Mobile Omnidirectional Monitoring System for Remote Personal Security

Tomoya Ishikawa[1], Kazumasa Yamazawa[2], Takeshi Kurata[1], Naokazu Yokoya[2]

[1]National Institute of Advanced Industrial Science and Technology (AIST), Japan

[2] Nara Institute of Science and Technology (NAIST), Japan

{tomoya-ishikawa, t.kurata}@aist.go.jp, {yamazawa, yokoya}@is.naist.jp

Abstract

This paper describes a mobile omnidirectional monitoring system that enables an observer such as a parent to keep watch on a subject such as a child walking around city-wide outdoors easily from a remote site. The subject is equipped with an omnidirectional camera, GPS (Global Positioning System) receiver, and microphone. Receiving those data via nationwide broadband mobile network, the remote observer cannot only confirm the subject's position measured by GPS on the maps, but also can watch the subject's view as well as the scenery around the subject generated from omnidirectional video images in different ways. Thus, the system allows the subject and remote observer to talk to each other while sharing information on the situation of the subject and the surroundings. In experiments, we demonstrated that the proposed system can be applied for subject-centered remote surveillance and monitoring with simultaneous presentation of auditory, visual, and positional information.

1. Introduction

Auditory, visual, and positional information is quite important for situational awareness. By transferring such information taken with sensors in real time to remote sites, there will emerge a lot of applications on telepresence, communication, navigation, CSCW [1,2], security, etc. The emergence had been prevented by the bottle neck of network bandwidth of which uplink is about 14 kbps or lower in 2G cell phone networks that is not enough for transferring rich information like video streams. In recent years, however, 3G or later cell phone networks realize wireless high-speed data communication with 384 kbps uplink or more. We therefore come to be able to deploy situation-aware anywhere.

The widespread wireless broadband network has prompted companies to release a number of networked applications [3,4,5]. For instance, Google Maps [3] with GPS allows users to see their own locations by maps or satellite images on the users' phone or other mobile devices. Video Share [4] presents live or recorded videos transferred from remote users. The communication using videos is effective for understanding situations of remote sites.

In Japan, several IT services such as *imadoco* search [5] (*imadoco* means "Where are you now?") have been released taking into account the recent social context including many crimes that involve children. By employing those services, parents can search for the location of their children who are bringing a GPS-compatible cell phone and confirm their positions on maps. However, the representation of child's location is insufficient to understand the details of current situations such as "What do you do?" and "Who are with you?". Furthermore, the system that uses only GPS does not work well in indoor environments and underground facilities.

In this paper, we propose a subject-centered remote monitoring system so that the parents can understand the situation of their children intuitively. In our scenario, a subject such as a child wears an omnidirectional camera, microphone, handy GPS receiver, and mobile PC. The mobile PC acquires omnidirectional videos around the subject, voice, and positional information. They are transferred to an observer such as the parent via broadband cell phone network. The observer cannot only check the position of the subject on maps but also can watch sceneries around him/her in arbitrary view directions interactively. The subject and the observer can also converse with each other. The concept of our system is to enhance understanding of the child's situation by the correspondence among the position, view, and the conversation in an intuitive way.

In the following, section 2 gives the details of the proposed system and the implementation. Section 3 presents experimental results, and finally, section 4 concludes with the summary and future works.

