The Potential of Virtual Real World for Usability Test of Location-Aware Apps

Tomoyo Sasao[1(✉)], Ruochen Si[2], Mitsutoshi Shibuya[1], and Min Lu[3]

[1] Tokushima University, Minamijosanjimacho 1-1, Tokushima City 770-8502, Japan
sasao@tokushima-u.ac.jp
[2] The University of Tokyo, Kashiwanoha 5-1-5, Kashiwa City 277-8568, Japan
[3] Kyushu University, Motooka Nishi-Ku Fukuoka City 819-0395, Japan

Abstract. Recently, many kinds of location-aware applications of smartphone are developed. However, it is difficult to test them for usability in the same way as the non-location-aware ones because of the test environment. For example, in case of supportive applications for emergency situations, the usability test can be carried out only in normal period of outdoor fields. Key competitive technologies beyond the issues will be Augment Reality and Virtual Reality. We focus on advantages of laboratory experiment and present a usability test environment of location-aware apps based on virtual real world using Google Street View. In this paper, we aim at understanding the weaknesses and strengths of this approach. We developed an initial prototype of the usability test environment and a simple map viewer app. The usability test has two conditions from a designated start point: (c1) go to a designated place and (c2) explore the area freely. We observed two participants' behaviors: (b1) movement in the virtual real world, (b2) body motion, and (b3) operation of the map viewer app. From the results, we discuss the better ways to collect these data as well as the weakness and strengths of laboratory-based usability test environment. These outcomes spotlight laboratory survey's strengths of fine grain spatiotemporal data.

Keywords: Smartphone · Usability · Virtual city · Google Street View · Map · Spatial perception

1 Introduction

In recent years, the number of users of location-based applications is increasing with the popularization of mobile devices [1]. The application designers and developers have to consider the ordinary users' ease. User-centered design includes the idea of involving the target users to take part in the design processes to create suitable designs for them. This idea incorporates designs of different aspects and fields, not only application development. To guarantee the diversity of users and the accuracy of the recorded data, user experiments evaluating the applications are often conducted in the laboratories. However, such experiment environments are insufficient for the recent mobile applications, whose operations are responded to the context of ambient environment. It has been proposed that such applications are better to be experimented in

N. Streitz and S. Konomi (Eds.): HCII 2019, LNCS 11587, pp. 172–186, 2019.
https://doi.org/10.1007/978-3-030-21935-2_14

the real-world environment [2, 3]. Operating only in the laboratorial environment is difficult to reveal the problems of interactions and usability issues and their reasons, which are strongly related to the context of ambient environment. With the recent smartphones equipped with Global Positioning System (GPS) and various high-performance sensors, conducting real-world user experiments becomes much easier than before. For example, by overlaying the estimated damage of earthquakes and tsunami to the real world, researches of supporting disaster drills by presenting the emergency situations on smartphones [4] as well as using Augmented Reality (AR) technologies [5, 6] have been conducted.

On the other hand, there are still problems that are difficult to solve in the real-world experiments. One of them is to gather enough target users, whose number can meet the requirements of quantitative evaluations. For example, tests of the applications supporting disaster evacuations often invite local residents near the experimental fields, while usually difficult to gather other target users, such as visitors, strangers, or even foreigners.

The target of this research is to realize the evaluation of the location-based applications by the potential users who are difficult to attend the real-world experiments. For such purpose, we propose a test environment of a Web-based virtual space of the real world to synchronize the location information to the user's smartphone. The users can test the location-based application in the virtual space with their virtual movement reflected on their smartphones in real-time. With this environment, the usability evaluation experiments for certain geographical areas can be conducted anywhere in a laboratory, and can involve more participants with diverse viewpoints to improve the applications. It can contribute to creating easy-to-use location-based applications for everyone, which is also connected to the exploration of better interactions between human and urban environments.

Of course, information and context in the virtual space is not as rich as that of the real world. Understanding how the users will behave in the virtual environment and what kinds of data can be generated is necessary at the first place. Thus, the current stage of this research has developed the prototype of the proposed environment and a map application for tests, and conducted the preliminary experiments with them. In particular, the users' movement in the virtual space, physical information of body motions and operation logs of smartphones are collected, in order to reveal its feasibility of acquiring sufficient data that are helpful to the application evaluation and understanding the ambient context of the smartphone operations, as well as the necessary functions for conducting the usability evaluation.

