Chapter XVIII
Virtual Cities for Simulating Smart Urban Public Spaces

Hideyuki Nakanishi
Osaka University, Japan

Toru Ishida
Kyoto University, Japan

Satoshi Koizumi
Osaka University, Japan

ABSTRACT

Many research projects have studied various aspects of smart environments including smart rooms, home, and offices. Few projects, however, have studied smart urban public spaces such as smart railway stations and airports due to the lack of an experimental environment. We propose virtual cities as a testbed for examining the design of smart urban public spaces. We developed an intelligent emergency guidance system for subway stations and used the virtual subway station platform to analyze the effects of the system. This experience allows us to argue that simulations in virtual cities are useful to pre-test the design of smart urban public spaces and estimate the possible outcome of real-life scenarios.

INTRODUCTION

Virtual cities are three-dimensional graphical representations of digital cities (Ishida, 2002a). This chapter describes virtual cities for testing smart environments installed in large-scale crowded urban public spaces such as airports and railway stations. Smart environments are living spaces with embedded abilities to perceive what their inhabitants are doing and support their lives. Since living spaces vary with respect to scale, smart environments of various sizes have been
developed: smart rooms (Bobick et al., 1999), smart classrooms (Brotherton & Abowd, 2004), smart homes (Kidd et al., 1999), smart offices (Addlesee et al., 2001), and smart conference sites (Sumi & Mase, 2001). There have been few attempts, however, to develop smart environments in urban public spaces such as airports and railway stations, even though they are an essential part of our everyday life. A manifest reason for this is the sheer vastness of such spaces. Many researchers have built their own room (Bobick et al., 1999) and home (Kidd et al., 1999) in which to conduct their projects, but it is almost impossible to build urban public spaces in research laboratories. Conference sites are too large to be purpose-built as laboratories, too. However, there is no need to do so, since researchers prefer to take advantage of actual conference events as testbeds for evaluating their systems (Sumi & Mase, 2001). Such deployment is preferred also for testing smart classrooms (Brotherton & Abowd, 2004) and smart offices (Addlesee et al., 2001). Deployment of smart urban public spaces is challenging, because it is difficult not only to attach sensors to the spaces but also to involve the visitors in situ; it can be awkward or prohibited to ask them to participate in experiments. Furthermore, it is usually impossible to shut visitors out of the space and occupy it in order to conduct experiments with study participants. We propose virtual cities as a solution. We present a user testing method that utilizes virtual cities populated with scenario-controlled software agents developed by us (Ishida, 2002b).

Figure 1. Evacuation simulation in the virtual Kyoto station
More than 300,000 passengers pass through Kyoto station, the main railway station in Kyoto City, every day. In this station we installed a guidance system that tracks passengers to help their navigation based on their current positions (Nakanishi et al., 2004). Beyond conventional navigation systems, which passively present route information, our system proactively sends instructions to the individuals’ mobile phones to control their routes and avoid congestion. The system’s primary application is crowd control in emergency situations. Fortunately, we were permitted to attach positioning sensors to the station’s subway platform and install the system, though we were not allowed to conduct experiments that would employ many subjects playing the role of an escaping crowd. To conduct experiments without occupying urban public spaces, we developed a virtual city simulator integrating a large number of software agents and humans into the same crowd. This simulator can produce complex group behaviors such as escaping crowds. This simulator enables the agent-based user testing described in this chapter.

The next section explains how agents, humans, and avatars are integrated in the agent-based user testing. The third section presents an experiment, which was conducted to see how the user testing method can work on our guidance system. In the fourth section we discuss implications obtained from the experiment. The fifth section summarizes related work. The sixth section concludes this chapter.

