Setup-Independent Full Projector Compensation

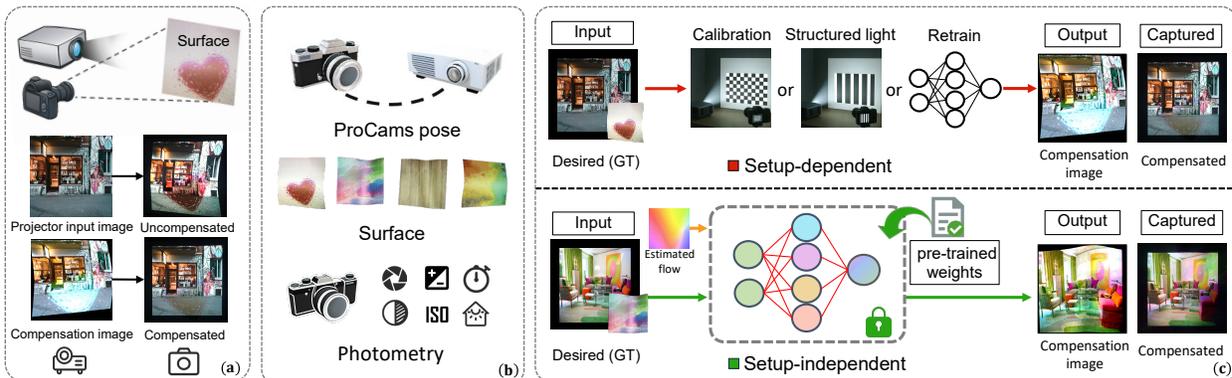Haibo Li (iD), Qingyue Deng (iD), Jijiang Li (iD), Haibin Ling (iD), and Bingyao Huang (iD)

Fig. 1: Setup-independent full projector compensation. (a) General projector compensation pipeline: If a projector directly projects an image onto a nonplanar and textured surface, the user may see a distorted (uncompensated) result, and the goal is to generate a compensation image such that when projected, it results in a visually pleasing (compensated) target appearance. (b) Traditional setup-dependent projector compensation methods cannot generalize to untrained new setups with various projector-camera poses, surface geometries, and photometric properties, such as exposure parameters (e.g., ISO, shutter speed) and ambient lighting conditions. (c) The top path depicts setup-dependent methods, which require per-setup calibration, reconstruction, or retraining for each new setup. In contrast, our setup-independent method (bottom path) is trained once on a diverse dataset; it only requires the estimated flow (detailed in Fig. 2) and pre-trained weights to generalize to unseen setups without further retraining or fine-tuning.

Abstract— Projector compensation seeks to correct geometric and photometric distortions that occur when images are projected onto nonplanar or textured surfaces. However, most existing methods are highly setup-dependent, requiring fine-tuning or retraining whenever the surface, lighting, or projector–camera pose changes. Progress has been limited by two key challenges: (1) the absence of large, diverse training datasets and (2) the entanglement of geometric correction within trained networks. We introduce **SIComp**, the first **S**etup-**I**ndependent framework for full projector **Comp**ensation, capable of generalizing to unseen setups without fine-tuning or retraining. To enable this, we construct a large-scale real-world dataset spanning 277 distinct projector–camera setups. SIComp adopts a co-adaptive design that decouples geometry and photometry: a carefully tailored optical flow module performs online geometric correction, while a novel photometric network handles photometric compensation. To further enhance robustness under varying illumination, we integrate intensity-varying surface priors into the network design. Extensive experiments demonstrate that SIComp consistently produces high-quality compensation across diverse unseen setups, substantially outperforming existing methods in terms of generalization ability and establishing the first generalizable solution to projector compensation.

Index Terms—Projector compensation, projection mapping, spatial augmented reality, model generalization, optical flow

✦

1 INTRODUCTION

Projectors are essential components in modern AR/VR systems, supporting a variety of applications ranging from interactive displays to projection mapping [1, 5, 7, 10, 12, 33, 34, 41, 49, 51, 53, 58, 71]. However, when projecting onto surfaces that are nonplanar and textured, the geometric and photometric distortions caused by surface geometry and reflectance significantly compromise visual quality. Projector compensation techniques aim to modify the projector input image to compensate for these distortions [3, 8, 16, 20, 30, 35, 38, 56, 63, 72]. As illustrated in Fig. 1(a), the camera-captured uncompensated image reveals significant geometric and photometric distortions, which are corrected in the compensated result.

- *Haibo Li, Qingyue Deng, Jijiang Li and Bingyao Huang are with Southwest University. E-mail: haiboli@email.swu.edu.cn, qingyue@email.swu.edu.cn, jijiangli@email.swu.edu.cn, bhuang@swu.edu.cn.*
- *Haibin Ling is with Westlake University. E-mail: linghaibin@westlake.edu.cn*
- *Bingyao Huang is the corresponding author.*

A typical projector compensation system consists of a projector-camera pair that operates on a textured projection surface, as shown in Fig. 1(a). Throughout this paper, we use the term *setup* to refer to a specific scene configuration, defined by the projection surface, the photometric conditions, and the poses of the projector–camera pair, as illustrated in Fig. 1(b). Currently, numerous compensation methods [15, 17, 20, 22, 42, 47, 66, 73] achieve effective results. However, their reliance on a fixed setup creates a significant bottleneck. The core reason for setup dependency is that most previous methods rely on structured light to perceive scene geometry [15, 20, 47, 73] or require explicit calibration of the projector and camera [21, 37]. Although some methods [19, 22, 66] do not require calibration or structured light for geometric correction, their dependency on retraining or fine-tuning prevents the model from generalizing to new setups. The setup-dependent nature, which requires costly recalibration or retraining for any change in projector-camera pose, surface, or photometric condition, limits their scalability and practicality in dynamic environments.

