Efficient Distortion-Free Neural Projector Deblurring in Dynamic Projection Mapping

Yuta Kageyama *Student Member, IEEE*, Daisuke Iwai, *Member, IEEE*, and Kosuke Sato, *Member, IEEE*

Abstract—Dynamic Projection Mapping (DPM) necessitates geometric compensation of the projection image based on the position and orientation of moving objects. Additionally, the projector's shallow depth of field results in pronounced defocus blur even with minimal object movement. Achieving delay-free DPM with high image quality requires real-time implementation of geometric compensation and projector deblurring. To meet this demand, we propose a framework comprising two neural components: one for geometric compensation and another for projector deblurring. The former component warps the image by detecting the optical flow of each pixel in both the projection and captured images. The latter component performs real-time sharpening as needed. Ideally, our network's parameters should be trained on data acquired in an actual environment. However, training the network from scratch while executing DPM, which demands real-time image generation, is impractical. Therefore, the network must undergo pre-training. Unfortunately, there are no publicly available large real datasets for DPM due to the diverse image quality degradation patterns. To address this challenge, we propose a realistic synthetic data generation method that numerically models geometric distortion and defocus blur in real-world DPM. Through exhaustive experiments, we have confirmed that the model trained on the proposed dataset achieves projector deblurring in the presence of geometric distortions with a quality comparable to state-of-the-art methods.

Index Terms—Projector deblurring, Geometric compensation, Dynamic projection mapping, Deep neural networks,

1 Introduction

Dynamic Projection Mapping (DPM) is an advanced technique for rapidly mapping images onto moving objects using a projector. Recent hardware advancements have significantly enhanced the practicality of DPM [1]–[6]. In contrast, Projection Mapping (PM) on static objects has already achieved significant progress [7], [8], finding applications in various fields such as medicine [9], industrial design [10], online conferencing [11], office work [12], [13], and entertainment [14], [15]. Consequently, the development of DPM is anticipated to augment the practicality of PM applications in these established areas and in other domains still under exploration. However, defocus blur poses a significant challenge for DPM. This issue arises because a projector typically has a very shallow Depth of Field (DoF) due to the large lens aperture required for emitting high-brightness images. Consequently, even a slight movement of the projected object in DPM may cause it to move out of the DoF, resulting in defocus blur in the projected image.

Numerous studies have focused on projector deblurring in PM, with most of them grounded in the understanding that defocus blur is modeled by the convolution of the projection image and the Point Spread Function (PSF). Prior research has successfully addressed defocus blur through PSF deconvolution [16]–[18], necessitating precise PSF estimation. PSF measurement methods involve projecting dot patterns [16], [17] or a natural image [18], [19] and capturing the projected light distribution with a camera. However, these techniques are limited to static projected objects. While some studies target dynamic scenes [20]–[22], they often involve the installation of special optics, such as an Electrically Tunable Lens (ETL), into the projector. This can result in significant image quality degradation, including color distortion due to optical aberrations. Kageyama et al. recently proposed a method to compensate for defocus blur in DPM without requiring special optics [23]. Instead, it utilizes camera-based feedback and a Deep Neural Network (DNN). While this technique demonstrates the feasibility of defocus blur compensation in DPM, it lacks consideration for geometric distortion of projected results and computation time, making it impractical for real-world DPM scenarios. DPM necessitates geometric registration of the projector to the projection target whenever the target is moved to visually align the projected image onto the moving target. Geometric registration typically involves finding pixel correspondence between the projector and the camera. The simplest method is projecting a sequence of Structured Light (SL) pattern images encoding the projection image coordinates. Although accurate, this technique is limited to static setups due to the need for multiple pattern projections. Therefore, imperceptible geometric registration techniques utilizing feature matching in projected natural images [24]–[26], tracking markers [4], [5], and spatio-temporal embedding techniques [27], [28] are essential in DPM. It is important to note that all geometric registration methods relying on the projected image cannot obtain exact pixel correspondence between the projector and the camera if the image is obscured by defocus blur.

This paper introduces an almost real-time approach to alleviate defocus blur and geometric distortion in DPM, obviating the necessity for specialized optical devices and aiming to achieve an all-in-focus DPM. For geometric compensation, we employed a state-of-the-art optical flow estimation network, GMA [29]. The estimated optical flow establishes pixel correspondence between the projector and camera images. To address projector deblurring, we introduced two lightweight sub-networks: PSFNet and SharpenNet. PSFNet is responsible for estimating the parameters associated with image quality degradation in the projected result. On the other hand, SharpenNet is tasked with optimizing the projection

Y. Kageyama, D. Iwai, and K. Sato are with the Graduate School of Engineering Science, Osaka University, Japan.
E-mail: see https://www.sens.sys.es.osaka-u.ac.jp

image by leveraging the estimated parameters. The projection of the optimized image effectively eliminates geometric distortion and defocus blur, as illustrated in Fig. 8.

The critical factor in achieving projection compensation in DPM through a neural framework lies in the training methodology of the network. The patterns of geometric distortion and defocus blur observed in DPM are notably diverse, contingent upon the position, posture, and characteristics of the projector, camera, and projection target. It is not feasible to train networks capable of compensating for these myriad degradation patterns from scratch within a real-world environment. Consequently, pre-training the network becomes imperative, necessitating a substantial dataset that encompasses a wide spectrum of degradation patterns. Collecting such extensive data in a real-world environment is impractical, and currently, publicly available data is nonexistent. To address this challenge, we have concentrated on recognizing the effectiveness of synthetic data in training networks across various domains [30], [31]. In other words, we propose an innovative method for generating realistic datasets within the virtual PM that incorporate both geometric and radiometric distortions.

Our novel dataset provides three significant advantages. Firstly, our network is fully trained using synthetic data and its parameters are optimized. Consequently, there is no need to adjust the network's parameters even when the projection target moves in DPM. Secondly, our dataset-trained GMA demonstrates robustness to image quality degradation in a PM environment. PM introduces complex image quality degradation due to a combination of nonlinear geometric distortions, luminance changes resulting from light attenuation and camera gain, projector gamma characteristics, and defocus blur. A pre-trained GMA, publicly available, is not equipped to handle such image quality degradation, leading to errors when tested in an actual PM environment. Conversely, GMA trained on our dataset, accurately reproducing the described image quality degradation, exhibits fewer errors. Finally, we can train GMA using data augmentation that assumes geometric compensation in DPM. The projection image in DPM with geometric compensation is not a natural image, as illustrated in the upper (b) of Fig. 3, but a non-linearly deformed image filled with black pixels, as illustrated in the lower (b) of Fig. 3. GMA must effectively perform the challenging task of estimating pixel-wise optical flow from the deformed projection image and its captured image. To address this challenge, we propose geometric data augmentation that non-linearly distorts the projection image used to train the GMA and fills the background of that image with black pixels.

Another contribution to realizing practical DPM is the lightweight of the deblurring network. The state-of-the-art method [23] for online deblurring requires more network parameters than necessary, making real-time compensated image generation impossible. In contrast, the proposed network has only 1% of the parameters of that method, significantly reducing the computation time required to generate compensated images. Through extensive experiments, we have verified that our method can effectively compensate for defocus blur in practical scenes, even with a significantly reduced number of parameters.

To summarize, our primary contributions are as follows:

- To the best of our knowledge, this study represents the first attempt to compensate for defocus blur in nearly real-time without special optical devices, even in the presence of geometric distortion.
- To accomplish this, we devised a geometric compensation network that demonstrates robustness to image quality degradation in DPM. Additionally, we created a projector deblurring network with a minimal number of parameters.
- We tackled the challenging task of collecting datasets that cover a multitude of patterns of geometric distortion and defocus blur in real-world DPM by replicating these degradation patterns within a virtual PM environment.
- We showcased our compensation's performance in terms of image quality and computational time through experiments involving various static and dynamic projection targets.

2 Related Work

Two major research topics related to our study are geometric compensation and projector deblurring, employing projector/camera pairs known as ProCams (projector-camera systems). We will first introduce previous works on these topics. Subsequently, we will describe "combined compensation" that performs these compensations simultaneously. Finally, we will outline our contributions in comparison to previous research.

2.1 Individual compensation

While geometric distortion and defocus blur often coincide in most PM scenarios, simultaneous compensation for these two distortions poses a significant challenge. As a result, many researchers have opted for independent compensation for each.

2.1.1 Geometric compensation

Geometric compensation can be achieved by establishing correspondence between projector and camera coordinates and warping the projection image accordingly to be appropriately superimposed on the target surface. The most well-known method is projecting a series of SL patterns, such as gray code patterns, encoded with the projector coordinates. The correspondence between projector and camera coordinates is determined by capturing and decoding these patterns with a camera. However, this method necessitates the projection of multiple artificial patterns, disrupting the PM application and limiting its use in a static environment. Alternatively, several techniques leverage feature point matching in natural images instead of artificial pattern images [24]–[26]. While this technique is less visually intrusive than projecting artificial patterns, feature detection is contingent on image contents and tends to fail when high-frequency feature points, such as edges in the projected image, are lost due to degradation factors, including defocus blur.

The challenge of geometric registration significantly intensifies in DPM, where the pose and location of the projection target rapidly change. The difficulty arises because the geometric calibration process must remain imperceptible to humans, requiring seamless integration into the system without causing noticeable disruptions or artifacts. One solution is to employ a coaxial ProCam system [36]–[38]. This specialized system ensures that the projector and camera pixels coincide regardless of the position and orientation of the projection target. Other studies have utilized the calibration pattern projection that is detectable by the camera but imperceptible to humans. For instance, humans do not perceive light sources flickering above 60 Hz, known as the Critical Flicker Fusion (CFF) frequency. Focusing on this physical characteristic, some researchers achieved geometric calibration by projecting an image embedded with a gray code at high speed [27], [28].

TABLE 1
Comparison of the proposed method with conventional defocus blur compensation algorithms using a single projector.

