Pacific Graphics (2017)
J. Barbic, W.-C. Lin, and O. Sorkine-Hornung (Editors)

Short Paper

Real-time Shadow Removal using a Volumetric Skeleton Model in a Front Projection System

Jaedong Kim[1], Hyungoog Seo[1], Seunghoon Cha[1,2] and Junyong Noh[1]

[1]KAIST [2]KAI Inc.

Abstract

When a person is located between a display and a projector in operation, a shadow is cast on the display. The shadow on a display may eliminate important visual information and therefore adversely affects the viewing experiences. There have been various attempts to remove shadows cast on a projection display by using multiple projectors. We propose a real time novel approach to removing shadow cast by the person who dynamically interacts with the display making limb motions in a front projection system. The proposed method utilizes a human skeleton obtained from a depth camera to track the posture of the person which changes over time. A model that consists of spheres and conical frustums is constructed based on the skeleton information in order to represent volumetric information of the person being tracked. Our method precisely estimates the shadow region by projecting the volumetric model onto the display. In addition, employment of intensity masks that is based on a distance field helps suppress the afterimage of shadow that appears when the person moves abrubtly and smooth the difference of the brightness caused by different projectors at the boundary of the shadow region.

Categories and Subject Descriptors (according to ACM CCS): I.4.0 [Image Processing and Computer Vision]: General Image displays, Human-centered computing Mixed/augmented reality—

1. Introduction

A projection system allows a user to easily obtain displays of any size on any arbitrary surfaces. The projection system can be installed in two different ways: front projection and rear projection. In practice, the front projection is often preferred over the rear projection for the following two reasons. First, installation is relatively easy. Second, unlike the rear projection system that a projector is located behind the display, the front projection system does not require any additional space for installation of the projector. These two advantages result from the sharing of the same side space between the projector and the viewers. This sharing of the same side space, however, makes the front projection system very vulnerable to loss of visual information displayed on the screen. For example, if any object, an example being a presenter or another viewer, accidentally steps in the path of the ray from the projector to the screen, shadow is cast on the screen creating an unpleasant viewing experience to the viewer [SACR05].

There have been studies that employ multiple projectors to eliminate the shadow created in a front projection environment. A common strategy of such studies is to compensate for the intensity of an occluded region with unobstructed projectors. In order to fill a shadow region with the unobstructed projector, the region should be identified first. There are two approaches to estimating a shadow region. The first approach employs a camera that observes the display area to detect the moment of shadow appearance [CS01, JWS*01, CRSS03, JWS04, SMK10, TIK17]. Upon the detection of the shadow, the occluded region is compensated accordingly. The second approach predicts a shadow region by observing the occluder [AC07, INS14]. By projecting the tracked occluder onto the display, the shadow region is estimated. The current systems that employ either of these approaches fail to achieve real-time performance or to quickly respond to movements of the person.

In this paper, we propose a novel approach to removing the shadow of a moving person who steps in a path of the ray from the projector in real-time in a front projection environment. Our method employs a single depth camera to track skeleton information for detecting the movement of a person. Based on the skeleton, our approach constructs a model that consists of spheres and conical frustums to represent the volume of a person. The shadow region can be estimated with precision by projecting the volumetric model onto the display. Our method automatically generates a mask that determines the intensity of each pixel based on the distance from the shadow region. The use of an intensity mask makes our system robust against movements of the person and allows to smooth the boundary of a shadow region effectively when the brightness difference is noticeable. Our method works fast making it applicable to dynamic content such as games that require various limb motion inputs, as well as static content such as presentations.

DOI: 10.2312/pg.20171318

2. Related Work

To remove shadow cast in a front-projection system, various techniques utilizing multiple projectors have been proposed. One notable approach removes the shadow formed on a display surface by compensating for the shadow area with unoccluded projectors after the camera detects the position and shape of the shadow [CS01,JWS*01,CRSS03,JWS04,SMK10,TIK17]. Requiring at least one camera to identify unoccluded areas on the display, this method handles complex shadow and obtains a quality image on the display.

Another type of approach employs an infrared light source and infrared cameras [TP02, FSR05, SFC*07]. This approach detects the occluders instead of shadows by reflecting the infrared light from the display surface to the infrared camera. Because the infrared light illuminates the display surface while not illuminating the occluder, the occluded region that creates shadow can be reliably estimated by using a simple background subtraction technique.

