Multi-layered see-through movie in diminished reality

Uematsu, Yuko, Hashimoto, Takanori, Inoue, Takuya, Shimizu, Naoki, Saito, Hideo

Yuko Uematsu, Takanori Hashimoto, Takuya Inoue, Naoki Shimizu, Hideo Saito, "Multi-layered see-through movie in diminished reality," Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII, 828825 (27 February 2012); doi: 10.1117/12.910076

Event: IS&T/SPIE Electronic Imaging, 2012, Burlingame, California, United States

Multi-Layered See-through Movie in Diminished Reality

Yuko Uematsu, Takanori Hashimoto, Takuya Inoue, Naoki Shimizu, Hideo Saito

Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan

ABSTRACT

This paper presents generating a multi-layered see-through movie for an auto-stereoscopic display. This work is based on Diminished Reality (DR), which is one of the research fields of Augmented Reality (AR). In the usual AR, some virtual objects are added on the real world. On the other hand, DR removes some real objects from the real world. Therefore, the background is visualized instead of the real objects (obstacles) to be removed. We use multiple color cameras and one TOF depth camera. The areas of obstacles are defined by using the depth camera based on the distance of obstacles. The background behind the obstacles is recovered by planar-projection of multiple cameras. Then, the recovered background is overlaid onto the removed obstacles. For visualizing it through the auto-stereoscopic display, the scene is divided into multiple layers such as obstacles and background. The pixels corresponding to the obstacles are not visualized or visualized semi-transparently at the center viewpoints. Therefore, we can see that the obstacles are diminished according to the viewpoints.

Keywords: Diminished Reality (DR), Auto-stereoscopic display, TOF depth camera, multiple cameras

1. INTRODUCTION

Diminished Reality (DR) is one of research fields in Augmented Reality (AR). In contrast with usual AR that overlays some virtual objects on the real world, DR removes real objects from the real world. Actually, the real objects are replaced with images captured at different viewpoints and seem to be diminished. Therefore such DR is usually visualized through a 2D display.

Several researches have been proposed for removing real objects from input videos. Mann et al.[5] proposed a method for tracking a planar object and replacing it with another texture. Wang et al.[2] divided a video into several layers and rendered the video without one layer. The object's region in the removed layer was replaced with different frames in the video sequence. Therefore the objects seemed to be diminished from the video. Zokai et al.[4] used three images that captured by calibrated cameras. A region of an object to be removed is replaced with an appropriate background image that is recovered by a paraperspective projection of two reference images. Jarusirisawad et al.[3] proposed a calibration method of multiple hand-held cameras for DR, which are calibrated by Projective Grid Space. The scene was segmented into foreground and background, and then unwanted objects are removed from the images. Enomoto et al.[1] also used multiple cameras. As same as Zokai's method, a region of an obstacle in the center image was replaced by a background image generated by the other two cameras. In this research, the background is approximated as a plane and is recovered by planar-projection. Each camera can move around to achieve a real time DR application.

In these related works, the result images of DR are displayed in a normal 2D display. Users watch the scene where the target object is removed and the background is appeared through the display. Therefore our DR results are displayed in a 3D display for achieving different viewing style of DR. By using an auto-stereoscopic display, we would like to achieve the feeling of looking into the display over the obstacles instead of just overlaying background image onto the obstacles. The obstacle is not removed when seen from side viewpoints, and then it is diminished and the background is appeared when seen from the center viewpoints as shown in Fig. 1.

For such display style, the obstacle and background should be segmented into multiple layers. In the related approaches, the target object or region is segmented by manual definition[4] or semi-automatic method.[6] Since our system uses multiple stable cameras but the target object can move in real time, these previous approaches is not available. Graph Cut algorithm or background subtraction are also typical and useful for such segmentation purpose. However, such color-based segmentation is not enough when the background is similar in color to the obstacles. Therefore we use a depth camera that can measure the distance to a target object in real time to define the obstacle area automatically. Of course, the distance of the obstacle is specified at first. Even if multiple

Stereoscopic Displays and Applications XXIII, edited by Andrew J. Woods, Nicolas S. Holliman,
Gregg E. Favalora, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8288, 828825

