



Investigating Effects of Users' Background in Analyzing Long-Term Images from a Stationary Camera

Koshi Ikegawa^(✉), Akira Ishii, Kazunori Okamura, Buntarou Shizuki, and Shin Takahashi

University of Tsukuba, Tsukuba, Japan
{ikegawa,ishii,kokamura,shizuki,shin}@iplab.cs.tsukuba.ac.jp

Abstract. Images recorded over a long term using a stationary camera have the potential for revealing various facts regarding the recorded target. We have been developing an analyzing system with a heatmap-based interface designed for visual analytics of long-term images from a stationary camera. In our previous study, we experimented with participants who were recorded in the images (*recorded participants*). In this study, we conducted a further experiment with participants who are not recorded in the images (*unrecorded participants*) to reveal the discoveries that participants obtain. By comparing the results of participants with different backgrounds, we investigated the difference between discoveries, functions used, and analysis process. The comparison suggests that *unrecorded participants* could discover many facts about environment, and *recorded participants* could discover many facts about people. Moreover, the comparison also suggests that *unrecorded participants* could discover many facts comparable to *recorded participants*.

Keywords: Data visualization · Big data management
Information presentation · Heatmap · Surveillance system
Visual analytics · Lifelog

1 Introduction

Images recorded over a long term using a stationary camera have the potential for revealing various facts regarding the recorded target. For example, if a department store installs a stationary camera that produces aerial images of a floor, the recorded images could contain useful data for evaluating the layout of the floor. However, it is difficult to view such images in their entirety, because the speed at which the images are replayed must be sufficiently slow for the user to comprehend them, and thus it is difficult to obtain valuable information from the images quickly.

To address this problem, we developed an analyzing system [7], which has a heatmap-based interface designed for visual analytics [11, 13] of long-term images



Fig. 1. An image from our omni-directional camera.

from a stationary camera (Fig. 1). This system provides heatmaps, each of which summarizes the movement of people and objects during a specific term, allowing the user to analyze the recorded target by comparing heatmaps of two different intervals. In our previous study, we experimented with participants who were recorded in the images (*recorded participants* [4]).

In this study, we conducted a further experiment with participants who are not recorded in the images (*unrecorded participants*) to reveal the discoveries that participants obtain. Furthermore, we compared the results with those of our previous study. The comparison suggests that *unrecorded participants* could discover many facts about environment, and *recorded participants* could discover many facts about people. Moreover, the comparison also suggests that *unrecorded participants* could discover many facts comparable to *recorded participants*.

2 Related Work

Heatmap-based visualization for analyzing images from a stationary camera have been explored. Romero et al. proposed Viz-A-Vis [8] and evaluated it [9]. While their visualization is different, we use their evaluating method as a reference. Heatmaps were also used in Viz-A-Vis, which are 3D heatmaps in which time is represented in the third axis to provide a spatial and temporal abstraction of a video. Kubbat et al. proposed TotalRecall [5] that focused on transcribing and adding annotations based on audio and video recorded at the same time for a 100,000 h. While the visualization of the above systems is similar to ours, our system focuses on discovering by comparing two different terms using 2D heatmaps, each of which summarizes the movement of people and objects during a specified term.

Various systems for analyzing images from a stationary camera have explored visualization techniques other than heatmaps. DeCamp et al. proposed HouseFly [2], a system that visualizes the entire floor, consisting of several rooms, as one 3D representation by mapping the image of the camera attached to the ceiling of each room as a texture on the floor plan. In addition to video browsing, this system serves as a platform for visualizing patterns of a multi-modal data over time, including person tracking and speech transcripts. Chiang and Yang [1] proposed a browsing system to help the users quickly locate desired targets in surveillance videos. Their basic idea is to collect all moving objects, which carry the most significant information in surveillance videos to construct a corresponding compact video. Shin et al. proposed VizKid [10] that visualizes the position and orientation of an adult and a child as they interact with one another in one graph.

Many studies have explored interfaces for browsing videos from non-stationary cameras. Nguyen et al. proposed Video Summagator [6] that visualizes a video in 3D, allowing a user to look into the video cube for rapid navigation. Tompkin et al. [12] proposed a video browsing system that embeds videos into a panorama, allowing the users to comprehend the videos within its panoramic contexts across both the temporal and spatial dimensions. Higuchi et al. proposed EgoScanning [3], a video fast-forwarding interface that helps users find important events from lengthy first-person videos continuously recorded with wearable cameras.

