Augmented Environment Mapping for Appearance Editing of Glossy Surfaces

Takumi Kaminokado
School of Engineering Science
Osaka University
Toyonaka, Japan
Email: kaminokado@sens.sys.es.osaka-u.ac.jp

Daisuke Iwai and Kosuke Sato
Graduate School of Engineering Science
Osaka University
Toyonaka, Japan
Email: {daisuke.iwai, sato}@sys.es.osaka-u.ac.jp

***Abstract*—We propose a novel spatial augmented reality (SAR) framework to edit the appearance of physical glossy surfaces. The key idea is utilizing the specular reflection, which was a major distractor in conventional SAR systems. Namely, we spatially manipulate the appearance of an environmental surface, which is observed through the specular reflection. We use a stereoscopic display to present two appearances with disparity on the environmental surface, by which the depth of the specularly reflected visual information corresponds to the glossy surface. We refer to this method as augmented environment mapping (AEM). The paper describes its principle, followed by three different implementation approaches inspired by typical virtual and augmented reality approaches. We confirmed the feasibility of AEM through both quantitative and qualitative experiments using prototype systems.**

***Keywords*-augmented reality; projection mapping; environment mapping;**

I. Introduction

Spatial augmented reality (SAR) realizes AR environments by optically superimposing computer generated images onto real world surfaces by projected imagery. Compared to other approaches (i.e., video see-through and optical see-through AR), SAR allows users to observe augmentations without bounding them with cumbersome display devices such as smartphones, tablets, and head-mounted displays. Following the pioneering work by Raskar et al. [1], SAR has been integrated in many fields, such as education [2], art creation [3], industrial design [4], daily life support (e.g., searching everyday objects [5]), makeup [6], virtual restoration of historical objects [7], and entertainment (e.g., games [8] and theme parks [9]). As basis of the applications, researchers proposed various technologies, such as geometric registration of projectors to arbitrarily shaped surfaces [10] and radiometric compensation of projected images to visually cancel surface textures [11]. Furthermore, they have overcome fundamental technical limitations of projector devices by developing novel projection technologies, such as high-speed and low-latency projection, high dynamic range projection, and extended depth-of-field projection, which have allowed novel SAR experiences [12].

However, an unsolved limitation still remains. Namely,

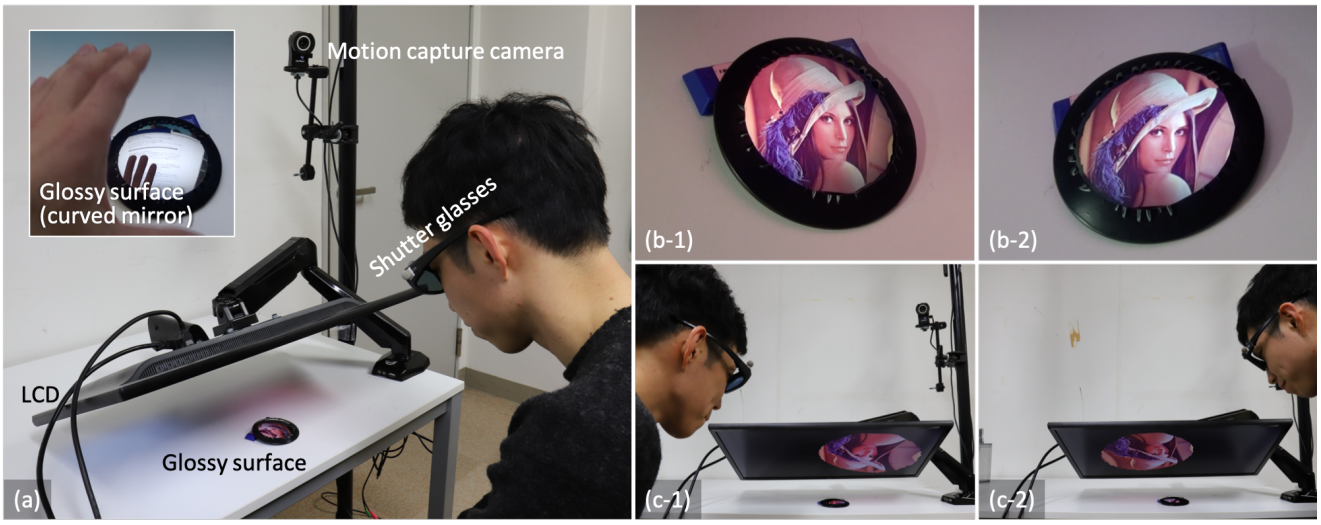


Figure 1. Augmented environment mapping for manipulating the appearance of a glossy surface (a curved mirror): (a) an overview of a system using a flat panel LCD (three more motion capture cameras are used but out of the capturing camera's field of view), (b) appearance manipulation results from different observation positions, (c) another view of (b) where a displayed image on the LCD is changed according to the observation position. Note that the displayed image was rendered for monoscopic vision for a reader's better visibility.

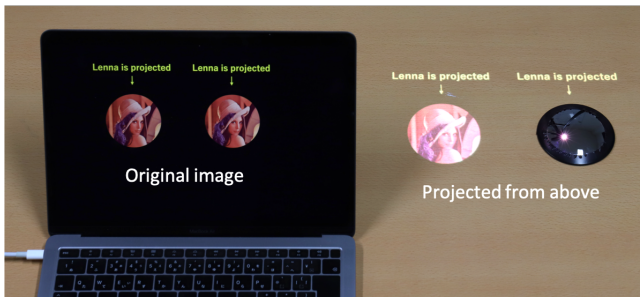


Figure 2. SAR on a wooden tabletop and a glossy surface (the curved mirror in Figure 1).

SAR cannot edit the appearance of glossy surfaces. Because glossy surfaces reflect projected light mostly to the specular direction, only a small portion of a projected image can be observed from a single viewpoint as a specular highlight, which is generally extremely bright (Figure 2). Therefore, it is hard to display a desired image on glossy surfaces by a projector. Prior works tackled this problem by applying multiple projectors [13], [14]. They place the projectors at different locations, and select one for each area on a surface, from which the specular reflection of projected pixels does not hit observer's eyes. However, these methods are effective only when the surface reflectance property contains both specular and diffuse components. More critically, the observer still sees dazzling specular highlights, because the current projectors cannot completely turn off the light on a per pixel basis (i.e., black offset).