Inspired by the strategy in CompenNeSt++ [22], which decouples the compensation problem into its geometric and photometric components, our work focuses on resolving their respective setup dependencies.

For geometric dependency, a seemingly straightforward approach is to leverage optical flow [44] to directly estimate the projector-camera pixel mapping online. However, this is not feasible because optical flow

assumes brightness constancy and small motion, which is not the case for projector compensation, as a significant discrepancy exists between the projector input image and the uncompensated image captured by the camera, as shown in Fig. 1(a). To address this issue, we employ a pre-trained learning-based optical flow model, FlowFormer [24], and carefully fine-tune it together with the pre-trained photometric compensation module on our projector compensation dataset. Consequently, our method eliminates the need for per-setup structured light or explicit calibration.

For photometric dependency, we find that typical CNN-based full projector compensation [19, 22, 66] approaches use a single camera-captured image of the projection surface to implicitly encode the properties of the surface and environmental lighting. However, we argue that relying on a single surface image representation is often insufficient, particularly for complex or unseen setups with varying lighting conditions, as it necessitates retraining or fine-tuning. Our subsequent analysis quantitatively reveals the limitations of this approach. Therefore, we use additional surface images under varying projector illumination levels. This enables the photometric module to more accurately model the dynamic range of the projector-camera system (ProCams) and the surface reflectance. Furthermore, we introduce an adaptive photometric module (IVPCNet) that integrates the Convolutional Block Attention Module (CBAM) [70] and the Swin Transformer [43]. This design enables the architecture to effectively capture latent features of varying surface and photometric properties.

Another obstacle to achieving setup independence is the limited scale and diversity of the existing training dataset. A robustly generalizable model must be trained on data that represents a wide spectrum of real-world variations. To this end, we construct a large-scale dataset that specifically varies these critical factors, including ambient lighting, surface materials and reflectivities, and the geometric configuration of the ProCams.

Finally, we integrate the modules above and jointly optimize them using a co-adaptive training strategy on our large-scale and diverse training set. By working together, the modules constantly refine one another during training. Therefore, our SIComp can generalize to unseen projector-camera poses, photometric conditions, and projection surfaces without retraining or fine-tuning. Our experimental results demonstrate that our setup-independent method, SIComp, shows highly competitive performance and even outperforms some setup-dependent approaches. Our key contributions are summarized below:

- We propose SIComp, the first setup-independent full projector compensation framework capable of generalizing to unseen setups without retraining or fine-tuning.

- Our SIComp consists of a tailored optical flow module for online geometric correction under unseen setups and a photometric compensation module that leverages intensity-varying surfaces priors for more robust feature learning.

- We construct a large-scale and diverse dataset for full projector compensation comprising 277 unique setups. Experimental results demonstrate that our method significantly outperforms some setup-dependent baselines.

For reproducibility and to facilitate future research in this domain, we will release all training and testing data (277 projector-camera setups), pre-trained models, and source code upon paper acceptance.

2 RELATED WORK

2.1 Setup-dependent projector compensation

Projector compensation techniques are designed to compensate for the geometric and photometric distortions that arise from projecting onto non-ideal surfaces. Most research focuses on *setup-dependent* methods, which are intrinsically tied to a specific setup. Consequently, any change to the setup requires a recalibration or retraining. These methods are broadly classified into three categories: geometric correction, photometric compensation, and full compensation.

Table 1: Comparison of full projector compensation methods. SI indicates Setup-Independent methods (✔).

Methods	SI	Pre-calibration?	Setup time
Matrix-based [47, 69]	�’	Yes	Medium
CNN-based [22, 66]	✗	No	Medium
Optical Flow-based [40]	✗	No	Medium
Differentiable rendering-based [39, 48]	✗	Yes	Long
Radiance field-based [9, 11]	✗	Yes	Medium
SIComp (ours)	✔	No	Rapid

2.1.1 Geometric correction

Geometric correction aims to eliminate spatial distortions caused by nonplanar surfaces, ensuring that the projected image appears undistorted from the viewer's perspective. Traditional strategies rely on structured light (SL) to establish camera-projector correspondences. Tardif *et al.* [59] propose an SL-based mapping system for multi-projector setups that bypasses traditional calibration by directly recovering camera-projector mappings. Several studies have enhanced SL-based correction by improving projection accuracy, usability, and robustness [45, 52, 65]. These methods achieve geometric correction by projecting and decoding a sequence of coded patterns. This reliance on pattern projection is often time-consuming and sensitive to environmental interference.

Marker-based approaches reduce calibration overhead by leveraging visual features. For instance, Audet *et al.* [2] use printed markers, while Narita *et al.* [46] employ infrared (IR) markers for dynamic surfaces. To further enhance interaction, Kagami *et al.* [31, 32] develop "Animated/Interactive Stickies" for low-latency mapping onto markerless planes. However, these are primarily optimized for planar targets and specific hardware.

For complex deformations, Ibrahim *et al.* develop dynamic registration using Bézier patches [28, 29] and self-calibrating extensions for jitter and occlusion [25–27]. These still rely on continuous ArUco marker visibility, which limits robustness on textured or rapidly stretching surfaces. To achieve markerless registration, Tehrani *et al.* [61] develop an automated system for arbitrary 3D shapes, but it requires a discrete re-estimation phase when projectors move.

Another geometric correction approach uses natural feature matching to bypass projected patterns. While the method by Takahashi *et al.* [57] enables correction from a single captured image, its reliance on 2D feature alignment limits its application to planar surfaces, as it cannot model 3D surface deformation.

All these approaches remain setup-dependent: they require recalibration when the projector-camera pose or projection surface changes.