In contrast to the above work, our visualization is designed to help the users analyze the images from one stationary camera on the ceiling by providing the summarization of the movement of people and objects in the images as a heatmap.

3 Implementation

This section describes the system used in this experiment. Our system consists of a recording system and an analyzing system. The recording system obtains images from a stationary camera that was mounted on the ceiling of the authors' laboratory room, preprocesses the images for the generation of a heatmap, and stores the images to a network attached storage (NAS). The analyzing system generates heatmaps using the images stored in the NAS and presents the heatmaps to users.

3.1 Recording System

Our recording system stores the images, with a 608×608 pixel spatial resolution at 1 fps, from an omni-directional camera (Sharp Semiconductor LZ0P3551) mounted on the ceiling of our laboratory as shown in Fig. 2. This frame rate is the one frequently used in the video archives of surveillance systems and produces 86,400 frames per day. The images are stored in a NAS (QNAP TS-859

and TS-859 Pro+). The recording system (Fig. 3) runs on a laptop computer (MacBook Pro 13-in. Late 2011 and MacBook Pro Retina 15-in. Early 2013).

We installed two sets of the above system in our laboratory's two rooms (Room-A and Room-B) with the sizes of approximately $7.50\text{ m} \times 7.75\text{ m}$ (58 m^2) and $7.5\text{ m} \times 15.0\text{ m}$ (113 m^2), and heights of approximately 2.5 m and 2.7 m, respectively.

3.2 Analyzing System

Our analyzing system generates heatmaps using the images stored on NAS and presents the heatmaps to users. Figure 4 shows our analyzing system. It consists of Image-Presenting Panel, Time-Operation Panel, and Heatmap-Operation Panel.

Image-Presenting Panel displays a camera image (Fig. 4A). The users can select a part (Fig. 4B) of the image for further analysis. Time-Operation Panel allows the users to specify the date (Fig. 4C) and time (Fig. 4D) of the camera image. The system colors the calendar (Fig. 4C) and the date slider (Fig. 4D) blue with the depth which shows the amount of the movement in the selected part. Using this function, the users can identify the terms to be analyzed. Heatmap-Operation Panel controls two colors of heatmaps; our system can display two different heatmaps with two different colors, each of which can be turned on/off using the two checkboxes (Fig. 4E) on Heatmap-Operation Panel. Turned-on heatmaps are overlaid on the camera image. The users can specify the terms for the two heatmaps using the term selectors (Fig. 4F).

Our heatmap summarizes the movement of people and objects in the specified term based on the pixel changes of the camera images: the more movement there



Fig. 2. Omni-directional camera mounted on the ceiling of our laboratory.

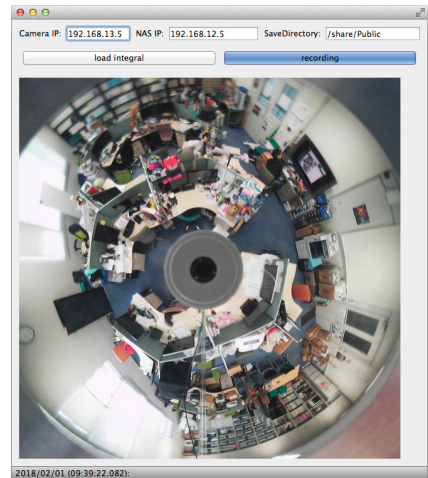


Fig. 3. Recording system.

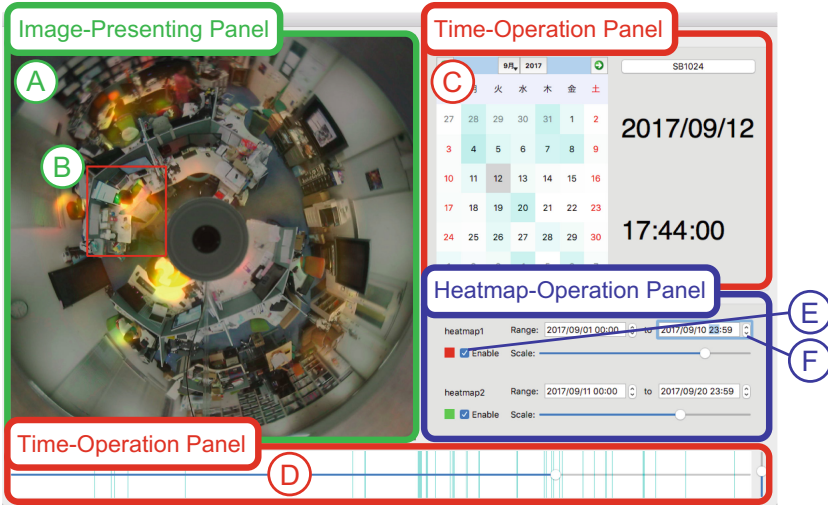


Fig. 4. Analyzing system using heatmaps. (Color figure online)