In this paper, we propose a novel SAR framework to edit the appearance of physical glossy surfaces (Figure 1). The key idea is utilizing the specular reflection, which was a major distractor in conventional SAR systems. Namely, we spatially manipulate the appearance of an environmental surface, which is observed through the specular reflection. We use a stereoscopic display to present two appearances with disparity on the environmental surface, by which the depth of the specularly reflected visual information corresponds to the glossy surface. We refer to this method as augmented environment mapping (AEM). The paper describes its principle and three different implementation approaches inspired by typical virtual and augmented reality approaches. It also describes how to geometrically calibrate AEM systems, which is a common technique for all the implementation approaches. We build prototype systems to validate the proposed framework and the calibration technique through both quantitative and qualitative experiments. Finally, we implement two different applications—(1) visual design support of glossy product and (2) projection mapping—on the prototypes to demonstrate the high applicability of AEM.

We envision a SAR system where users do not have to remove specular objects from a target scene, which was required in conventional SAR systems. In other words, SAR systems with our technique can augment not only diffuse surfaces but also glossy surfaces. Therefore, our technique contributes to expanding the application fields of SAR, and thus, the number of practical applications is theoretically more than that of conventional SAR. To summarize, this paper makes the following three prime contributions:

- We propose a novel SAR framework, AEM, in which we can edit glossy surface appearance based on spatially controlling the specular reflection.
- We introduce various implementation approaches of AEM and a common geometric calibration technique for them.
- We demonstrate the feasibility and applicability of AEM by quantitative and qualitative experiments on prototypes and multiple application examples, respectively.

Note that although some of our implementation approaches are not within the scope of traditional SAR, we interpret SAR in a broader sense in this paper, i.e., we regard AR technologies without any see-through mechanisms between an observer and a target glossy surface as SAR.

II. Related Works

SAR technologies so far assumed that projection targets are Lambertian surfaces. For example, a room augmentation system allowed users to interact with virtual characters projected onto the diffuse walls and floor of the room by texturing the surfaces with projected imagery [8]. Not just changing the surface appearances, several researchers focused specifically on making matte surfaces glossy in projection mapping. Okazaki et al. showed that a free-form plaster surface is altered to a glossy surface of metallic material [15]. Such technique can be applied in industrial design process where designers and customers can discuss and sophisticate a product's design in situ by visually changing its surface material properties (e.g., specular reflectance) on the fly [4], [16], [17]. The surface material editing from matte to glossy was recently applied even to a dynamic target object [18]–[20]. Increasing the maximum luminance of specular highlight and supporting multiple observers were achieved by combining optical see-through head-mounted displays with projection mapping [21].

However, the inverse manipulation, i.e., from glossy to matte, is still technically difficult. Because a normal projector illuminates a glossy surface approximately from a small aperture, only a small part of a projected image on a glossy surface can be observed from each viewpoint. If the diffuse reflectance of the glossy surface is sufficiently large, we can observe the projected image from arbitrary viewpoints, but the specular highlight still exists depending on the relation of the surface normal, the projector's position, and the observer's position. For such setups, previous works suppressed specular highlights by applying multiple projectors such that projectors causing no specular reflections for

an observer are selected to display an image for each surface area [13], [14]. Wetzstein and Bimber realized the same principle by solving inverse light transport problem [22]. However, although these methods are applied, a specular reflection is still visible due to the black offset of a projector, which significantly distract the projected image quality. In addition, these methods completely fail to form images when the surface has almost no diffuse reflection such as a mirror.

Specular reflection was also a serious problem in computer vision for the shape measurement of a glossy surface. Conventional active stereo based on a projector-camera system does not work due to the same reason why SAR is not suitable for a glossy surface. Knauer et al. solved this problem by spatially modulating the luminance of an area light rather than using a projector regarded as a point light due to a small aperture [23]. Their phase measuring deflectometry method measures the 3D shape of free-form specular surfaces by capturing a series of specular reflections of fringe patterns on a flat panel display. Because pixels of a flat panel display reflect on a large area of a glossy surface, they can be observed by a camera. We extend this principle to SAR. We control the appearance of a glossy surface by modulating the appearance of a large environmental surface, which is specularly reflected on a large area of the glossy surface to an observer.

There are some SAR systems assuming the projected results are observed through a mirror [24], [25]. However, these systems focused on augmenting a surface such as an observer's body in the mirror, not the mirror itself. Yamamoto et al. proposed an optical mechanism based on a special aerial imaging device to display a floating image on a glossy surface [26]. The system requires a much larger space than typical AR systems (including ours) due to the special optics, and is not designed to manipulate the appearance of a glossy surface. To the best of our knowledge, our work is the first attempt to modify a mirror image of environment for manipulating the glossy surface appearances.

III. Augmented Environment Mapping

This section describes our AEM technique that controls the appearance of a glossy surface. First, we explain the principle of AEM. Then, three different approaches of AEM implementations are introduced. Finally, we describe our calibration method of AEM systems.

A. Principle

AEM controls the appearance of a glossy surface by manipulating that of environment using either a flat panel display or projection mapping (Figure 3). The shape of the glossy surface S is represented as a mesh, and each vertex of the mesh is denoted as s. We trace back the ray of a specular reflection from the view position of an observer v to the display surface D. The inverse ray of the specular reflection $r(v, s)$ is computed with the view position v,

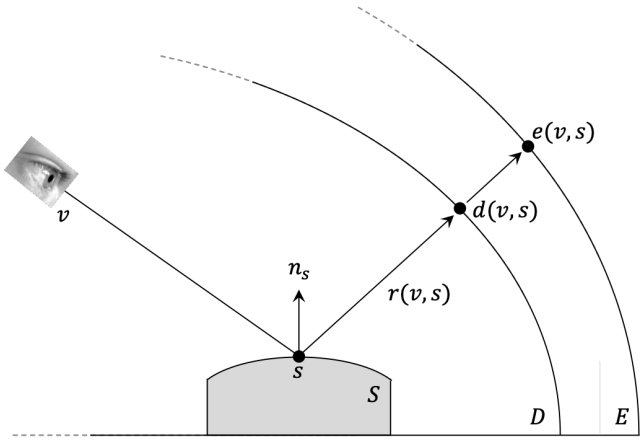


Figure 3. The principle of AEM.

the glossy surface position s, and its normal n_s. We then compute the intersection of $r(v, s)$ and D to obtain the corresponding position on the display surface, denoted as $d(v, s)$. An image to be displayed on D is computed by deforming a target appearance on the glossy surface using a mapping $f(v) : s \rightarrow d(v, s)$. In the case of fully convex or fully concave object, the image D can be computed uniquely because the mapping $f(v)$ is an injective function. In the case of bumpy object, however, some pixels in the image D corresponds to different points on the surface s.

