NICOGRAPH International 2012, pp. 94 - 101

Forensic CG Video Generation with Augmented Reality Technology

Atsushi Sugiura Masahiro Toyoura Xiaoyang Mao

University of Yamanashi

{g12dhl02, mtoyoura, mao} (at) yamanashi.ac.jp

Abstract

This paper presents a technology for supporting in-house forensic CG video generation by police and prosecution. For confidentiality, it is usually not allowed to generate criminal CG video through subcontracting, whereas existing CG modeling and video editing systems are too complicated to use for police or prosecution. The proposed technique solves this problem with Augmented Reality (AR) technology. We propose to deploy multiple cameras and AR markers on criminal site. While a user of the police or the prosecution plays a role of the criminal wearing a video see-through head mount display, video is automatically generated. CG objects, such as the virtual victim, are displayed on the head mount display and rendered in the video. Our main contributions in this research are concurrent displaying of first-person and third-person views for visualizing spatial relationship between the user and its surroundings, and complementary use of marker information from multiple cameras for solving the occlusion problem inherently caused by operating at the actual criminal site.

1. Introduction

The lay judge system has been introduced all over the world. The lay judges are chosen from citizens in general. They discuss about guilt or innocence and the appropriate punishment with professional judges. The system makes the judge to follow the sense of ordinary people. It also makes the period much shorter. Since the lay judges are amateur, the materials for them are required to be easily understandable without expert knowledge.

We focus on the demonstration CG video as a part of forensic materials, which enables the judges to understand the criminal situation intuitively. Although it is already generally used in TV programs, it has not been frequently used in the court. This is because subcontracting is inappropriate for confidentiality, while police or prosecution is not expertise for generating movie. We propose to support in-house CG video generation with Augmented Reality (AR) technology.

AR technology support to display CG objects in video captured real world. Supplementary information can be added in the video. We suppose multiple cameras and multiple markers are deployed on the criminal site. A user of the police or the prosecution plays the role of the criminal wearing a video see-through head mounted display. CG objects are rendered in the video captured criminal site and displayed to the head mounted display of the user. All things the police or the prosecution should do for generating CG video is playing a role of the criminal on site. Modeling of human motion and actual environment can be omitted in this situation. Although the AR markers may be occluded by the user and real objects, complementing marker information among multiple cameras can solve the problem. Displaying multiple views from the multiple cameras will help the user to grasp the spatial relationship between the user and its surroundings, even if the user is not well trained for using AR system.

In the next section, we first introduce related works. Section 3 describes how the proposed system generates CG video. The results of a subject study are described in Section 4, and Section 5 concludes the paper.

2. Related works

Several research groups have proposed to use CG models as forensic materials. Section 2.1 introduces the existing work on creating CG models based on the investigation data of criminals.. Section 2.2 briefly reviews the medical application of AR.

We use multiple views to help the user understand the spatial relationship between himself/herself and surroundings. The system consists of multiple cameras and multiple PC connected through a client/server network. An existing work on multiple view AR is described in Section 2.3. The research work which using AR technology in a network environment is introduced in

2.4.

2.1 CG Modeling

Creating the 3D CG model of criminal site requires technical skills. It also costs much time. March et al. proposed a technique for converting forensic pathology descriptions into 3D CG models [1]. Thali et al. generated individual models of corpuses using markers [2]. Captured images, MRI (Magnetic Resonance Imaging) and CT (Computed Tomography) data are integrated to create the 3D CG model with the markers. Compared to the corpuses, criminal site and the animation depicting the process of a crime are much more difficult to model since they varies case by case and there are no standard models are available. We solve the problem by using AR technology. In our system, the CG model of victim, which can usually be obtained by modifying a standard model, is rendered into the captured image of crime site, and a video depicting the process of crime can be automatically generated simply by having a user playing the role of criminal. CG objects can be arranged in the crime site intuitively by using any markers.

2.2 Medical applications of AR

Augmented Reality technology has been adopted to an endoscope surgery supporting system [3][4]. A small hole on the surface of a patient's body is made for the endoscope surgery. An endoscope and forceps are inserted into the hole. The operation is done with a narrow view of the endoscope. The surgery is skill demand and requires advanced training. AR technology supports the surgery. Virtual internal organs are generated from CT images of the patient before the operation. They are projected on the surface of the patient's body during the operation. The deformable registration between the patient's body and the virtual organs is given with optical markers. Since human body is a non-rigid object, the registration is usually very difficult.