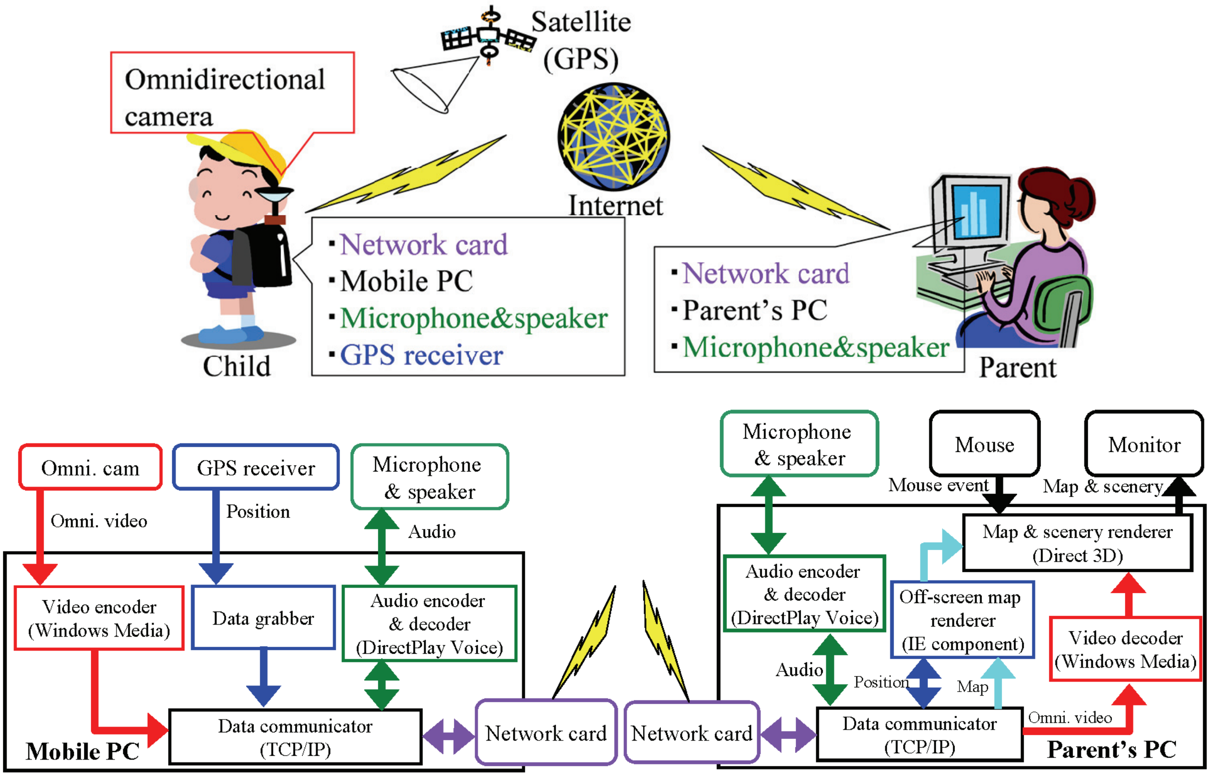


Figure 1. System overview and data flow.

2. GPS positioning and omnidirectional monitoring

2.1. System overview

In our scenario, the proposed system consists of the child-side (subject-side) and the parent-side (observer-side) equipments as well as widespread infrastructures (see Figure 1). As described above, the child-side equipments include a mobile PC, omnidirectional camera, microphone, speaker, GPS receiver, and network card. On the other hand, the parent-side equipment setting is quite conventional such as a PC with a mouse, monitor, microphone, speaker, and network card. In this system, a broadband cell phone network is employed for data transmission so that it can work nationwide and transfer rich contents.

Omnidirectional video, GPS, and audio data are acquired by the mobile PC connected to a camera and sensors. The video and audio are encoded by Windows Media Encoder [6] and the DirectPlay Voice [7] functions. The encoded data and GPS data are transferred to the parent's PC after they are divided into packets.

The parent's PC receives transferred data and presents them to the user. The handling of the audio data is the same as in the mobile PC. The received omnidirectional video is decoded and a view renderer transforms it into planar perspective video images according to the mouse operations (the detail is described in Section 2.2). The received positional data are used for getting map images from Google Maps [3] (the detail is described in Section 2.3). The maps and the views are arranged onto the monitor in various ways. Our system presents auditory, visual, and positional information to the parent simultaneously and interactively to prompt understanding of the details of the child's situation.