Because of the nature of specular reflection, the mapping $f(v)$ is view dependent. Therefore, we need to render different images to the left and right eyes of an observer considering the binocular disparity. Specifically, we render two images by deforming each target appearance using the mappings of viewing positions of the left and right eyes, $f(v_l)$ and $f(v_r)$, respectively. To render these images without crosstalk, we use stereoscopic displays applying an active shutter 3D technique in both flat panel- and projection-based systems.

To allow an observer to move around, we track the observer's viewing position and update the mappings of $f(v_l)$ and $f(v_r)$ by computing the correspondences from s to $d(v_l, s)$ and $d(v_r, s)$ at every frame. The deformation and map update must be conducted at faster than 120 Hz for active shutter 3D systems. To this end, we apply a small number of correspondences (e.g., 700 in the current implementation) to the mappings and interpolate them using a simple texture mapping technique in the deformation process. These processes are parallelized on GPU.

B. Implementation approaches

We consider three approaches of implementing AEM. They are inspired by typical virtual reality (VR) and AR variations—immersive VR, see-through AR, and SAR. Figure 4 shows the illustrations of the approaches.

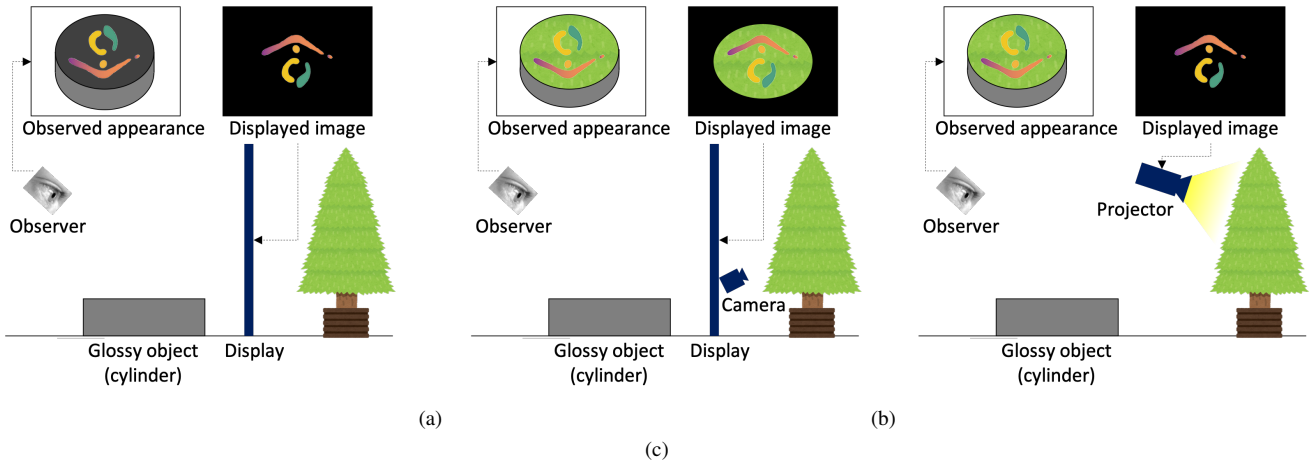


Figure 4. Conceptual illustrations of the proposed implementation approaches: (a) immersive AEM, (b) see-through AEM, and (c) spatial AEM.

Immersive AEM: In this approach, we place a glossy object in an immersive VR display system such as CAVE. Specifically, projection screens or flat panel displays are installed around the object. The surface appearance is edited by the method described in Section III-A without considering the environment outside the system. We call this type of implementation as immersive AEM (I-AEM). I-AEM can accurately control the appearance of the glossy surface. However, real-world contexts outside the system are missing in this setup.

See-through AEM: The display setup of this approach is same as the immersive AEM, i.e., projection screens or flat panel displays are placed around a glossy object. On the other hand, this approach takes into account the color of the environment outside the display system. Specifically, the displayed image is determined by overlaying two layers of images. The first layer, referred to as surface layer, represents a new appearance of the glossy surface, which is computed by the method described in Section III-A. The second layer, referred to as environment layer, represents the environmental scene, which creates an illusion that the display is transparent. There are two methods to compute the environment layer. The first method computes it by tracing the ray $r(v, s)$ to an environmental surface E and picking a color at the intersection point $e(v, s)$ (Figure 3). The shape and color of the environmental surface are captured by a depth camera. The second method is based on a light field reconstruction technique. We capture the environmental scenes reflected on the target surface using a normal camera from various viewpoints in advance, and stores them in a database before the display is placed. In online, the method computes an environment layer by interpolating stored images which were taken from positions close to the observer's current viewing position. Consequently, in the edited appearance on the target surface, the surface layer does not move according to the viewing position movement, while the environment layer moves. We call this type of implementation as see-through AEM (ST-AEM). ST-AEM can accurately change the surface appearance and convey real-world contexts. However, the screens or displays occlude a part of environment from an observer.

Spatial AEM: In this approach, the surface appearance of a glossy object is edited by projected imagery on environmental surfaces. In contrast to the other approaches, the display and environmental surfaces are identical. It is expected that this approach may not work when the environment surfaces are extremely rough. We believe this can be solved by placing multiple projectors at different locations to cover the whole surface area. We call this type of implementation as spatial AEM (S-AEM). In S-AEM, real-world contexts can be conveyed and there is no problem in occlusion by the displays. However, accurately controlling the appearance of the glossy surface by projected imagery is sometimes difficult due to textures of the environmental surfaces which cannot be perfectly compensated even with the latest radiometric compensation technique [27].

C. System calibration

As mentioned in Section III-A, AEM requires the shape of environmental surface, that of a target glossy object, and the viewing position of an observer to calculate the mapping $f(v)$. The shape of the environmental surface can be measured by using either an RGB-D camera (i.e., active stereo) or a multi-camera system (i.e., passive stereo), both of which are widely employed in modern AR systems. However, these depth sensing cameras do not support accurate surface normal acquisition, though both the shape and normal of the glossy surface must be accurately measured to compute the correct mapping $f(v)$. Furthermore, these cameras are obviously not suitable for measuring the shape of the glossy surface due to the specular reflection. Therefore, in this paper, we develop a surface reconstruction method, which

can accurately measure both the shape and normal of a glossy surface, based on a deflectometry method [23].