Portable system is desired for our system. Easy calibration of multiple cameras is very important in case of in-house movie generation. We achieve this by adopting ARToolKit markers [5] for calibration and registration. Our system is portable and easy to handle.

2.3 Displaying multiple views for AR

How two or more views are displayed in AR has been actively discussed. Sport is a good application of AR. The 3D soccer observation system developed by Inamoto and Saito generates free viewpoint video from images captured with a large number of cameras [6]. A viewer can specify an arbitrary viewpoint and the system generates the corresponding image by interpolating multiple images from different cameras. Cameras are calibrated in advance.

The video generated by interpolating multiple views gives virtual views, which is not suitable for being used as forensic materials since the reliability is most important for the judge. Our system generates third-person views from real cameras. Viewers can change the camera, not the viewpoint, freely. They can examine the situation from the viewpoints of different cameras.

2.4 Practical use of AR and network

There is a research [7] which combines AR technology and a network. It transmits and receives the information on the real world between remote places. It aims at deleting physical restriction, and can perform all operations in real-time.

Our system also performs the data communications between personal computers using a network. To be immersive, real time processing is important. However, there is a delay problem by a network. In order to solve the problem, the data volume for communication was reduced by sending thumbnails of images.

3. Proposed method

Our method generates a video for investigating a crime scene with AR technology. The video consists of the images obtained by rendering the virtual objects, such as the 3D model of a victim, weapon, over the real camera captured images of criminal site. The knowledge about CG modeling is not needed.

When using the actual criminal site as the background, we cannot expect large equipment such as position sensors or a good lighting condition. We should detect AR markers in such situation. The problem is solved by the complement of multiple marker information with multiple cameras. The multiple cameras also support a user to act in the space. The user wears a HMD, and multiple views of the cameras are displayed on the monitor of HMD.

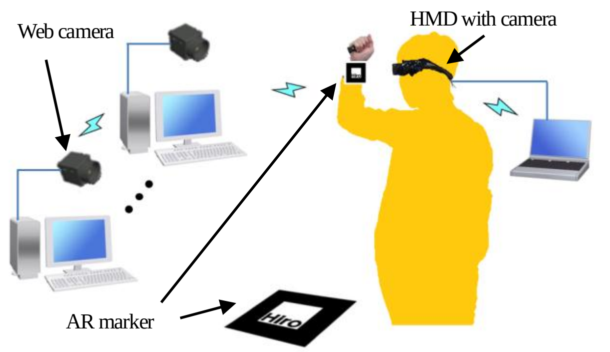
Section 3.1 describes the overview of our system, and Section 3.2 describes the complement of multiple marker information. Section 3.3 is about the user support by displaying multiple views, and Section 3.4 introduces the configuration of server/client network for transferring multiple views.

3.1 Overview

We arrange multiple cameras and multiple AR markers on the actual crime site. The lightings and the position sensors are not available since the space is not a managed studio. The markers are often occluded by the user or the other objects in real world. When multiple cameras are placed in the space, we can expect some of them can detect the marker. Then, marker information can be complemented by integrating the information between cameras. In addition, the multiple cameras provide multiple views for the user to grasp the spatial relationship of the virtual objects and real objects including the user himself/herself.

The configuration of our proposed system is shown in Figure 1. As shown in Figure 1(a), the user wears a head mounted display with cameras (HMD). The video of first-person view is displayed on the HMD. Multiple cameras which are connected to the network are arranged in the space. They capture the video from the third-person viewpoints. The video is transferred to the user's HMD. A composite video with both of the first-person view and the third-person views is provided to the user.

Multiple markers are deployed in the space. The individual markers correspond to individual virtual objects. The marker for a virtual victim is set on the floor, and one for a virtual weapon grabbed in the user's hand. The virtual victim and the virtual weapon are displayed on the positions of the markers in the video. It means the video captured in the actual crime site is augmented with the virtual objects by AR technology.



(a)Before overlaid. The user wears HMD. The multiple cameras and markers are deployed in the space.