2 Related Work

2.1 Supportive Environment for Field Tests

The existing supportive environments for user tests of location-based applications emphasize the importance of the real-world context, and developed a lot of supportive systems for field experiments. For example, to record the actual movement, operations and speeches of the participants, they were asked to walk with head cameras and transceivers in addition to the mobile devices with the target application installed, while the experimenters could supervise the processes in the laboratory [7]. Another research comparing the results of user experiments of social applications used in large-scale events between in lab-based and field-based environments has revealed that real-world environment is much better for user evaluations as the quality of context information is dominantly different [8]. On the other hand, it is reported that 71% of the usability tests of the applications were actually conducted in laboratorial environments [9]. In recent years, with the improvement of the sensors' performance in the smartphones and the development of the technologies estimating users' actions from the sensor data, the feasibility of out-door experiments has been dramatically improved. However, the current systems have not been focusing on the issues of participants' diversity. Nakanishi et al. pointed out the difficulty of repeating the field experiments in the case of researching on crowd behaviors in disaster evacuations, and developed a simulator to represent the crowd behaviors in the virtual space from the records of the field experiments [10]. This research intends to take the important context in the experiments into account, which can be missing in the above experiments in virtual spaces, to realize better user test environment.

2.2 "Realness" of Virtual Reality Based on Real World

The "Realness" of virtual space that modeling the real world has been discussed in the fields of architecture and urban design, in which designing with 3D modeling is studied for many years [11]. Accordingly, there are a lot of researches and developments pursuing the "realness", such as the collaborative design environment in virtual space emphasize 3D sound sources, light, shadows, clouds, plants and so on [12, 13], with the revealed importance of the impressions influenced by the context of environment. However, the recent researches also indicate that, "realness" is not necessary in drawing the participants' opinions and creativities. For example, PlacePulse [14] is a Web-based crowdsourcing system to evaluate the impressions to streets. With two images of the streets randomly captured from Google Street View displayed side-by-side, the system collects answers of the questions like "which is safer?" and "which is more beautiful", and then uses the location-based data for quantitative evaluation and mapping the results of safety, beauty and liveliness on maps. Although the context like ambient sounds, crowds, durations and so on, are not included in Google Street View, the reliability of the results could be guaranteed by the conditions in common with certain restrictions. The "Block to Block" project being promoted by United Nations represents the slumified public areas in some developing countries in the virtual space of a popular video game "Minecraft", in which the users can assemble blocks freely to create their unique spaces. Female and young residents around these areas are able to build their ideal space together within the game. Although it may be far from the "realness" when using the highly abstracted blocks, it can be considered as a successful example of involving the participants with the attributes seldomly covered and drawing opinions from them [15].

In a survey of the supportive environments utilizing virtual space for human-centered design in the field of urban design [11], it is proposed that a suitable degree of "realness" is required according to the purpose and context of an experiment. From this discussion, this research is conducted with the viewpoint that, it is possible to construct a user-participating test environment on basis of comprehending the limitation of virtual space and focusing on specific goals.

3 System Design

3.1 Concept

This research proposes a usability test platform for mobile applications, which synchronizes the location information in the virtual real world with real smartphones. From the related works, the importance of represent the context of using the target application in the virtual space is made clear. Thus, this platform is designed to contain the following three processes (as shown in Fig. 1):

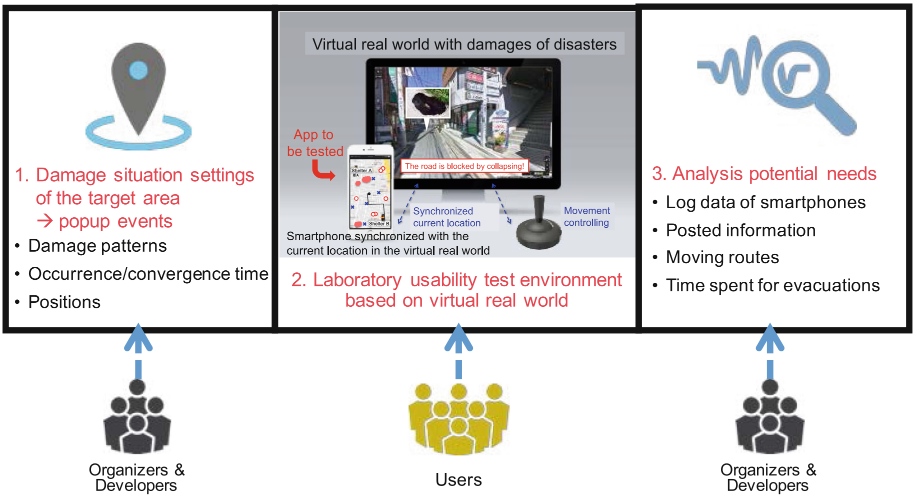

Fig. 1. The user test process of our proposed platform (a case of usability test of a mobile application for disaster)

1. Context Setting: experiment organizers and/or application developers set up the context expected when using the application in real-world environment, and represent it in virtual space. For example, to set up the context of earthquakes, popup events of the damaged places can be set at the blocked roads, which are generated from the past earthquake damage data, to create barriers in the virtual space.
2. Experiment environment: the participants as potential users of the application use a controller to move in the virtual space, which is built based on Google Street View and displayed on the screens. The location of the participant in the virtual space is synchronized to the smartphone with the application to be tested. The organizers

and/or developers inform the scenario, tasks and duration of the experiment. The functions of the location-based application to be tested will work according to the participant's location information in the virtual space. The participant's speech, body motions, movement in the virtual space and smartphone operations are recorded by sensors and cameras.