2.1.2 Photometric compensation

Photometric compensation aims to mitigate photometric distortions caused by textured projection surfaces and ambient lighting, typically assuming prior geometric correction. Early works [4, 13, 47, 64, 73] often rely on linear models to simulate light interactions, where the contribution of each projector channel to each camera channel is approximated by a per-pixel (3×3) matrix. Fujii *et al.* [13] demonstrate rapid frame-level adaptation to dynamic environmental changes, but their approach is limited in simultaneously handling complex reflectance and illumination variations, and its adaptability to new setups remains constrained. Yoshida *et al.* [73] incorporate ambient light effects into photometric compensation; however, linear models struggle to represent the complex nonlinear photometric transformations encountered in real-world scenes.

Addressing shortcomings of linear approaches above, nonlinear methods [14, 15, 42] based on lookup tables (LUTs) enable more accurate modeling of complex optical interactions. However, these methods require extensive calibration data and computational resources. For multi-projector setups, Pjanic *et al.* [50] leverage the RLab color model to improve perceptual seamlessness, though it requires time-consuming LUT generation. Similarly, Tehrani *et al.* [62] introduced 3D gamut morphing for color uniformity, but it lacks the capability to compensate

for surface textures and suffers from high computational complexity in multi-projector setups.

More recently, deep learning approaches [20] show progress in learning complex color mappings without explicit calibration. Despite improved accuracy and reduced manual effort, these methods still require retraining for every new setup, making adaptation a time-consuming process.

2.1.3 Full compensation

Full compensation methods aim to correct geometric and photometric distortions simultaneously. Previous methodologies are characterized by specific technical assumptions and operational requirements. For instance, the light transport matrix method by Wetzstein *et al.* [69] is designed for offline processing and involves significant computational time. For real-time applications, the system developed by Siegl *et al.* [56] operates based on the assumption that projection surfaces are Lambertian surfaces. Similarly, the approach by Shahpaski *et al.* [54] requires a pre-calibrated camera and a specific planar target, while its photometric compensation capabilities are constrained.

More recently, advanced data-driven and learning-based frameworks address geometric and photometric compensation problems jointly. These frameworks leverage techniques such as differentiable rendering, neural rendering, and deep neural networks.

Deep convolutional neural networks-based methods [22, 66] learn corrections directly from data, avoiding the need for explicit calibration. However, their applicability is constrained to static setups, and even with rapid fine-tuning, the process takes several minutes, which impedes real-time deployment. Recently, some methods [40,67] leverage optical flow to directly estimate the dense pixel mapping between the projector and the camera images, without baking the geometric correction in the network parameters. However, they still require a multi-minute adaptation phase for new setups.

Differentiable rendering frameworks are proposed to decouple light, material, and shape from the compensation process. DeProCams [21] introduces a neural rendering framework that simultaneously performs ProCams relighting, compensation, and shape reconstruction. Park *et al.* [48] leverages differentiable rendering for iterative refinement-based compensation. DPCS [39] uses physically-based path tracing to model the light transport between the projector and the camera, and leverages differentiable rendering to compensate for projector images. While powerful for modeling complex light transport and scene details, these methods require high computational costs and typically rely on accurate initial calibration.

Radiance field-based frameworks are also applied to projector compensation. Erel *et al.* [11] propose the first neural reflectance fields-based ProCams model, named NepMap. Although powerful, it requires massive GPU memory and training time. Deng *et al.* [9] introduce the first Gaussian splatting-based ProCams model. By representing the scene with 2D Gaussians [23], GS-ProCams enables efficient, view-agnostic projection mapping. However, both methods still require a per-scene optimization from multi-view captures.

In summary, prior works (Tab. 1) suffer from a rigid dependency on system configuration. Because any change to the hardware requires an adaptation, recalibration or retraining cycle, we focus on developing a projector compensation method that is truly setup-independent.

2.2 Techniques used in our work

To the best of our knowledge, our work, SIComp, is the first setup-independent full projector compensation method. We achieve this by designing a novel framework that integrates a fine-tuned optical flow module for online geometry correction, and a pre-trained photometric compensation network with intensity-varying surface priors and attention mechanisms.

Optical flow. Recent advances in optical flow methods, such as RAFT [60] and FlowFormer [24], enable highly accurate and robust pixel-level correspondence estimation without predefined patterns. This capability presents a promising alternative to structured light for establishing the geometric mapping between the projector and camera, forming a solid foundation for calibration-free geometric correction.

Attention mechanisms. Modules like the Swin block [43] and CBAM [70] have become a cornerstone of deep learning due to their ability to dynamically weigh the importance of different spatial and channel-wise features. This adaptivity empowers a single projector compensation model to better learn diverse surface textures and photometric conditions without being explicitly trained for each setup.

3 METHOD

3.1 Problem formulation

The primary goal of projector compensation is to project light in such a way that the resulting image, as perceived by an observer in the real world, appears geometrically correct and photometrically faithful, regardless of the surface's shape or texture.