(b)After overlaid. Video captured with multiple cameras. The virtual victim and the virtual weapon are displayed on the positions of the markers in the video.

Figure 1.Configuration of the proposed system.

3.2 Complement of multiple marker information

ARToolKit is a library for creating AR environment easily. It detects a marker with one camera, estimates the position and posture of the marker, and displays a virtual object on the position indicated by the marker. However, when a part of the marker is absent from the field of view of the camera or when the mark is in an inappropriate angle relative to the camera, the marker becomes undetectable. We solve the problem by using the information from multiple cameras complementally.

Figure 2 shows the flow of the complement of marker information.

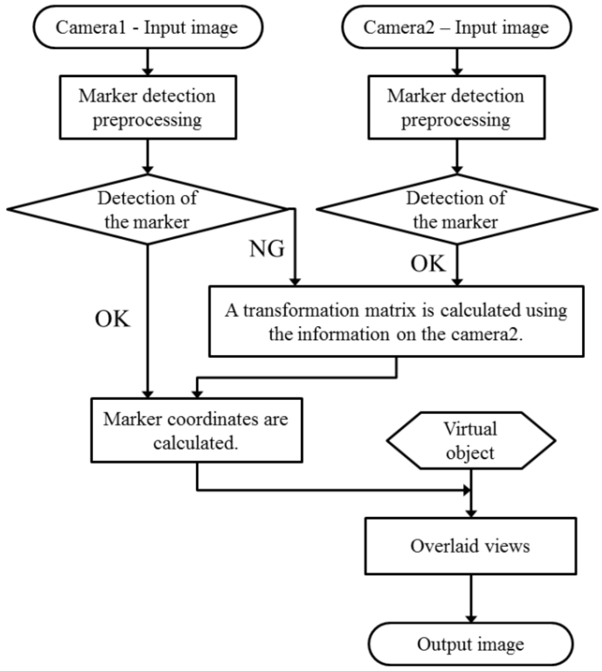


Figure 2. Complement of undetectable marker information. When camera 1 fails to detect a mark, a geometric transformation matrix is calculated using the information from camera 2.

When a marker is detectable, a geometric transformation matrix P can be calculated with the function of ARToolKit for computing the marker coordinate system. Then a virtual object can be displayed on the marker in the marker coordinate system.

Without loss of generality, let`s consider a system consists of two cameras and two markers as shown in Figure 3. Denote the geometric transformation matrix of marker A and B detected from camera 1 are P_{1A} and P_{1B}, the geometric transformation matrix of marker A and B detected from camera 2 are P_{2A} and P_{2B}. Assume camera 1 can detect both marker A and B, but camera 2 can detect marker B only. Because marker A is undetectable from camera 2, geometric transformation matrix P_{2A} cannot be calculated. Assume the camera coordinate system of camera 1 viewed from camera 2 is $\Phi_{2\to1}$, than the transformation matrix of marker B viewed from camera 2 via camera 1 becomes $\Phi_{2\to1} \cdot P_{1B}$. That is, we can obtain the following equations:

$$P_{2B} = \Phi_{2\to1} \cdot P_{1B} \qquad (1)$$

Similarly, the transformation matrix for maker A viewed from camera 2 via camera 1 is

$$P_{2A} = \Phi_{2\to1} \cdot P_{1A} \quad (2)$$

Combining equation (1) and (2) we obtain the following equation:

$$P_{2B} = \Phi_{2\to1} \cdot P_{1B} \quad (3)$$

Since P_{1A}, P_{1B}, and P_{2B} are available, P_{2A} can be computed with equation (3) and the marker coordinate system of marker A viewed from camera 2 can be calculated from P_{2A}.

The principle behind the above calculation is that we can first obtain the spatial relationship between the two cameras through the marker which is detectable by both camera (Equation(1)) and then compute the coordinated system of the occluded marker from its coordinate viewed from another camera. Therefore, to an occluded marker, at least one more marker which is not occluded from both cameras is required. Moreover, there is no way to complement a marker which is occluded from all cameras.

In the proposed system, a network of client/server model is constructed for communicating among cameras. Marker information is detected by the client and sends to a server. A server receives data from a client, when there is a marker which was undetectable, the server complements marker information by using the data from other clients.