There has been an attempt to employ multiple cameras to extract the 3D information of an occluder that cast shadow. Audet and Cooperstock [AC07] utilized stereo cameras to extract simple 3D information of the occluder. With the 3D information of the occluder and the geometric relationship among the camera, the projector, the occluder, and the display surface, the shadow region is estimated. Assuming that a person is standing vertically on the floor, this approach generates a shadow mask of a simple bounding box. Recently, a visual hull reconstruction technique has been introduced to remove shadow formed in a projection system [INS14]. To remove the shadow and afterimage, the intensity of the projected pixels are adjusted according to the distance from the projection of an occluder in a voxel space using multiple cameras.

Our goal is to remove shadow of a person freely moving in a front projection environment. A conventional camera based approach that detects shadow on the display is unsuitable for our purpose because at least one frame of delay is inevitable with that approach before the removal of the shadow after the detection. Our method is similar to the previous work that detects an occluder instead of shadow itself [AC07, INS14]. Our method is different in that we can handle the limbs of a person unlike Audet and Cooperstock [AC07]. Unlike Iwai et al. [INS14] who perform heavy 3D computation for the creation of an intensity mask, our method utilizes a 2D distance field for the mask, achieving real time performance.

3. System Overview

We assume a front projection environment and try to remove the shadow using multiple projectors all of which are responsible for creating one coherent display. For brevity of the explanation, we further assume the environment created by two projectors. The basic idea for shadow removal lies in compensating brightness of the shadow region using an unoccluded projector. The overall procedure of our method is illustrated in Figure 1. In order to deal with the shadow detection and removal in real-time, our approach first performs a pre-processing including display registration and calibration of the projectors and the camera. Next, a simplified 3D model approximated by spheres and conical frustums is constructed using the skeleton information obtained from a depth camera. The shadow region formed on a display is estimated by the constructed model projected onto the display from the 3D position of the occluded projector. Based on the estimated shadow region, our method creates an intensity mask for each projector, which contains an appropriate alpha value for transparency at each pixel. Finally, the shadow free display is obtained by blending the colors from each projector using the intensity masks.

4. Shadow Region Estimation

Our approach transforms the point cloud into the volumetric model with spheres and conical frustums. The depth camera captures a depth image along with a skeleton of the person. Then, the point cloud of the person is obtained by removing the background of the depth image. Once skeleton information is obtained from a depth camera, our method places spheres at each joint and conical frustums at the mid-point of neighboring joints. The torso is represented with an elliptic conical frustum. The radius of the spheres and the conical frustums are determined as follows. Each side of the conical frustum takes the value of the radius of each of the neighboring spheres. The width of the frustum is linearly interpolated from both sides. The height of the conical frustum is determined by the distance between the neighboring joints. The radius of each sphere is determined by iteratively reducing the value from an initial radius until the smallest radius that contains all of the points in the point cloud is reached.

The volumetric model initially defined in the depth camera coordinate needs to be transformed to the coordinate of each projector. After the transformation, the 3D position of the mesh vertices in the model is identified from the projector point of view. Each projected 2D face with respect to each projector implies a shadow patch corresponding to the 3D face of the model. By integrating all of the projected faces, our algorithm estimates the shadow region corresponding to the occluding person.

5. Intensity Mask Generation

Although our method successfully identifies a shadow region defined as S, estimating an intensity mask that contains an alpha value at each pixel directly from S does not work well due to unpredictable movements of a fast moving person and the difference in the brightness of the projectors. Because the inferred shadow region tightly matches the actual shadow, abrupt motion may produce shadow while the algorithm attempts to estimate a new shadow area for an updated pose.

Inspired by Iwai et al. [INS14], our method computes an unsigned distance field to construct an intensity mask for each projector to achieve smooth transition of the brightness emitted by different projectors. Unlike their approach that calculates the distance between projector rays and the occluder in the 3D domain, our method measures the distance from the boundary of the estimated shadow area to each pixel in the 2D image. This allows our algorithms to operate in real time. Note that prior to the intensity mask generation, our method first transforms the shadow region S