3. Analysis: the organizers and/or developers analyze the collected data according to the evaluation axis. As the platform focuses on exploring the relations of the spatio-temporal context and the operations of the application, all the data are tagged with time and location, therefore can be visualized on maps or as time-series graphs.

3.2 Prototype Components

This research at first focuses on investigate the effectiveness of the data collected in the preliminary experiments. Therefore, a prototype of the experiment environment with the functions of context setting is developed. At the same time, an experimental location-based application expected to be evaluated with the prototype is developed.

Virtual Real-World Simulator. Based on Google Maps API, a virtual-space environment is implemented on Web browsers for operating and logging in the space represented with Google Street View. JavaScript and PHP are used for the development. As human's viewing angle is 10 to 20 in text reading, 5 to 30° in symbol recognizing, 30 to 60° in color distinguishing and 62° with both eyes [16], three screens are set as shown in Fig. 2. The movement and viewing angle in the virtual space are operated by mouse clicks and drags. Operation logs, including locations (latitude, longitude) and viewing angles (pinch, yaw), are recorded and saved on the server with timestamps and location information at the timing when their values are changed.

Test Sample of Location-Based Application. iPhone A simple map navigation application is developed on iPhone (as shown in Fig. 3). When moving in the virtual space, the update of location recorded in the server will invoke a real-time update of the current location represented in the map application. The viewing range and scale can be operated by pan and pinch gestures. Operation logs in the map application, including map center (x, y), rotation (angle), zoom level (1 to 19) are recorded and saved on the server with timestamps at the timing when their values are changed.

Motion and Eye Sensing. To record the physical information of the participants, JINS MEME ES [17] is applied to measure the head motions (accelerations in XYZ directions, pitch, roll, yow) and eye movements (up, down, left, right, blinking strength). The data are saved on the server with location information.

Voice Sensing. The experiment participants are asked to practice Think Aloud Protocol [18], which need them to speak out all the ideas raised when operating the application. The voices are recorded with videos by the Web camera mounted on top of the screen. The content and timing of the speeches labeled manually for analysis.

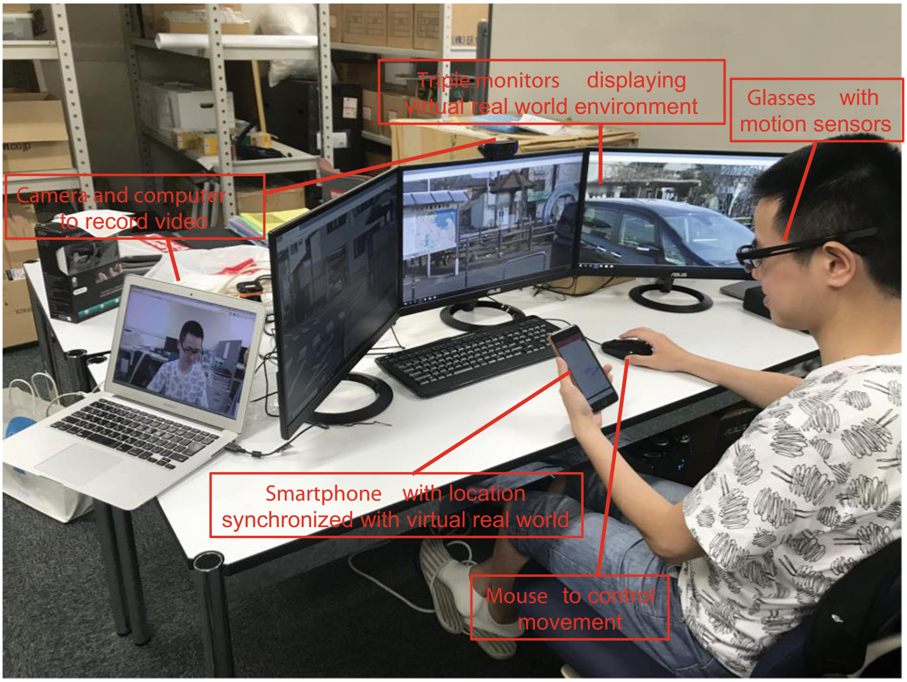

Fig. 2. User test environment

Fig. 3. Test sample of location-based application

4 Initial Experiment

4.1 Aims

The initial experiment aims to investigate the advantages and limits of the proposed experiment environment for future improvement through executing the application evaluation processes, which observe how the users are supported by the map application.