AGENT-BASED USER TESTING

Augmented Experiments by Agents and Humans

We developed an AR (Augmented Reality) based user interface for the virtual city simulator. Even in a crowded urban public space, it is not difficult to test smart environments that support individuals (e.g. normal pedestrian navigation systems (Abowd et al, 1997), since an experiment in which just one person or a group is taking part does not disturb the space. In contrast, it is extremely
intrusive to test smart environments that support crowds (e.g. crowd navigation systems such as our emergency guidance system). To solve this problem, we contrived multi-agent crowd simulations that can be overlaid onto physical spaces. A large number of agents in the simulation augment an on-site small-scale experiment. We call this kind of experiment an “augmented experiment” (Ishida et al., 2007). Augmented experiments enable us to conduct large-scale experiments in an urban public space with minimum interference with its daily operation. In the experiment performed on the subway platform at Kyoto Station we overlaid an evacuation simulation in which a hundred agents escaped from the platform to the upstairs concourse through the central staircase. Figure 1 presents a couple of screenshots of this simulation visualized in the virtual Kyoto Station. To avoid disrupting the station’s operation, only three subjects escaped in each evacuation trial. The AR based user interface was necessary for the subjects to experience such a simulation on the physical platform.

The AR based user interface was a mobile phone which displays four symbols: a cross, a triangle, a circle, and a double circle. See-through head-mounted displays are not suitable for presenting the simulation of augmented experiments, since it is unsafe to mask the field of view of a walking person with a wide-field image such as a virtual crowd. Semitransparent images are safer but it is harder for users to recognize features. As Figure 2 shows, we used mobile phones, because there are always people looking at their phone’s screen, reading and writing text messages while walking around in crowded places. Mobile phones are a simple means of enabling AR. Since small images of a 3D virtual space are difficult to understand, instead of displaying visual simulations, the mobile phones displayed symbols that directly represented what the subjects of the experiment needed to recognize. To produce a situation in which the subjects were often blocked by a surrounding virtual crowd and could not advance freely, we used four symbols representing the following four different degrees of density and slowness of the crowd: When the subjects were facing a crowd too dense to advance through and they had to stop walking immediately, a cross was presented on the screen. If they found a triangle or a circle on the screen, they could move ahead slowly for one or two meters, as they were approaching the crowd. A double circle was presented whenever it was possible for them to walk freely. Before starting each evacuation trial we asked the subjects to keep to these rules. Note that their mobile phones displayed a symbol and also a guidance message (described later). Which symbol and which message a subject’s phone displayed was determined by the simulation and his or her position, which was tracked by the guidance system. The direction of subjects’ move was estimated by the predefined evacuation route in the experiment.

**Participatory Simulations by Agents and Avatars**

Since it is conceivable that augmented experiments are not the best solution, we formed another group of subjects, who sat in a laboratory room and controlled their avatars in Figure 1’s virtual Kyoto Station. This kind of human-in-the-loop multi-agent simulation is called a “participatory simulation” (Guyot et al., 2005). As Figure 3 shows, the subjects held a gamepad in one hand to control their avatars and a mobile phone in the other hand to receive the guidance messages (they did not receive the symbols described in the previous section). We designed our participatory simulations (VR mode) to contrast with the augmented experiments (AR mode): AR subjects walked around a physical space, whereas VR subjects controlled their avatars just like they would play a videogame. AR subjects received the symbols representing the current situation, while VR subjects could see the visual simulation from their avatars’ viewpoint.
User Testing by Agents, Humans, and Avatars

In general, multi-agent simulations of group behaviors (Noda & Stone, 2003) show a different movement in each trial. If the AR and VR subjects could share the same simulations, we could exclude unnecessary variance in the analysis results of questionnaires, recorded data, and so on. Thus, we designed our simulator to be able to create both symbols and 3D animations from the same simulation, and to use both the sensors and the gamepads to control avatars. Figure 4 depicts agent-based user testing, which is a combination of augmented experiments and participatory simulations. The virtual space contains the VR and AR subjects’ avatars, and the agents. The AR subjects and also passengers inhabit the real-life space. The laboratory room includes the VR subjects and an experimenter who monitors the simulation displayed on a large screen and administrates the experiment.

VR and AR subjects, and agents, have different input and output means in order to take part in the simulations. To integrate these heterogeneous participants, we attached a wrapper to the simulator’s walking and seeing modules. VR subjects manipulate their gamepads to input the direction which they wish to advance. AR subjects’ movements are tracked by sensors that inform the system of their current positions. Agents are controlled by their assigned simulation scenarios, in which their next destinations are specified. The wrapper converts these heterogeneous input data into changes in velocity and orientation. According to these changes, the walking module determines the following positions based on the pedestrian model (Okazaki & Matsushita, 1993) and the gait model (Tsutsuguchi et al, 2000). The wrapper also presents the same data managed by the seeing module in different forms. The data are positions and other parameters of the agents and the avatars. Agents can directly access the data to perceive the situation within the limitation of their visual power and visual field. VR subjects can see the 3D animations drawn according to the data, while AR subjects use their mobile phones to receive notifications of situational changes.