Our glossy surface reconstruction method consists of two processes: (1) camera calibration, and then (2) shape and normal estimation. We assume the shape of the environmental surface (i.e., display surface) is known in the world coordinate system in advance. In the first process, we place a flat mirror surface at a position where a target glossy object is placed after the calibration. Two cameras (cam_1 and cam_2) are installed such that they observe the display surface through the mirror. Their intrinsic parameters as well as the poses in the world coordinate system are calibrated by a combination of existing techniques [28], [29]. In particular, we display structured-light patterns on the environmental surface, and capture them by the cameras through the mirror. We repeat this process five times with different mirror poses, and then estimate the intrinsics and extrinsics of the cameras and the mirror poses through an iterative optimization. More specifically, we estimate the initial solution of the camera parameters and the relative pose of the mirror to the environmental surface [29]. Finally, all estimated parameters are integrated into an optimization problem and refined through an iterative process [28].

In the second process, we replace the flat mirror with the target glossy object. Then, we display structured-light patterns on the environmental surface, and capture them by the cameras through the object. The stereo deflectometry method [23] is applied to estimate an initial guess of the position of each glossy surface point s as the distance from cam_1. The initial guess is denoted as d_s^{init}. It also estimates the normal vector of s from the geometry of specular reflection, which is denoted as n_s^r. We smooth the initial guess, which generally forms a noisy point cloud, by making two normal vectors of s consistent. The first normal vector is n_s^r and the other is derived from 4-neighboring estimated surface points. The smoothed distance d_s is computed by solving the following optimization problem using a quasi-Newton method (L-BFGS):

$$\text{minimize} \quad \sum_{s \in S} \{\|n_s^p - n_s^r\|^2 + \alpha \|d_s - d_s^{init}\|^2\}, \qquad (1)$$

where n_s^p represents the normal vector computed by the neighboring points in each iteration. The second term penalizes the displacement from the initial guess. We set the balancing parameter α as 1,000. Through actual measurements of various glossy surfaces, we confirmed that the optimized surface shape is accurately enough to compute the mapping $f(v)$. The accuracy is formally evaluated in Section IV-C1.

The viewing position of an observer is measured using a motion capture (mocap) system. In this paper, we apply a mocap system consisting of multiple infrared cameras and retroreflective markers. We calibrate the 6 degree-of-freedom (DOF) relationship between the mocap coordinate system and the world coordinate system as follows. We attach retroreflective markers onto predefined positions of a checkerboard, which is then captured by cam_1. From the captured image, the pose of the board in the world coordinate system is estimated by solving PnP problem. The positions of the markers on the board are measured by the mocap system, from which the pose of the board in the mocap coordinate system is estimated. Finally, from these two poses, we obtain the 6DOF relationship between the coordinate systems.

The system calibration technique described in this section is common for all the implementation approaches of AEM introduced in Section III-B.

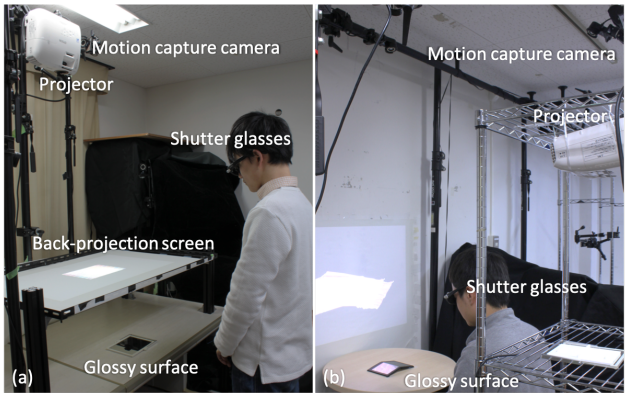


Figure 5. Prototype systems: (a) projection-based I-AEM system, (b) S-AEM system.

IV. Experiment

We built prototype systems to validate our AEM framework. This section introduces the implementations of prototype systems, followed by proof-of-concept experiments. Then, we evaluate the systems both quantitatively and qualitatively. Finally, we show a couple of application examples to demonstrate the potential applicability of AEM.

A. Prototype system

We built three prototype systems to validate the concept of AEM (Figures 1(a) and 5). These systems employed the same optical motion capture system (Optitrack Flex 13, 4 units) for tracking an observer's viewing position. They also used the same RGB cameras (XIMEA MQ013CG, 1280×1024, 2 units) and lenses (12.5mm) for the system calibration. Although the original deflectometry method [23] adopts sinusoidal patterns as structured-light patterns, we choose graycode patterns in our calibration process because we found that gamma values of our liquid crystal display (LCD) are not uniform according to the viewing angle. The textures of target appearances were prepared as 512×512 pixels images.

Two of the prototypes were implemented based on the I-AEM approach. The first one applied a LCD (ASUS SWIFT

PG248Q), which illuminated a target glossy surface from above (Figure 1(a)). An LCD holds the advantages of a high spatial resolution, a high dynamic range, and a wide color space. However, due to the directional dependency of the LCD panel, desired appearances are not always correctly observed in the reflected image on the glossy surface. This system can be used as an ST-AEM system once an additional camera (either normal or RGB-D) is applied. The second I-AEM-based system employed a back projection display consisting of a projector (EPSON EH-TW5350) and a semi-transparent screen (a sheet of paper) (Figure 5(a)). The back projection display is flexible in size, thus can be larger than the LCD-based system. In addition, it does not suffer from the directional dependency problem. However, the spatial resolution, dynamic range, and color space are generally worse than the LCD-based system. The third prototype was implemented based on the S-AEM approach (Figure 5(b)). It employed the same projector as the projection-based I-AEM system that illuminates a wall. The projected image on the wall was reflected on a target glossy surface and observed by a user.

B. *Proof-of-concept experiment*

To confirm the feasibility of our proposed concept of AEM, we conducted two experiments. In these experiments, we particularly checked if AEM solved the essential technical issue—a displayed texture on a target glossy surface is stabilized while the viewing position is moving.