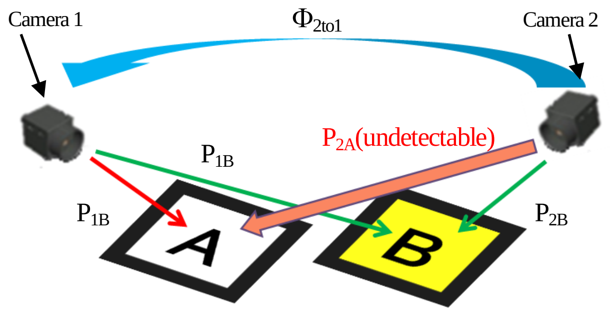


Figure 3. Marker information complement.. The projection metrics by the two cameras and two markers can be calculated by the complement of marker information.

3.3 Displaying multiple views to the user

When creating a synthetic image, the spatial relationship of reality and an virtual object becomes important. It may be difficult for a performer to grasp the whole spatial relationship only with the image of a first person viewpoint. We improve the situation by providing the performer the image of a third person viewpoint. Since the performer can see the image of various viewpoints, it becomes easy for him/her to understand the spatial relationship among virtual objects and actual criminal site, and to set up optimal positions and numbers of cameras and markers, which is usually very difficult for un-specific environment.

3.4 Network by socket connection

A network is used in order to transmit images of two or more viewpoint between personal computers. A server/client model using TCP socket connection is used. The image of the third person view mode acquired with another camera is sent to a server using a network. However, the images captured with the cameras usually are large, and the system can become slow depending on the traffic of network. However, real time processing is very important in case of an AR system. We solve the problem by using small images for the third-person view. The downsizing is executed before the images are transferred.

4. Experiments

We examined how much subjects grasp the spatial relationship between the subjects and virtual objects by displaying third-person views. Complement of marker information between multiple cameras and multiple markers is also examined through another subject study.

4.1 Enviroment

Experimental environment is shown in Figure 4. Three PCs are deployed in a space. A HMD with cameras (Wrap920AR, Vuzix Corporation) is connected to a server PC. The resolution of cameras is 640x480 and frame rate is 30fps. The resolution of display monitors is 800x600.

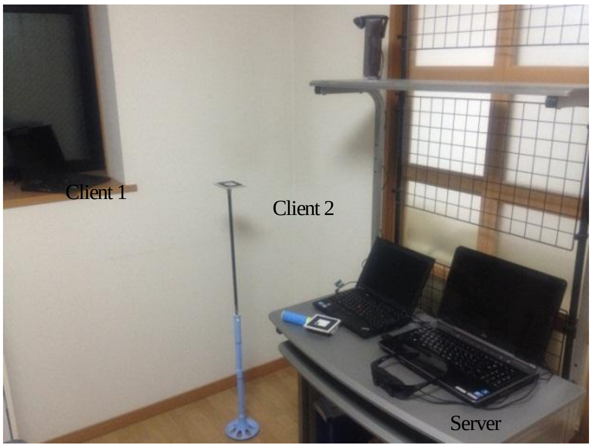


Figure 4. Environment for experiments. A server PC with HMD and two clients PCs with web camera are deployed in a space. The cameras capture the video. The marker information are acquired from the cameras.

The HMD provides the video of the first-person view to a subject. The remaining two PCs are clients with web cameras (Logicool C920 and QCam-200SX) connected. They capture video from the third-person views. The video were 640x480 in resolution and 30fps in frame rate.

Two markers are deployed in the space. One of them represents the position of a virtual victim, and the other the position of a virtual weapon. Figure 5(a) shows the marker for the virtual victim, and Figure 5(b) the marker for the virtual weapon. The size of both markers is 80 square mm. When the marker for the virtual victim was put on the floor, it rarely came into sight of the subject. Therefore, the marker was put on a bar as shown in Figure 5(a). We designed the weapon as shown in Figure 5(b) to correctly calculate occlusion via the virtual victim. The edge of the weapon was rendered as a virtual object.

(a) Victim marker (b) Weapon marker

Figure. 5 Markers for experiments. The size of both markers is 80 square mm. The height of the victim marker is 1.1m. A handle is attached to the weapon marker.