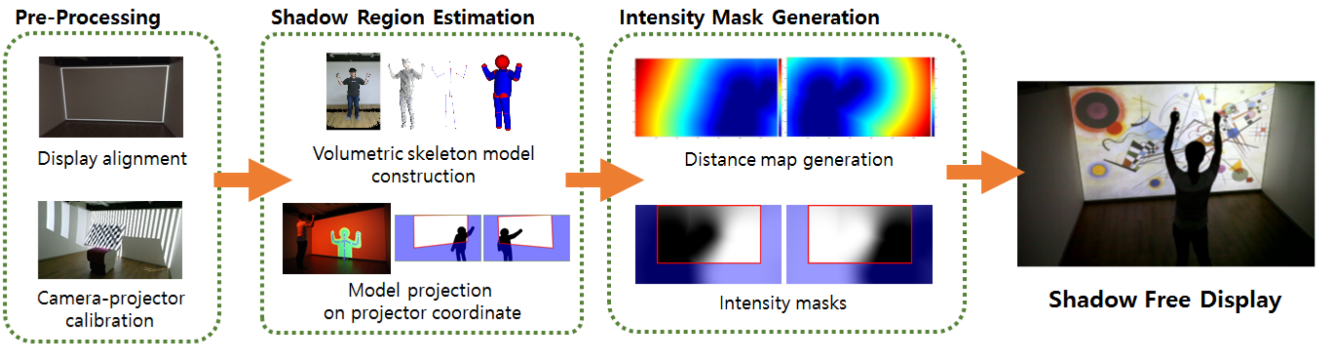

Figure 1: *System overview.*

	Position (m)			Orientation (degree)		
	x	y	z	Roll	Pitch	Yaw
Projector (1)	0.44	2.18	4.52	-171.7	-15.3	-2.0
Projector (2)	3.76	2.23	4.68	-176.1	16.3	-1.3
Camera	2.06	2.27	0.04	-33.4	-0.8	179.4

Table 1: *External parameters for the devices in our system.*

estimated in Section 4 which is represented in the projector coordinate to the display coordinate to obtain a warped shadow region S'.

Given a warped shadow map associated with the j-th projector S'^{P_j}, pixels in a non-shadow region are filled with the distances to the shadow region. The constructed map is a distance field D^{P_j} for projector P_j.

$$M^{P_j}(p) = \frac{(D^{P_j}(p))^{\alpha}}{\sum_{n=1}^{N}(D^{P_n}(p))^{\alpha}}, \quad (1)$$

where $D^{P_j}(p)$ represents the distance from each point p on the display to the shadow region associated with each projector P_j and N is the number of projectors. In order to handle both of the scenarios, α is included as a user parameter in Equation (1). The α determines the degree of intensity variation along the distance. A larger α means a larger area is compensated by the unoccluded projectors, causing the reduction of a blending area.A high value of α can be used for a dynamic scenario such as game playing while a low value can be used for a static scenario such as presentations. The intensity mask goes through a gamma correction step ($\hat{M}_l(p) = M_l(p)^{\gamma}$) for all pixels p to compensate for the linear representation of intensity [AC07]. We use $1/2.2$ for gamma. The final intensity mask which is going to be applied to content is created by removing the dummy pixels around the display region.

6. Results

To verify the performance of the proposed method, we used interactive content as a display medium and applied various conditions. Robustness and practicality of the system is demonstrated by a comparison of the results from our method and from previous methods. Our projection system consists of two projectors (ViewSonic PJD7820HD Projector) and a RGB-D camera, i.e., KINECT for window v2, with white wall surface whose height and width is 207cm by 395cm. Our approach only uses skeleton information from KINECT. Table 1 shows the calibrated position of each device. All the experiments were performed on an Intel i7-5820K 3.30GHz CPU with 32GB of memory. We set the lower-left corner of the display surface as the origin of the standard coordinate in which the x-axis and the y-axis coincide with the horizontal and the vertical axis of the screen, respectively. The resolution of the projector is 1920×1080 full HD.

The proposed algorithm is implemented in C++. The resolution of the intensity mask is 960×540 and the size of the dummy pixels is 200. We installed each device at a reasonable place and performed the calibration process proposed by Moreno et al. [MT12]. Our volumetric model consists of triangles with 946 vertices and 1764 faces. We prepared a raw skeleton that consists of 25 joints, and removed 6 redundant joints such as the tip of the hands, the thumbs, and each side foot. Total processing time of our system is approximately 24 milliseconds achieving real time performance. Each part of our system was parallelized using openMP [Ope08].

Figure 2 shows examples produced using two interactive content. One utilizes a scenario of presentation and the other utilizes a scenario of game playing that requires limb motion inputs from the user. The results clearly show that our algorithm successfully removes the shadow region caused by the motion of the person. The depth camera was installed as camera in Table 1. Because people do not move dynamically during their presentation, we use the α value of 2 which expands a blending area to make the brightness at the boundary of the shadow region transition smoothly. On the other hand, people playing a motion-based game tend to move dynamically. Thus, we use 4 for the value of α.