The first experiment was conducted using the projection-based I-AEM system (Figure 5(a)). The target object in this experiment was a flat mirror plate, and the target appearance was a colored geometric pattern. We prepared two experimental conditions, without and with AEM. In the first condition, we projected the target image onto the semi-transparent screen without applying any deformations. In the second condition, we deformed the target image by our AEM technique (see Section III-A). We recorded the appearance of the target object using a camcorder as an observer on which motion capture markers were attached to track its pose. Figure 6 shows the appearance manipulation results of both conditions. In the first condition (without AEM), the target object was just appeared to be a normal glossy surface, i.e., the displayed image reflected on the surface was moved according to the camera's movement. On the other hand, we confirmed that the AEM technique successfully altered the appearance of the glossy surface to the target one and made it stable for different viewing positions

The second experiment was conducted to see how our ST-AEM approach works using the LCD-based I-AEM system. As described in Section III-B, ST-AEM displays two layers of visual information, the view-independent surface layer and view-dependent environmental layer, on a glossy surface. We render the environmental layer using the second method described in Section III-B. In particular, we used

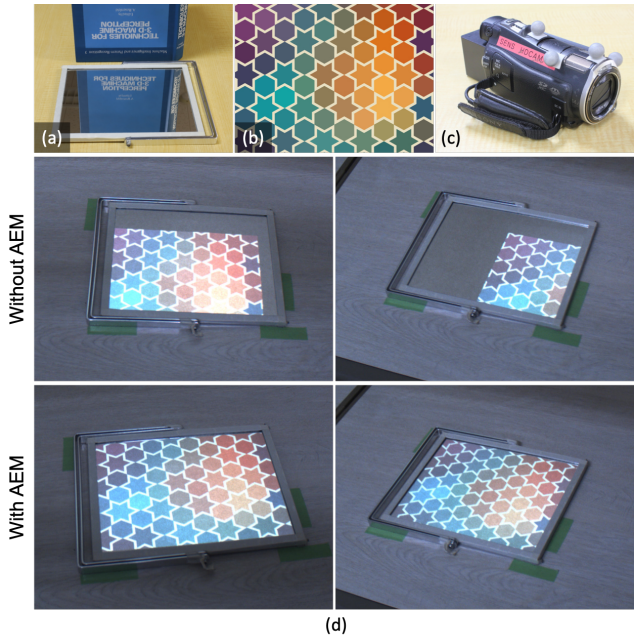


Figure 6. Result of the first proof-of-concept experiment (d), where the appearance of a flat mirror surface (a) was manipulated to be a colored geometric pattern (b) in two conditions (with and without AEM). Two images were captured from different viewpoints by a camcorder (c) in each condition.

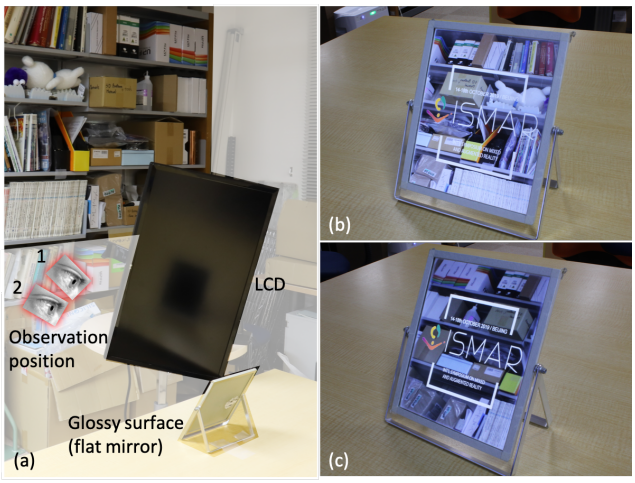


Figure 7. Result of the second proof-of-concept experiment, where the S-AEM approach was implemented: (a) system overview, (b) appearance manipulation result from the observation position 1 and (c) that from the position 2.

a camera to record the environmental scenes from different viewing positions in advance. After placing the LCD display, we displayed a recorded image which was captured from a position which is closest to the current observation position. Figure 7 shows the result using the same flat mirror plate as the target glossy surface. The logo of the ISMAR conference was used as the target appearance for the surface layer. We

confirmed that the displayed logo was static on the glossy surface, while the appearance of the environment displayed around the logo moved according to the camera's movement, which yields an impression that the LCD display except for the logo's area is transparent.

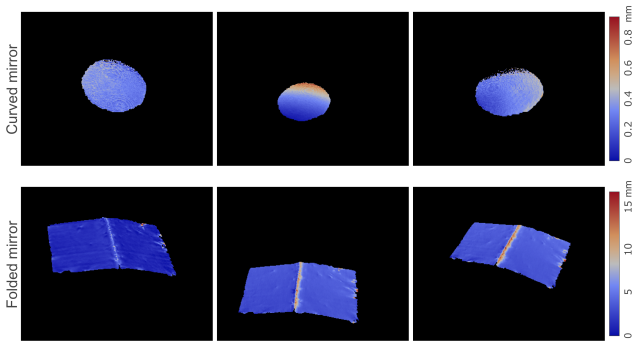
C. Evaluation

We formally evaluated our AEM systems through both quantitative and qualitative experiments.

1) Accuracy of appearance editing: The accuracy of appearance editing was evaluated by the error in $f(v)$, the mapping from each surface point s to the corresponding display point $d(v, s)$ (see Section III-A). We placed a camera at different positions. At each camera position v, we estimated the corresponding display point $d_e(v, s)$ for each surface point s by the proposed method. The ground truth of $d_t(v, s)$ was acquired by capturing the graycode pattern displayed on the display surface. Then, the error can be evaluated using the distance between $d_e(v, s)$ and $d_t(v, s)$ in the display image coordinate system. However, the error value is up to the area of a display region on which a texture is mapped. Therefore, we evaluated the error in the texture coordinate system. We converted $d_e(v, s)$ and $d_t(v, s)$ to the texture coordinate system ($t_e(v, s)$ and $t_t(v, s)$, respectively) and evaluated the error as the distance of these two points (i.e., the unit of the distance is [texel]). Consequently, we could know how far a displayed texel is from the correct texel.

We used one of the calibration cameras in the evaluation, to which motion capture markers were attached. We evaluated the accuracy using two experimental setups with different glossy surfaces. The first setup and glossy surface were the S-AEM system and the folded mirror surface whose width is 15 cm (Figure 9(b)). The second setup and glossy surface were the LCD-based I-AEM system and the curved mirror surface whose width is 6 cm (Figure 1(a)). We measured the error for three different camera locations in each setup. Figure 8 visualizes the errors converted texels to length. This result indicates that our method manipulates the appearance of the glossy surface at a high geometric accuracy on average (from 2.5 to 15.0 texels in 512×512 texel texture), while large errors caused at some small areas of the glossy surface. We will discuss about the error distribution in Section V.

2) User perception: Even when the appearance editing works geometrically correctly, an undesirable situation possibly occurs. Namely, an observer may still perceive that an AEM system manipulates the appearance of the reflected environmental surface rather than that of the glossy surface. In other words, the manipulated appearance may not be perceptually attributed to the target surface. Therefore, we conducted a user experiment to investigate if such undesirable situation occurs in an AEM system.