2.2. Interactive monitoring using omnidirectional video

Visual information is one of the most important factors to understand what is going on in our daily lives. Such information has been used for remote cooperative works in the field of CSCW [1,2,8]. In particular, video conferencing for inspection and guidance purposes in industries uses standard perspective cameras [8] because they have to shot a specific part as clear as possible. But, in the purpose of personal security, we are considering that surrounding information is more important than a specific part of the scene. Ogawa et al. [2] proposed the telepresence system that enables the observer to control the viewpoint in the virtual environment textured from transferred live videos captured by a standard perspective camera. Their system needs a 3D scene model generated in advance and can not take current surrounding visual information except for view-direction of the camera. In the proposed system, such visual cues are provided efficiently using omnidirectional videos. The omnidirectional videos capturing wide field-of-view (FOV) sceneries help parents to know where the child is heading to and what he/she is doing. Moreover, the parents can observe the sceneries in desired view-directions freely and interactively.

In this system, we use HyperOmni Vision [9] for capturing omnidirectional videos and apply our proposed technique [10] to rendering the arbitrary views from omnidirectional videos. The technique enables real-time view rendering using hardware acceleration of a GPU (Graphics Processing Unit). Figure 2 shows an example of planar perspective transformation of an omnidirectional image. The transformation can be carried out at video rate on a standard PC. The parents can change the view-direction interactively using the mouse operation of dragging on the displayed planar perspective images. Because of omnidirectional videos, the parent can foresee risks comprehensively and prevent the child from taking the risks by voice communication.

2.3. Presentation of maps and omnidirectional videos for intuitive remote monitoring

Presentation of maps pointing to the child's location enables the parent to check it in a wide range depending on the accuracy of GPS. Our system compensates for the measurement error of GPS by displaying both maps and views from the camera.

In the parent-side, this system first calls a web browser such as Internet Explorer internally after receiving the position data from the child-side. Then, the system inputs the URI of Google Maps including place information (latitude, longitude) and map attributes (map type, zoom level) to the browser as a query. According to the query, the browser renders the map into an off-screen buffer and the system captures the rendered map as an image. The captured image is used in map-and-scenery renderer as a texture of the map.

The map-and-scenery renderer displays the map and the planar-perspective scenery image transformed from the omnidirectional video onto the monitor. We have implemented three view modes consisting of bird's-eye view, top view, and child's view modes. Examples of each mode are shown in Figure 3.

The bird's-eye view enables the parent to observe both a map and a moderately-sized scenery image. The system initially displays this view mode. When the parent clicks on the map, the bird's-eye view is changed to the top view mode. The top-view mode makes it easy to confirm the child's location in a wide range, but the scenery from the camera becomes

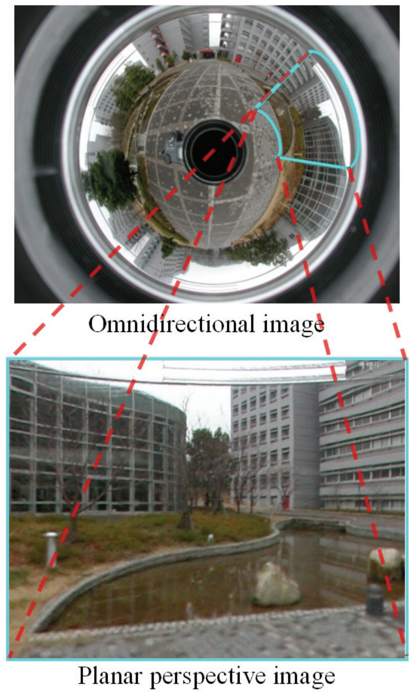


Figure 2. An example of transforming an omnidirectional image into a planar perspective image.