Methods	Geometric calibration	Online deblurring	Additional devices	Fine-tuning with real data
Non-DNN methods [16]–[18]	Required	No	No requirement	N/A
Non-DNN methods w/ ETL [20]–[22]	Required	Yes	Required	N/A
ProDebNet [19]	Required	No	No requirement	No requirement
OnlineProDeb [23]	Required	Yes	No requirement	No requirement
CompenNet++ family [32]–[35]	No requirement	No	No requirement	Required
Ours	No requirement	Yes	No requirement	No requirement

Another possibility is geometric calibration through infrared light projection [39], [40]. Although the methods mentioned above are capable of online geometric calibration, none supports projector deblurring.

2.1.2 Projector deblurring

There are two categories of defocus blur compensation techniques: the single-projector and multiple-projector approaches. In the single-projector approach, the projection image undergoes pre-sharpening to approximate the target appearance when projected, thereby compensating for defocus blur. The parameters for this pre-sharpening process rely on pixel-dependent PSFs. Therefore, the pixel-wise PSF is measured by projecting dot patterns [16], [17] or a target image [18]. The pre-sharpening process is achieved by deconvolution of the image and the estimated PSFs. Although sharpening with a Wiener filter is the most straightforward and fastest method [16], [18], this technique often suffers from ringing artifacts. While constrained iterative optimization can provide high-quality compensation [17], the computational complexity becomes a bottleneck. This trade-off between computational complexity and image quality can be addressed by using a coded aperture [41] or a DNN-based technique [19], [42]. All the mentioned methods share a common drawback: they require projecting calibration images and estimating the PSF each time the projection setup changes, making them unsuitable for DPM. This problem can be mitigated by using an additional optical component, such as ETL [20]–[22]. Kageyama et al. recently introduced an online deblurring technique [23] with the goal of achieving all-in-focus DPM without the need for special optical devices. While this method is closely related to ours, it lacks consideration for the geometric registration of the projector and the projection surface. Moreover, the method is associated with a significant drawback in terms of its long computation time, making it far from practical for real-world DPM applications.

In the multiple-projector approach, various projectors are focused on different positions. This arrangement enables the selection of the projector with the best focus at a specific point on the projection target, allowing the image to be projected from that projector as much as possible. As a result, the projected result avoids suffering from defocus blur [43], [44]. Additionally, the accuracy of defocus blur compensation can be further enhanced by optimizing the projection images to closely resemble the target image when projected by multiple projectors [3], [45]. Thus, the multiple-projector approach achieves projector deblurring in a manner impossible with the single-projector approach. However, complex geometric and radiometric calibration is required to achieve this goal.

While numerous methods have been developed for projector deblurring, all are based on the assumption that the pixel correspondence between the projector and camera is known or that the relationship between depth and PSF is known. This implies that geometric calibration is necessary before projector deblurring can be effectively implemented.

2.2 Combined compensation

Several researchers have tackled the challenging task of simultaneously compensating for geometric distortion and radiometric distortion, including defocus blur. The most robust but computationally expensive method is to use the full Light Transport Matrix (LTM) [46]. The full LTM can represent light transitions between all pixels of the projector and all pixels of the camera, thereby compensating for various image quality degradations beyond geometric distortion and defocus blur. However, its matrix is enormous (i.e., the product of the number of projector pixels and the number of camera pixels), making it computationally expensive. Additionally, the method requires sampling pattern projections to determine those elements, which is far from realizing practical DPM.

In recent years, some DNN-based combined compensation methods have been proposed. CompenNet++ and similar networks are pioneer works in this field, achieving end-to-end combined compensation [32]–[35]. However, full-scratch training of these networks requires the projection and capture of hundreds of sampling images in an actual PM setup. Therefore, these approaches share a common drawback: they necessitate data collection and network training every time the PM environment changes. To address this problem, Li et al. proposed a fast physics-based optimization method [47], which achieves combined compensation with the same accuracy as CompenNeSt++ [33] with less training data. Nevertheless, it still requires around ten pattern projections for optimization, making it unsuitable for practical DPM applications.

2.3 Our contribution

One of our main contributions is that we have addressed two unresolved challenges in the state-of-the-art deblurring method, OnlineProDeb [23], enabling practical DPM. Firstly, our technique can compensate for defocus blur even in the presence of geometric distortion. To achieve this, our network consists of two parts: one for geometric compensation and the other for projector deblurring. While one might assume our two-part network is similar to existing approaches achieving combined compensation [32], [33], [35], it surpasses them by not requiring sampling image projections for network training in an actual PM setup, allowing for the use of pre-trained models in arbitrary setups. Secondly, our deblurring network is significantly lighter, enabling real-time defocus blur compensation. OnlineProDeb used networks with an excessively large number of parameters for deblurring, which did not meet the image generation time requirement of a common projector refresh rate (i.e., $\geq$ 60 Hz). In contrast, we achieved qualitatively and

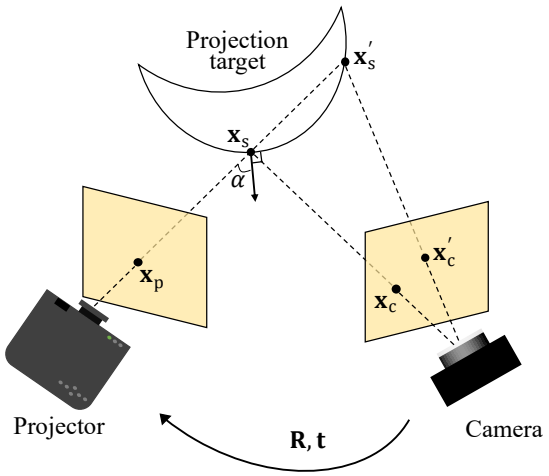


Fig. 1. Description of our virtual PM setup. This configuration comprises a projector, a camera, and a projection target that is uniformly white and entirely diffuse. It is crucial to note that when points on the surface are projected to the same coordinates of the projector, such as $\mathbf{x_s}$ and $\mathbf{x'_s}$, only the point closest to the projector is illuminated.

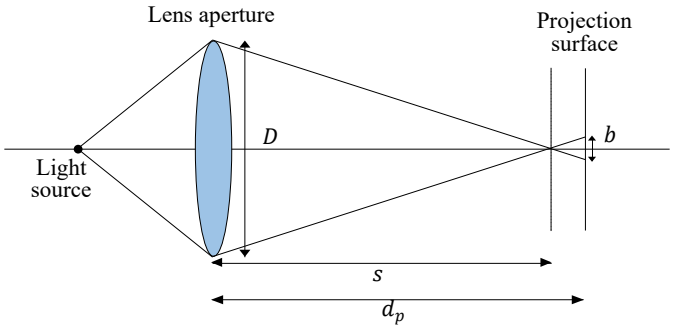


Fig. 2. Thin-lens model for computing the PSF of a projected pixel.

quantitatively better deblurring, even though we used a network with 1% of the parameters compared to OnlineProDeb.

Our other main contribution, which helps realize the two advantages mentioned above, is that we addressed the lack of datasets needed for training the network to achieve all-in-focus DPM. In DPM, the projection object of arbitrary shape moves rapidly, and the degree of geometric distortion and defocus blur changes each time it moves, resulting in myriad degradation patterns. Therefore, it is not feasible to collect these countless patterns in an actual PM setup. To tackle this problem, we propose a realistic synthetic data generation method by numerically modeling geometric distortion and defocus blur in real-world PM.

Since our primary focus is projector deblurring with a single projector, Table 1 compares this study with previous single-projector techniques that compensate for defocus blur.

3 Dataset Synthesis

This section outlines the rendering process involved in capturing an image projected onto a target in a virtual PM environment. Subsequently, we outline the assumptions guiding the parameter selection for data generation. Finally, we introduce the unique data augmentation techniques designed to address geometric compensation in DPM using our network.

3.1 Numerical modeling of PM

3.1.1 Geometric correspondence

As illustrated in Fig. 1, our setup in a virtual space consists of a projector, a camera, and a white, diffuse, arbitrarily shaped projection object. To render the captured image, it is essential to identify the pixel correspondence between the projector and camera images. To achieve this, we first determine the correspondence between each pixel of the camera $\mathbf{x}_c \in \mathbb{R}^2$ and the surface point $\mathbf{x}_s \in \mathbb{R}^3$ using the following equation.

$$\mathbf{x}_s = \mathbf{K}_c^{-1} \bar{\mathbf{x}}_c d_c, \tag{1}$$

where, $\mathbf{K}_c \in \mathbb{R}^{3\times3}$ is the camera's intrinsic parameter, $\bar{\mathbf{x}}$ is the homogeneous coordinate of $\mathbf{x}$, and d_c is the depth of $\bar{\mathbf{x}}_c$ in the camera coordinates. Let $\mathbf{K}_p \in \mathbb{R}^{3\times3}$ be the projector's intrinsic parameter and $\mathbf{R} \in \mathbb{R}^{3\times3}$, $\mathbf{t} \in \mathbb{R}^3$ be the relative rotation and translation of the camera from the projector. Then, the point $\mathbf{x}_s$ is illuminated by the point $\mathbf{x}_p$ in the projector image, which is obtained using the following equation.

$$\bar{\mathbf{x}}_p = \mathbf{K}_p[\mathbf{R}|\mathbf{t}]\bar{\mathbf{x}}_s. \tag{2}$$

It is important to note that several points in the camera image may map to the same projector coordinates. For example, in Fig. 1, two different camera coordinates $\mathbf{x}_c$ and $\mathbf{x}'_c$ map to the same projector coordinate $\mathbf{x}_p$. In such a case, we employ a z-buffer to ensure that only the surface point $\mathbf{x}_s$, closest to the projector, is illuminated by the projector.