Ave. (max.) error [mm]	Left	Center	Right
Folded mirror	1.55(17.06)	4.49(16.14)	4.95(17.23)
Curved mirror	0.33(0.94)	1.65(2.77)	0.38(1.09)

Figure 8. Quantitative evaluation result: (top) heat maps of the errors in the mapping $f(v)$ (i.e., distance between $t_e(v, s)$ and $t_t(v, s)$) of the folded and curved mirrors that are evaluated at three different camera locations, and (bottom) their statistical information.

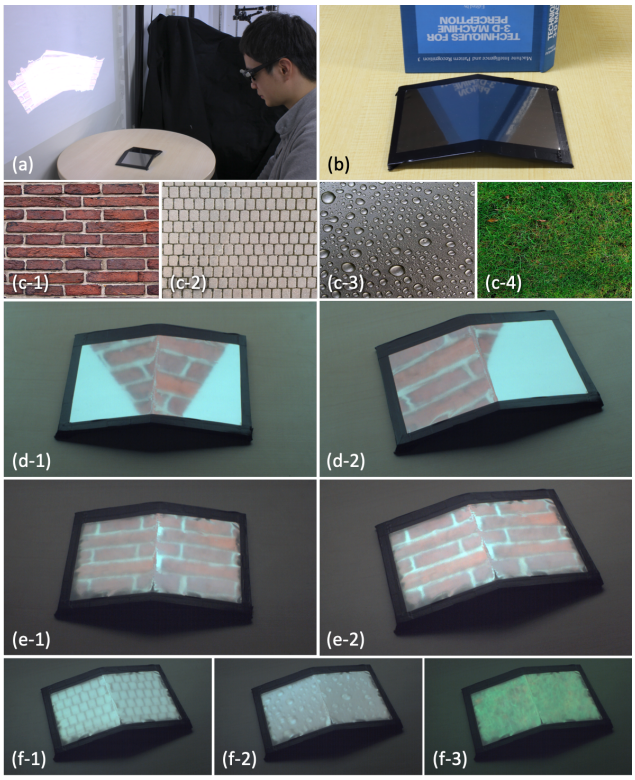


Figure 9. User experiment: (a) an overview of the experiment, (b) the target glossy surface (a folded mirror), (c) target textures to be appeared on a glossy surface, (d) appearance results from different viewpoints (captured by a tracked handheld camcorder), (e) appearance results of the same viewpoints of (d), and (f) appearance results of the other textures.

We conducted the experiment with an error-prone setup which has higher possibility to cause the undesirable situation than perfectly aligned setups. Particularly, we used the folded mirror surface used in Section IV-C1, which was found to contain some areas where large errors in the

mapping $f(v)$ occurred. The S-AEM system was applied. Four different textures were randomly projected at intervals of 10 seconds (Figure 9). There were two experimental conditions—with and without AEM. The textures were deformed by our AEM method and projected onto the environmental wall in the first condition, while they were not geometrically manipulated in the second condition. Twelve participants were recruited from a local university. In each condition, each participant was asked to watch the glossy surface as long as s/he decided the answer to a question whether or not s/he perceived the appearance manipulation was occurring on top of the glossy surface. As a result, eleven participants answered 'yes' in the first condition (with AEM), while all answered 'no' in the second condition (without AEM). From our informal interview, the participant who answered 'no' responded that he did not perceive the appearance manipulation occurred on top of the glossy surface because his eyes were attracted to unnatural deformations around an edge of the folded surface. Therefore, we consider the reason why this participant didn't prefer our solution is mainly due to the poor shape estimation accuracy.

This result indicates that the manipulated appearance is perceptually attributed to the target surface. Consequently, the result is an evidence proving that an observer perceives that an AEM system manipulates the appearance of the glossy target surface rather than that of the environmental surface reflected on the target surface. Therefore, we confirm that the undesirable situation discussed in the first paragraph of this section does not occur in AEM systems.

D. Application

We deployed AEM to a couple of application systems. The first application is a visual design support system for a glossy product. This application assumes a situation where a product designer explores different colors, patterns, and surface materials of the product. We implemented this application on the S-AEM system (Figure 10). The target glossy product was a lid of a cylindrical tea case. Tiny specular flakes are spread over the top, which are covered by a transparent specular layer (Figure 10(b)). As shown in Figure 10(f), conventional SAR systems cannot properly manipulate the surface appearance of the lid, due to the strong specular reflections of the flakes and the transparent layer. In contrast, the S-AEM system allowed a user to observe different visual designs of the product from arbitrary viewing positions (Figures 10(c)-(e)). From this result, we believe that AEM can be useful for the appearance design of glossy products.

As the second application example, we combined AEM with a projection mapping system. Projection mapping changes the appearances of surfaces of everyday objects such as buildings and rooms by projected imagery mainly for entertainment purposes. Projection mapping exhibitions

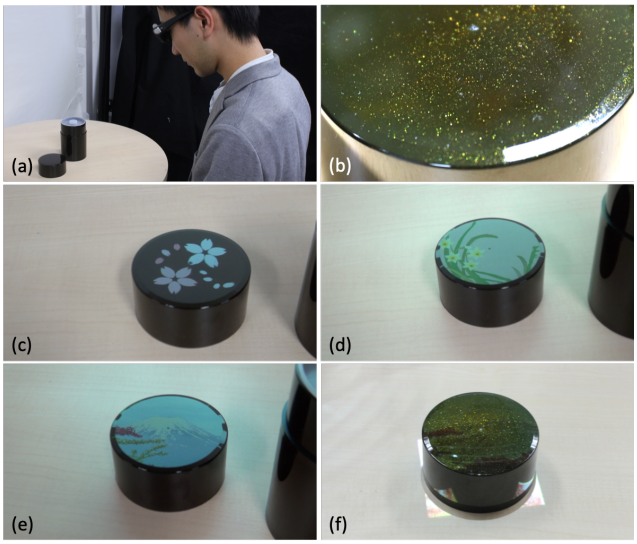


Figure 10. Glossy product design support application: (a) a user and a target glossy product (tea case), (b) close-up view of the lid of the tea case, (c)-(e) different appearance designs are displayed by AEM (captured by a tracked handheld camcorder), and (f) projection result of the same texture of (e) in a typical SAR system.