4.2 Complement of marker information

In this experiment, server and client processing is correctly performed among two or more PCs. We examined whether the complement of marker information was realized.

An operation scene is shown in Figure 6. The user wore a HMD and grabbed the weapon marker in one's hand. The cameras on the HMD captured the video of first-person view. The video of the third-person view from the other cameras were transferred to the server PC.

Figure 6. Operation scene. The user wore a HMD and grabbed the weapon marker in one's hand. Multiple cameras capture the video.

The victim marker was arranged in the detectable space from all cameras. In several frames, the weapon marker was out of sight in first-person view. The result of marker complement is shown in Figure 7.

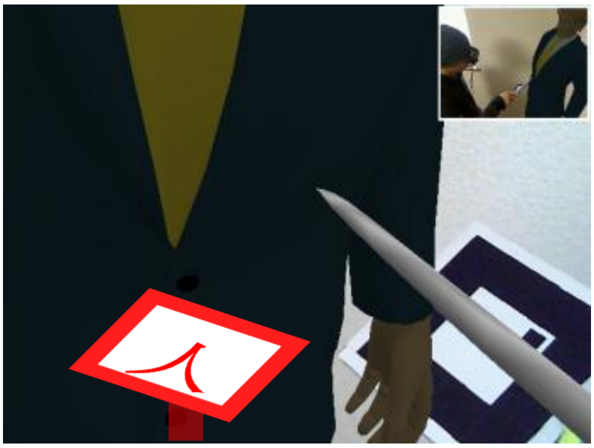
Figure 7(a) shows the result without marker complement process. Since the weapon marker went out of sight of HMD cameras, the virtual weapon could not be displayed. Figure 7(b) shows the result with marker complement process. The virtual weapon could be displayed even if the marker went out of the view. The marker information was transformed to the server PC. The information of undetected weapon marker was complemented by integrating marker information detected in other cameras.

Figure 8(a) represents the first-person view, and Figure 8(b) the third-person view. Based on the victim marker, which could be detected in both two cameras, the weapon marker is complemented in its position although the marker could not be detected in the camera of the first-person view. The information of the relationship between the victim marker and the weapon marker is transferred through the network. As the result, the virtual weapon could be displayed in the first-person view.

(a) Result without complement of marker information. Since the weapon marker went out of sight of HMD cameras, the virtual weapon could not be displayed.

(b) Result with complement of marker information. The information of undetected weapon marker was complemented by integrating marker information detected in other cameras. The virtual weapon could be displayed using the information even if the marker went out of the view.

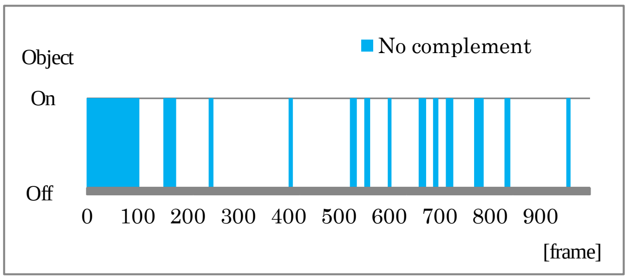
Figure 7. Marker information complement.



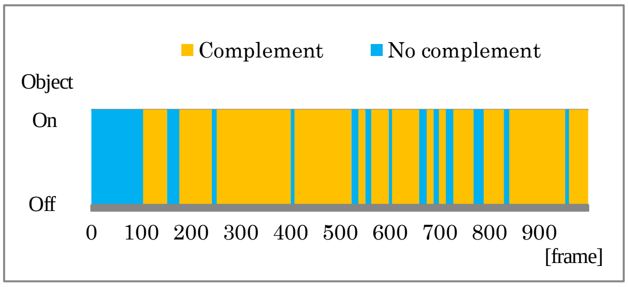
(a) First-person view.

(b) Third-person view.

Figure 8. Complement of marker information based on the victim marker. The marker information calculates using the victim marker of both first-person view and third-person view.



(a) The result without the complement of marker information.



(b) The result with the complement of marker information.

Figure 9. The result of marker detection with/without complement of marker information through the network.