USER TESTING IN KYOTO STATION

The goal of the experiment was a trial use of the agent-based user testing. We tried to confirm that
our method was useful to pre-test the design of smart urban public spaces and estimate the possible outcome of real-life scenarios.

**Smart Environment Tested in the Experiment**

General emergency guidance is usually offered through public address systems which announce general information that is meaningful to the whole crowd. An example of this general guidance is: “There is a fire. Please use the nearest staircase to exit.” Our emergency guidance delivered through mobile phones disseminates site-specific information that is suitable for each person. An example of this location-based guidance would be: “Please do not use the nearest staircase because it is too crowded.” Location-based guidance systems need sensors to know the location of the addressed person and also the movement of the surrounding crowd.

We attached a vision sensor network to the station. We attached twelve sensors to the concourse and sixteen sensors to the platform. Figure 5(a) is the floor plan, on which the black dots show the sensors’ positions, and Figure 5(b) shows how they have been installed. The vision sensor network can track passengers between the platform and the ticket gate. In Figure 5(c), you can see a CCD camera and a reflector with a special shape (Nakamura & Ishiguro, 2002). If we could expand the field of view (FOV) of each camera, we could reduce the number of required cameras. However, a widened FOV causes minus (barrel) distortion in the images taken by conventional cameras. The reflector of our vision sensor can eliminate such distortion. The shape of the reflector can tailor a plane that perpendicularly intersects the optical axis of the camera to be projected perspective to the camera plane. As shown in Figure 5(d), this optical contrivance makes it possible to have a large FOV without distortion. From the images taken by the cameras, the regions of moving objects are extracted using the background subtraction technique. The position of each moving object is determined based on geographical knowledge, including the position of the cameras, the occlusion edges in the views of the cameras, and the boundaries of walkable areas. Figure 5(e) shows AR avatars synchronized with the retrieved positions of AR subjects.

*Figure 5. Positioning sensors installed in Kyoto Station*
In Figure 2, the AR subjects wore a cap with a halogen lamp attached. The lamp and infrared filters covering the sensors were necessary to avoid errors in tracking the movements of the subjects on the platform, which was crowded with passengers. Passengers did not keep their distance from the subjects because they apparently ignored our experiment. Our location-based guidance system can work without this trick when the platform is sparsely populated.

The system needs the subjects’ email addresses to send them guidance messages. In the experiment we registered the addresses before the evacuation began. We suppose that a real-life system would automatically register the addresses via the smart cards that people use to pass through the ticket gates. The system cannot work unless delay in email delivery is short. In the experiment the delay was about several seconds, which was short enough. We kept the email messages short to avoid distracting the subjects. Our previous prototype (Nakanishi et al., 2004) supported vocal instructions instead of guiding email messages, and we did not evaluate ways of conveying the guidance but did evaluate the psychological responses the guidance induces. Thus, the results of this experiment can be applied to both previous and current prototypes.

Hypotheses

Among other things, emergency guidance must be trustworthy in order to safely lead escaping people, who tend to lose their composure and recklessly follow others around them (Helbing et al., 2000; Sugiman & Misumi, 1988). A study on interpersonal trust in remote communication suggests that there are two independent factors in trust (Greenspan et al., 2000). One is the emotion-oriented attribute, which builds supportive impressions. In the case of emergency guidance, this is the degree to which the guide seems to be willing to help people to escape safely. The other factor is the cognitive-oriented attribute, which increases confidence that a task will be successfully completed. In other words, the guide is believed to be able to grasp the situation perfectly and manage the evacuation efficiently. We supposed that location-based guidance is better than general guidance in terms of cognitive-oriented trust. Since location-based guidance takes into account the guided person’s position and his/her surrounding situation, we expected that the guidance could give a feeling that the system is monitoring the scene of evacuation and issuing optimal instructions, with the guided person finally feeling at ease. Our hypotheses are that location-based guidance: 1) would be perceived as trustworthy; 2) would be useful as a navigation aid; and 3) might induce more calmness, which is important to prevent panic. We did not pressure the subjects into a panic, since even just a few people could cause serious trouble if they were to panic in a crowded station. Instead, we merely asked the subjects to arrive at the destination as soon as they could. The previous research showed that this sort of moderate pressure does work to a certain extent for investigating evacuations (Sugiman & Misumi, 1988).