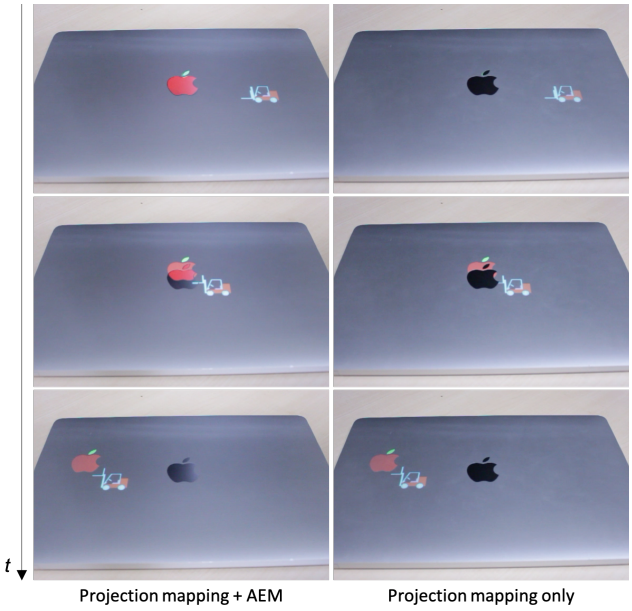


Figure 11. Projection mapping application.

are currently very popular in all over the world. However, conventional projection mapping systems have been limited to a scene where only diffuse surfaces exist. We overcome this limitation by combining our AEM framework. Figure 11 shows the time series a projection mapping application in two conditions, 'projection mapping and AEM' and 'projection mapping only'. We applied a surface (Apple MacBook) consisting of specular (Apple's logo) and diffuse

materials in this application. In the projection mapping and AEM condition, we applied an LCD-based I-AEM system. As shown in the right column of the figure, projected image contents on the specular surface (a colorful apple) is not observable in the projection mapping only setup. On the other hand, the combination of projection mapping and AEM can consistently display image contents as shown in the left column.

V. Discussion

So far, SAR researchers and developers have either unconsciously or intendedly removed glossy objects from their systems. On the other hand, the user test (Section IV-C2) showed that our proposed method, AEM, can successfully provide a human observer with the perception of manipulating the appearance of a glossy surface. This paper also shows that AEM works in actual application systems (Section IV-D). Therefore, AEM allows users to build their own SAR systems without the constraint of glossy objects. In addition, as discussed in Section III-B, our method can be implemented in various display setups. We believe that this flexibility is a strong point of AEM, which motivates the researchers and developers to apply it in their systems.

Although the user test shows the prototype system achieves our goal, there is still room for improvement in the current calibration method. Here, we discuss to reduce the errors in the mapping $f(v)$ visualized in Figure 8. Taking into account the size of the texture (512×512), the averaged error (from 2.5 to 15.0 texels) is already sufficiently small. We consider to further reduce the error. There are two major factors causing the error spreading over the surface. They are the inaccuracy of the camera pose estimation and that of the surface shape estimation. We could improve the error due to the first factor by using a motion capture system with a higher accuracy. We tested with less tracking cameras (three) and found the accuracy degradation was not significant. Therefore, the number of tracking cameras would not significantly reduce the errors. Although it was not examined, we believe that the error might be significantly reduced by adopting higher spatial resolution cameras in the motion capture system. The error due to this factor could be improved theoretically by an order of magnitude by acquiring camera-to-display pixel correspondences with subpixel accuracy, which can be achieved by exploiting a phase-shift pattern rather than a gray-code pattern. In addition, we noticed that the error becomes large at a surface area where the curvature is significantly large. This would be due to the error of shape estimation of such area. The density of observed display pixels in such area is high, and thus, the gray-code pattern becomes too dense to be correctly captured by the current camera in the shape estimation. This type of error could be alleviated by using a camera with a higher spatial resolution than the current one.

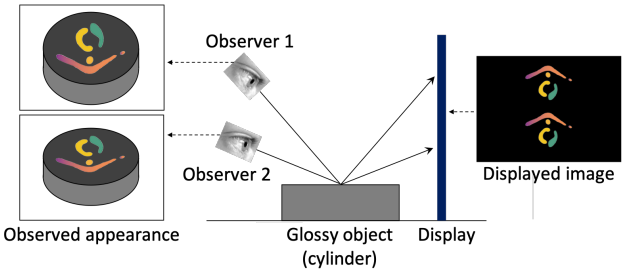


Figure 12. Extension for a multi-user system.

There are a few technical limitations of our current implementation. First, our system cannot modify the appearance of the surface whose normal vector is directed to an observer. Second, it is difficult to control a bumpy glossy surface, because there is a possibility that the same display pixel is observed at multiple surface positions. The PMD method (i.e., shape measurement technique for glossy surface) holds the same limitation. A potential solution would be to use a light field display that can emit directionally dependent light from each pixel [30]. Third, the number of observers is limited to one in the current setup. However, in the condition where target glossy surface is slightly curved, we can simply extend our system to multiple users at different viewing positions. Because they observe different display areas through the same glossy surface, we can show the same surface appearance for them by displaying the appearance at the observed display areas (Figure 12). This can be regarded as an application of IllusionHole mechanism, which realizes multi-user stereoscopic display [31]. Fourth, an observer is not allowed to move around quickly in our S-AEM setup due to high latency of the display. We could solve this problem by installing a stereoscopic display of law latency.

VI. Conclusion

This paper presents a novel SAR framework where the appearance of a glossy surface can be manipulated by augmenting the surrounding environments. Our proposed principle and calibration technique has been validated through subjective and objective experiments in various implementation methods. Although the current system contains errors in the mapping from each surface point to the corresponding display point, subjects in the user study satisfied with the appearance editing. We showed a couple of application systems to demonstrate the effectiveness of the appearance editing of glossy surfaces. We discussed our technical limitations of the current implementations and introduced the potential solutions such as phase-shift pattern projection for more accurate pixel correspondence acquisition.

A future work would be to improve the image quality of our system by adopting these solutions. Another interesting future direction is to deploy the principle of AEM to broader application fields.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP15H05925, Japan.

REFERENCES

[1] R. Raskar, G. Welch, K.-L. Low, and D. Bandyopadhyay, "Shader lamps: Animating real objects with image-based illumination," in *Proceedings of Eurographics Workshop on Rendering Technique*, 2001, pp. 89–102.

[2] D. Iwai, R. Matsukage, S. Aoyama, T. Kikukawa, and K. Sato, "Geometrically consistent projection-based tabletop sharing for remote collaboration," *IEEE Access*, vol. 6, pp. 6293–6302, 2018.

[3] M. Flagg and J. M. Rehg, "Projector-guided painting," in *Proceedings of ACM symposium on User interface software and technology*, 2006, pp. 235–244.