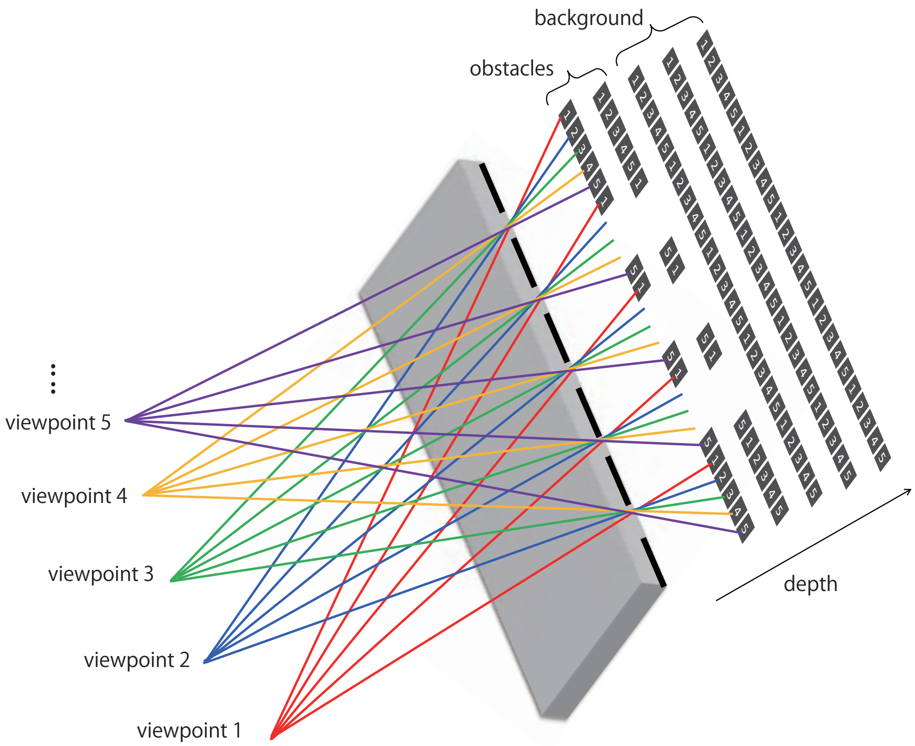
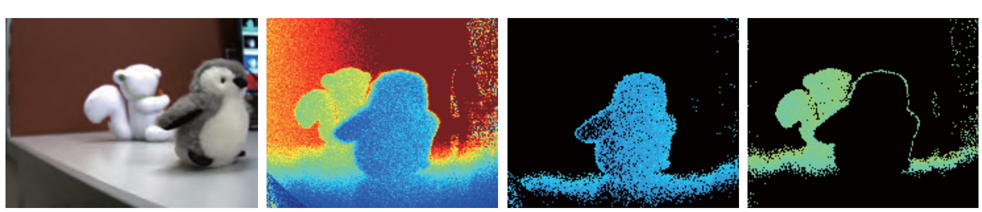


Figure 1. Example of multi-layered Diminished Reality for auto-stereoscopic display

obstacles appear, they can be detected as long as they are in the specified distance. Fig. 2 shows example images captured by a color camera and a depth camera. There are two objects at slightly different positions in the scene, and the area of each object can be extracted from depth data.

The algorithm of DR is similar as Zokai's method and Enomoto's methods. We use three high-definition cameras that are available for broadcast. In the center camera image, the obstacle is detected and removed by replacing it with background. The detection is achieved with the depth camera by extracting the object in pre-defined distance. Therefore the depth camera is placed side-by-side with the center camera. The background is recovered by synthesizing the images of the other two cameras. In Zokai's method, the structure of background is modeled in advance. Since we aim to create a real time application, we approximate the background by planes as well as Enomoto's method. Then we extend to use multiple planes for approximation of the background because the background also moves.



(a) color camera (b) depth camera (c) object 1 (d) object 2

Figure 2. Images captured by color camera and depth camera. According to the distance, objects can be extracted in the depth image.