Procedure

We recruited nine VR and eight AR subjects for a total of seventeen subjects. All of them were undergraduate students. We paid them for their participation. In each evacuation trial, three VR and three AR subjects escaped together with a hundred agents. To make the evacuation as simple as possible, the system guided everyone to the central staircase that soon became crowded as shown in Figure 1. In an actual evacuation on the subway platform more than a thousand of people would escape. A hundred agents, however, were enough to produce a crowded situation around a single staircase.

The subjects took part in an evacuation where they were guided with either the location-based method or the general way, and answered a ques-
tionnaire about the trustworthiness and usefulness of experienced guidance and calmness during the evacuation. Each subject repeated this twice in random order to experience both guidance methods. We analyzed the data of eight VR and five AR subjects, because one VR and three AR subjects were unable to experience one of the two kinds of guidance due to system problems.

In both location-based and general guidance evacuation, the system sent the subjects five guidance messages based on each subject’s current location as follows. The first message was sent when the subject started escaping. The second, third, and fourth messages were sent when the subject passed through the spots that are fifteen, ten, and five meters away from the staircase respectively. Delay in email delivery did not become a problem since the subjects walked according to the messages that told them to move ahead slowly or stop walking. The final message was sent when the subject began climbing the staircase.

The messages used in the location-based guidance included information on the direction in which to proceed and the crowdedness around the staircase. At the moment the evacuation started, the subjects received the message “Please escape through the front staircase.” Then, after they began walking and had advanced a certain distance, they were told to “Please keep going toward the staircase.” When they arrived at the tail of the crowd jostling at the bottom of the staircase, they were guided with “Please use this staircase even though it is crowded.” When they were about to finish passing through the crowd, they were told “Please do not hurry because you will pass the crowd soon.” Finally, they were advised to “Please go up the staircase calmly,” when halfway up the staircase to the concourse. The general guidance messages, on the other hand, were the plain emergency announcements: “Please escape through a staircase close to you,” “Please do not hurry when fleeing from here,” “Please choose the staircase nearest to you,” “Please keep calm during the evacuation,” and, “Please use a staircase nearby for evacuation.”

Results

After each evacuation, all the VR and AR subjects answered the same questionnaire, which contained three items about trustworthiness, three items about usefulness, six items related to the calmness, and twelve items prepared for hindering the subjects’ interpretation of the experimenter’s intention. All the items were measured using a nine-point Likert scale ranging from 1 to 9. We analyzed the VR subjects’ data and the AR subjects’ data separately. We used a two-sided paired t-test to analyze differences in the impressions of the two guiding methods.

From the VR subjects’ data, we found that location-based guidance was more trustworthy ($t(7)=3.1$, $p<.05$). This is the result of comparing the means of the “TRUST” index (general: 13.1, location-based: 19.6, 3 was the lowest, 27 was the highest). We obtained this index by summing the scores of these three items: “How much were you willing to follow the guidance?”, “How trustworthy was the guidance?”, and “How persuasive was the guidance?” (Cronbach’s $\alpha$: .84). We also found that location-based guidance was more useful ($t(7)=2.8$, $p<.05$). This was obtained from the “USEFUL” index (general: 15.3, location-based: 20.2, 3 was the lowest, 27 was the highest), which was made from these three items: “How useful were the guidance messages?”, “How easily could you understand the messages?”, and “How clearly was the guidance service?” (Cronbach’s $\alpha$: .70). Figure 6 summarizes these results.