To formalize this challenge, we model the physical process within a projector-camera system, where a camera acts as an effective proxy for a human observer. Consider a setup where the intended projector input image, denoted by x, is sent to the projector and is directly projected onto the surface without any compensation. The camera typically captures a distorted version of this input, which we label as the uncompensated image $\tilde{x}$ (see Fig. 1 (a)). This distortion stems from two primary sources in the image formation process: a **photometric transformation** R, which accounts for surface reflectance and ambient lighting E interactions, and a **geometric mapping** G, which models the spatial warping caused by the surface's shape and the projector-camera alignment. Let S denote the combined properties of the surface (geometry and reflectance). The entire process that maps the ideal input x to the captured distorted image $\tilde{x}$ can be expressed as:

$$\tilde{x} = G(R(x;E,S)) \tag{1}$$

Consequently, the objective of projector compensation is to find an adjusted compensation image, $\hat{x}^*$, for projection. This image should be designed to counteract the distortions that occur when it is projected onto a surface and subsequently captured by a camera. We assume there exists a perfect compensation image, x^*, when subjected to the same degradation projection-capture pipeline, would completely neutralize all distortions, leading to the following theoretical relationship, reproducing the original projector input image x. This ideal compensation process is defined by the inverse relationship:

$$x = G(R(x^*;E,S)) \tag{2}$$

However, directly measuring the complex interplay of surface properties S and ambient illumination E is often impractical. Following [22], we instead utilize a camera-captured image $\tilde{s}$ (surface image in Fig. 2(a)) to implicitly encapsulate the combined effects of E and S. This surface image $\tilde{s}$ is obtained by projecting a uniform gray image (x_0 in Fig. 2(a)) onto the surface under both ambient and projector illumination:

$$\tilde{s} = G(R(x_0;E,S)), \tag{3}$$

where x_0 is a plain gray image providing a reference illumination.

To derive the compensation image x^*, we aim to approximate the inverse of the composite function in Eq. (2). However, due to the inherent nonlinear characteristics of both the projector and the camera's color response functions, obtaining an exact inverse of R is generally infeasible. Therefore, we introduce a pseudo-inverse operation $R^\dagger$, which represents the approximation of the inverse photometric transformation. Then the ideal compensation image x^* is expressed as:

$$x^* = R^\dagger\left(G^{-1}(x),\ G^{-1}(\tilde{s})\right), \tag{4}$$

where $R^\dagger$ denotes the pseudo-inverse of R, and G^{-1} represents the inverse geometric mapping that aligns images from the camera view back to the projector's coordinate space.
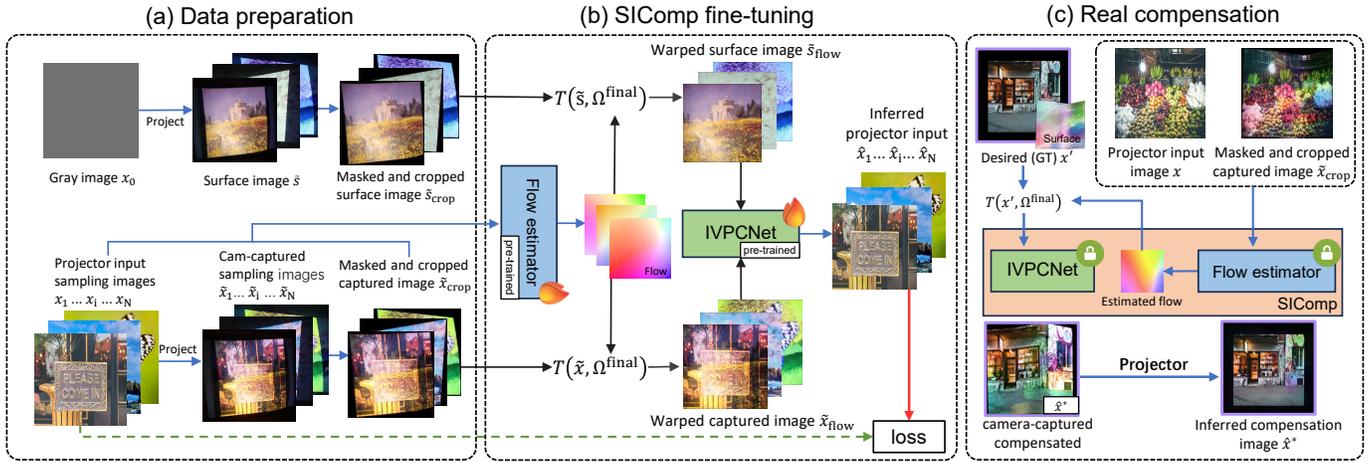
Fig. 2: SIComp pipeline. (a) Data Preparation phase, including the acquisition of surface images and various captured projection images collected under diverse setups, followed by masking and cropping. (b) SIComp fine-tuning pipeline. It comprises a flow estimator module for geometric correction, which is pre-trained on the Sintel [6] dataset. A pre-trained IVPCNet performs photometric compensation, inferring the projector input image based on optical flow warped images, whose pre-training process is detailed in Fig. 3. (c) Real compensation application, where the trained SIComp takes a desired image x' and the optical flow (pre-estimated from an initial reference projection) as input to infer the compensation image, which is then physically projected and captured by a camera to approximate the desired image x'. The symbol T in (b) denotes the warping operation applied using the estimated optical flow.

3.2 Definition of unseen setup

In this work, a *setup* denotes a specific configuration characterized by three core factors: (1) the projection surface, encompassing its geometry and reflectance properties; (2) the spatial relationship (relative pose) between the projector and the camera; and (3) the photometric conditions during capture, which include ambient illumination and camera imaging parameters (e.g., exposure, white balance). A key challenge in projector compensation is that these factors are coupled in the captured image. Our method is *setup-independent* as it is trained to disentangle these factors and can generalize to *unseen setups*, i.e., novel, previously unseen combinations of the above elements, without requiring any per-setup retraining or fine-tuning.

3.3 SIComp: setup-independent compensation

Distinct from prior methods, our SIComp is specifically designed to jointly correct for both geometric and photometric distortions in a setup-independent manner, allowing it to adapt to new setups without retraining. We implement Eq. (4) using two modules, each dedicated to a specific part of the decomposition.