3.1.2 Intensity of reflected light

Now that the correspondence between the projector coordinate $\mathbf{x}_p$ and the camera coordinate $\mathbf{x}_c$ has been established, the subsequent step involves calculating the luminance on the projected result captured by the camera when the projector displays the image. Let $\mathbf{I}_p(\mathbf{x}_p) \in \mathbb{R}$ be the pixel value at coordinate $\mathbf{x}_p$ in the projection image $\mathbf{I}_p$. Then, the light emitted from the projector undergoes a non-linear transformation, influenced by the display characteristics (i.e., Gamma characteristics), expressed as follows:

$$L(\mathbf{x}_p) = \{\mathbf{I}_p(\mathbf{x}_p)\}^\gamma, \tag{3}$$

where γ is the Gamma value of the projector. When the emitted light reaches the point $\mathbf{x}_s$, its irradiance is attenuated by two primary factors: the distance d_p from the projector to the surface and the incidence angle α of the light ray with respect to the surface. This attenuation term is referred to as the *form factor* [48], and by utilizing it, the irradiance at the surface point $\mathbf{x}_s$ due to light emitted from the point $\mathbf{x}_p$ can be computed using the following equation:

$$R(\mathbf{x}_c) = \frac{cos(\alpha)L(\mathbf{x}_p)}{d_p^2}. \tag{4}$$

3.1.3 Reproduction of defocus blur

Next, we will delve into the reproduction of defocus blur. Similar to certain prior studies [20], [23], we characterize the projector's lens as a thin lens. As depicted in Fig. 2, when the light emitted from a source point reaches a plane at a distance d_p, the light is perceived as a circle. The diameter of the blur circle b can be calculated using geometrical similarity as follows:

$$b = |D(\frac{d_p}{s} - 1)|, \tag{5}$$

TABLE 2
Description of the parameters randomly sampled during data generation.

Description	Camera intrinsic	Projector intrinsic	Rotation matrix		Transition vector	Gamma	Focal distance	Lens aperture	Camera gain
Mark	$\mathbf{K}_c$	$\mathbf{K}_p$	$\mathbf{R}$		$\mathbf{t}$	γ	s [m]	D [m]	G
	$f_c (f_x = f_y)$ [px]	$f_p (f_x = f_y)$ [px]	r_x, r_y [°]	r_z [°]	t_x, t_y, t_z [m]				
Value range	[1500, 2000]	[1000, 1200]	[-3.0, 3.0]	[-30, 30]	[-0.3, 0.3]	[1.0, 2.2]	[0.5, 4.0]	[1.5, 2.5]	[1.0, 2.0]

where D is the diameter of the projector's lens aperture, and s is the distance between the lens and the focusing point. It is important to note that the light source is not an ideal point source but rather a pixel on the projector's display. Consequently, the PSF of the light emitted from there is approximated by following an isotropic Gaussian function, rather than a pillbox function [16], [18], [19], [23]. This approximation is expressed as follows:

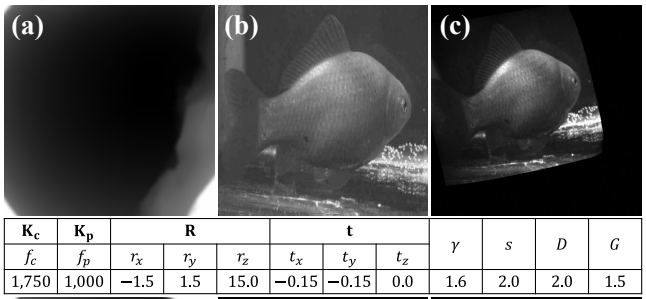
$$PSF(r,b) = \frac{2}{\pi b^2} \exp(-\frac{2r^2}{b^2}), \tag{6}$$

where r is the distance from the blur center. In accordance with this PSF model, the light calculated in Equation 4 will spread values to the surrounding area, representing defocus blur. Finally, the captured image is rendered by multiplying a camera gain G by the light emitted from the scene and adding slight Gaussian noise. It is important to note that the camera's DoF is sufficiently wide compared to that of the projector. Therefore, the defocus blur of the camera is disregarded in this study. Additionally, we assume that the camera's gamma characteristic is linear. This assumption is based on the fact that the gamma characteristic of most projectors is not user-adjustable, while most cameras provide users with the option to adjust the gamma characteristic.

3.2 Assumptions in parameter selection

This section outlines the assumptions made in selecting parameters for data synthesis. Since our PM environment is entirely virtual, we possess the flexibility to assign arbitrary values to the parameters of the projector and camera. This flexibility enables us to generate data spanning various PM setups. However, to maintain the realism of the rendered captured image and improve the training efficiency of the network, we impose constraints on the range of values for each parameter, as detailed in Table 2. The establishment of these parameter ranges is grounded in the following conditions.

- The resolution of the projection image is set to 256×256 pixels, while the captured image is configured at 600×600 pixels.
- The depth range over which the projection object exists is defined as $[1.5\ m,\ 2.5\ m]$, and the depth image values are normalized to fall within these ranges.
- The focal lengths (f_x and f_y) of the intrinsic parameter $\mathbf{K}$ are identical.
- The optical axes of the projector and camera intersect at the center of each image.
- Light that does not hit the projection target diverges to infinity and remains unobservable.
- It is ensured that at least a portion of the projection target is within the projector's field of view. Specifically, when the camera's optical axis intersects with the projection target at point $\mathbf{X}_s$, the projector is rotated after translation so that its optical axis intersects with point $\mathbf{X}_s$. The sampled rotation angles are then used to appropriately adjust the projector's orientation.



$\mathbf{K_c}$	$\mathbf{K_p}$	$\mathbf{R}$			$\mathbf{t}$			γ	s	D	G
f_c	f_p	r_x	r_y	r_z	t_x	t_y	t_z				
1,750	1,000	−1.5	1.5	15.0	−0.15	−0.15	0.0	1.6	2.0	2.0	1.5

$\mathbf{K_c}$	$\mathbf{K_p}$	$\mathbf{R}$			$\mathbf{t}$			γ	s	D	G
f_c	f_p	r_x	r_y	r_z	t_x	t_y	t_z				
1,900	1,200	1.0	−0.5	−5.0	0.1	0.1	0.1	2.0	1.5	2.0	2.0

Fig. 3. Two examples of our virtual PM. (a) represents the depth image of the projection target, while (c) illustrates the outcome of projecting the image in (b) onto this projection target using the parameters provided below. It is important to highlight that the projection image in the example below is deformed, a result of applying the geometric data augmentation method proposed in Sect. 3.3.

It is important to note that the projector is initially oriented toward the projection target even before applying the sampled parameters, in adherence to the six conditions outlined above. Consequently, the absolute values of the ranges for r_x and r_y in Table 2 are relatively smaller compared to r_z. Two examples of data synthesis under the specified conditions are illustrated in Fig. 3.

3.3 Geometric data augmentation for DPM

In the context of PM with a static projection target, it suffices to determine the pixel correspondence between the projector and the camera once. However, in DPM, this correspondence must be recalculated each time the projection target moves. In this scenario, the projection image used to find the pixel correspondence is a geometrically compensated image, which is non-linearly deformed, and the background is filled with black pixels. To train GMA to account for this situation, we propose a geometric data augmentation that fills the background with black pixels after a non-linear transformation of the projection image by combining Affine and Thin-Plate Spline (TPS) transforms. The projection image transformed by our data augmentation is depicted in the bottom (b) of Fig. 3, and its appearance resembles a geometrically compensated image.

4 Compensation Algorithm

We introduce a neural framework designed to generate a projection image that compensates for defocus blur, even in the presence of

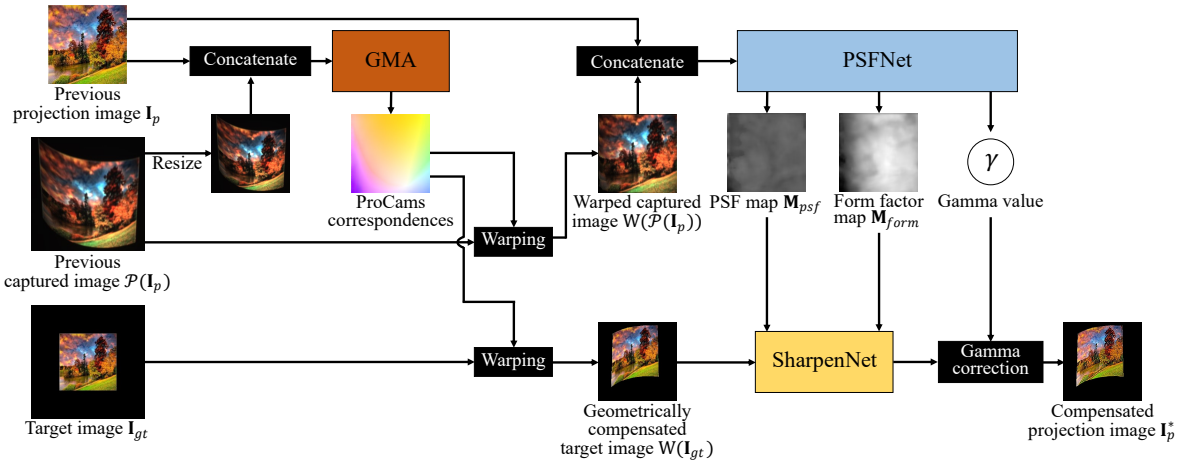


Fig. 4. An overview of our network, comprising two main components: the geometric compensation part and the defocus blur compensation part. In the former, GMA estimates the optical flow between the projection image and the captured image of the previous frame, establishing pixel correspondence between the projector and the camera. The latter includes PSFNet and SharpenNet. Initially, PSFNet calculates the pixel-wise PSF map, the pixel-wise form factor map, and the projector's gamma value from the projection and warped captured images. Based on the estimated parameters, SharpenNet sharpens the geometrically compensated target image.