4.2 Method

Setup. The prototype shown in Sect. 3 was set up in a laboratory. Except for the log data of the virtual space and smartphone operations, video records of the screens were also taken for confirmation.

Experimental Scenarios. The experiment area was set at Komatsushima City, Tokushima Prefecture in Japan, the start point was Minami-Komatsushima Station. The following two scenarios were prepared:

1. Support of moving to a destination: walk towards Komatsushima Minato Communication Center, and say "arrived" when arriving the destination.
2. Support of strolling: walk freely and say "finished" when the participant likes.

Participants. One participant (P1) often visited the experiment area and one participant (P2) visited the area for the first time (1 female and 1 male, 25–35 years old) attended the experiment. It was still difficult to make quantitative evaluation from the results of the experiment. In this time, the feasibility of observing the interactions with real world through smartphones was focused through analysis of individual cases.

User Experience Measurement. The users' speeches, body motions, movement in virtual space, operations of smartphone were measured and recorded with spatio-temporal information labeling.

Procedure. The experiment organizer explained the purpose, tasks and scenarios of the experiment to the participant, and then a practice in a place other than the experiment area was made for about 1 min. The Think Aloud Protocol was also instructed in the practice. After presenting the start point in the virtual space and starting data recording, "start" was called, and then the participant began to operate the application and virtual space. All the data were collected after the experiments were finished.

Analysis of Data. The recorded data were labeled with timestamps and locations (in latitude and longitude). Time-series graphs and mappings on maps were generated to observe the characteristics of first-visit and often-visit participants, as well as movement to a destination and free strolling, to find out the differences between them.

4.3 Results

Walking Speed. The time spent by the participants in the scenarios (a) moving to a destination and (b) strolling is shown as Fig. 4 (left), and the average is 9 min and 35 s. In scenario (a), the actual shortest path from the start point to the destination is 900 meters (about 12 min on foot), but the average time spent for moving in virtual space is 5 min and 30 s, which means 54% time saved in the experiment. However, the possibility of context information loss comparing to field experiments should be considered. Therefore, it is important to investigate the context information that can or cannot be recorded in this experiment environment. The average number of clicks for moving in Google Street View is 53.5 in scenario (a) and 102 in scenario (b), as shown in Fig. 4 (right). The average speed (number of clicks per minute) in (a) is 9.7 as a baseline for comparisons, while in (b), it becomes 6.4 for P1 and 8.97 for P2, which means the participants tended to have more time between moving actions (clicks) to take other actions. However, in the initial experiment, the number of samples was only 2 for comparing the moving speed, which limited the reliability. More samples are needed in the future to get more accurate baseline moving speed.

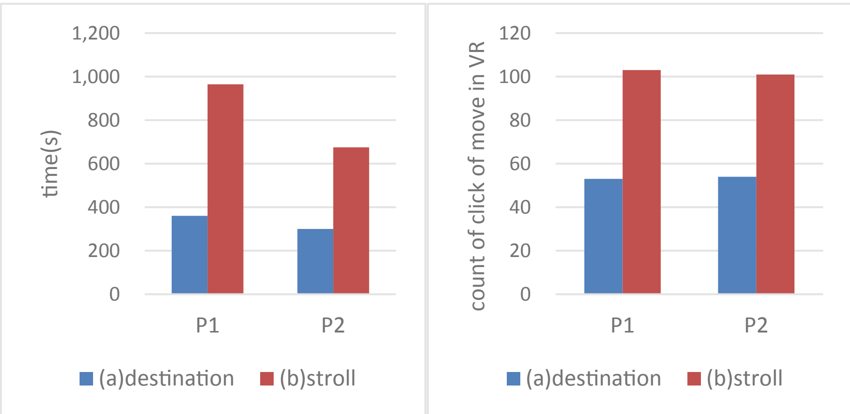

Fig. 4. (left) The time required for the experiments, (right) The number of movements in our VR environment

Operation of the Map Application. Regarding the zooming functions of the map in the smartphone application, the zoom level to view the start point and the destination of (a) on the screen at the same time was 17 (as shown in Fig. 5), but P1, who knew well of the area, kept the zoom level at 19 to display buildings' names and shapes from the start to the end, while never confirmed the overall map. In (b), P1 mainly used zoom level 18 that can show information within about 500 meters when moving. P2, who had never visited the area, tended to use zoom level 15 to 17 more to have an overview in both (a) and (b). P2 also tended to use more different zoom levels in (b) than in (a). The initial experiment revealed that, there was a relation between the experience to the area and the tendencies in operating the map application. Although more samples are needed, the proposed environment can be easier to conduct the experiments investigating behaviors of participants unfamiliar to the place, compared to field experiments.