For geometric correction, Inspired by previous works [40, 67], we leverage an online optical flow estimator. Its ability to predict dense pixel correspondences without relying on explicit projector-camera poses is ideal for our setup-independent goal. This module's primary role is to compute the inverse geometric mapping G^{-1} by predicting a flow field (or displacement field) Ω. This design directly addresses a key limitation of previous approaches like WarpingNet [22] and GANet [66], which cannot generalize to new setups.

For photometric compensation, we take the geometrically-corrected images from the previous stage as input, this module's task is to learn and approximate the pseudo-inverse operation $R^\dagger$, compensating for the complex photometric distortions caused by textured surfaces and photometric conditions.

By integrating these two components, SIComp directly models geometric and photometric inverse operations. This process can be described as:

$$x^* = R^\dagger(T(x, \Omega), T(\tilde{s}, \Omega)), \qquad (5)$$

where T represents the operation of warping an image using the flow field Ω.

3.3.1 Surrogate training for compensation

A challenge in learning-based compensation methods is the lack of ground truth for the ideal compensation image x^*. To address this, CompenNet [20] proposed a surrogate training strategy: instead of directly learning the true mapping $(x \rightarrow x^*)$, the network learns to recover the original input image x from the uncompensated image $\tilde{x}$, i.e., $(\tilde{x} \rightarrow x)$. This approach allows for easily collected training pairs $(\tilde{x}, x)$, enabling effective supervision and ensuring that the network can perform the forward compensation task during inference.

Following the surrogate training principle above, we train SIComp to perform the inverse task of recovering the original projector input image. We define the overall SIComp network as $O_{g,r}$, parameterized by its geometric (g) and photometric (r) components. The network $O_{g,r}$ takes an uncompensated image $\tilde{x}$ and surface information $\tilde{s}$ as input to predict the original projector image, denoted as $\hat{x}$:

$$\hat{x} = O_{g,r}(\tilde{x}, \tilde{s}) \qquad (6)$$

Accordingly, the trainable parameters g and r are optimized by minimizing a loss function $\mathcal{L}$ that measures the discrepancy between the network's prediction $\hat{x}$ and the ground truth input x. The overall training objective for a mini-batch of size B is formally expressed as:

$$\min_{g,r} \frac{1}{B} \sum_{i=1}^{B} \mathcal{L}\left(O_{g,r}\left(\tilde{x}^{(i)}, \tilde{s}^{(i)}\right), x^{(i)}\right), \qquad (7)$$

where $\mathcal{L}$ represents a composite reconstruction loss, which will be detailed in Sec. 4.1.3.

3.3.2 Detailed training strategy

While the surrogate training objective is well-defined, optimizing it directly on raw camera-captured, uncompensated images is challenging. These inputs suffer from significant background clutter and photometric inconsistencies, which can destabilize training, particularly for the optical flow subnet. To address this, we introduce a three-part strategy involving input pre-processing, data augmentation, and two-stage optimization.

Input masking and cropping. The first step in our strategy is to eliminate irrelevant background regions from the captured images, which act as a major source of noise for geometric alignment. We employ a two-step process to isolate the valid projector Field of View (FOV). First, a binary mask is applied to the uncompensated image $\tilde{x}$
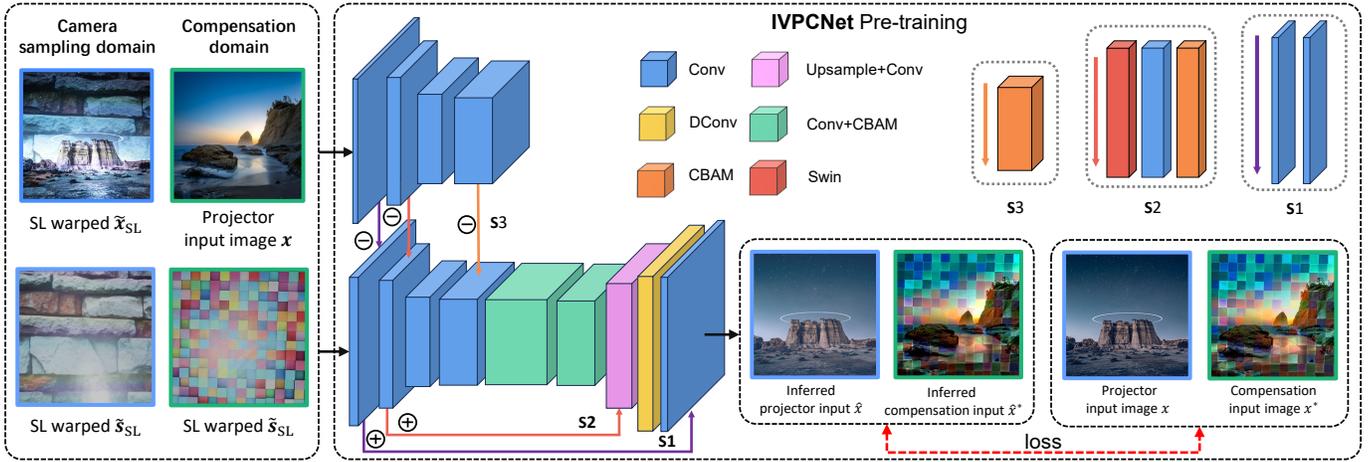
Fig. 3: Pre-training pipeline for IVPCNet. IVPCNet, a siamese U-Net architecture, is pre-trained using two different domains. The first is the **camera sampling domain** (solid blue border). It takes a camera-captured image warped by structured light and the corresponding surface image as input, predicting an output $\hat{x}$. The second is the **compensation domain** (solid green border), where the network learns to predict a compensation image. It uses the projector input image x and the same warped surface image to predict $\hat{x}^*$, with the loss computed against a target compensation image x^*. This target image x^* is derived for the specific hardware setup using Nayar's iterative refinement method [47]. This pre-training process yields the IVPCNet model for subsequent fine-tuning within the SIComp framework (see Fig. 2). The purple, red, and orange lines denote the skip connections at stages s1, s2, and s3, respectively.