is within the specified term, the more densely the pixel is colored. Therefore, the heatmap allows the users to recognize areas with little movement and areas with much movement within the specified term at a glance. Moreover, the users can compare movements in different terms by using two heatmaps.

4 Experiment

We conducted an experiment to examine what and how users obtain discoveries by each function provided by our analyzing system. In addition, we used the experimental results of *recorded participants* in the previous study for comparison.

4.1 Participants

Four new *unrecorded participants* (three males, one female) aged 18 to 22 years were recruited for the experiment. The age of *recorded participants* (three males, one female) in the previous study was 22 to 23 years. None of the participants had previously used our system, nor did they have prior knowledge regarding our system.

4.2 Experimental Environment

We used a MacBook Air 13-in. Mid 2011 (CPU: 1.7 GHz Intel Core i5, RAM: 4 GB, OS: macOS 10.12.6) as the computer for running our analyzing system. For high-speed communication, the computer and the NAS connected to the local

network using a wired LAN cable. The experiment was conducted in a calm office environment. We recorded the whole experiment with a video camera, voice recorder, and screen capture software (QuickTime Player 10.4). In the experiment, two authors were sitting near the participants. One author recorded the participants' remarks and the other author addressed the question of the participants.

4.3 Procedure

First, we informed the participants of the purpose and the procedure of the experiment. We then informed the participants that the reward for participation in the experiment included not only a basic reward (820 JPY, 7.78 USD) but also a bonus depending on the number of discoveries (30 JPY per discovery). After the explanation, we explained the use of our analyzing system to the participants. We then asked them to engage with the system until they felt that they completely understood how to use it, as a practice. We used the Room-A images for the practice, in which *recorded participants* were not recorded. For reference, we showed the seating chart of the two rooms to the participants (Fig. 5).

In the analyzing task, we asked the participants to use our analyzing system for 30 min, during which they should attempt to make as many discoveries as possible and inform the experimenters of each of these using think aloud protocol. In addition, we asked the participants to inform the experimenters of the facts that led them to each discovery (e.g., the color of the heatmap is dense in certain areas, as described in Sect. 5.3). In this analyzing task, we used the images of Room-B that were recorded over six months (July 1, 2014 to December 31, 2014; 4,416 h; approximately 32 million images) using the stationary camera.

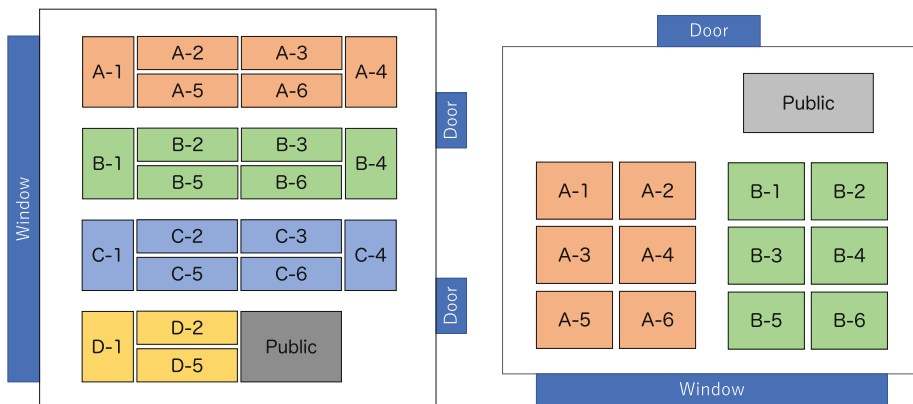


Fig. 5. Seating chart: (left) Room-A, (right) Room-B.

After the analyzing task was completed, we asked the participants to answer a questionnaire related to our system. The experiment took approximately 60 min in total.

5 Results

In this section, we show the results of the experiment and discuss them. First, we classify discoveries by five properties and five functions and compare them based on participants' background. Next, we examine the results of questionnaires. Finally, we compare the analysis process based on participants' background.