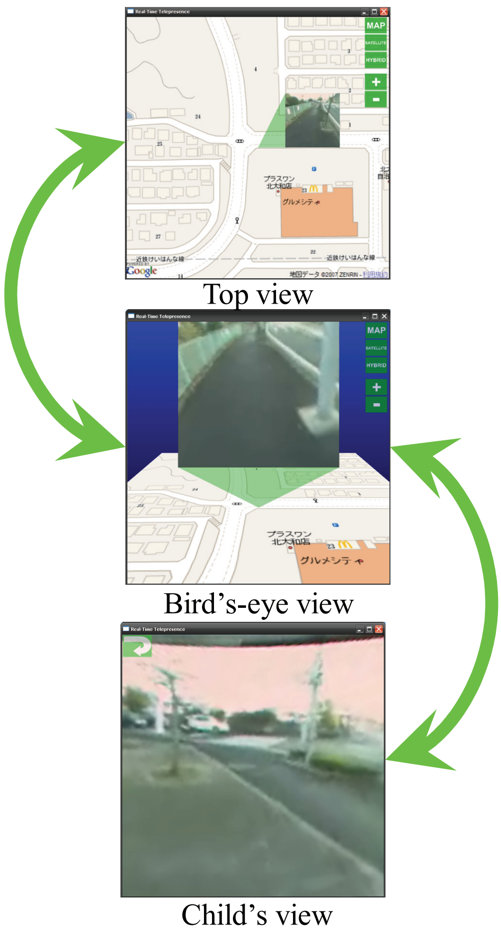


Figure 3. Examples of three view modes.

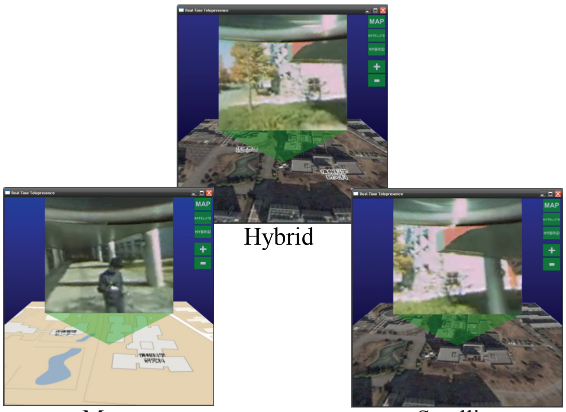


Figure 4. Selectable map types.

smaller. In the bird's-eye-view mode and top-view mode, the parent can change the map type by Google Map APIs. There are three types: map, satellite image, and hybrid (See Figure 4). When the parent clicks on the displayed scenery from the camera in the bird's-eye-view mode, it is changed to the child's-view mode. This mode fully displays the scenery generated from omnidirectional videos. In this mode, the parent can control the view-direction interactively by the mouse operation. This mode enables the parent to watch the scenery which the child looks at, like being together.

2.4. Voice communication

Voice communication is the best way to draw someone's attention directly to potential risks and appropriate routes. For realizing voice communication, we have to encode, transfer, and decode voice data in real time without long delay. We realize it using DirectPlay Voice [7] functions. They automatically detect users' voice and transfer the voice to the other user after encoding it. We can reduce the network load and latency compared with using Windows Media Encoder, because the DirectPlay Voice is for live communication rather than just media streaming, and also the functions transfer voice data only when the user is talking.

3. Experiments

We have implemented a prototype system and carried out experiments. In the experiments, we assume that a child walks around city-wide outdoors and the parent monitors the situation of the child in a remote indoor site.

Figure 5 shows a subject taking the role of a child and a remote observer taking the role of the parent. We assume that a child is equipped with an omnidirectional camera on a school bag as illustrated in Figure 1, but the subject was equipped with the camera on the head in the experiments. The omnidirectional camera made by Suekage Inc. has

Figure 5. Subject and remote observer.

VGA resolution and can capture videos at 15fps. A mobile PC (IBM: ThinkPad X31, Pentium M 1.4GHz) with a wireless network card (E-MOBILE: downlink 3.6Mbps, uplink 384kbps) and a Bluetooth GPS receiver (GlobalSat: BT-338) were in a bag or in a pocket. The subject was also equipped with a Bluetooth headset (SONY: VGP-BRM1) as a microphone and speaker. The remote observer in the indoor site uses a laptop PC (SONY: VGN-SZ90PS, Core Duo 2300 1.6GHz) as a parent's PC with a wireless LAN card (IEEE802.11g), mouse, microphone, and speaker.

The subject walked for a 2 km route, which was about 30 minutes long. Figure 6 shows the displayed views on the laptop PC in the indoor site and the corresponding appearances of the walking subject.

The prototype system worked successfully during the experiment. In spite of the highly compressed omnidirectional video (about 300kbps), the displayed image quality was enough to understand the situation. But, when the subject looked around, the observer got a feeling of seasickness by watching the child's view. Because of the coordinate system of presented view-directions depending on the subject's head, the views had been changed without the observer's intention. To avoid the seasickness, we should use a stabilization technique for child's views such as the technique proposed in [1] and should equip with a gyro for video stabilization.