Fig. 4: Projector and camera field of view (FOV). In the projector FOV mask, the white region represents the projector's FOV. The camera FOV encompasses the broader scene, represented by the blue region, and the optimal viewing area is highlighted by the red boundary. Due to the projector's FOV and potential geometric distortions of the surface, an affine transformation is necessary to apply to the projected input image, resulting in the ideal visualization after accounting for the projector's FOV limitations and the surface geometric distortions, i.e., desired (GT).

and the surface image $\tilde{s}$ to remove all areas outside the projector's FOV as illustrated in Fig. 4. Then, we tightly crop the masked images to the bounding box of the projector's FOV. This yields clean, content-focused inputs, $\tilde{x}_{\text{crop}}$ and $\tilde{s}_{\text{crop}}$, which significantly improve the accuracy of flow prediction and accelerate network convergence. The result of this process is visualized in Fig. 2(a).

Optimal visualization area   During the training phase, the network is supervised using captured uncompensated images $\tilde{x}$, which inherently account for the surface's geometric distortion and are not limited by the projector's field of view. In contrast, at real compensation inference time, directly inputting the projector input image x to obtain the compensation image $\hat{x}^*$ will result in a suboptimal compensation result. To prevent clipping or content loss during projection, we apply an affine transformation to the projector input x, producing a transformed version x' (Desired (GT)) that fits within the largest inscribed rectangle of the projector's field of view (Fig. 4). This ensures that the predicted compensation image $\hat{x}^*$ is fully visible when projected. In other words, during inference, we use x' as the input to generate the compensation image $\hat{x}^*$.

Intensity-varying surface prior.   To improve the generalization ability of SIComp to varying surface illumination levels, we introduce an intensity-varying surface prior. Unlike previous methods [22,66] that provide the network with a single surface image ($\tilde{s}_{\text{crop}}$), we augment the input with a sequence of surface captures that profile the surface's response to different projector illumination intensities.

We project uniform grayscale images at a set of five discrete intensity levels: $\{0, 64, 128, 191, 255\}$, spanning the projector's dynamic range (Fig. 5). This yields five fixed surface images per setup. We experiment with three specific input configurations defined by a hyperparameter K, each using a fixed set of intensities throughout training and evaluation: for $K = 1$ only intensity 64; for $K = 3$ intensities $\{0, 128, 255\}$; and for $K = 5$ all five intensities. These intensity-varying surface priors allow the network to learn a more robust compensation model.

Pre-training and fine-tuning.   Training SIComp from scratch is challenging due to the complexity of jointly learning geometric and photometric transformations. We address this with a two-stage optimization strategy. In the first stage, the geometric module is pre-trained on the Sintel [6] dataset, while the photometric module is pre-trained on geometrically-corrected image pairs from real-world setups to model photometric inconsistencies. These pre-trained subnets provide a robust foundation for the second stage, where both modules are fine-tuned jointly on real-world geometrically-uncorrected data. This fine-tuning process allows both modules to adapt together to the final full projector compensation task. During this phase, SIComp takes the pre-processed inputs defined in the Sec. 3.3.2 to predict the original projector image. The final input is given by:

$$\hat{x} = O_{g,r}\left(\tilde{x}_{\text{crop}}, \mathcal{P}_K\right), \quad K \in \{1,3,5\}, \tag{8}$$

where $K \in \{1,3,5\}$ specifies the number of intensity-varying surface priors included in the input set $\mathcal{P}_K$.

During real compensation, the desired GT x' and the pre-computed optical flow are input to the trained SIComp to obtain the predicted compensation image $\hat{x}^*$. Then $\hat{x}^*$ is projected onto the surface. The camera-captured result of this projection effectively compensated both geometric and photometric distortions, resulting in a visually corrected image that closely matches the desired GT x'. The full process is illustrated in Fig. 2(c).

For each new setup, a one-time preparation is performed: a reference image is captured to pre-compute the optical flow, followed by the acquisition of the five intensity-varying surface priors. This entire process takes approximately 2-3 seconds. During inference, the model utilizes the pre-computed flow field and the subset $\mathcal{P}_K$ (where $K \in \{1,3,5\}$) corresponding to its trained configuration. Since these components are stored and reused for all subsequent compensations on the same
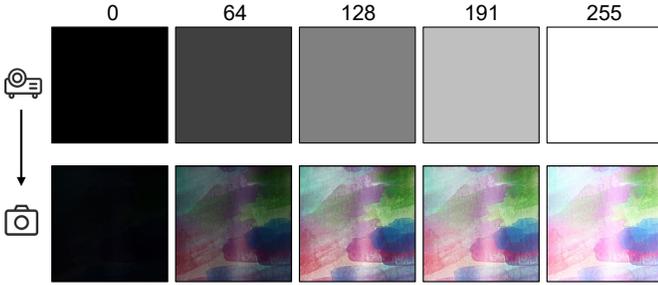
Fig. 5: Intensity-varying surface priors. (Top) Uniform gray images of varying intensities (0, 64, 128, 191, and 255) are projected onto a surface. (Bottom) The corresponding camera captures reveal the surface's distinct reflectance properties under different illumination levels. This set of captured images provides intensity-varying surface priors, enabling SIComp to more effectively learn complex surface properties.

setup, this one-time data capture overhead is excluded from the reported per-image inference times.

3.4 Network architecture

The SIComp framework consists of two main modules: an optical flow estimator module based on FlowFormer [24], and a photometric compensation module based on a modified siamese U-Net, which is named Intensity-Varying Photometric Compensation Network (IVPCNet).