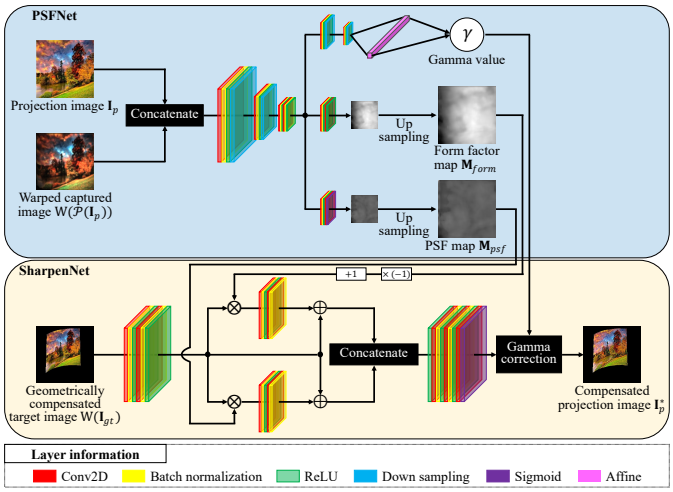


Fig. 5. PSFNet and SharpenNet are detailed as follows: PSFNet begins by extracting low-resolution features through convolution and downsampling of the concatenated input images. Subsequently, three heads are employed to output the projector's gamma value, the 1/16 size form factor map, and the 1/16 size PSF map. Bicubic interpolation is then utilized to resize the maps to match the size of the projection image. SharpenNet combines the resized maps with the geometrically compensated target image in feature space. This process yields an intermediate image, which undergoes gamma correction to produce the final compensated projection image.

geometric distortion. This section delves into the specific details of the proposed framework. It is important to note that in this study, as in several previous studies [23], [32], [33], [47], the user's viewpoint is replaced by the camera viewpoint. Therefore, our research goal is for the projected image captured by the camera to match the target image.

4.1 Network design

In DPM, all frames of the projection image need to be compensated while the object is in motion. For this reason, our framework aims to generate a projection image that compensates for geometric distortion and defocus blur based on both the projection image and its captured image of the previous frame. However, accomplishing this goal within a single network is a complex and challenging task. Therefore, we have determined that the three processes of geometric warping, radiometric degradation estimation, and image sharpening are essential to achieving our objective. Consequently, we have divided the entire network into these three distinct parts, as illustrated in Fig. 4. The details of each part are explained in the following sections. It is worth noting that although the images in Fig. 4 and Fig. 5 are visualized in RGB color spaces, the network processing is conducted solely on the value component of the HSV-transformed image.

4.2 Geometric warping

As depicted in Fig. 4, the initial step in our framework is warping the captured image to a projector viewpoint and applying geometric compensation to the target image. To achieve this, it is crucial to establish pixel correspondence between the projector and the camera. For this purpose, we opted to detect the optical flow of all pixels from the projection image to the captured image. Optical flow is typically detected when the same scene is captured from two different viewpoints. However, in this research, it is necessary to detect the optical flow from two images in different domains: the projection image and the captured image with geometric distortion and defocus blur. Therefore, we utilized GMA [29], a DNN-based optical flow estimation network, as classical optical flow detection methods such as the Lucas-Kanade method [49] are unsuitable for our task.

The input/output of GMA can be described by the following equation:

$$\mathbf{U}_{p2c} = GMA(\mathbf{I}_p, \mathscr{P}(\mathbf{I}_p)_{\downarrow}), \tag{7}$$

where $\mathbf{I}_p \in \mathbb{R}^{H_p \times W_p}$ and $\mathscr{P}(\mathbf{I}_p) \in \mathbb{R}^{H_c \times W_c}$ are the projection image and the captured image of the previous frame (i.e., $\mathscr{P}$ is the projection and capture function), and $\downarrow$ indicates that the captured image is downsampled to the same size as the projection image. $\mathbf{U}_{p2c} \in \mathbb{R}^{2 \times H_p \times W_p}$ represents the pixel-wise optical flow from the projection image to the captured image. H_p and W_p represent the height and width of the projection image, while H_c and W_c represent those of the captured image. Using the optical flow, the captured

image warped to the projector viewpoint $\mathscr{W}(\mathscr{P}(\mathbf{I}_p)) \in \mathbb{R}^{H_p \times W_p}$ can be obtained from the following equation:

$$\mathscr{W}(\mathscr{P}(\mathbf{I}_p))(x,y) = \mathscr{P}(\mathbf{I}_p)_{\downarrow}(x+u_x, y+u_y), \quad (8)$$
$$u_x, u_y = \mathbf{U}_{p2c}(x,y). \quad (9)$$

Similarly, the same warping process can be applied to the target image $\mathbf{I}_{gt} \in \mathbb{R}^{H_c \times W_c}$ to obtain a geometrically compensated image $\mathscr{W}(\mathbf{I}_{gt}) \in \mathbb{R}^{H_p \times W_p}$ as follows:

$$\mathscr{W}(\mathbf{I}_{gt})(x,y) = \mathbf{I}_{gt_{\downarrow}}(x+u_x, y+u_y), \quad (10)$$
$$u_x, u_y = \mathbf{U}_{p2c}(x,y). \quad (11)$$

The loss function for GMA is described by the following equation:

$$\mathscr{L}_{mse}(\mathbf{U}_{p2c}, \mathbf{U}^{gt}_{p2c}), \quad (12)$$

where, $\mathbf{U}^{gt}_{p2c}$ is the Ground-Truth (GT) P2C map, and $\mathscr{L}_{mse}$ measures the mean squared error between each element in the estimated and GT P2C maps. It is important to note that there are points in the projection image that did not illuminate the projection object or were not captured by the camera due to self-occlusion. In the calculation of the loss, these coordinates are excluded.

4.3 Radiometric degradation estimation

After geometric warping, the subsequent task involves estimating pixel-wise radiometric distortion in the projection image. As discussed in Sect. 3.1, in addition to geometric distortion, the primary factors influencing changes in the appearance of the projected result's captured image include the projector's gamma characteristics, the form factor, and defocus blur. To estimate these three factors, we propose a three-headed PSFNet, denoted as $\mathscr{N}_{PSF}$. The detailed structure of PSFNet is illustrated in the upper row of Fig. 5.

The projection image and the warped captured image are concatenated into the channel dimension and then input into the convolution block. The input undergoes multiple convolutions and downsampling operations within the network, followed by splitting into three heads. The first head performs two convolutions and downsampling of the branched features, finally estimating the gamma value γ of the projector through an affine layer. The second head estimates a form factor map $\mathbf{M}_{form} \in \mathbb{R}^{H_p \times W_p}$, representing how much the luminance of the captured image has changed relative to the projection image. The last head estimates the PSF map $\mathbf{M}_{psf} \in \mathbb{R}^{H_p \times W_p}$, representing pixel-wise PSFs. Here, as indicated in Equation 6, we approximate the PSF using an isotropic Gaussian distribution. Thus, each pixel in the estimated PSF map contains the standard deviation of the Gaussian distribution.

To reduce the number of trainable parameters in the network compared to OnlineProDeb [23], we implemented two design improvements to streamline the network. Firstly, we structured the network into a three-head architecture, enabling common weights for feature extraction up to branching. Secondly, we reduced the resolution of the map output from the convolution block to 1/16 of the projection image by employing bicubic interpolation at the end of the head. Given that the majority of projection surfaces in a typical PM exhibit smoothness, up-sampling the maps through bicubic interpolation is a reasonable approach. These simplifications significantly decrease the parameters of our network to 1% of those in OnlineProDeb, as illustrated in Table 4.

In summary, PSFNet is represented by the following equation:

$$\gamma, \mathbf{M}_{form}, \mathbf{M}_{psf} = \mathscr{N}_{PSF}(\mathbf{I}_p, \mathscr{W}(\mathscr{P}(\mathbf{I}_p))). \quad (13)$$

We define the loss function for PSFNet using the following equation:

$$\mathscr{L}_{mse}(\gamma, \gamma^{gt}) + \mathscr{L}(\mathbf{M}_{form}, \mathbf{M}^{gt}_{form}) + \mathscr{L}(\mathbf{M}_{psf}, \mathbf{M}^{gt}_{psf}), \quad (14)$$
$$\mathscr{L} = \mathscr{L}_{mse} + \mathscr{L}_{tv}, \quad (15)$$

where, γ^{gt} and $\mathbf{M}^{gt}_{psf}$ are the GT values of the gamma and the PSF map obtained from Equation 3 and Equation 6, respectively. $\mathbf{M}^{gt}_{form}$ is the GT value of the form factor map and is obtained by the product of the attenuation term described in Equation 4 and the camera gain G. $\mathscr{L}_{tv}$ is the total variation [50] loss used to regularize the estimated form factor map and PSF map.

4.4 Image sharpening

Our final task is to sharpen the geometrically compensated target image $\mathscr{W}(\mathbf{I}_{gt})$ according to three parameters estimated by PSFNet. To achieve this, we introduce SharpenNet, denoted as $\mathscr{N}_{Sharp}$. We utilize the PSF map and the form factor map as attention maps in the feature space to enhance image sharpness. Thus, we model SharpenNet with the following equation:

$$\mathbf{I}'_p = \mathscr{N}_{Sharp}(\mathscr{W}(\mathbf{I}_{gt}), [\mathbf{M}_{psf}], [\mathbf{1} - \mathbf{M}_{form}]), \quad (16)$$
$$\mathbf{I}^*_p = (\mathbf{I}'_p)^{1/\gamma}, \quad (17)$$

where, Equation 17 corresponds to gamma correction applied to linearize the luminance of the projector. $\mathbf{I}^*_p \in \mathbb{R}^{H_p \times W_p}$ is a projection image designed to compensate for geometric distortion and defocus blur simultaneously, and the brackets indicate the injection of the estimated maps into the middle layers. It is worth noting that the form factor map is input into the network after being subtracted from the matrix $\mathbf{1} \in \mathbb{R}^{H_p \times W_p}$ where all elements are 1. This subtraction is performed because the form factor map reflects the magnitude of luminance change due to image projection and capturing. The further the value is from 1, the more compensation is needed in that region.