In the AR subjects’ data, the TRUST and USEFUL indexes did not show significant differences, but we instead observed that the subjects were calmer under location-based guidance ($t(4)=3.1$, $p<.05$). This comes from the “CALM” index (general: 9.8, location-based: 14.6, 2 was the lowest, 18 was the highest), which was the sum of two items: “How unhurriedly did you escape?” and “How calm did you feel during the evacuation?” (Cronbach’s $\alpha$: .70). Figure 7 shows this finding. In the VR subjects’ data there was no significant
difference in the CALM index.

Note that all the items that had significant difference are presented in the above analysis. We could not find any result that indicated superiority of the general guidance.

**IMPLICATIONS**

**Analyzing the Findings of the Experiment**

Interestingly, the location-based guidance elicited different responses from the VR and the AR subjects. The VR subjects appraised the method—the location-based guidance was trustworthy and useful, and the AR subjects became aware of their feelings—the location-based guidance kept them calm. Through analysis of the recorded videos, we explored what caused these different reactions.

First, we found that the VR subjects felt that the location-based guidance system was context-aware and tracking the current situation to make decisions. The VR subjects were watching the PC’s screen drawing a visual simulation from their avatars’ viewpoint. When the guidance’s message informed them of the crowded staircase, they could actually see it. On the other hand, the AR subjects were just signaled to slow down or stop when they received the same message. It was not easy for them to know whether the system was context-aware or not.

Next, we observed that the AR subjects physically experienced the evacuation. The physical experience means that the subjects actually moved their bodies on the platform. When the virtual crowd around the staircase forced the AR subjects to stop walking and wait for the cross symbol—the “pause” instruction—to disappear, they received a different message for each guidance method. The location-based message, “Please do not hurry because you will pass the crowd soon,” was predictive, which might have helped to keep the subjects calm. In contrast, the general message, “Please keep calm during the evacuation,” did not predict anything, thus the subjects could not estimate how long they had to keep waiting, something that might have made them uneasy. If the AR subjects could see the virtual crowd as the VR subjects could do, they would have felt more uneasy and consequently the ‘calm’ effect of the location-based message could be observed more clearly. On the other hand, the VR subjects held down the “up” key to keep going forward before reaching the crowd and also after colliding with it. They did not need to release the key, since their avatars would automatically stop walking once the avatars hit other agents or avatars. This kind of operation is also observed.

![Figure 6. VR subjects’ response](image)

![Figure 7. AR subjects’ response](image)
in videogame playing. The difference in guidance messages barely affected the VR subjects because whether they were walking or pausing did not matter to them.

**Understanding Agent-Based User Testing**

A lack of visual representations prevented the AR subjects from recognizing the crowd of agents and avatars around the staircase, so they were unable to notice the context-awareness of the guidance system. However, physical reality like what they felt when they were blocked by a surrounding crowd led them to indirectly evaluate the context-aware guidance system highly. Since the VR subjects’ experience was deliberately differed from that of the AR subjects, it may be possible to generalize this as follows: In the physical space, software agents have difficulty in appearing visually, but they can interact with humans via mobile notification devices in terms of physical movement, even though they do not have a physical body. In a virtual space the agents can easily display their appearance and simulate social interaction with humans by means of 3D computer animations but can barely make them feel physical interaction. The kind of results that can be obtained from agent-based user testing depends significantly on the user interface for interacting with the agents. The VR and the AR interfaces, then, could provide results based on social and physical interactions. The results were different but both of them showed the advantages of the location-based guidance as we hypothesized in the previous section. If we had used a user interface that combined the two kinds of interactions, we might have different results. The development of such a user interface is a future work.

**RELATED WORKS**

Toward the goal of substituting agents for humans, researchers have been trying to develop various social agents (Nagao & Takeuchi, 1994), such as sales agents (Cassell et al., 1999) and trainer agents (Rickel et al., 2002). These agents are, however, still far from being used in practice. Meanwhile, agents have been used as movie extras (Macavinta, 2002). To realize the dream, it seems to be a much shorter way to develop agents with the ability to follow a scenario describing what to do than agents with complete autonomy. As described in this paper, the user testing of smart environments exemplified how such scenario-controlled agents could substitute for humans.