The sequences of center points of the map displayed in the smartphone application and the moving trajectories in the virtual space were overlaid on maps to find their differences in both scenarios (shown in Figs. 7 and 8). The following three patterns of map operations were found: searching bit by bit along the moving route; moving far away from the route and then return; following the road across the moving route. If adding a visualization of time series variation on a map, it might possible to analyze more deeper relations between the map operations and actual moving routes. In future, we try to figure out the relations between users' mental context and the map operations when users are walking around a real world through virtual testing environment (Fig. 6).

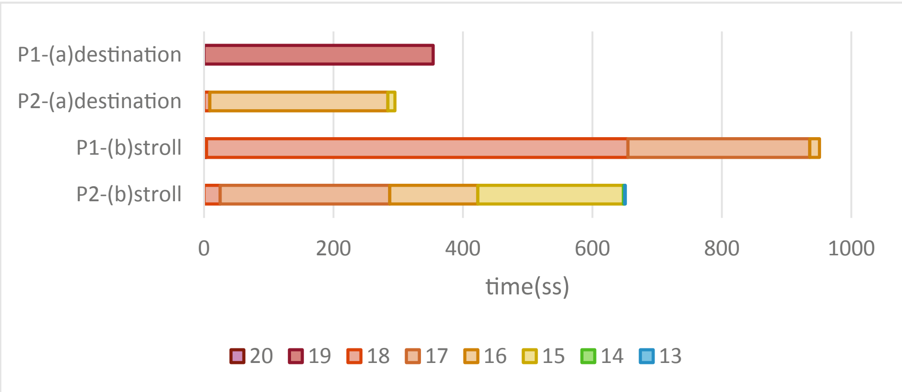

Fig. 5. Sojourn time of each zoom level.

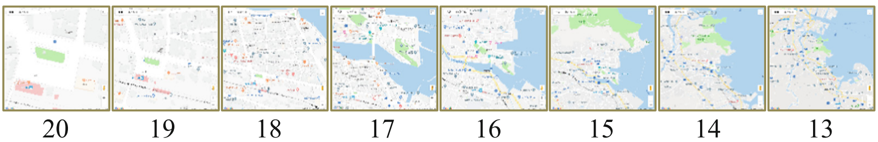

Fig. 6. Map samples of zoom level based on Google Maps API

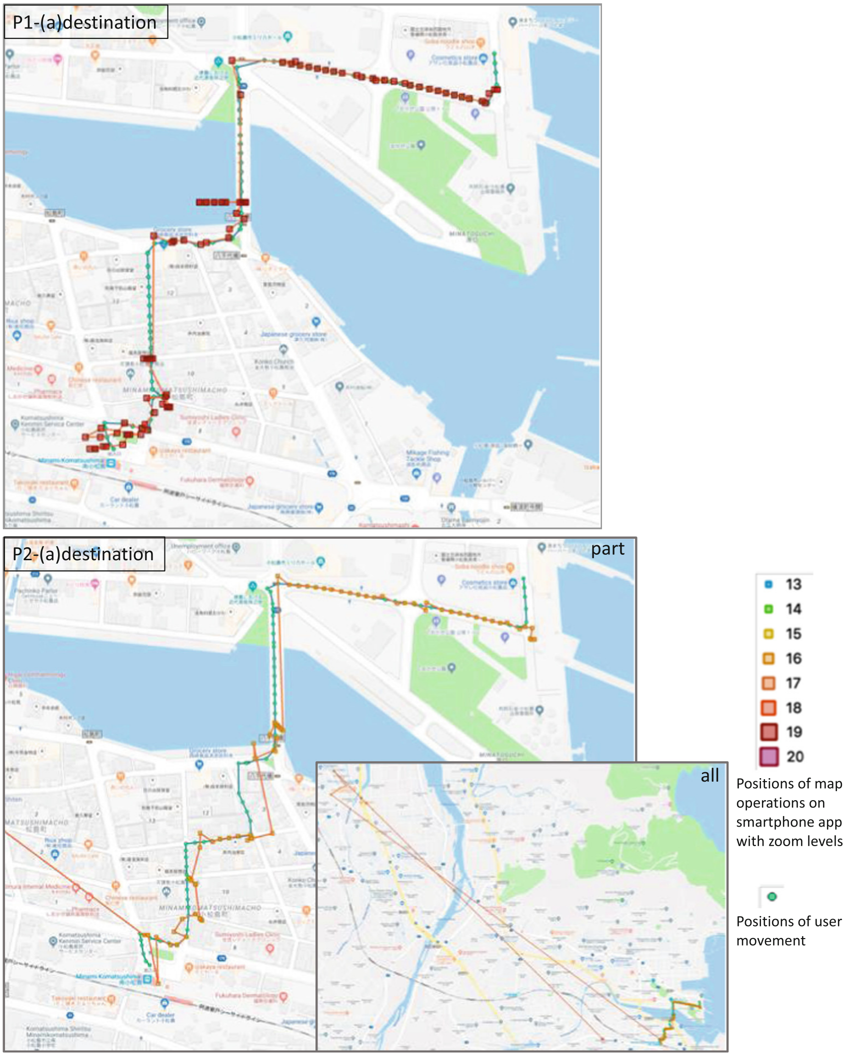

Fig. 7. Mapping movement locus and movement of the center of the map application with zoom levels in (a) destination scenario