The detailed flow of SharpenNet is depicted in the lower part of Fig. 5. Initially, convolutional blocks extract features from the geometrically compensated target image $\mathscr{W}(\mathbf{I}_{gt})$. Subsequently, we compute the Hadamard product of the extracted features with the form factor map and the PSF map, respectively. After passing through the convolution block, the features are added to the original features. Finally, these features are concatenated in the channel direction and subjected to several convolution operations to output a single image. The resulting image is then gamma-corrected to obtain the final compensation image $\mathbf{I}^*_p$.

The loss function for training SharpenNet is as follows:

$$\mathscr{L}_{lpips}(\mathscr{P}(\mathbf{I}^*_p), \mathbf{I}_{gt}) + \mathscr{L}_{mse}(\mathscr{P}(\mathbf{I}^*_p), \mathbf{I}_{gt}), \quad (18)$$

where, $\mathscr{L}_{lpips}$ measures LPIPS [51], a DNN-based image quality metric, between the two given images. Consequently, the training of SharpenNet advances such that the projected result of the compensation image approaches the target image. It is worth noting that the loss function is differentiable since the projection function $\mathscr{P}$ is implemented in our virtual PM setups.

5 Experiment

We evaluated the proposed network in physical setups. This section begins with a detailed description of the experimental setup. In the following sections, we discuss the efficacy of the proposed method in both static and dynamic scenes, comparing it with state-of-the-art methods.

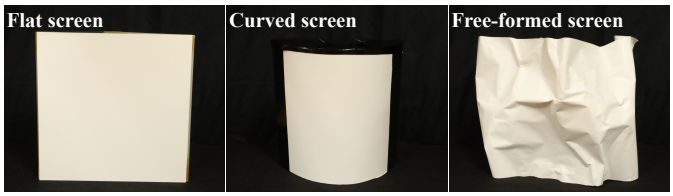


Fig. 6. Three types of screens used in the experiment: flat, curved, and free-formed.

TABLE 3
Quantitative comparison between the proposed method and state-of-the-art methods. It is important to note that the metrics for uncompensated results are as follows: PSNR = 10.44, SSIM = 0.186, LPIPS = 0.721, and DISTS = 0.344.

Metrics	SL	OnlineProDeb w/ SL	CompenNeSt++	Ours w/o deblur	Ours
PSNR (↑)	16.48	13.52	19.71	15.52	17.55
SSIM (↑)	0.500	0.462	0.599	0.398	0.511
LPIPS (↓)	0.561	0.562	0.466	0.558	0.477
DISTS (↓)	0.289	0.299	0.244	0.284	0.251

5.1 Experimental setup

5.1.1 Training details

To train our network, we utilized 201,000 images from ImageNet [52] for the projection images, reserving 1,000 for validation. Concurrently, we acquired 51,000 depth samples of the projection target from the OmniObject3D dataset [53]. Among these, 50,000 were allocated for training and 1,000 for validation. Notably, we excluded objects unsuitable for PM, such as toothbrushes and asparagus, from the OmniObject3D dataset.

All trainable parameters in the proposed network were determined using the method by He et al. [54]. Optimization was performed using the Adam optimizer [55] with a learning rate of 1e-3, and momentum parameters were set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The training procedure involved initially training only GMA. Once GMA was trained, its weights were frozen, and PSFNet and SharpenNet were trained simultaneously. The batch size and the number of epochs are 8 and 5, respectively. The training times were approximately 19 hours for GMA and about 31 hours for the combined training of PSFNet and SharpenNet. The source code is implemented using PyTorch, and the network was trained on a shared workstation equipped with a GPU (NVIDIA RTX A6000, GPU memory: 48 GB) and CPU (Intel Xeon Platinum 8260, CPU memory: 768 GB).

5.1.2 Testing details

Since our network does not require fine-tuning in any experimental setting, we utilized frozen weights trained under the previously described conditions in all experiments. We consistently employed the same DLP projector (Optoma ML1050ST+) and industrial CMOS camera (FLIR FL3-U3-13S2C-CS) across all subsequent experiments. The number of pixels in the projection image is 256×256, and the number of pixels in the captured image is 600×600. The conditions for source code implementation are consistent with those used for training.

5.2 Comparison with state-of-the-art methods

In comparison to the proposed method, we considered two state-of-the-art methods with publicly available source codes. The first method is OnlineProDeb, a state-of-the-art defocus blur compensation method [23]. OnlineProDeb eliminates the need for training the network for each projection environment and can be applied immediately after downloading publicly available weight parameters. However, it requires a separate geometric registration, involving the projection of 42 gray-code SL patterns to establish pixel correspondence between the projector and the camera. The second method is CompenNeSt++, which concurrently performs geometric and radiometric compensation [33]. Unfortunately, CompenNeSt++ necessitates the projection of images to train the network for each projection environment. Consequently, we conducted this comparison experiment using a fixed setup of the screen and the ProCam system. For training CompenNeSt++, we projected and captured 125 sampling images, with a training time of five minutes. Additionally, we compared our method's results with those achieved by solely performing geometric compensation (denoted as "Ours w/o deblur") and those obtained by combining both geometric and defocus blur compensation (denoted as "Ours"). To assess these techniques in various setups, we utilized three types of screens: flat, curved, and free-formed, as illustrated in Fig. 6.

Figure 7 presents results from the comparison experiment. The uncompensated results exhibit severe geometric distortion, attributed to the complex geometry of the projection surface. This distortion is effectively mitigated by projecting SL patterns. Notably, in column six, "Ours w/o deblur" demonstrates comparable geometric compensation to SL, despite not being explicitly trained on these setups. However, the image quality is inferior to the target image due to defocus blur. OnlineProDeb addresses defocus blur, albeit with a slight reduction in brightness. This reduction is attributed to the training approach of OnlineProDeb, which considers the projector's dynamic range, leading to lower luminance in the output image [23]. CompenNeSt++ outperforms OnlineProDeb as it exhibits a more profound understanding of light transport in the experimental setups, facilitated by utilizing a large number of sampling images during training. Results from "Ours" in the rightmost column demonstrate effective geometric and defocus blur compensation, similar to CompenNeSt++, despite our network not being trained with sampling images in these specific setups.

We also conducted a quantitative evaluation of the results obtained in this experiment. For each of the three setups depicted in Fig. 7, we projected 100 evaluation images as presented in [33]. Subsequently, we assessed the similarity between these projected results and target images using four metrics: PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity) [56], LPIPS [51], and DISTS [57]. LPIPS and DISTS are DNN-based metrics, offering a more human-sensitive evaluation than PSNR and SSIM.

Table 3 presents the evaluation values. Similar to the qualitative evaluation results, we verified that "Ours" outperforms all other methods, except CompenNeSt++, across all evaluation measures. When comparing "Ours" with CompenNeSt++, a noticeable distinction arises in the context of PSNR and SSIM. This disparity stems from our compensation being solely applied to the value component in the HSV space of the image, whereas CompenNeSt++ compensates for the entire image in the RGB space. Consequently, disparate values were obtained for PSNR and SSIM, metrics designed to gauge image differences in RGB space. In contrast, LPIPS and DISTS, metrics more rooted in human perception, did not exhibit significant differences. This underscores that our method is comparable to CompenNeSt++ in terms of human perception, despite our networks not being fine-tuned specifically for these scenes.

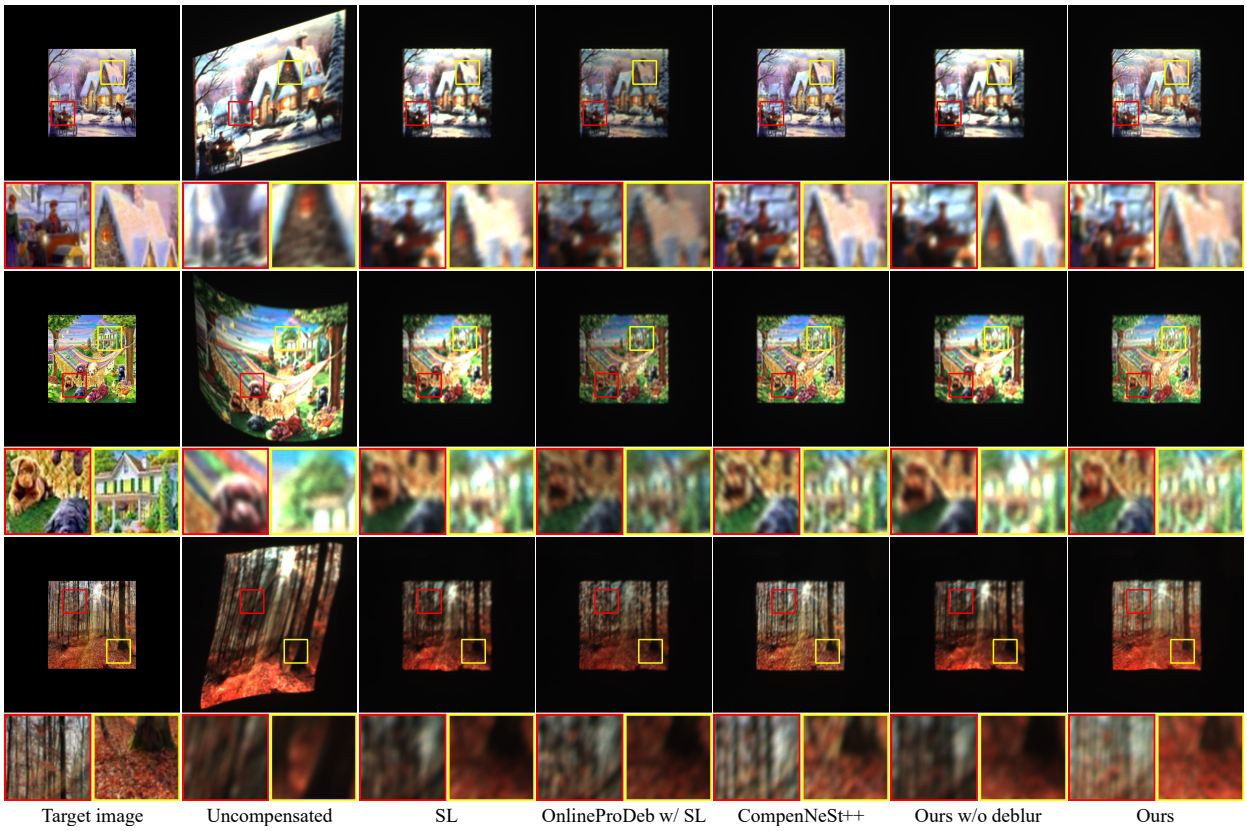


Fig. 7. Comparison of compensation results between the proposed method and state-of-the-art methods on three surfaces: (upper) a flat surface, (middle) a curved surface, and (lower) a free-formed surface. The first and second columns represent the target and uncompensated captured images, respectively. The subsequent columns illustrate the projected results compensated by various methods.