[4] R. Koch and C. Menk, "Truthful color reproduction in spatial augmented reality applications," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 2, pp. 236–248, 02 2013. [Online]. Available: doi.ieeecomputersociety.org/10.1109/TVCG.2012.146

[5] D. Iwai and K. Sato, "Limpid desk: See-through access to disorderly desktop in projection-based mixed reality," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '06. New York, NY, USA: ACM, 2006, pp. 112–115. [Online]. Available: http://doi.acm.org/10.1145/1180495.1180519

[6] A. H. Bermano, M. Billeter, D. Iwai, and A. Grundhöfer, "Makeup lamps: Live augmentation of human faces via projection," *Computer Graphics Forum*, vol. 36, no. 2, pp. 311–323, 2017.

[7] D. G. Aliaga, A. J. Law, and Y. H. Yeung, "A virtual restoration stage for real-world objects," *ACM Transactions on Graphics*, vol. 27, no. 5, p. 149, 2008.

[8] B. Jones, R. Sodhi, M. Murdock, R. Mehra, H. Benko, A. Wilson, E. Ofek, B. MacIntyre, N. Raghuvanshi, and L. Shapira, "Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units," in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '14. New York, NY, USA: ACM, 2014, pp. 637–644. [Online]. Available: http://doi.acm.org/10.1145/2642918.2647383

[9] M. Mine, D. Rose, B. Yang, J. van Baar, and A. Grundhöfer, "Projection-based augmented reality in disney theme parks," *IEEE Computer*, vol. 45, no. 7, pp. 32–40, 2012.

[10] O. Bimber and R. Raskar, *Spatial augmented reality: merging real and virtual worlds*. CRC Press, 2005.

[11] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer, "The visual computing of projector-camera systems," *Computer Graphics Forum*, vol. 27, no. 8, pp. 2219–2245, 2008.

[12] A. Grundhöfer and D. Iwai, "Recent advances in projection mapping algorithms, hardware and applications," *Computer Graphics Forum*, vol. 37, no. 2, pp. 653–675. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13387

[13] H. Park, M.-H. Lee, S.-J. Kim, and J.-I. Park, "Specularity-free projection on nonplanar surface," in *Advances in Multimedia Information Processing - PCM 2005*, Y.-S. Ho and H. J. Kim, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 606–616.

[14] H. Park, M.-H. Lee, B.-K. Seo, H.-C. Shin, and J.-I. Park, "Radiometrically-compensated projection onto non-lambertian surface using multiple overlapping projectors," in *Advances in Image and Video Technology*, L.-W. Chang and W.-N. Lie, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 534–544.

[15] T. Okazaki, T. Okatani, and K. Deguchi, "A projector-camera system for high-quality synthesis of virtual reflectance on real object surfaces," *IPSJ Transactions on Computer Vision and Applications*, vol. 2, pp. 71–83, 2010.

[16] M. K. Park, K. J. Lim, M. K. Seo, S. J. Jung, and K. H. Lee, "Spatial augmented reality for product appearance design evaluation," *Journal of Computational Design and Engineering*, vol. 2, no. 1, pp. 38 – 46, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2288430014000050

[17] M. R. Marner, R. T. Smith, J. A. Walsh, and B. H. Thomas, "Spatial user interfaces for large-scale projector-based augmented reality," *IEEE Computer Graphics and Applications*, vol. 34, no. 6, pp. 74–82, Nov 2014.

[18] C. Siegl, M. Colaianni, L. Thies, J. Thies, M. Zollhöfer, S. Izadi, M. Stamminger, and F. Bauer, "Real-time pixel luminance optimization for dynamic multi-projection mapping," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 237:1–237:11, Oct. 2015.

[19] Y. Watanabe, T. Kato, and M. Ishikawa, "Extended dot cluster marker for high-speed 3d tracking in dynamic projection mapping," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Oct 2017, pp. 52–61.

[20] L. Miyashita, Y. Watanabe, and M. Ishikawa, "Midas projection: Markerless and modelless dynamic projection mapping for material representation," *ACM Trans. Graph.*, vol. 37, no. 6, pp. 196:1–196:12, Dec. 2018. [Online]. Available: http://doi.acm.org/10.1145/3272127.3275045

[21] T. Hamasaki, Y. Itoh, Y. Hiroi, D. Iwai, and M. Sugimoto, "Hysar: Hybrid material rendering by an optical see-through head-mounted display with spatial augmented reality projection," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1457–1466, April 2018.

[22] G. Wetzstein and O. Bimber, "Radiometric compensation through inverse light transport," in *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, Oct 2007, pp. 391–399.

[23] M. C. Knauer, J. Kaminski, and G. Hausler, "Phase measuring deflectometry: a new approach to measure specular free-form surfaces," *Proc. SPIE*, vol. 5457, pp. 1–11, 2004. [Online]. Available: https://doi.org/10.1117/12.545704

[24] D. Iwai and K. Sato, "Optical superimposition of infrared thermography through video projection," *Infrared Physics & Technology*, vol. 53, no. 3, pp. 162 – 172, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350449509001650

[25] D. Saakes, H.-S. Yeo, S.-T. Noh, G. Han, and W. Woo, "Mirror mirror: An on-body t-shirt design system," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 6058–6063. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858282

[26] H. Yamamoto, H. Kajita, N. Koizumi, and T. Naemura, "Enchantable: Displaying a vertically standing mid-air image on a table surface using reflection," in *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, ser. ITS '15. New York, NY, USA: ACM, 2015, pp. 397–400. [Online]. Available: http://doi.acm.org/10.1145/2817721.2823476

[27] A. Grundhöfer and D. Iwai, "Robust, error-tolerant photometric projector compensation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5086–5099, Dec 2015.

[28] H. Ren, F. Gao, and X. Jiang, "Iterative optimization calibration method for stereo deflectometry," *Optics express*, vol. 23, no. 17, pp. 22 060–22 068, 2015.

[29] T. Zhou, K. Chen, H. Wei, and Y. Li, "Improved system calibration for specular surface measurement by using reflections from a plane mirror," *Applied optics*, vol. 55, no. 25, pp. 7018–7028, 2016.

[30] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 80:1–80:11, Jul. 2012. [Online]. Available: http://doi.acm.org/10.1145/2185520.2185576

[31] Y. Kitamura, T. Konishi, S. Yamamoto, and F. Kishino, "Interactive stereoscopic display for three or more users," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '01. New York, NY, USA: ACM, 2001, pp. 231–240. [Online]. Available: http://doi.acm.org/10.1145/383259.383285