The current system requires long latency, about 10 seconds between capturing videos in subject's side

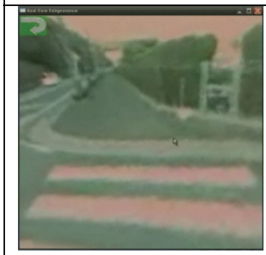
Displayed views on the parent PC	Appearances of the walking subject
	
	
	

Figure 6. Displayed views and appearances of the subject.

and displaying the video images, so that the observer had to forecast the subject's present position during talking with the subject. The latency was derived from Windows Media Encoder and the decoding function in Direct Show. We must consider the change of the video compression method.

We also conducted the same experiment as a live demonstration in a PRMU/CVIM/SIG-MR joint workshop held in Ryukoku University in Japan. In both experiments, we were able to demonstrate that the remote observer is capable of keeping watch on the situation of the subject with our prototype system.

4. Conclusions

In this paper, we proposed a real-time positioning and monitoring system using omnidirectional videos and wireless broadband network for remote-personal security by integrating maps, omnidirectional videos, and voice. The future work includes miniaturizing the mobile devices [11], and evaluating the system with actual test subjects of parents and children. In addition, we will support the function that enables an observer to observe two or more subjects for a purpose of CSCW.

One thing that we have to mention for future improvements is the latency of video streaming. As explained in section 2.4., live audio communication is realized with DirectPlay Voice. However, the total latency of video streaming often comes up to around 10 seconds mainly because the encoding/decoding delay is considerably long. Since the wide FOV of the omnidirectional videos made it easy to observe the remote site, voice communication was well organized even with such long latency of video steaming in each experiment described above. Although it resulted in proving the advantage of the omnidirectional videos, the latency issue should be alleviated with live video communication technologies.

References

[1] N. Sakata, T. Kurata, T. Kato, M. Kourogi, and H. Kuzuoka, "WACL: Supporting Telecommunications Using Wearable Active Camera with Laser Pointer", *In Proc. ISWC2003*, pp.53-56, 2003.

[2] T. Ogawa, K. Kiyokawa, and H. Takemura, "A Hybrid Image-based and Model-based Telepresence System Using Two-Pass Video Projection onto a 3D Scene Model", *In Proc. IEEE and ACM Int. Symp. On Mixed and Augmented Reality*, pp.202-203, 2005.

[3] Google Inc., "Google Maps", http://www.google.com/gmm/index.html.

[4] AT&T Inc., "Video Share", http://www.wireless.att.com/learn/messaging-internet/media-entertainment/attvideoshare.jsp.

[5] NTT DoCoMo, Inc., "Imadoco search", http://www.nttdocomo.co.jp/english/service/gps/imadoco.

[6] Microsoft Inc., "Microsoft Windows Media", http://www.microsoft.com/japan/windows/windowsmedia/9series/encoder/default.aspx.

[7] Microsoft Inc., "DirectX resource center", http://msdn2.microsoft.com/engb/xna/aa937781.aspx.

[8] B. Hestnes, S. Heiestad, P. Brooks, and L. Drageset, "Real situations of wearable computers used for video conferencing – and implications for terminal and network design", *In Proc. ISWC2001*, pp.85-93, 2001.

[9] K. Yamazawa, Y. Yagi, and M. Yachida, "Omni-directional imaging with hyperboloidal projection", *In Proc. Int. Conf. on Intelligent Robots and Systems*, vol.2, pp.1029-1034, 1993.

[10] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya, "Telepresence by real-time view-dependent image generation from omnidirectional video streams", *Computer Vision and Image Understanding*, vol.171, no. 2, pp.154-165, 1998.

[11] Y. Mukaigawa, Y. Nishioka, and Y. Yagi, "Development of mobile omni-alarm using compound omnidirectional sensor", *In Proc. MIRU2006*, pp.722-727 , 2006 (in Japanese).