TABLE 4
Comparison of the number of parameters and computation time between OnlineProDeb and our network. The generated image is of size 256×256 pixels.

	Parameters		Computation time (ms)	
Methods	OnlineProDeb	Ours	OnlineProDeb	Ours
Warping	N/A	5,867,329	N/A	49.70
Deblurring	12,307,211	130,264	111.1	5.941
Total	12,307,211	5,997,593	111.1	55.64

5.3 Dynamic PM

We proceeded to validate the effectiveness of the proposed method in the context of DPM. The designed setup is illustrated in Fig. 8(a), where a curved white screen is affixed to the robot arm (UFACTORY xArm 7). In this experiment, we compared the projected results under three conditions: "Uncompensated," "Ours w/o deblur," and "Ours." The projection target underwent the same pre-determined movement controlled by the robot arm in each condition. Notably, the images projected at each position of the projection target were not identical due to the projector displaying a movie in this experiment.

Figure 8(c) illustrates the uncompensated results, which exhibit degradation from geometric distortion and defocus blur, causing their appearance to deviate significantly from the target images of each frame. Moreover, the position and orientation of the screen vary in each frame, resulting in distinct degradation patterns. We verified that the severe geometric distortion was notably alleviated by "Ours w/o deblur." This indicates that the geometric compensation introduced by the proposed method performs effectively in the context of DPM. Although the image quality suffered from defocus blur, the projection with "Ours" demonstrated an improvement in image quality. This implies that our method successfully achieved geometric compensation and projector deblurring in DPM.

5.4 Validation of geometric data augmentation

This section demonstrates the impact of the proposed geometric data augmentation, as explained in Sect. 3.3. We trained the network under the exact conditions as the proposed method, but without the geometric data augmentation, referring to this network as "Ours w/o aug." To assess the compensation accuracy of this network compared to the proposed network, we conducted video projection in a static setup.

Figure 9 illustrates the outcomes of the first three frames of the projected video. In both conditions, the projection of the first frame is distorted because an uncompensated image is projected. The result for the second frame is the projection of the compensated image generated by each network from the projected image of the first frame and its captured image. Similarly, the result for the third frame is the outcome of projecting the compensated image generated by each network from the projected image and its captured image to which the compensation of the second frame has been applied. We observed that while the results for the second frame are similar for both conditions, the results for the third frame exhibit distortion in "Ours w/o Aug." (indicated by red arrows). This distortion arises because the second frame projection image used to generate the third frame projection image was a geometrically compensated image, and there was no such input

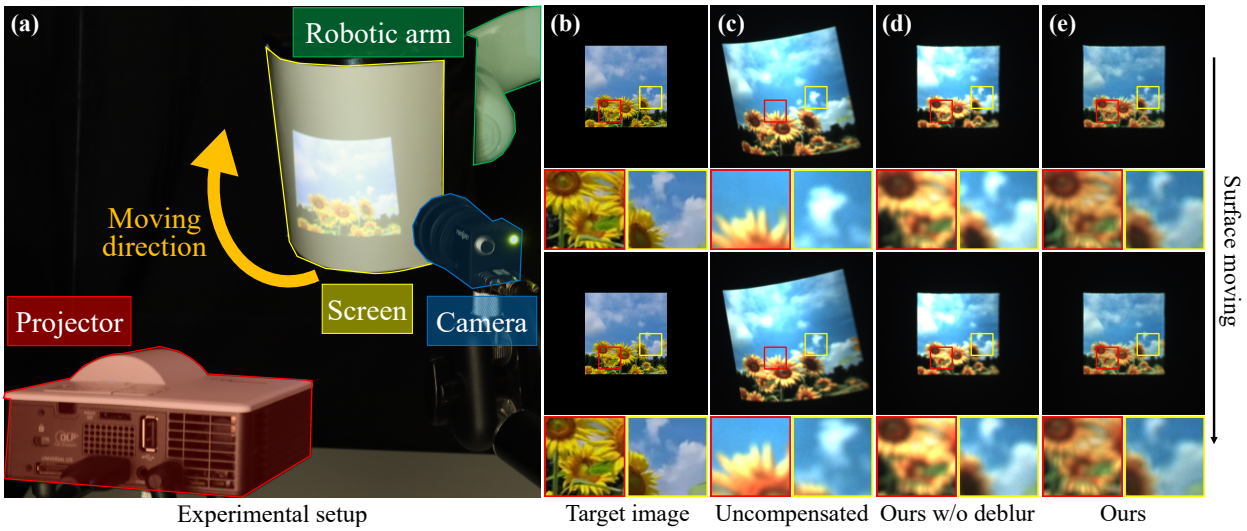


Fig. 8. Our compensation result for defocus blur in DPM, even in the presence of geometric distortion. (a) Our experimental setup, where a robotic arm moves the curved white screen, and (b) target images. (c) Without compensation, when the projector displays the images on the moving screen, the uncompensated results suffer from severe geometric distortion and defocus blur. (d) GMA, trained on our novel dataset, effectively eliminates geometric distortion. (e) Our lightweight deblurring network, comprising PSFNet and SharpenNet, also real-time compensates for defocus blur.

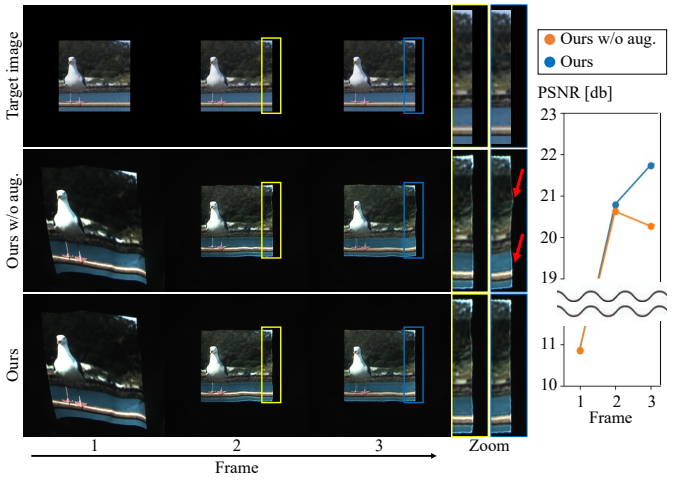


Fig. 9. Validation of geometric data augmentation in the static ProCam setup. The projection results for the second frame exhibit approximately the same compensation accuracy under the conditions of "Ours w/o aug." and "Ours." However, for the third frame, differences in compensation accuracy emerge between the two conditions, contingent on whether the proposed data augmentation is utilized or not.

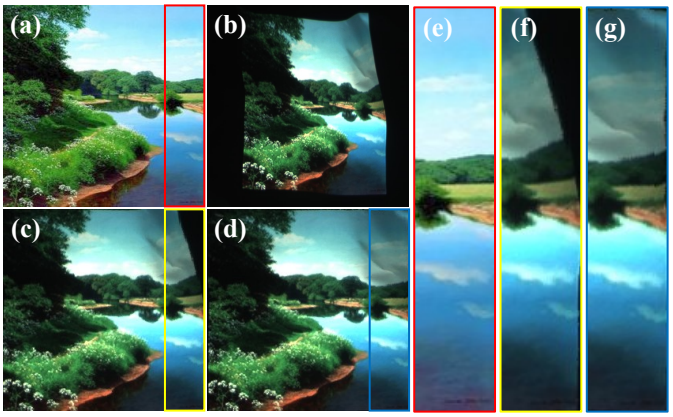


Fig. 10. Comparison of GMA pre-trained on the Sintel dataset and GMA trained on our proposed dataset. (a) The projection image and (b) the distorted projected result. The warped images of (b) by GMA pre-trained on the Sintel dataset and our proposed dataset are represented in (c) and (d), respectively. Additionally, (e), (f), and (g) present enlarged views of a portion of (a), (b), and (c), respectively. Upon comparing (f) and (g), it becomes evident that GMA trained on the proposed dataset performs more accurate warping.

during the training of "Ours w/o aug." On the other hand, during the training of "Ours," the input was a deformed projection image due to the proposed data augmentation. This enables the results of the third frame to maintain the same quality as the results of the second frame.

We present these results quantitatively as well. The right graph in Fig. 9 plots the PSNR between the target image and the projection result in each frame. The results for the second frame are similar for each condition, reflecting the removal of distortion. However, in the third frame, the differences between conditions become more pronounced, underscoring the efficacy of the proposed data augmentation.

5.5 Comparison with pre-trained GMA

GMA [29], employed in the geometric compensation component, is an exceptional network for estimating optical flow. Consequently, the publicly available pre-trained model may suffice for geometric compensation in PM. A previous study successfully demonstrated geometric compensation using a pre-trained model [47]. This section compares the pre-trained model and the GMA trained on the proposed dataset. It is important to note that the dataset used to pre-train GMA is the Sintel dataset [58].