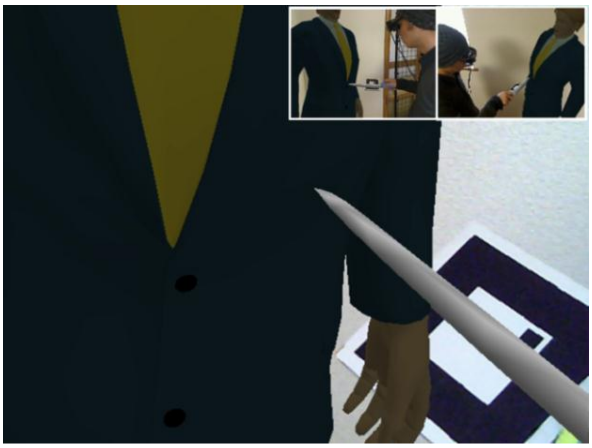
We examined the scene that the subject stabbed the virtual victim with the virtual weapon. Whether the virtual weapon was detected or not is shown in Figure 9. The marker could not be detected in 765 frames without complement of marker information as shown in Figure 9(a). Comparing with the result, the marker could be detected in all frames with complement of marker information as shown in Figure 9(b). We conclude that the complement was effective for displaying the marker for CG video generation in our proposed system.

4.3 User support by displaying multiviews

In this experiment, the ease of grasping spatial relationship between the virtual and real objects by displaying video of additional third-person views was examined.

This system was used by three male subjects and a female subject in their 30's. Figure 10 shows captured images of our system. The third-person views are not displayed in Figure 10(a), and they are displayed in Figure 10(b).

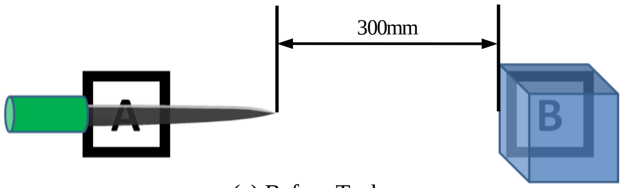
(a) No third-person view. Grasping spatial relationship is difficult with first-person view only.



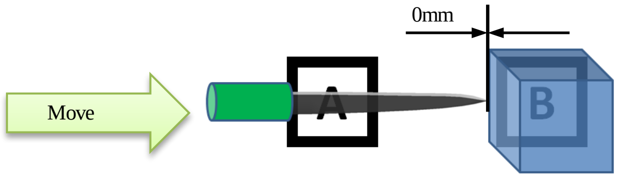
(b) With third-person views. The third-person views are added to the upper right of the image displayed to the HMD. Grasping spatial relationship become easier by referring to the third-person views.

Figure 10. Images displayed to the HMD of a user.

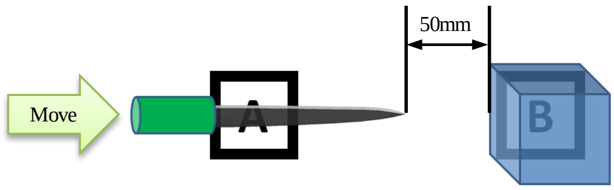
From the comments of all subjects, displaying third-person views helped to easily grasping the spatial relationship between the virtual objects and real objects including the subjects themselves. We had the subjects put the virtual weapon at a specified point to evaluate how they could grasp the spatial relationship. A cube was displayed on the victim marker. The first task was to put the end of edge of the virtual weapon on the surface of the cube. The second task was to put the end at 50mm distance away from the surface of the cube. We examined the accuracy and time to complete the tasks with and without third-person views displayed. When the virtual objects are placed on 300mm distance from each other, we start to measure the time for the tasks. We stop the time when the subjects stopped the motion of the virtual weapon. The scene of the tasks is shown in Fig. 12