5.1 Discoveries and Classification

As a result, *unrecorded participants* (P1–P4) had 32.25 discoveries on average. In the previous study, *recorded participants* (P5–P8) had 24.00 discoveries on average. As the normalities of the results were satisfied with Shapiro-Wilk tests, we compared the number of discoveries using t-test. As a result of t-test, there was no significant difference ($t = 1.02$; $df = 6$; $p = 0.38 > 0.05$). This result suggests that *unrecorded participants* could discover many facts comparable to *recorded participants* when using our system.

We classified the discoveries by the following five properties (Figs. 6 and 7):

1. **Overviewing.** Discoveries obtained by paying attention to the entire image.
2. **People.** Discoveries obtained by paying attention to the people in recorded images.
3. **Environment.** Discoveries obtained by paying attention to the environment (e.g., objects in recorded images and changes in the appearance of the room).
4. **Suggestion.** Opinions/discoveries that are related to the analyzing system (e.g., requests to extend functionality, proposals of a new function, and ideas to improve the system).
5. **Other.** Additional opinions/discoveries (e.g., suggestions for applications of our system).

We also classified the discoveries by the five functions of our analyzing system by examining which function the participants used when they obtained discoveries (Figs. 8 and 9):

1. **Heatmap/all (HM/all).** Discoveries obtained by paying attention to the whole heatmap (Fig. 4A).
2. **Heatmap/part (HM/part).** Discoveries obtained by paying attention to a part of the heatmap (Fig. 4B).
3. **Calendar.** Discoveries obtained by paying attention to the color of the calendar (Fig. 4C).
4. **Time slider.** Discoveries obtained by paying attention to the color of the time slider or comparing different images by operating the time slider (Fig. 4E).
5. **Camera image.** Discoveries obtained by paying attention to the camera image (Fig. 4A).

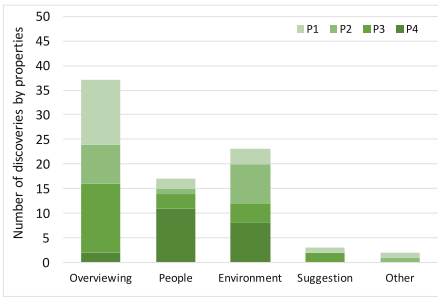


Fig. 6. Results of classifying discoveries of *unrecorded participants* by properties.

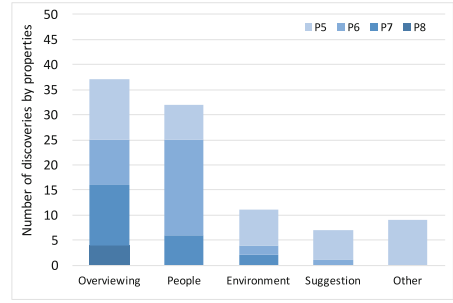


Fig. 7. Results of classifying discoveries of *recorded participants* by properties.

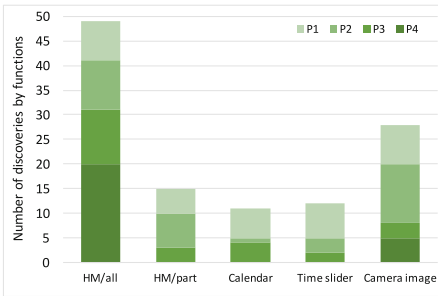


Fig. 8. Results of classifying discoveries of *unrecorded participants* by functions (breakdown by participants).

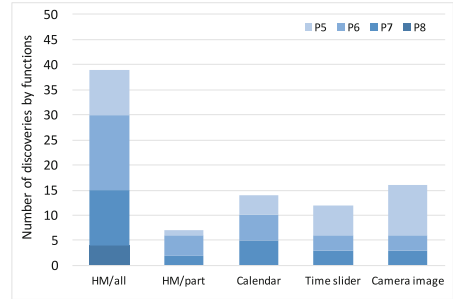


Fig. 9. Results of classifying discoveries of *recorded participants* by functions (breakdown by participants).

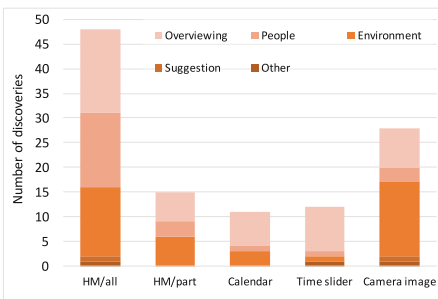


Fig. 10. Results of classifying discoveries of *unrecorded participants* by functions (breakdown by properties).