Figure 10 presents the comparison results. The projection results exhibit severe geometric distortion due to the image being projected onto a free-formed screen. Figures 10(c) and (d) showcase the outcomes of warping the captured images from these projections using the pre-trained GMA and our GMA, respectively. While the pre-trained GMA produces near-precise warping, we observed

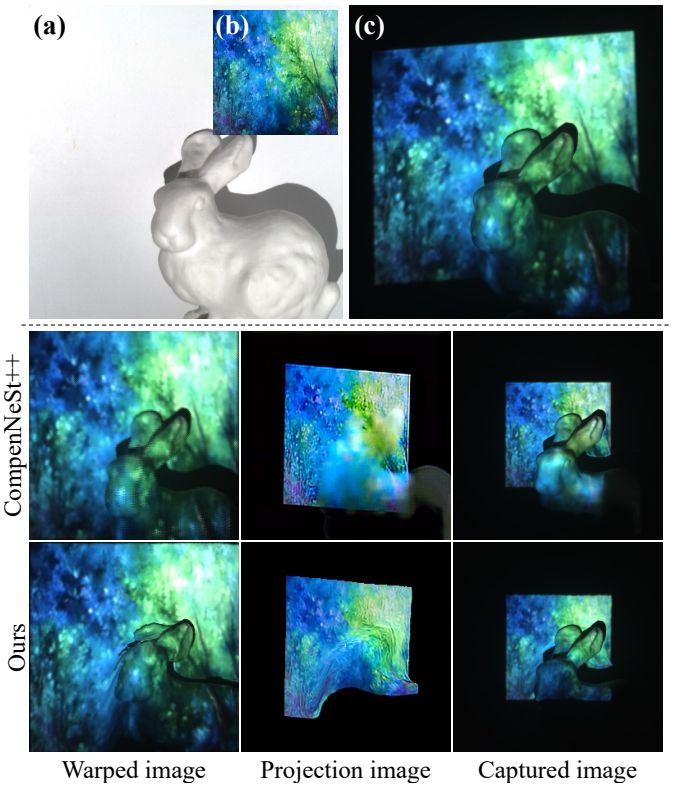


Fig. 11. Compensation results of the proposed method and CompenNeSt++ in the setup with occlusion. (a) Our projection setup. (b) The projection image and (c) the uncompensated projected result. Notably, the compensated result with CompenNeSt++ exhibits artifacts influenced by the bunny, whereas our compensated result is free from such artifacts.

significant errors (see Fig. 10(f), a close-up of (c)). In contrast, the GMA trained on our dataset does not exhibit such errors. This outcome suggests that our dataset realistically reproduces the image quality degradation in PM, thereby enhancing the robustness of GMA.

5.6 Robustness to occlusion

When conducting PM on a screen with hard edges and/or in a setup featuring obstructions in front of the projection target, the projected results may encounter issues related to occlusion. To assess the robustness of the proposed network under such circumstances, we designed experiments using the setup depicted in Fig. 11(a), incorporating the bunny positioned in front of a flat screen. The image projected onto this scene includes a non-negligible occlusion area, as illustrated in Fig. 11(c). For comparison, we also included CompenNeSt++ [33] as a technique. Similar to Sect. 5.2, we generated the training dataset by projecting 125 images onto this screen and trained CompenNeSt++ over five minutes.

In the results presented in Fig. 11, it is evident that the projection image generated by CompenNeSt++ exhibits artifacts, likely stemming from the bunny's presence. This occurs because the warping process of CompenNeSt++ relies mainly on affine and TPS transformations, making it challenging to handle non-contiguous regions like occlusion. Consequently, the warped imaging results contain numerous shadow areas. On the contrary, our method warps the image by computing pixel-wise optical flow, which proves robust in occluded scenes. We verified that our warped images exhibit minimal shadows, resulting in artifact-free generated projection images.

6 Limitation

6.1 Computational cost for image generation

The computation time for generating the projection image is crucial to achieving delay-free DPM. Ng et al. [59] conducted a case study on delay in video interaction, indicating that people cannot perceive the projected image and touch interaction if the image delay is less than 6.04 ms, which is an important benchmark in DPM [60]. However, the generation speed of OnlineProdeb [23] is significantly slower than this threshold, as demonstrated in Table 4. This is attributed to the large number of parameters in the network. In contrast, our lightweight deblurring network has only 1% of the parameters of OnlineProDeb, enabling it to generate each image in a swift 5.941 ms. This suggests that our deblurring network can be seamlessly integrated with a coaxial ProCam system [36]–[38], which obviates the need for online geometric registration, to achieve complete real-time DPM. On the other hand, accomplishing projection, imaging, and data processing within approximately 0.1 ms is not realistic. Therefore, there is a need to further decrease the computational complexity of the deblurring network. One potential approach to reduce computational time is to restrict the processing area to the high-frequency components in the image. This is because, even if defocus blur is present in regions of the projected image with high low-frequency components, the degree of degradation is minimal and not easily perceptible to humans.

Our method also has a limitation in terms of the time required for both warping and deblurring, which amounts to 55.64 ms. Given that the refresh rate of a typical commercial projector is 60 Hz, our method operates at approximately 18 Hz, slightly below real-time. However, it is noteworthy that the target and our refresh rate are within the same order of magnitude. We anticipate that this slight difference can be addressed by replacing GMA with a more efficient optical flow estimation network [61].

6.2 Complex PSF modeling

This study approximates the PSF with a simple isotropic 2D Gaussian model, as in many previous methods [16], [18], [19], [23]. This simplification allows the PSFNet to estimate only one standard deviation of the Gaussian model, facilitating efficient learning convergence. On the other hand, if the PSF can be approximated by a more complex model, such as an anisotropic two-dimensional Gaussian model [44], this could potentially lead to improved accuracy in compensating for defocus blur. Therefore, our next direction is to design a framework that incorporates more complex PSFs. Furthermore, we believe it would be valuable to evaluate the extent to which the simulated PSFs approximate the actual PSFs. Additionally, analyzing the improvement in the accuracy of defocus blur compensation by the proposed network as the simulated PSFs approach the real PSFs is also essential.

6.3 Other image quality degradation

While our method successfully compensated for geometric distortion and defocus blur, it is important to note the presence of various image quality degradations in the PM, such as specular reflection and sub-surface scattering [7], [8]. Moreover, although this study assumed no impact of image quality degradation during image capture, real-world scenarios necessitate consideration of factors

like color conversion between the projector and the camera. The proposed network is not equipped to handle these degradations. Additionally, the warping performance by GMA may deteriorate when these degradations are more pronounced.

On the other hand, we posit that incorporating these image quality degradations into the dataset generation process or substituting GMA with a more robust optical flow estimation network [62] can mitigate the aforementioned issues. Consequently, developing a network capable of compensating for a broader range of image quality degradations, extending beyond geometric distortion and defocus blur in DPM, remains an intriguing avenue for our future research.

6.4 Shadowed regions

In Sect. 5.6, we confirmed that our network is robust against shadowy scenes. However, as evident from the bunny shadows in Fig. 11, our method has a limitation in that it cannot eliminate shadowed regions, which is an inherent challenge when employing a single projector. Generally, the mitigation of shadows is accomplished through the use of multiple projectors, but the geometric and radiometric calibration of multiple projectors poses significant complexity [1], [44], [63]. Our forthcoming research goal is to streamline this intricate calibration process by extending our virtual PM environment to accommodate multi-projection scenarios.

7 Conclusion

This paper addressed the challenging task of compensating for defocus blur in DPM using only one projector and one camera. The key to addressing this task lies in (1) geometric registration of the ProCam coordinates based on the movement of the projection target and (2) fast compensation. To fulfill these requirements, we proposed a neural technique that combines two sub-parts for geometric compensation and deblurring. Additionally, we introduced a realistic data synthesis method with geometric data augmentation in the virtual PM setup. We validated the proposed network through extensive experiments and established three significant findings. Firstly, our method provides compensation comparable to the state-of-the-art method [33], achieving combined compensation, even without fine-tuning in actual PM setups. Secondly, the parameters of our deblurring network are approximately 1% of those in the state-of-the-art technique [23], enabling practical online deblurring. Thirdly, training the geometric compensation network on the proposed dataset enhances its robustness in DPM environments. In our future studies, we aim to enhance our method and extend it to more real-time compensation systems, going beyond geometric distortion and defocus blur. This involves replacing GMA with other state-of-the-art networks [61], [62] and incorporating more complex image quality degradation in the dataset generation. Additionally, we plan to address the challenge of simplifying complex calibration in multi-projection scenarios using the proposed framework.

Acknowledgement

This work was supported by JSPS KAKENHI grant numbers JP20H05958 and JP23KJ1455, Japan.

References

[1] T. Nomoto, W. Li, H.-L. Peng, and Y. Watanabe, "Dynamic multi-projection mapping based on parallel intensity control," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2125–2134, 2022.

[2] L. Miyashita, Y. Watanabe, and M. Ishikawa, "Midas projection: Markerless and modelless dynamic projection mapping for material representation," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–12, 2018.

[3] 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, Oct. 2015.

[4] Y. Kitajima, D. Iwai, and K. Sato, "Simultaneous projection and positioning of laser projector pixels," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 11, pp. 2419–2429, 2017.

[5] D. Tone, D. Iwai, S. Hiura, and K. Sato, "Fibar: Embedding optical fibers in 3d printed objects for active markers in dynamic projection mapping," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 2030–2040, 2020.

[6] S. Kagami and K. Hashimoto, "Sticky projection mapping: 450-fps tracking projection onto a moving planar surface," in *SIGGRAPH Asia 2015 Emerging Technologies*, ser. SA '15. New York, NY, USA: Association for Computing Machinery, 2015. [Online]. Available: https://doi.org/10.1145/2818466.2818485

[7] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer, "The visual computing of projector-camera systems," *ACM SIGGRAPH 2008 classes*, pp. 1–25, 2008.

[8] A. Grundhöfer and D. Iwai, "Recent advances in projection mapping algorithms, hardware and applications," *Computer Graphics Forum*, vol. 37, no. 2, pp. 653–675, 2018.

