 m. Moreover, all photos were taken at approximately the same camera angle, which could be sufficiently used as the reference images for "Window to the Past." If the appearance of the uploaded images was significantly different from the corresponding past pictures such as in Pattern 2, they were unable to judge accurately because there was very little information for making a decision. Moreover, users were unable to evaluate the correctness of the present pictures accurately if the appearance of the images was significantly different from the corresponding past ones. We should incorporate functions to judge the accuracy of the pictures.

In addition, we distributed this application, "Crowd-Cloud Window to the Past" through the Internet. We held workshops using this application in cooperation with many local communities and evaluated the interest and feasibility of this application. Here, we found that this application was slightly complex and time-consuming, but it had many interesting and acceptable aspects. For example, our application for looking for photographed positions provides users with entertainment, like games. Although we should sophisticate its user interface and user-motivating design, these results suggest that the proposed system could work well in gathering valuable user-generated contents to provide a richer AR experience.

References

1. Nakasugi, H., Yamauchi, Y.: Past viewer: development of wearable learning system for history education. In: Proceedings of the International Conference on Computers in Education, 2002, pp. 1311–1312. IEEE (2002)
2. Wither, J., Allen, R., Samanta, V., Hemanus, J., Tsai, Y.T., Azuma, R., Carter, W., Hinman, R., Korah, T.: The westwood experience: connecting story to locations via mixed reality. In: IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities (ISMAR-AMH 2010) (2010)
3. Okada, N., Imura, J., Narumi, T., Tanikawa, T., Hirose, M.: Manseibashi reminiscent window: on-site AR exhibition system using mobile devices. In: Streitz, N., Markopoulos, P. (eds.) DAPI 2015. LNCS, vol. 9189, pp. 349–361. Springer, Heidelberg (2015)
4. Nakano, J., Narumi, T., Tanikawa, T., Hirose, M.: Implementation of on-site virtual time machine for mobile devices. In: Virtual Reality (VR), March 2015, pp. 245–246. IEEE (2015)
5. Bae, S., Agarwala, A., Durand, F.: Computational rephotography. ACM Trans. Graph **29**(3), 1–15 (2010)
6. Kasada, K., Hayashi, O., Narumi, T., Tanikawa, T., Hirose, M.: On-site virtual time machine with navigation to past camera position. In: Proceedings of ASIAGRAPH 2010, Tokyo, Japan, vol. 4, no. 2, pp. 35–40 (2010)
7. Nakano, J., Kasada, K., Nishimura, K., Tanikawa, T., Hirose, M.: On-site virtual time machine-navigation to past camera position and past picture superimpose on present landscape. Trans. Inf. Process. Soc. Jpn. **52**(12), 3611–3624 (2011)
8. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. ACM Trans. Graph. (TOG) **25**(3), 835–846 (2006)
9. Beeton, S.: Film-Induced Tourism, vol. 25. Channel View Publications, Clevedon (2005)
10. Gyimóthy, S., Lundberg, C., Lindström, K.N., Lexhagen, M., Larson, M.: Popculture tourism. In: Tourism Research Frontiers: Beyond the Boundaries of Knowledge. Tourism Social Science Series, vol. 20, pp. 13–26. Emerald Group Publishing Limited (2015)
11. Webb, R.H.: Repeat Photography: Methods and Applications in the Natural Sciences. Island Press, Washington, D.C. (2010)
12. Von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, pp. 319–326. ACM, New York (2004)
13. Friedman, B.: Value-sensitive design. Interactions **3**(6), 16–23 (1996)
14. Kuznetsov, S.: Motivations of contributors to Wikipedia. ACM SIGCAS Comput. Soc. **36**(2), 1 (2006)
15. Nov, O.: What motivates wikipedians? Commun. ACM **50**(11), 60–64 (2007)
16. 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). IEEE (2011)