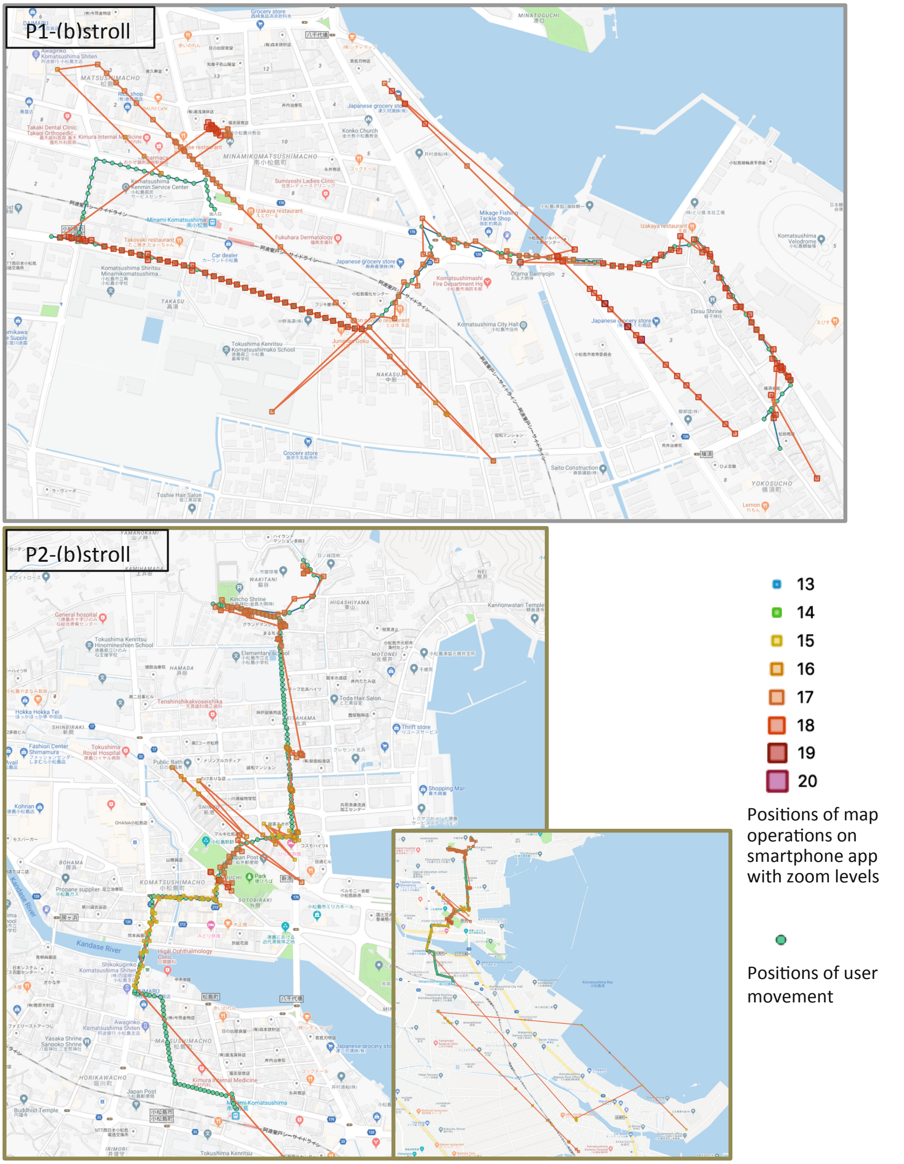

Fig. 8. Mapping movement locus and movement of the center of the map application with zoom levels in (b) stroll scenario

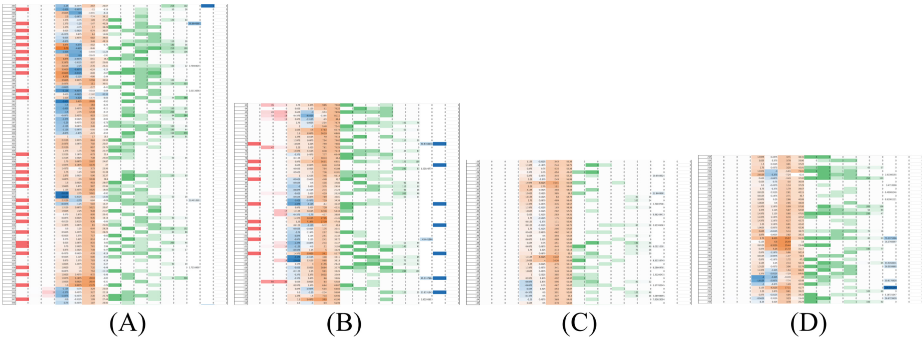

Fig. 9. Parts of the timeline of P1's (a)destination. From the left row: time(ss), GoogleStreetViewOperation(red rows){movement, heading, pitch}, GlassesOperation1(orange rows){accX, accY, pitch, yaw}, GlassesOperation2(green rows){eyeMoveUp, eyeMoveDown, eyeMoveLeft, eyeMoveRight, blinkSpeed, blinkStrength}, SmartphoneOperation(blue rows){length of move, map rotation, change of zoom level} (Color figure online)