(a) Before Task



(b) Task 1

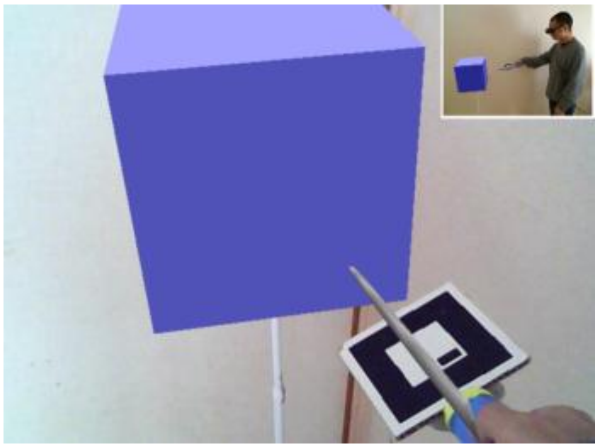


(c) Task 2

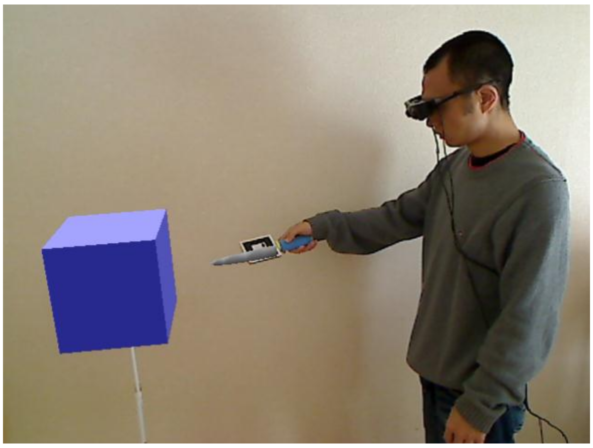
Figure 11. Tasks for evaluating the accuracy and time by displaying third-person views.

Results are shown in Figure 13 and Figure 14. The error of distance between virtual objects became smaller by displaying the third-person views. The time for the tasks also became shorter. Visual support of the third-person views was confirmed to be valid for the tasks.

Based on the problems found through experiments, the system has been improved. The data amount of captured images was large to transfer through the network in real time. The original size of captured images is 640x480. The image size of 160x120 is used for The third-person views. Before transferring, the image size of 640x480 is reduced to 160x120. We solve the problem by transferring only small images (160x120) to the server PC. It reduced the data amount on the network and real time display of multiple views was realized.



(a) With third-person view

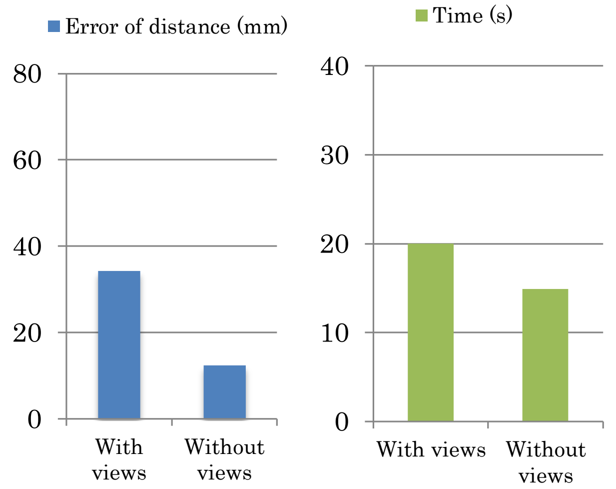


(b) A example of generated image of third-person view

Figure 12. Experiment scene. We evaluated how they could grasp the spatial relationship. The virtual weapon is brought close to a cube. The lapsed time and distance at that time are measured.

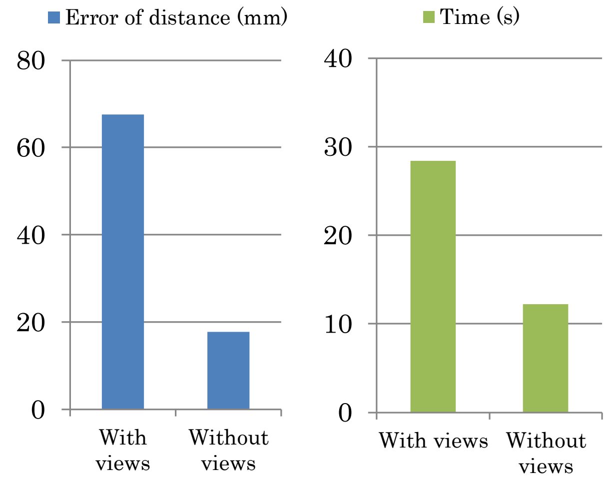
Another problem was that the planer weapon marker as shown in Figure 15(a) could not be detected from many directions. When the marker was rotated, it was often missed from third-person view cameras. We improved the marker as shown in Figure 15(b). Five markers are arranged on a cube. It was robustly detected from cameras when the relative position between the markers and cameras was changed. The subject could play a role of the criminal without special attention to the marker direction.