3.4.1 Geometric correction via FlowFormer

FlowFormer [24] estimates an optical flow field Ω between the cropped camera image $\tilde{x}_{\text{crop}}$ and the projector input image x. This flow field represents the pixel-wise displacements from the camera viewpoint to that of the projector.

The geometric correction process within FlowFormer is performed in two stages: (1) A Transformer-based cost volume encoder embeds a 4D cost volume into a latent space, forming a global cost memory. (2) A decoder retrieves information from this memory using cost queries and iteratively predicts residual flows $\Delta\Omega(t)$ via recursive attention and a ConvGRU module. Starting from an initial estimate $\Omega(0) = 0$, the flow estimate is updated recursively as $\Omega(t+1) \leftarrow \Omega(t) + \Delta\Omega(t)$. After N iterations, the final accumulated flow field is:

$$\Omega^{\text{final}} = \sum_{t=1}^{N} \Delta\Omega(t) \quad (9)$$

We then use this final flow field, Ω^{final}, to warp the cropped camera image $\tilde{x}_{\text{crop}}$, producing the geometrically corrected image $\tilde{x}_{\text{flow}}$. The same flow field is then applied to the cropped surface images $\tilde{s}_{\text{crop}}$ to yield $\tilde{s}_{\text{flow}}$, as illustrated in Fig. 2(b).

3.4.2 Photometric compensation via IVPCNet

The IVPCNet architecture is shown in Fig. 3, and it corrects photometric distortions in the flow-warped image $\tilde{x}_{\text{flow}}$ to produce the compensated output $\hat{x}$. To do so, it takes the $\tilde{x}_{\text{flow}}$ and the corresponding aligned surfaces $\tilde{s}_{\text{flow}}$ as input. The design incorporates three key components to achieve generalized photometric compensation: (1) **Intensity-varying surface priors input design**, which enables the network to learn different surface photometric variations and improve generalization across diverse surfaces. (2) **Swin [43] Transformer Attention**, applied in skip connections with shifted windows. This mechanism captures long-range dependencies through Query-Key-Value (QKV) similarity, enhancing the network's ability to model complex photometric conditions. (3) **Dynamic Feature Gating** via CBAM [70], applied after each encoder and decoder block. This module adaptively combines channel and spatial attention to emphasize critical regions for photometric correction.

4 EXPERIMENT

To evaluate the generalization ability of SIComp, we conduct two types of experiments under unseen setups: real compensation and surrogate compensation. Real compensation involves physically projecting the inferred compensation image onto a surface and comparing the camera-captured result with the desired (GT). This protocol facilitates a more realistic and comprehensive assessment of the model's effectiveness under practical, unconstrained conditions.

We benchmark SIComp against several projector compensation methods, broadly categorized into two main groups based on their adaptability to unseen setups: setup-dependent and setup-independent.

Setup-dependent methods: This group includes DPCS [39], DePro-Cams [21], CompenNeSt++ [22], and CompenHR [66]. These models require retraining or extensive per-setup optimization for each new surface, photometric condition, or projector-camera pose.

Setup-Independent (SI) methods: This group includes FF+PANet, FF+CompenNeSt, and our proposed SIComp variants, which are trained only once on a diverse dataset and can be directly applied to unseen setups without any retraining or fine-tuning. To create strong setup-independent baselines, we combine a shared FlowFormer (FF) with the photometric networks from CompenNeSt++ and CompenHR, forming FF+CompenNeSt and FF+PANet, respectively. To ensure a fair comparison, we evaluate these baselines under two settings: (1) their original, published configurations (denoted as (origin)), and (2) variants retrained with our unified training setup (see Sec. 4.1.2) to isolate the impact of the training strategy.

Our ablation study includes three SIComp variants, differing only in the number of input surface priors (1, 3, and 5), to evaluate the impact of the number of surface priors on model generalization.

4.1 Datasets and implementation details

4.1.1 Datasets

To pre-train IVPCNet and fine-tune SIComp, we constructed a large-scale real-world dataset consisting of 277 distinct projector-camera setups. Among them, 223 setups were collected by us through a structured data acquisition process that involved real projections and camera captures. The remaining 54 setups were obtained from existing public datasets: nine from CompenNeSt++ [22], 21 from CompenHR [66], and 24 from CompenNet [20].

Each setup contains 500 training images and 200 validation images, totaling approximately 138,500 training samples and 55,400 validation samples. All training and validation image resolutions are 256×256 and these setups were specifically designed to include a wide range of variations in surface texture (e.g., smooth, rough, patterned), reflectance properties (e.g., matte, glossy), projection angles and distances, and environmental lighting conditions (e.g., varying ambient light, exposure time, and ISO settings).

Each data sample includes three components: (1) the camera-captured images; (2) multiple surface images captured under pure-gray projections at varying brightness levels, serving as surface priors; and (3) the projector input images. These components together allow the network to learn both geometric and photometric compensation across diverse real-world setups. The comprehensive dataset provides sufficient diversity and scale to support robust training of setup-independent projector compensation models.

To evaluate the generalization ability of SIComp, we created two distinct evaluation sets with a resolution of 600×600 pixels, significantly higher than the training resolution (256×256), thereby assessing the models' ability to generalize across different image scales. Specifically, two datasets were established for real compensation evaluation:

(1) Set A: Same ProCams devices and unseen setups: It comprises 8 unseen setups (4 surfaces $\times$ 2 poses), captured using a Canon 600D camera and an EPSON CB-965 projector, with 5 test images per setup.