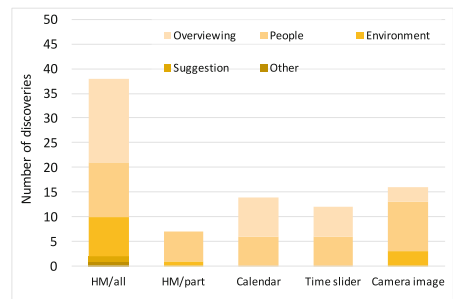


Fig. 11. Results of classifying discoveries of *recorded participants* by functions (breakdown by properties).

Comparison between Figs. 6 and 7 shows that *unrecorded participants* discovered many facts about Environment. On the contrary, *recorded participants* discovered many facts about People. In the experiment with *recorded participants*, most of the people in the images were acquaintances of the participants. This would draw the participants' attention to the people, resulting in more discoveries about People. Conversely, in the experiment with *unrecorded participants*, there were no acquaintances in the images. This would make the participants observe the whole images without focusing on specific people, resulting in more discoveries about Environment. We also examined the result using statistical analysis. As the normalities of the results were not satisfied with Shapiro-Wilk tests, we compared the number of discoveries by properties using Mann-Whitney U test. However, there was no significant difference (Overviewing: $p = 0.8 > 0.05$; People: $p = 0.69 > 0.05$; Environment: $p = 0.09 > 0.05$; Suggestion: $p = 1 > 0.05$; Other: $p = 1 > 0.05$); this would be because the number of participants in this experiment was small.

Moreover, there is not much difference between Figs. 8 and 9. Due to the normalities of the results were not satisfied with Shapiro-Wilk tests, we compared the number of discoveries by functions using the Mann-Whitney U test. As a result of test, there was no significant difference (HM/all: $p = 0.74 > 0.05$; HM/part: $p = 0.37 > 0.05$; Calendar: $p = 0.89 > 0.05$; Time slider: $p = 0.97 > 0.05$; Camera Image: $p = 0.31 > 0.05$). This result suggests that both participants could use the functions of our system regardless of the participants' background. The two figures also suggest that usage proportion of each function of our system has the same tendency regardless of the participants' background. Because the usage proportion has the same tendency, the cause of the difference in the previous paragraph might be the difference in participants' background.

As shown in Figs. 10 and 11, both participants discovered many facts about Overviewing, People, and Environment using the HM/all function. Compared to *recorded participants*, *unrecorded participants* discovered many facts about Environment using the HM/part and Camera Image functions. Conversely, *recorded participants* discovered many facts about People using the Calendar, Time slider, and Camera Image functions. Therefore, these results suggest that participants with different backgrounds tend to discover facts with different properties. The videos of the whole experiment support this observation. We found in the videos that *recorded participants* enjoyed looking back on and analyzing their research days since they often smiled and laughed sometimes, while *unrecorded participants* performed their tasks in a businesslike manner.

5.2 Questionnaire Results

We asked all participants to answer a questionnaire that consisted of two questions: “*did you use our system with ease?*” (Q1) and “*do you want to use our system in the future?*” (Q2). Each question included a five-point Likert scale form (1 = strongly disagree, 5 = strongly agree) and a comment form. The results of Q1 and Q2 for *recorded participants* and *unrecorded participants* are shown in Figs. 12, 13, 14 and 15.

As shown in Figs. 13 and 15, the scores provided by P6 were lower than the others. While P6 skillfully used all functions, he stated that “the analyzing was fun for me, but I cannot imagine application examples” in the questionnaire. Furthermore, P6 provided many requests for extending the functionality, proposals for new functions, and ideas to improve our system.

As shown in Figs. 12 and 13, *unrecorded participants* had similar results compared to *recorded participants*; this means that participants in both backgrounds were neutral to the usability of the analyzing system. In addition, P3 and P5 mentioned complaints about the processing speed of the analyzing system. Therefore, if the processing speed improves, the usability would improve.

As shown in Figs. 14 and 15, we found that there were many participants who wanted to use our system in the future. In addition, P2, who is a college student studying biology, commented that our system could be used to observe and analyze places where animals gather.

