AR Navigation System Using Interaction with a CG Avatar

Hirosuke Murata, Maiya Hori, Hiroki Yoshimura, and Yoshio Iwai

Graduate School of Engineering, Tottori University
101 Minami 4-chome, Koyama-cho, Tottori, 680-8550, Japan

Abstract. This paper describes a navigation system that is guided by a CG avatar using augmented reality (AR) technology. Some existing conventional AR navigation systems use arrows for route guidance. However, the positions to which the arrows point can be unclear because the actual scale of the arrow is unknown. In contrast, a navigation process conducted by a person indicates the routes clearly. In addition, this process offers a sense of safety with its expectation of arrival at the required destination, because the user can reach the destination as long as he/she follows the navigator. Moreover, the user can communicate easily with the navigator. In this research, we construct an AR navigation system using a CG avatar to perform interactively in place of a real person.

1 Introduction

Guidance systems based on display devices that are fixed in their environment, such as digital signage, can be found in public spaces. These systems require installation spaces and have high equipment costs. Guidance systems using mobile phones have been developed as one method to overcome these problems. As an example of the guidance systems that are offered by mobile devices, sophisticated applications using augmented reality (AR) technology are widely used, although they depend on the level of technological advancement of the device. AR is a technology that provides a digitally enhanced view of the real world. Conventional AR navigation systems[1][2] use arrows for guidance along a route. However, the positions to which the arrows are pointing can be unclear because the actual scale of the arrow is unknown. In contrast, when navigation is guided by a person, it is easy to understand the directions given intuitively. Furthermore, this navigation mode provides a sense of safety with the expectation of arrival at the desired destination, because the user can reach the destination as long as he/she follows the navigator. In addition, the user can communicate directly with the navigator.

In this research, we implement a CG avatar using AR technology to perform interactively with the user in place of a human navigator. It is easy for the user to understand the directions intuitively through interactions with the CG avatar. Additionally, the reliability can be improved by giving the CG avatar a favorable image.

R. Shumaker and S. Lackey (Eds.): VAMR 2014, Part I, LNCS 8525, pp. 274–281, 2014.

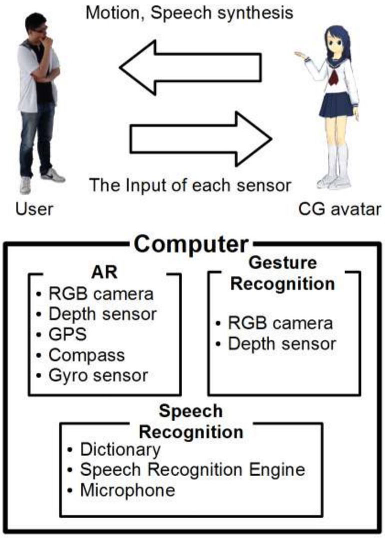


Fig. 1. The concept of the proposed method

2 AR Navigation System Based on Interaction with a CG Avatar

Figure 1 shows the concept of the proposed method. The user can communicate by speech and gestures with a CG avatar that has been superimposed on the real world. These communications are realized using speech recognition, gesture recognition and AR technologies. We use various sensors to monitor the user's behavior. Movements and speech are generated for the CG avatar that correspond to this behavior. The user, wearing a head-mounted display (HMD), is then guided by the CG avatar in real time. The CG avatar can perform several actions, including walking, waiting, pointing, gazing and greeting. It is also necessary to show that the CG avatar is standing on the ground for realism and we must consider the occlusions that are derived from real objects such as walls and buildings. These occlusions are realized by detecting the ground and measuring the depths of the objects using a depth sensor.

2.1 Speech Recognition

Speech recognition technology is required to conduct a conversation and provide explanations using the CG avatar. The user's speech can be recognized using a large vocabulary continuous speech recognition engine, such as Julius[3]. Various set speeches are generated in advance using a speech synthesis technique[4]. After the speech recognition process, the corresponding required speech is then spoken by the avatar.

2.2 Gesture Recognition

Gestures play an important role in communication. The user's gestures are recognized using a 3D hand pose estimation method, such as [5]. The avatar's various actions to be taken when guiding the user are generated in advance, similar to the method used for the avatar's speech. After the user's gesture is recognized, the corresponding required actions are performed by the avatar using AR techniques.

2.3 AR Technology

To overlay CG objects on the real world using AR technology, the application needs to know both the camera position and the camera orientation. There are several methods that can be used to estimate the camera position and orientation, including vision-based methods[6] and sensor-based methods. Because our system is intended to be used in both indoor and outdoor environments, the camera position and orientation are estimated using various sensors. A Global Positioning System (GPS) device, a gyroscope, and a compass are used for the estimation process in outdoor environments. In indoor environments, a vision-based method, a wireless LAN-based system and an Indoor MEssaging System (IMES)[7] are used for the estimation process. By switching between these methods, depending on the environment, the system can be operated over a wide area. The CG avatar can be superimposed on specific locations using the estimated camera position and its orientation.