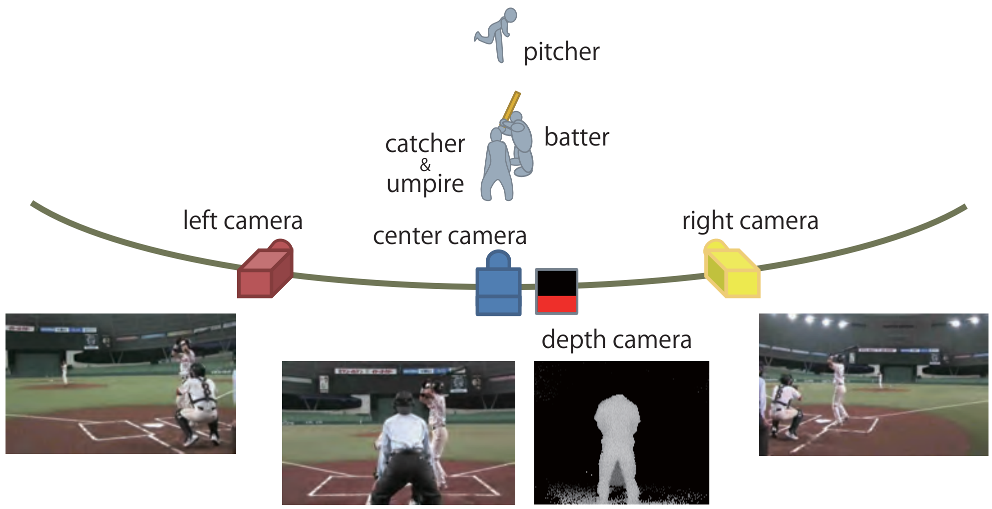


Figure 3. overview

As an actual application, we make a DR system using at baseball game scene. We are motivated by creating a new viewpoint movie of baseball game especially for the pitching scene. This is because the pitching scene is almost captured from back of the pitcher (around center field screen) in the normal broadcasting but viewers want to watch the scene from various viewpoints. We think that the catcher's viewpoint is one of the attractive but quite rare viewpoints in the pitching scene. For watching the pitching scene from the catcher's viewpoint, it is necessary to put the camera in front of the catcher or on the head of the catcher (umpire). Of course it is impossible in the official game. Only in the particular games, Nippon Television (NTV) has broadcasted some games with putting a camera on the head of the umpire. It was very exciting and developed a reputation. However, the viewpoint is swinging according to the umpire's head, so it might induce visually motion sickness. Therefore, we put the cameras at the fence behind the catcher, which preserves normal camera settings, remove the catcher and the umpire, and recover the pitcher's appearance by using DR technique. In this case, the catcher and the umpire are obstacle to be removed and the pitcher is background to be recovered. Since the pitcher exists far enough from the cameras in our camera settings, it can be approximated by a plane. According to the pitching motion, the plane also moves to appropriate place. Fig. shows the overview of DR system in the baseball game.

2. DEPTH CAMERA BASED DIMINISHED REALITY

We use a Time Of Flight (TOF) depth camera, which emits and receives a laser or LED light to calculate the distance from some objects by the time taken for the round-trip. We also use high-definition color cameras with the depth camera because such depth camera cannot get color textures except for a reflected intensity image. For using multiple types of cameras, there are lots of issues to be considered such as temporal synchronization and geometrical registration (calibration).

2.1 temporal synchronization

Normal high definition cameras can mechanically control the synchronization, however, most of depth camera do not have such a mechanics. Moreover the frame rate of the depth camera is not kept constant because it works via a PC 's CPU. Therefore we have to get frame correspondence between the depth camera and color cameras. When running our system in real time capturing, all images are simultaneously captured by a PC, so we do not care the temporal synchronization issue. When running it with off-line mode (use pre-captured images), we get correspondence between the cameras by detecting motion of the captured objects.

Our research utilizes optical flow or vertical displacement for temporal synchronization. Optical flow is computed from each frame of the color image and depth image and totalized at every frame. If we capture ball sports, vertical motion of the ball (position in Y axis) is detected at every frame. These are represented as 1D signal. The signals of the color and depth cameras have similar shapes, however, the number and rate of frames are different each other. Therefore we get correspondence by applying DP matching.