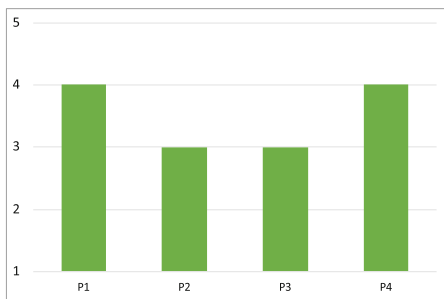


Fig. 12. Answer to the question “*did you use our system with ease?*” by *unrecorded participants* (mean = 3.50; SD = 0.50).

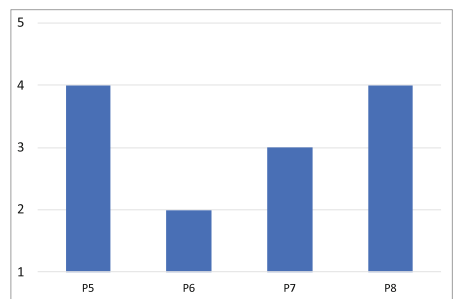


Fig. 13. Answer to the question “*did you use our system with ease?*” by *recorded participants* (mean = 4.00; SD = 0.71).

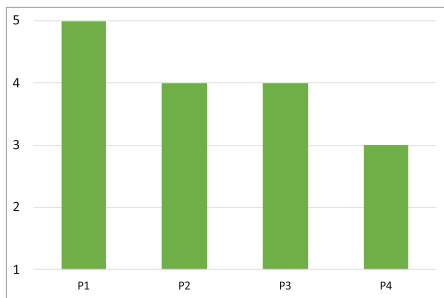


Fig. 14. Answer to the question “*do you want to use our system in the future?*” by *unrecorded participants* (mean = 3.25; SD = 0.83).

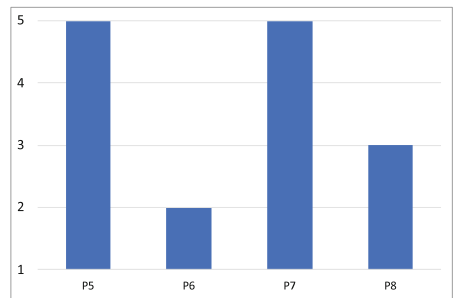


Fig. 15. Answer to the question “*do you want to use our system in the future?*” by *recorded participants* (mean = 3.75; SD = 1.30).

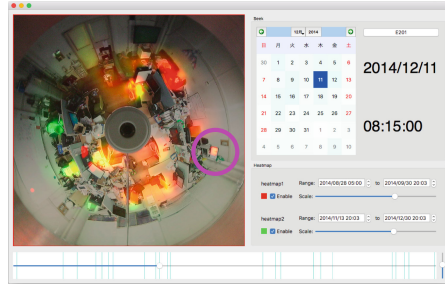


Fig. 16. Discovery about desktop monitor by P2.

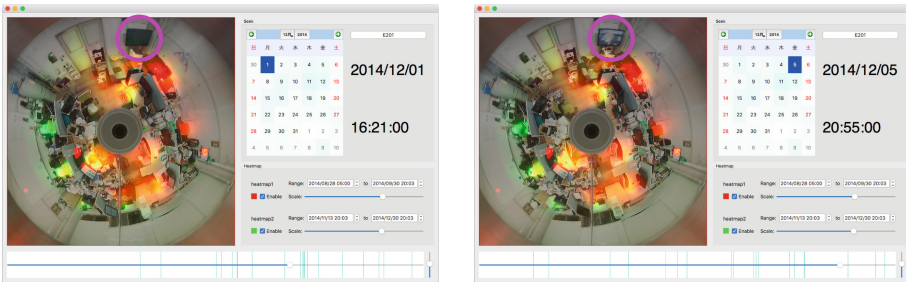


Fig. 17. Discovery about large screen monitor by P2.

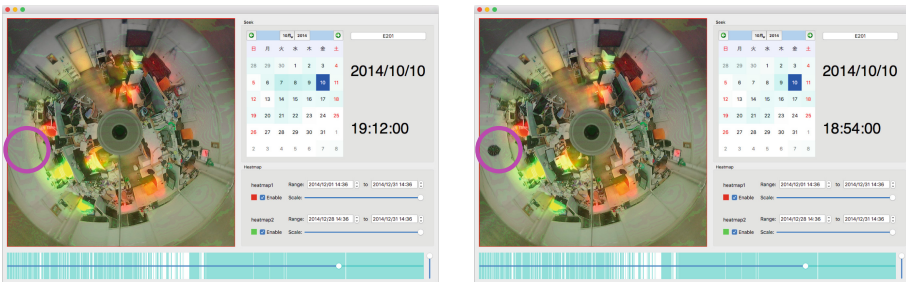


Fig. 18. Discovery about dartboard by P3.