To realize the interactions with the CG avatar, it is necessary to implement various avatar motions. The required motions are as follows. The movements that are necessary for the avatar itself are, "walking", "run", "go upstairs", "go downstairs", and "waiting". As motions that are required for the guidance process, "greeting", "hand gesture", "eye contact", and "representation of emotions" must all be implemented. Additionally, to make the user feel as if the CG avatar exists in the real world, we need further techniques such as detection of the ground and occlusion culling. These functions are achieved by using a depth sensor to obtain depth information for objects in the real world.

3 Experiment

In our experiments, we show the results of the construction of our navigation system using the CG avatar in both the indoor and outdoor environments. In the indoor environment, the CG avatar gave a poster presentation instead of a real person, as shown in Fig. 2. Occlusion culling is considered based on use of the available depth information, as shown in Fig. 3. The camera position and the posture are estimated using a magnetic sensor (3SPACE FASTRAK, Polhemus). The occlusion culling process is implemented using a depth sensor (Xtion Pro Live, ASUS).

(a) Turning the avatar's gaze on the user (b) Explaining the description of the poster

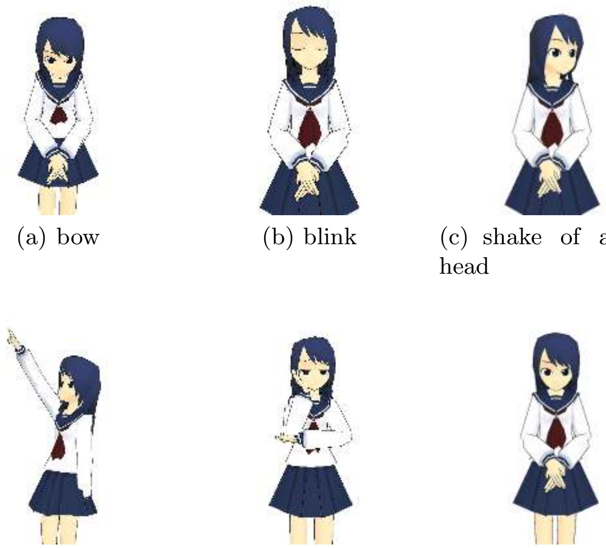
Fig. 2. Examples of a poster presentation

(a) Depth image (b) Rendered image

Fig. 3. Example of the CG avatar when considering occlusions

The interactions of the CG avatar are conducted using the Wizard of Oz (WoZ) method[8]. The operator controls the CG avatar in response to the user's behavior and speech. The CG avatar contains 120 bones, and various avatar movements can be implemented as shown in Fig. 4. These movements are played along with the speeches that were generated previously, such as a greeting and an explanation of the presentation. The CG avatar can also cope with the user's questions by generating speech interactively via a keyboard input. In a demonstration, the CG avatar was able to give a presentation coupled with interactions with the users. From assessment of some of the users' opinions, we found that the users could feel like the CG avatar actually existed in front of them.

In the outdoor environment, the CG avatar acts as a guide. Figure 5 shows the user's appearance in the outdoor environment. The camera posture is acquired using a compass and a gyroscope. The CG avatar is superimposed in such a position that the avatar appears to be leading the user. In this experiment, the navigation processes are conducted by an operator using the WoZ method. The user looks at the walking avatar and decides on the direction of movement at some forks in the road. The navigation was carried out using five paths that were determined in advance, and these paths are unknown to the users. One path



(a) bow (b) blink (c) shake of avatar's head

(d) pointing (e) thinking (f) nodding

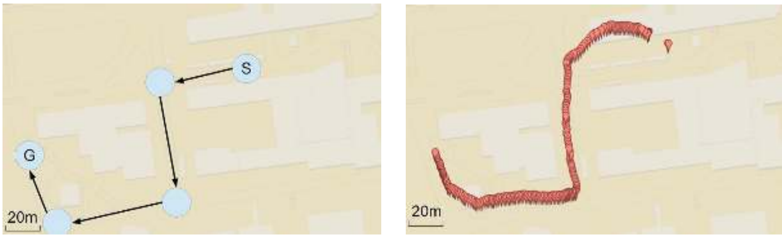
Fig. 4. Movements of the CG avatar

is selected in random order for each user. There are three forks in each path. Figure 6(a) shows an example of the ground truth that was predetermined by the operator. The trajectory of the user's movement acquired by GPS is shown in Fig. 6(b). From this figure, we see that the navigation has been conducted correctly. Examples of the user's view are shown in Fig. 7. The numbers of branches were three at points A and B, and two at points C and D. Users can arrive at the desired destination by simply following the CG avatar, as shown in Fig. 7.