7. Conclusion

We proposed a real time approach to removing the shadow of a person formed in a front projection environment. Our method estimates a shadow region faithfully using a volumetric model, which is constructed from the skeleton of a person, projected onto the display from the 3D position of the occluded projector. An unsigned distance field is also utilized to create an intensity mask which

Figure 2: *Results from a presentation scenario (first row,* $\alpha = 2$*) and from a game playing scenario (second row,* $\alpha = 4$*). Results with shadow (first column), results from our method (other columns)*

makes our system robust against the afterimage effect caused by the latency from the camera to the projector. The entire process of our algorithm works in real-time making our system useful for practical.

Although our method works well in most cases, our approach has some drawbacks as well. Since our method utilizes the skeleton information which only contains the position of body, our algorithm fails to compensate for the shadow when the person is holding an object or wearing a skirt. Another limitation of this study includes a situation of multiple people stepping into a shadow casting area. Our system can work in real time upto four projectors. Currently because our implementation runs on CPU, we expect that more projectors can be handled using GPU parallelization. A separate process of projector color which reduces the brightness difference between the projectors would also be helpful for smooth transition at the boundary of the shadow region.

8. Acknowledgements

This work was partly supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.2016-0-00349,(R&D) Development of Realistic Media and Presentation System for Large Auditorium) and the authors thank KOLON Corporation, Korea, for providing funding for this research through the KOLON-KAIST Lifestyle Innovation Center Project. <LSI13-ITNJY0001>

References

[AC07] AUDET S., COOPERSTOCK J. R.: Shadow removal in front projection environments using object tracking. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on* (2007), IEEE, pp. 1–8. 1, 2, 3

[CRSS03] CHAM T.-J., REHG J. M., SUKTHANKAR R., SUKTHANKAR G.: Shadow elimination and occluder light suppression for multi-projector displays. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (2003), vol. 2, IEEE, pp. II–513. 1, 2

[CS01] CHAM R. S. T.-J., SUKTHANKAR G.: Dynamic shadow elimination for multi-projector displays. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)* (2001). 1, 2

[FSR05] FLAGG M., SUMMET J., REHG J. M.: Improving the speed of virtual rear projection: A gpu-centric architecture. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on* (2005), IEEE, pp. 105–105. 2

[INS14] IWAI D., NAGASE M., SATO K.: Shadow removal of projected imagery by occluder shape measurement in a multiple overlapping projection system. *Virtual Reality 18*, 4 (2014), 245–254. 1, 2

[JWS*01] JAYNES C., WEBB S., STEELE R. M., BROWN M., SEALES W. B.: Dynamic shadow removal from front projection displays. In *Proceedings of the conference on Visualization'01* (2001), IEEE Computer Society, pp. 175–182. 1, 2

[JWS04] JAYNES C., WEBB S., STEELE R. M.: Camera-based detection and removal of shadows from interactive multiprojector displays. *Visualization and Computer Graphics, IEEE Transactions on 10*, 3 (2004), 290–301. 1, 2

[MT12] MORENO D., TAUBIN G.: Simple, accurate, and robust projector-camera calibration. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on* (2012), IEEE, pp. 464–471. 3

[Ope08] OPENMP ARCHITECTURE REVIEW BOARD: OpenMP application program interface version 3.0, May 2008. URL: http://www.openmp.org/mp-documents/spec30.pdf. 3

[SACR05] SUMMET J., ABOWD G. D., CORSO G. M., REHG J. M.: Virtual rear projection: Do shadows matter? In *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (2005), ACM, pp. 1997–2000. 1

[SFC*07] SUMMET J., FLAGG M., CHAM T.-J., REHG J. M., SUKTHANKAR R.: Shadow elimination and blinding light suppression for interactive projected displays. *Visualization and Computer Graphics, IEEE Transactions on 13*, 3 (2007), 508–517. 2

[SMK10] SUGAYA Y., MIYAGAWA I., KOIKE H.: Contrasting shadow for occluder light suppression from one-shot image. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on* (2010), IEEE, pp. 96–103. 1, 2

[TIK17] TSUKAMOTO J., IWAI D., KASHIMA K.: Distributed optimization framework for shadow removal in multi-projection systems. In *Computer Graphics Forum* (2017), Wiley Online Library. 1, 2

[TP02] TAN D. S., PAUSCH R.: Pre-emptive shadows: Eliminating the blinding light from projectors. In *CHI'02 extended abstracts on Human factors in computing systems* (2002), ACM, pp. 682–683. 2