5.3 Analyzing Processes

We compared the analyzing processes of *unrecorded participants* with those of *recorded participants*. To examine each participant's analyzing processes, we analyzed the screen captures. As a result, we found that the processes of both participants were similar: all the participants first browsed the images recorded in July, and then browsed the images recorded in August and later in the sequence.

Examining the analyzing processes of *unrecorded participants* with those of *recorded participants* gives us a suggestion for future possible improvement of our system. P1, P2, and P4 discovered the facts about monitors. For example, P2

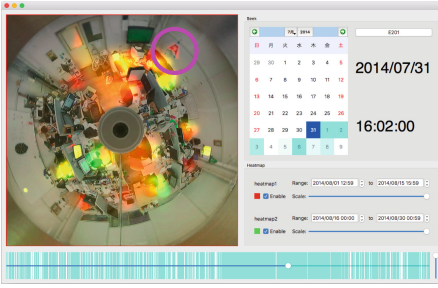


Fig. 19. Discovery about kite by P4.

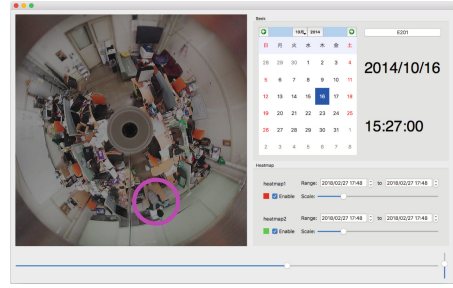


Fig. 20. Discovery about 3D printer by P5.

discovered that there was a movement behind the seat B-6 using the HM/all function (Fig. 16). Then, P2 used the Camera Image function and guessed there was a monitor; however, P2 could not identify what was displayed on that monitor. By contrast, P5, P6, and P7, who discovered similar facts about the monitors, could identify what the monitor displays as a clock since they had prior knowledge of Room-B. Moreover, we found another suggestive result. P2 discovered that the large monitor installed in the public space was used less frequently (Fig. 17 Left). Then, P2 used the HM/part function and discovered the detailed date and time that the large monitor was used. In addition, P2 discovered that an animation had been displayed on the large monitor (Fig. 17 Right). These results suggest that *unrecorded participants* would be able to discover more detailed facts if the resolution of the camera image is improved, because P2 succeeded in identifying the content displayed on the monitor since the monitor was large.

We also found how *unrecorded participants* discovered many facts about Environment while *recorded participants* discovered many facts about People when we observed the analyzing processes of both participants. P3, P4, and P5 discovered the facts about objects installed in Room-B. For example, P3 discovered that the dartboard was removed on October 12th using the Camera Image function (Fig. 18). P4 discovered that there was a kite in the public space (Fig. 19). In addition, P4 discovered that there was a movement around the kite using the HM/all function and guessed that someone was touching it. However, P4 stopped the analysis at this moment. By contrast, after P5 discovered that the 3D printer was installed on the shelf on October 16th by operating Time-Operation Panel (Fig. 20), P5 continued the analysis and discovered a person who installed the 3D printer using the Camera Image function. From the above, the focus of the analysis might depend on the background of the participants. Namely, it seems that P3 and P4 (*unrecorded participants*) focused on the object itself and thus discovered the fact about the object; the fact's property becomes Environment. Conversely, it seems that P5 (*recorded participants*) focused on the person and thus discovered the fact about the object; the fact's property becomes People.

In summary, even *unrecorded participants* can discover many facts. Furthermore, by improving the resolution of the camera image, *unrecorded participants* would be able to discover more detailed facts.

6 Discussion

In our experiment, we found that *unrecorded participants* can discover many facts using our system even if they had no prior knowledge of the room. This means that it is possible to crowdsource the task of analyzing images, which will lead to finding many facts in the images quickly.

We could not identify any difference in the tendency of functions to be used due to the difference in the background of participants. However, properties of the discovered facts tended differently depending on the background of the participants. Compared to *recorded participants*, *unrecorded participants* discovered several facts about Overviewing and Environment and few discoveries on People. However, this does not mean that *unrecorded participants* cannot discover facts about People; we are interested in how the discoveries of *unrecorded participants* will change by instructing *unrecorded participants* explicitly to try their best to discover facts about People before the analysis task.