To investigate the validity of the proposed method, subjective evaluations using questionnaires were conducted with 15 subjects. For comparisons, the following five methods were implemented.

- (a) Person
  A person guides the user, who is wearing the experimental apparatus.
- (b) CG avatar (the proposed method)
  A CG avatar guides the user, and the avatar exists at all times.
- (c) CG avatar only at forks
  A CG avatar appears only at forks in the road.
- (d) Arrows
  Arrows guide the user, and the arrows exist at all times.
- (e) Arrows only at forks
  Arrows appear only at forks in the road.

Fig. 5. User's appearance



(a) The ground truth (b) GPS data

Fig. 6. The trajectory of the user's movement

The questionnaires are used to rank the methods in terms of "Sense of safety", "Likability of the system", "Availability", "Reliability", "Visibility", "Responsiveness", and "Likability of the navigator". The evaluation was rated with scores from 1 (not felt at all) to 6 (felt extremely strongly).

The results of the questionnaires are shown in Fig. 8. From the results of the questionnaires, the proposed system exceeds the navigation by arrows method in terms of "Sense of safety", "Likability of the system", "Reliability", and "Likability of the navigator". With respect to "Likability of the system", the proposed system exceeds all other methods, including navigation by a person. The "Reliability" of the proposed method has a higher score than the methods where navigation guides appear only at forks in the road. In terms of "Visibility" and "Responsiveness", no significant differences can be observed between the methods.

(a) Point A

(b) Point B

(c) Point C

(d) Point D

Fig. 7. User views at each fork in the road

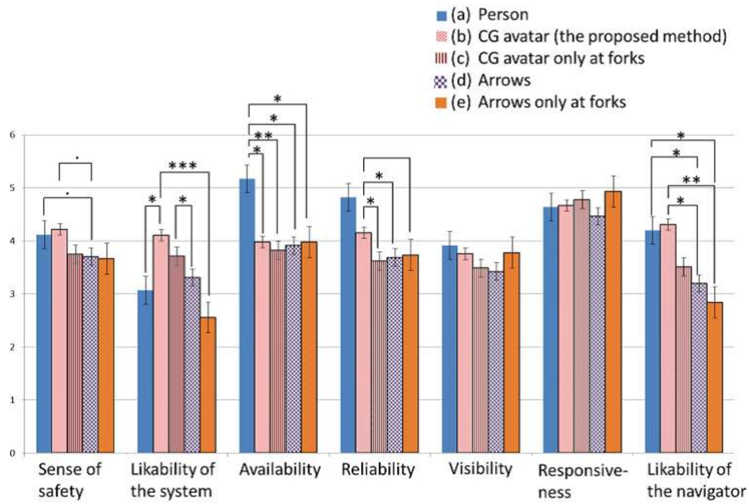


Fig. 8. Results of the questionnaires

4 Conclusion

This paper describes a navigation system where the user is guided by a CG avatar using AR technology. In the indoor environment, a poster presentation given by the CG avatar was demonstrated and the importance of user interactions has been shown. In the outdoor environment, the CG avatar acted as a guide and the navigation process was conducted correctly. From the results of subjective user evaluations, it was found that the proposed system has high likability and was preferred to the methods based on navigation by arrows in terms of their sense of safety and reliability. In our future work, we intend to automate the process that was carried out using the WoZ method in this work.

Acknowledgment. This work was supported by JSPS KAKENHI (Grant No. 24118705).

References

1. Kanbara, M., Nagamatsu, A., Yokoya, N.: Augmented reality guide system using mobile projectors in large indoor environment. In: Proc. Workshop on Personal Projection via Mobile and Wearable Pico Projection, pp. 16–19 (2010)
2. Yahoo!japan, `http://maps.loco.yahoo.co.jp/promo/spn/map/`
3. Julius, `http://julius.sourceforge.jp/en_index.php?q=index-en.html`
4. Open JTalk, `http://open-jtalk.sourceforge.net/`
5. Oikonomidis, I., Kyriazis, N., Argyros, A.: Efficient model-based 3D tracking of hand articulations using kinect. In: Proc. of the British Machine Vision Conference, pp. 101.1–101.11 (2011)
6. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: Proc. International Symposium on Mixed and Augmented Reality (2007)
7. Manandhar, D., Kawaguchi, S., Torimoto, H.: Results of IMES (indoor messaging system) implementation for seamless indoor navigation and social infrastructure platform. In: Proc. of Int'l Technical Meeting of the Satellite Division of the Institute of Navigation, pp. 1184–1191 (2010)
8. Fraser, N.M., Gilbert, G.N.: Simulating speech systems. In: Computer Speech and Language, pp. 81–99 (1991)