Our virtual city simulations appear to be similar to SpaceTag (Tarumi et al., 2000), which is location-based informational objects overlaid on the real world. Our simulator is, however, endowed with the gait (Tsutsuguchi et al., 2000), pedestrian (Okazaki & Matsushita, 1993), and multi-agent interaction models (Murakami et al., 2003) so that it can overlay complex group behaviors. Moreover, subjects can take part in our simulations from both mobile and desktop environments, whereas the SpaceTag system is an information-accessing mechanism for mobile users and has little meaning for desktop users. Social mixed reality systems (Crabtree et al., 2004; Okada et al., 2001) can enable mobile and desktop users share the same physical space, though they are not sufficient for the agent-based user testing. Our simulator, on the other hand, is equipped with a wrapper for integrating mobile and desktop users, and software agents. Furthermore, they can become the subjects of experiments, since the simulator is connected to the implemented smart environment, which is our guidance system.

When occupation and replication of the environment are impossible, there is an alternative to our method: exploitation of a similar environment. For example, one study uses a university building as a structural copy of a shopping mall.
Virtual Cities for Simulating Smart Urban Public Spaces

to evaluate a shopping guide system (Bohnenberger et al., 2002). To use this method, one has to be satisfied with a rough mock-up of the actual space unless an existing physical space is found whose structure imitates the space precisely. In our method, it is possible to use the actual space if the effort is made to construct a 3D model of the space and describe the scenario for group behavior. Our simulator FreeWalk/Q (Nakanishi and Ishida, 2004) is a combination of a virtual city “FreeWalk” and a scenario description language “Q” (Ishida, 2002b). Q is based on Scheme, which is a dialect of the Lisp programming language invented by Guy Lewis Steele Jr. and Gerald Jay Sussman. Q’s model is an extended finite state machine. According to the scenario written in Q, agents can walk, speak, and gesture in FreeWalk’s virtual space. Since FreeWalk/Q can simulate many kinds of group behaviors including evacuation, our method can be applied to various smart environments installed in crowded places.

CONCLUSION

We proposed agent-based user testing, that is a method for conducting experiments to test smart environments installed in large-scale crowded urban public spaces such as airports and central railway stations, since it is difficult to replicate such environments in a laboratory. Such places do not allow us to ask visitors to participate in our experiments or prevent visitors from entering so that we can conduct our experiments with selected subjects. For this method, we developed a virtual city simulator integrating agents, humans, and avatars. In testing the emergency guidance system installed in Kyoto’s central railway station, the simulator overlaid a virtual crowd consisting of a hundred agents, three humans, and three avatars on the station platform so that the experiment could be carried out with little interference to the station’s daily operation. The result of the experiment supported our hypotheses that our location-based guidance system installed in Kyoto station was superior in trustworthiness, usefulness, and inducing calmness. This motivated us to conduct a totally physical experiment to confirm these effects.

ACKNOWLEDGMENTS

We express our thanks to the Municipal Transportation Bureau and General Planning Bureau of Kyoto city for their cooperation. This work would have been impossible without the invaluable participation of Toyokazu Itakura, Ryo Watanabe, Shinji Konishi, Shunsuke Tanizuka, Takatoshi Oishi, and Armando Rubio Torroella. We thank Shigeyuki Okazaki, Toshio Sugiman, Ken Tsutsuguchi, CRC Solutions, Mathematical Systems, and CAD Center for their support in the development of the guidance system and the social interaction platform. The platform is available at http://www.ai.soc.i.kyoto-u.ac.jp/freewalk/ and http://www.ai.soc.i.kyoto-u.ac.jp/Q/.

REFERENCES


Virtual Cities for Simulating Smart Urban Public Spaces

Software and Technology, 97-104.


**KEY TERMS**

*Augmented reality*: overlapping virtual objects with a physical environment in order to provide additional information to people in the environment

*Avatar*: a graphically or physically embodied representation of a human user for virtual or augmented environments

*Augmented experiment*: an experiment in which virtual subjects participate

*Participatory simulation*: a simulation which interacts with human participants

*Pedestrian mode*: a model for determining the direction and velocity of human walking

*Gait mode*: a model for moving human legs

*Multi-agent interaction model*: a model for deciding next behavior based on perceptual information of other agents