Eye and Head Movement. The sensors in the glasses recorded the head movement as 6-axis accelerations, eye movement in 4 directions, blinking frequency and strength. The actions of looking at the smartphone were planned to be extracted from these data, but it was not so easy as expected. Fig. 9 shows parts of the log data recorded per 0.1 s for P1 in (a), which include Google Street View operations (red), head accelerations (light blue to orange), eye movement and blinkings (green), smartphone operations (blue). Fig. 9 (A) is in the cases of focusing on moving in the virtual space, and (B) is in the cases of operating the smartphone as well as the virtual space at the same time. In both cases, the eyes were moving frequently in all the directions, therefore it was difficult to distinguish the two cases from only the data recorded this time. Both (C) and (D) were in the cases of focusing on operating the smartphone, however, (C) showed less had movement compared to other cases. To distinguish between (A) and (B), as well as between (C) and (D), analysis using speech and video data with a larger number of samples are required in the future.

5 Discussion

Tang, L. et al. proposed eight open issues that should be covered in supporting rapid design and evaluation of pervasive applications [19]. This research selected five from them, which are effective for evaluations with user experiences of prototypes, to discuss the experiment results in Sect. 4.

5.1 Simulating Pervasive Environment

Tang, L. et al. suggested 3 roles of the simulators for pervasive environment: *"(1) simulating the input space of an application, including the explicit (e.g., mouse or keyboard events) and implicit input (e.g., location sensed input when user moves);*

(2) simulating the logical control flow that jumps between sensors, servers, handhelds (such as PDA) and any other kinds of networking appliances; and (3) simulating the output space of an application, which means to visualize the environment effects caused by the application behaviors.”

According to our initial experiment, our prototype seems to cover enough the functions of simulating (1) the input space and (3) the output space because of using real smartphone devices in the simulator. However, especially (2) the network conditions supported by our prototype are not real and it is usually better than outdoor fields. Therefore, it is thought that there is a limitation of the network simulation in our prototype. We might be better to prepare the function creating pseudo network latency in the virtual space.

5.2 Description of Context-Awareness

Tang, L. et al. mentioned *“Integrating the ability of associating digital data with the users’ context into a prototype would be a great help to promote easier retrieval.”*

From the results of our initial experiment, we could record body motions, speeches and operation logs labeled with spatio-temporal information in the virtual space, though they are still raw data, which are difficult to be analyzed to extract the context. The types of effective context data for retrieval of tested applications depend on the type of them. We will explore the useful context types for the location-based applications having a large number of users, such as applications for navigation, finding restaurants, finding friends nearby, checking public transportation, receiving special deals or offers from retailers [20]. Furthermore, as the reactions to the context in the virtual space can be different from those in the real world, experiments comparing the proposed environment and the real world are necessary.

5.3 Robust Debugging Environment

The debugging environment is important for inspecting the program architecture, data structure and communication flow. However, according to the investigations by Tang, L. et al., debugging of pervasive applications in distributive mode is very difficult. The proposed environment can set the timing of debugging, and is easier to gather necessary data in virtual space for debugging. In the future, the methods and functions of setting debug mode for experiment organizers can be explored.

5.4 Logging Test Data

Tang, L. et al. suggest that, *“When evaluating design, designers have to analyze the test data, especially to collect the feedback from a long-term, in situ test setting.”* The proposed environment takes the advantages of virtual space that, all kinds of log data can be automatically recorded with spatio-temporal labeling. The time needed for an experiment is usually short than that in the real world, therefore the burdens of experiment organizers and participants can be reduced. On the other hand, as discussed in Sect. 4, the task of free strolling ended in 15 min averagely, even if there is no time limit, which suggests that long-duration experiments may bring more burdens to participants. For example, it is possible to ask the participants to use certain devices in daily lives for several weeks in the real world, but which is not feasible in the virtual space like the proposed environment.

So far, the methods of extracting context information from the recorded raw data is not established yet, so it is still difficult to arrange and analyze the data efficiently. For example, the body motions and speeches are supposed to be useful to extract the phycological context like confusing in the virtual space. The analysis of eye and head movement in Sect. 4.3 also shows that it can be useful to playback the recorded spatio-temporal data for exploring new findings. Such functions should be implemented in the future.