2.2 geometrical calibration

Geometrical calibration is also important because the viewpoint of the depth camera should be coincident with the color camera. Usually, some objects whose shape is given are placed on the scene and obtain 2D-3D correspondences for calibration. In this case, since each pixel in the depth image has 3D information, we can get 2D-3D correspondences from the color image and depth image without known objects. Then a projection matrix is computed between the real 3D world and the color camera.

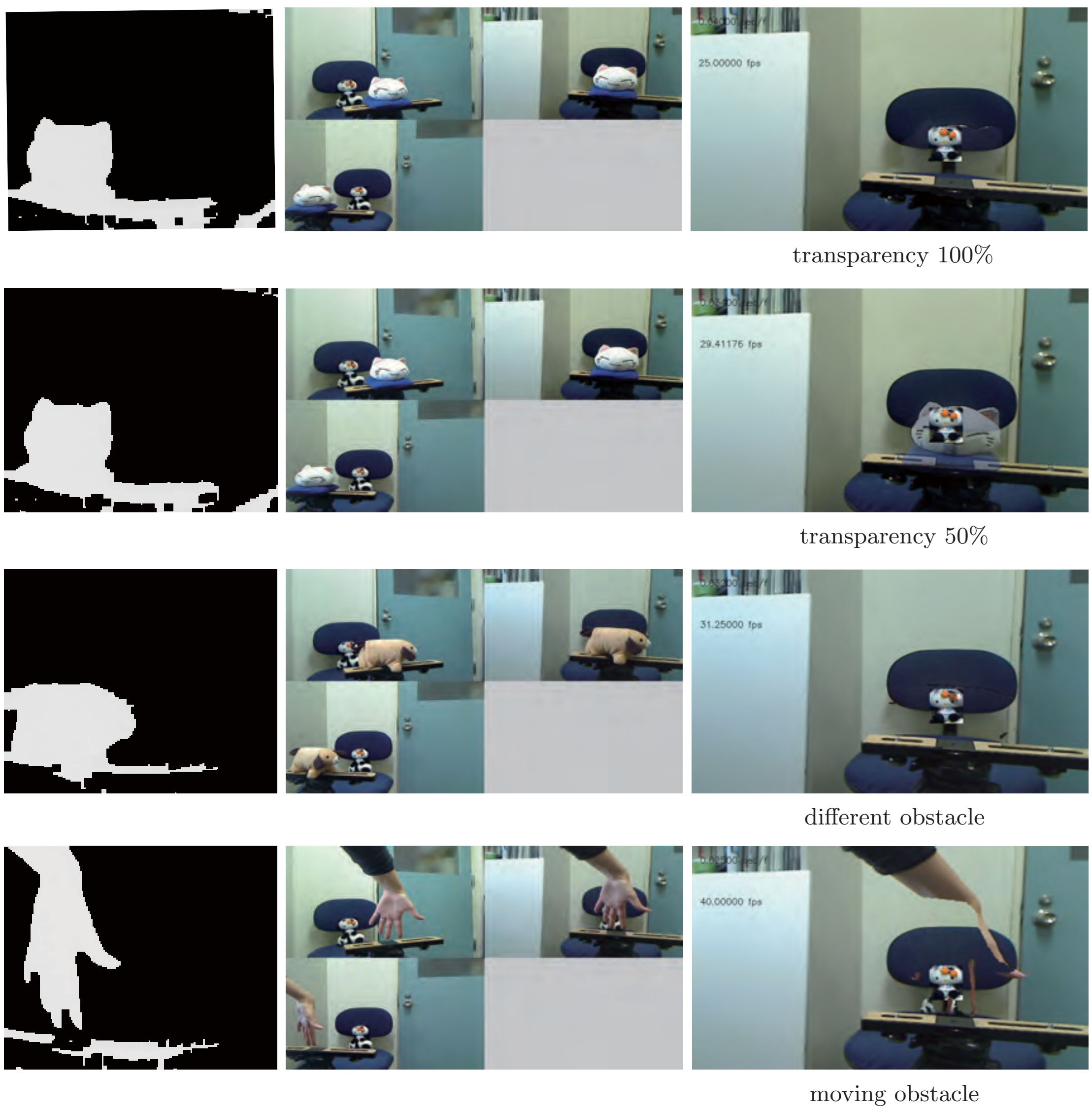


Figure 4. Result of Diminished Reality. Left: Mask image from depth camera. Center: Input images from three color cameras. Right: See-through movie.