7 Future Work

Currently, the recording system runs on a high-performance laptop PC (MacBook Pro). Therefore, we need to use a UPS with a large capacity battery in case of long-term power failure. We plan to modify our recording system to run on a low-power computer such as Raspberry Pi. Furthermore, we plan to set up another recording system with an omni-directional camera with higher resolution to examine whether *unrecorded participants* can discover more detailed facts.

There was an opinion that the processing speed of the analyzing system from participants was slow. We think that this problem is influenced by transferring images from NAS. We plan to implement an image prefetching algorithm based on the locality of reference. In addition, we will change the transfer protocol and compress image data before transfer.

In this experiment, we installed a recording system in our laboratory and used the images obtained from that system. In the future, we will use our system in a place other than our laboratory and conduct experiments with the images obtained from it. In addition, we will investigate the adaptability of our system for the observation of animals, plants, and natural phenomena.

8 Conclusions

In this study, we conducted a further experiment using the images in which the participants have not been recorded (*unrecorded participants*) to reveal the discoveries that participants obtain. By comparing the results of users with different backgrounds, we investigated the difference between discoveries, functions used, and analysis process. The comparison suggests that *unrecorded participants* could discover many facts about Environment, and *recorded participants* could discover many facts about People. Moreover, the comparison also suggests that *unrecorded participants* could discover many facts comparable to *recorded participants*.

References

1. Chiang, C.C., Yang, H.F.: Quick browsing and retrieval for surveillance videos. *Multimedia Tools Appl.* **74**(9), 2861–2877 (2015)
2. DeCamp, P., Shaw, G., Kubat, R., Roy, D.: An immersive system for browsing and visualizing surveillance video. In: *Proceedings of the International Conference on Multimedia, MM 2010*, pp. 371–380. ACM, New York (2010)
3. Higuchi, K., Yonetani, R., Sato, Y.: EgoScanning: quickly scanning first-person videos with egocentric elastic timelines. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017*, pp. 6536–6546. ACM, New York (2017). <http://doi.acm.org/10.1145/3025453.3025821>
4. Ishii, A., Abe, T., Hakoda, H., Shizuki, B., Tanaka, J.: Evaluation of a system to analyze long-term images from a stationary camera. In: Yamamoto, S. (ed.) *HIMI 2016*. LNCS, vol. 9734, pp. 275–286. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40349-6_26
5. Kubat, R., DeCamp, P., Roy, B.: TotalRecall: visualization and semi-automatic annotation of very large audio-visual corpora. In: *Proceedings of the 9th International Conference on Multimodal Interfaces, ICMI 2007*, pp. 208–215. ACM, New York (2007)
6. Nguyen, C., Niu, Y., Liu, F.: Video summagator: an interface for video summarization and navigation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012*, pp. 647–650. ACM (2012)
7. Nogami, R., Shizuki, B., Hosobe, H., Tanaka, J.: An exploratory analysis tool for a long-term video from a stationary camera. In: *Proceedings of the 5th IEEE International Symposium on Monitoring & Surveillance Research, ISMSR 2012*, vol. 2, pp. 32–37 (2012)
8. Romero, M., Summet, J., Stasko, J., Abowd, G.: Viz-A-Vis: toward visualizing video through computer vision. *IEEE Vis. Comput. Graph.* **14**(6), 1261–1268 (2008)
9. Romero, M., Vialard, A., Peponis, J., Stasko, J., Abowd, G.: Evaluating video visualizations of human behavior. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011*, pp. 1441–1450. ACM, New York (2011)
10. Shin, G., Choi, T., Rozga, A., Romero, M.: VizKid: a behavior capture and visualization system of adult-child interaction. In: Salvendy, G., Smith, M.J. (eds.) *Human Interface 2011*. LNCS, vol. 6772, pp. 190–198. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21669-5_23

11. Thomas, J.J., Cook, K., et al.: A visual analytics agenda. *IEEE Comput. Graph. Appl.* **26**(1), 10–13 (2006)
12. Tompkin, J., Pece, F., Shah, R., Izadi, S., Kautz, J., Theobalt, C.: Video collections in panoramic contexts. In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST 2013*, pp. 131–140. ACM (2013)
13. Wong, P.C., Thomas, J.: Visual analytics. *IEEE Comput. Graph. Appl.* **24**(5), 20–21 (2004)