We also considered about the position of the third-person views in the display monitor. Windows on the lower right often overlapped with the right hand of the users. Therefore, for the right-handed users, the virtual weapons may be occluded by the windows. We changed the position of the windows to the upper right of the monitor.



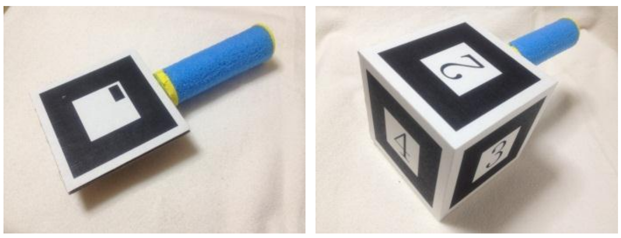
(a) Error of distance (b) Time

Figure 13. Result of task1.



(a) Error of distance (b) Time

Figure 14. Result of the task2.



(a) A single planar marker (b) Markers on a cube

Figure 15. Marker for the virtual weapon. A single planar marker was rotated, it was often missed from third-person view cameras. Five markers are arranged on a cube. The subject could play a role of the criminal without special attention to the marker direction.

5. Conclusions

In this paper, in-house 3D CG video generation supporting system using AR technology is proposed. The system focuses on the reenactment of the crimes causing injury or death, which are impossible to reenact actually. PCs connected with a Server/client network provide both of the first-person view and the third-person views on the HMD. It supports users to grasp the spatial relationship between themselves and surroundings. The marker information detected from multiple markers are integrated in the server PC. When a marker is occluded from a camera, the information of the marker can be given by another camera.

As a future work, the interface for augmenting additional information to the video is desired. It should support the visualization of information in the video. Recently, CT and MRI have been introduced to judicial autopsy. Such data will facilitate the generation of precise 3D models. We plan to provide the technologies for calibrating CG models to real objects.

As another potential application, the lay judges can not only watch the video, but experience the criminal site themselves by using the AR system themselves, so as to understand the situation of the crime better.

References

[1] J. March, D. Schofield, M. Evison, and N. Woodford, "Three-Dimensional Computer Visualization of Forensic Pathology Data," The American Journal of Forensic Medicine and Pathology, vol.25, pp.60-70, 2004.

[2] M. Thali, M. Braun, U. Buck, and E. Aghayev, "Scientific Documentation, Reconstruction and Animation in Forensic: Individual and Real 3D Data Based Geo-Metric Approach Including Optical Body/Object Surface and Radiological CT/MRI Scanning," Journal of Forensic Sciences, vol.50, No.2, 2005.

[3] H. Moriguchi, Y. Kuroda, H. Takiuchi, and O. Oshiro, "The calibration-less estimation of camera model of oblique-viewing endoscope for AR surgery," IEICE Technical Report, vol.110, no.108, pp.43-46, 2010.

[4] Y. Kuroda, H. Kanamori, H. Takiuchi, M. Tanooka, M. Imura, T. Kuroda, O. Oshiro, "Inside-body Registration with Pair-line Matching for AR Surgical Navigation," Journal of The Virtual Reality Society of Japan, vol.14, pp.435-444, 2009.

[5] H. Kato and M. Billinghurst, "Marker tracking nad hmd calibration for a video-based augmented reality conferencing system," International Workshop on Augmented Reality, pp.85-94, 1999.

[6] N. Inamoto and H. Saito, "Fly-Through Observation System for 3D Soccer Movie Based on Viewpoint Interpolation," Journal of The Institute of Image Information and Television Engineers, vol.58, No.4, pp.529-539, 2004.

[7] Y. Okamoto, I. Kitahara, Y. Ohta, "3D Human Body Display by Using Depth Information for Remote Shared Mixed Reality," IEICE Technical Report, vol.94, pp.830-838, 2011.