2.3 see-through image generation

In our approach, the background is recovered by planar projection. Therefore Homography matrices at the background position are computed between three color cameras by specifying corresponding points, at first. The area which the obstacles can move around is also defined. It is defined by the distance from the depth camera. After the obstacles coming in the scene, the area of the obstacles is detected by the depth camera. This depth image is used as a mask image for planar projection. Then the images of left and right cameras are transferred onto the center camera's viewpoint to recover the occluded background. It is overlaid onto the area of the obstacles. Therefore the obstacles seem to be diminished from the scene as shown in Fig. 4. By changing the transparency of the recovered background image and the obstacles, we can see the background through the obstacles.

3. DIMINISHED REALITY WITH AUTO-STEREOSCOPIC DISPLAY

Fig. 3 shows the overview of our DR in a baseball game. The distance from the depth camera to the umpire is measured at first. Since the position of each person is not changed in a baseball game, the area of the obstacles is extracted through the game. Three cameras are calibrated based on planar projection by computing Homography matrix at the pitcher's plate. When the pitcher moves from the pitcher's plate for throwing a ball, computed Homography is slightly changed according to the pitcher's position. Then, see-through movie is generated as well as Sec. 2.

input image with obstacles see-through

Figure 5. Results of Diminished Reality in auto-stereoscopic display

For representing these DR results in the auto-stereoscopic display, the original center camera's image (with obstacles) and the generated see-through image (without obstacles) are mixed into the input image. In this experiment, the image with the obstacles is set to first and second viewpoints (the most slanted viewpoint), and the image without the obstacles is set to the other viewpoint (almost center viewpoint).

As for the first row and second row in Fig. 5, the pitcher is throwing a ball and takes his foot in front of the pitcher's plate. However, the pitcher can be clearly recovered by adjusting the base plane for planar projection. In the second column, there are the displayed results seen from side viewpoints. The obstacle is normally visible in front of the background. On the other hand, the catcher and umpire are diminished seen from the different viewpoints. According to the viewpoints, therefore, we can see the Diminished Reality results by using auto-stereoscopic display.

4. CONCLUSIONS

In this paper, depth camera based Diminished Reality system has been presented. The obstacles to be diminished are detected from the depth image based on the distance of them. The background to be recovered is generated by planar projection of multiple cameras. Therefore the background is approximated as a plane. When the background moved from the base plane, the plane is adjusted according to the position of background. We also represented these DR results in the auto-stereoscopic display. According to the viewpoints, the obstacles are diminished or appear in the display. The goal of our DR in the baseball game is used in the real broadcasting. A TV set that has multiple viewpoints is now commercially available, so we can enjoy this DR effect by freely changing the viewpoint.

Acknowledgement

This research was partially supported by National Institute of Information and Communications Technology, Japan. This research was partially supported by Grant-in-Aid for Young Scientists (B).

REFERENCES

[1] H. Saito A. Enomoto. Diminished reality using multiple handheld cameras. *ACCV'07 Workshop on Multi-dimensional and Multi-view Image Processing*, pages 130–135, (2007).

[2] E. Adelsen J. Wang. Representing moving images with layers. *IEEE Transactions on Image Processing Special Issue, Image Sequence Compression*, 3(5):625–638, (1994).

[3] H. Saito S. Jarusirisawad. Diminished reality via multiple hand-held cameras. *First ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC '07)*, pages 251–258, (2007).

[4] Yakup Genc Julien Esteve Siavash Zokai, Nassir Navab. Multiview paraperspective projection model for diminished reality. *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003)*, pages 217–226, (2003).

[5] James Fung Steve Mann. Videoorbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene. *International Symposium on Mixed Reality (ISMR2001)*, pages 48–55, (2001).

[6] M. O. Berger V. Lepetit. A semi-automatic method for resolving occlusion in augmented reality. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2000)*, 2:225–230, (2000).