[9] H. Nishino, E. Hatano, S. Seo, T. Nitta, T. Saito, M. Nakamura, K. Hattori, M. Takatani, H. Fuji, K. Taura, and S. Uemoto, "Real-time navigation for liver surgery using projection mapping with indocyanine green fluorescence: Development of the novel medical imaging projection system," *Annals of Surgery*, vol. 267, no. 6, pp. 1134–1140, 2018.

[10] T. Takezawa, D. Iwai, K. Sato, T. Hara, Y. Takeda, and K. Murase, "Material surface reproduction and perceptual deformation with projection mapping for car interior design," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 251–258.

[11] 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.

[12] D. Iwai and K. Sato, "Document search support by making physical documents transparent in projection-based mixed reality," *Virtual Reality*, vol. 15, no. 2, pp. 147–160, Jun 2011.

[13] K. Matsushita, D. Iwai, and K. Sato, "Interactive bookshelf surface for in situ book searching and storing support," in *Proceedings of the 2nd Augmented Human International Conference*, 2011.

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

[15] B. R. Jones, H. Benko, E. Ofek, and A. D. Wilson, "Illumiroom: peripheral projected illusions for interactive experiences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 869–878.

[16] M. S. Brown, P. Song, and T.-J. Cham, "Image pre-conditioning for out-of-focus projector blur," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1956–1963.

[17] L. Zhang and S. Nayar, "Projection defocus analysis for scene capture and image display," in *ACM SIGGRAPH 2006 Papers*, 2006, pp. 907–915.

[18] Y. Oyamada and H. Saito, "Focal pre-correction of projected image for deblurring screen image," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[19] Y. Kageyama, M. Isogawa, D. Iwai, and K. Sato, "Prodebnet: projector deblurring using a convolutional neural network," *Optics Express*, vol. 28, no. 14, pp. 20 391–20 403, 2020.

[20] D. Iwai, S. Mihara, and K. Sato, "Extended depth-of-field projector by fast focal sweep projection," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 4, pp. 462–470, 2015.

[21] L. Wang, S. Tabata, H. Xu, Y. Hu, Y. Watanabe, and M. Ishikawa, "Dynamic depth-of-field projection mapping method based on a variable focus lens and visual feedback," *Optics Express*, vol. 31, no. 3, pp. 3945–3953, 2023.

[22] H. Xu, L. Wang, S. Tabata, Y. Watanabe, and M. Ishikawa, "Extended depth-of-field projection method using a high-speed projector with a synchronized oscillating variable-focus lens," *Applied Optics*, vol. 60, no. 13, pp. 3917–3924, 2021.

[23] Y. Kageyama, D. Iwai, and K. Sato, "Online projector deblurring using a convolutional neural network," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2223–2233, 2022.

[24] R. Yang and G. Welch, "Automatic and continuous projector display surface calibration using every-day imagery," 2001.

[25] T. Johnson and H. Fuchs, "Real-time projector tracking on complex geometry using ordinary imagery," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[26] C. Resch, P. Keitler, and G. Klinker, "Sticky projections—a new approach to interactive shader lamp tracking," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2014, pp. 151–156.

[27] S. Zollmann and O. Bimber, "Imperceptible calibration for radiometric compensation," in *Eurographics (Short Papers)*, 2007, pp. 61–64.

[28] A. Grundhofer, M. Seeger, F. Hantsch, and O. Bimber, "Dynamic adaptation of projected imperceptible codes," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 2007, pp. 181–190.

[29] S. Jiang, D. Campbell, Y. Lu, H. Li, and R. Hartley, "Learning to estimate hidden motions with global motion aggregation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9772–9781.

[30] A. Kar, A. Prakash, M.-Y. Liu, E. Cameracci, J. Yuan, M. Rusiniak, D. Acuna, A. Torralba, and S. Fidler, "Meta-sim: Learning to generate synthetic datasets," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4551–4560.

[31] E. Wood, T. Baltrušaitis, C. Hewitt, S. Dziadzio, T. J. Cashman, and J. Shotton, "Fake it till you make it: face analysis in the wild using synthetic data alone," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3681–3691.

[32] B. Huang and H. Ling, "Compennet++: End-to-end full projector compensation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7165–7174.

[33] B. Huang, T. Sun, and H. Ling, "End-to-end full projector compensation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[34] B. Huang and H. Ling, "Deprocams: Simultaneous relighting, compensation and shape reconstruction for projector-camera systems," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2725–2735, 2021.

[35] Y. Wang, H. Ling, and B. Huang, "Compenhr: Efficient full compensation for high-resolution projector," in *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, 2023, pp. 135–145.

[36] K. Fujii, M. D. Grossberg, and S. K. Nayar, "A projector-camera system with real-time photometric adaptation for dynamic environments," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 814–821.

[37] T. Sueishi, H. Oku, and M. Ishikawa, "Robust high-speed tracking against illumination changes for dynamic projection mapping," in *2015 IEEE Virtual Reality (VR)*. IEEE, 2015, pp. 97–104.

[38] K. Yamamoto, D. Iwai, I. Tani, and K. Sato, "A monocular projector-camera system using modular architecture," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–9, 2022.

[39] S. Hisaichi, K. Sumino, K. Ueda, H. Kasebe, T. Yamashita, T. Yuasa, U. Lippmann, P. Aswendt, R. Höfling, and Y. Watanabe, "Depth-aware dynamic projection mapping using high-speed rgb and ir projectors," in *SIGGRAPH Asia 2021 Emerging Technologies*, 2021, pp. 1–2.

[40] L. Uwe, A. Petra, H. Roland, S. Kiwamu, U. Kunihiro, O. Yoshihide, K. Hidenori, Y. Tohru, Y. Takeshi, and W. Yoshihiro, "High-speed rgb+ ir projector based on coaxial optical design with two digital mirror devices," in *Proceedings of the International Display Workshops*, 2021, p. 636.

[41] M. Grosse, G. Wetzstein, A. Grundhöfer, and O. Bimber, "Coded aperture projection," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 3, pp. 1–12, 2010.

[42] Y. Li, Q. Fu, and W. Heidrich, "Extended depth-of-field projector using learned diffractive optics," in *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2023, pp. 449–459.

[43] O. Bimber and A. Emmerling, "Multifocal projection: A multiprojector technique for increasing focal depth," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 4, pp. 658–667, 2006.

[44] M. Nagase, D. Iwai, and K. Sato, "Dynamic defocus and occlusion compensation of projected imagery by model-based optimal projector selection in multi-projection environment," *Virtual Reality*, vol. 15, no. 2-3, pp. 119–132, 2011.

[45] A. Bermano, P. Brüschweiler, A. Grundhöfer, D. Iwai, B. Bickel, and M. Gross, "Augmenting physical avatars using projector-based illumination," *ACM Trans. Graph.*, vol. 32, no. 6, Nov. 2013.

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

[47] Y. Li, W. Yin, J. Li, and X. Xie, "Physics-based efficient full projector compensation using only natural images," *IEEE Transactions on Visualization and Computer Graphics*, 2023.

[48] O. Bimber, A. Emmerling, and T. Klemmer, "Embedded entertainment with smart projectors," in *ACM SIGGRAPH 2005 Courses*, 2005, pp. 8–es.

[49] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI'81: 7th international joint conference on Artificial intelligence*, vol. 2, 1981, pp. 674–679.

[50] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[51] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018.

[52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[53] T. Wu, J. Zhang, X. Fu, Y. Wang, J. Ren, L. Pan, W. Wu, L. Yang, J. Wang, C. Qian, D. Lin, and Z. Liu, "Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 803–814.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13(4), no. 4, pp. 600–612, 2004.

[57] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Image quality assessment: Unifying structure and texture similarity," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 5, pp. 2567–2581, 2020.

[58] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12*. Springer, 2012, pp. 611–625.

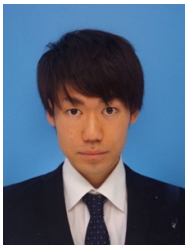
[59] A. Ng, J. Lepinski, D. Wigdor, S. Sanders, and P. Dietz, "Designing for low-latency direct-touch input," in *Proceedings of the 25th annual ACM symposium on User interface software and technology*, 2012, pp. 453–464.

[60] 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, Dec. 2018.

[61] S. Bai, Z. Geng, Y. Savani, and J. Z. Kolter, "Deep equilibrium optical flow estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[62] Z. Huang, X. Pan, W. Pan, W. Bian, Y. Xu, K. C. Cheung, G. Zhang, and H. Li, "Neuralmarker: A framework for learning general marker correspondence," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 6, pp. 1–10, 2022.

[63] C. Jaynes, S. Webb, and R. M. Steele, "Camera-based detection and removal of shadows from interactive multiprojector displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 3, pp. 290–301, 2004.



Yuta Kageyama (Student Member, IEEE) received the B.S. and M.S. degrees from Osaka University, Japan, in 2019 and 2021, respectively. He is currently pursuing a Ph.D. degree at Osaka University and is a JSPS research fellow (DC2). His research interests include spatial augmented reality, computer vision, and deep learning.

Daisuke Iwai (Member, IEEE) received B.S., M.S., and Ph.D. degrees from Osaka University, Japan, in 2003, 2005, and 2007, respectively. He was a Visiting Scientist at Bauhaus-University Weimar, Germany, from 2007 to 2008, and a visiting Associate Professor at ETH, Switzerland, in 2011. He is currently an Associate Professor with the Graduate School of Engineering Science, Osaka University. His research interests include spatial augmented reality and projector-camera systems.

Kosuke Sato (Member, IEEE) received B.S., M.S., and Ph.D. degrees from Osaka University, Japan, in 1983, 1985, and 1988, respectively. He was a Visiting Scientist at the Robotics Institute, Carnegie Mellon University, from 1988 to 1990. He is currently a Professor at the Graduate School of Engineering Science, Osaka University. His research interests include image sensing, virtual reality, and human interfaces.