5.5 Evaluation Criteria

Tang, L. et al. proposed that the evaluation criteria of prototypes should be field specific and distinctive in each phase. The criteria of "easy-to-understand" is suitable for the tests of getting feedbacks, while "latency" is suitable for the tests of the context-aware application's functions. In the former case, if the prototype's functions are easy to understand, the users are expected to have few problems even if the ambient environment is in virtual space. In the latter case, as Google Street View provides limited moving range and fixed context like time and weather, it may not meet the requirements of some evaluations. In the future, context other than location information should be considered, and methods of extending Google Street View to include them need to be investigated.

6 Conclusion

This research developed a prototype of user experiment environment based on the extension of Google Street View for evaluating location-based mobile applications aiming to reduce the costs of user tests compared to field experiments to involve more participants. The initial experiment investigated how the interactions with smartphone application in virtual space could be recorded and represented. The results showed that the proposed experiment environment is useful for user evaluations of location-based applications, and its limits and issues to be improved were also discussed. That means a fieldwork method, which is one of participatory design tools grows up into more collaborative platform and it has potential to influence future urban application design ways.

References

1. Number of location-based service users in the United States from 2013 to 2018 (in millions). Statista (2017). https://www.statista.com/statistics/436071/location-based-service-users-usa/. Accessed 10 Feb 2019
2. Johnson, P.: Usability and mobility; interactions on the move. In: Proceedings of the first workshop on human-computer interaction with mobile devices, GIST Technical report G98–1 (1998)

3. Elzakker, C.P.J.M., Delikostidis, I., van Oosterom, P.J.M.: Field-based usability evaluation methodology for mobile geo-applications. Cartogr. J. **45**(2), 139–149 (2008)
4. Nigetore. https://nigetore.jp/. Accessed 10 Feb 2019
5. Itamiya, T.: The virtual tsunami disaster situation experience system using a head-mounted display, media and communication for disaster risk reduction. In: Showcasing Innovation, An initiative for the Third UN World Conference on Disaster Risk Reduction in Sendai, pp. 2–3 (2015)
6. Kawai, J., Mitsuhara, H., Shishibori, M.: Game-based evacuation drill using augmented reality and head-mounted display. Interact. Technol. Smart Educ. **13**(3), 186–201 (2016)
7. Elzakker, V.C.P., Delikostidis, I., Oosterom, V.P.J.: Field-based usability evaluation methodology for mobile geo-applications. The Cartogr. J. **45**(2), 139–149 (2008)
8. Sun, X., May, A.: A comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices. Adv. Hum. Comput. Interact. 2 (2013)
9. Kjeldskov, J., Graham, C.: A Review of mobile HCI research methods. In: Chittaro, Luca (ed.) Mobile HCI 2003. LNCS, vol. 2795, pp. 317–335. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45233-1_23
10. Nakanishi, H., Yoshida, C., Nishimura, T., Ishida, T.: FreeWalk: A 3D virtual space for casual meetings. IEEE MultiMedia, **6**(2) (1999)
11. Portman, M.E., Natapov, A., Fisher-Gewirtzman, D.: To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. Comput. Environ. Urban Syst. **54**, 376–384 (2015)
12. Roussou, M., Drettakis, G., Tsingos, N., Reche, A., Gallo, E.: A user-centered approach on combining realism and interactivity in virtual environments. In: IEEE Virtual Reality 2004, pp. 251–252, Chicago, IL, USA (2004)
13. Drettakis, G., Roussou, M., Reche, A., Tsingos, N.: Design and evaluation of a real-world virtual environment for architecture and urban planning. Presence Teleoper. Virtual Environ. **16**(3), 318–332 (2007)
14. PlacePulse. http://pulse.media.mit.edu/. Accessed 10 Feb 2019
15. Block by Block, https://www.blockbyblock.org/. Accessed 7 Feb 2019
16. Royal Naval Personnel Research Committee: Human Factors for Designers of Naval Equipment. Medical Research Council, Swindon (1971)
17. JINS MEME, https://jins-meme.com/en/. Accessed 10 Feb 2019
18. Jaspers, M.W., Steen, T., Van Den Bos, C., Geenen, M.: The think aloud method: a guide to user interface design. Int. J. Med. Inform. **73**(11–12), 781–795 (2004)
19. Tang, L., Yu, Z., Zhou, X., Wang, H., Becker, C.: Supporting rapid design and evaluation of pervasive applications: challenges and solutions. Pers. Ubiquit. Comput. 15(3), pp. 253–269. Springer-Verlag, Heidelberg (2011)
20. Navigation is the most popular type of LBS app, http://www.tnsglobal.com/intelligence-applied/blessing-in-disguise. Accessed 10 Feb 2019