(2) Set B: Novel ProCams devices and unseen setups: This set was specifically designed to evaluate generalization to novel hardware and configurations (Tab. 2). It consists of 5 unseen setups across 3 diverse physical surfaces and 5 unique projector-camera poses, captured using a Nikon D3200 camera and a TOSHIBA TDP-T100C projector, with
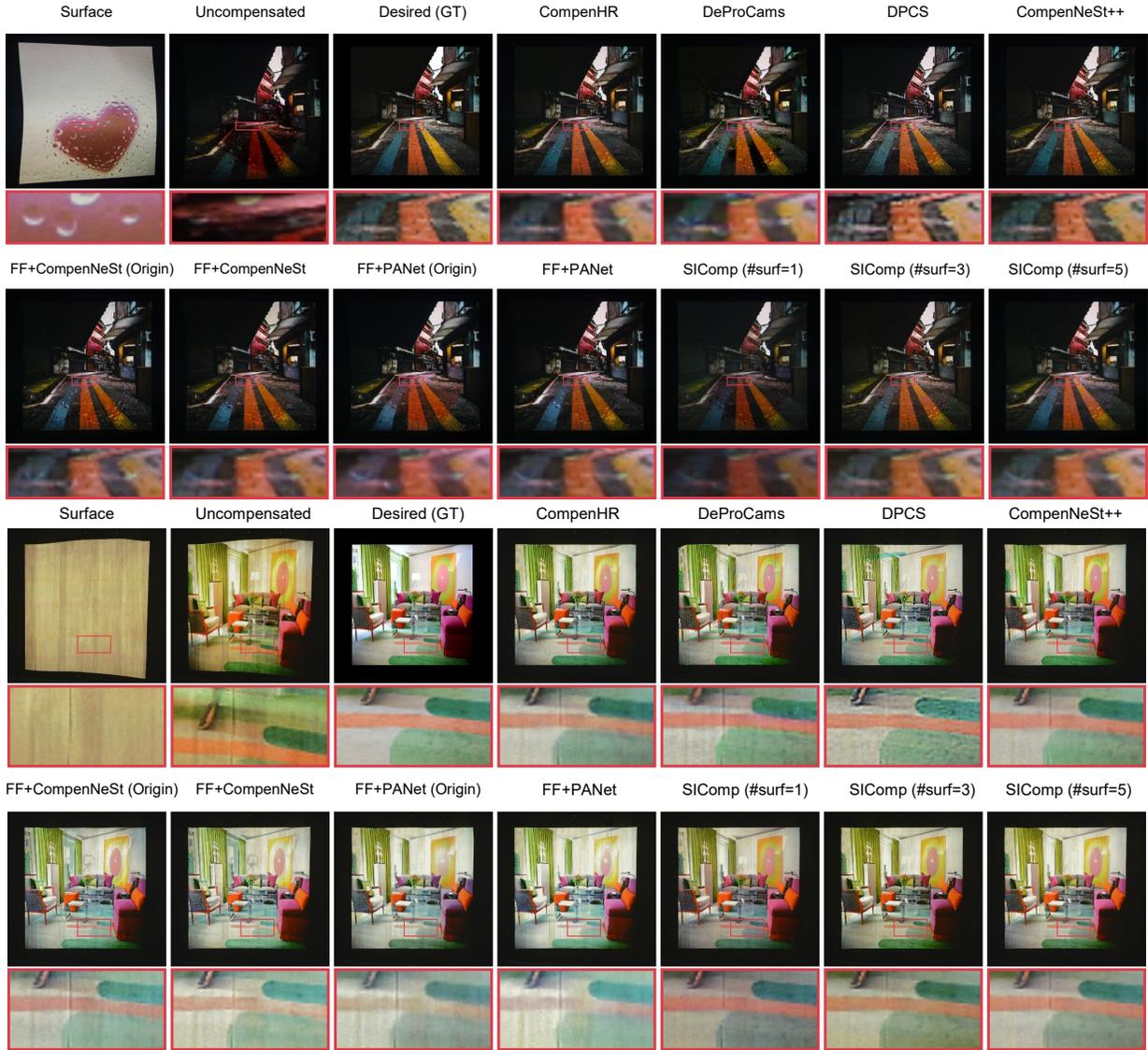
Fig. 6: Qualitative results of real compensation experiments. **Rows 1–2 (Set A)**: Same ProCams devices, unseen setups. **Rows 3–4 (Set B)**: Novel ProCams devices, unseen setups. For each pair of rows: the top row displays the surface, uncompensated image, ground truth, and setup-dependent results (CompenHR, DeProCams, DPCS, CompenNeSt++); the bottom row shows setup-independent results (FF+CompenNeSt, FF+PANet, and SIComp with 1, 3, and 5 surfaces). Red boxes indicate magnified insets for comparison. See more results in Supplementary Material.

5 test images per setup. Note that none of the training images were captured using this ProCams device.

4.1.2 Training configurations

All experiments were conducted on a workstation equipped with two NVIDIA RTX 3090 GPUs and one RTX 4080 GPU. Models were implemented using PyTorch 2.2.1. For all models, training was performed using the Adam [36] optimizer. The training process was set for a maximum of $12,000$ iterations. We adopted distinct learning rates and weight decays for the FlowFormer (our geometric compensation network) and the IVPCNet (our photometric compensation network). Specifically, the FlowFormer was trained with an initial learning rate of 3.5×10^{-5} and a weight decay of 1×10^{-5}. IVPCNet used an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. Learning rate scheduling for both networks followed a step decay strategy. For FlowFormer, the learning rate was decayed by a factor of 0.90 every $5,000$ iterations. For IVPCNet, the learning rate was decayed by a factor of 0.3 every $5,000$ iterations. A consistent batch size of 6 was used throughout the training.

4.1.3 Evaluation metrics and loss function

To quantitatively assess the performance of each projector compensation method, we employ several widely recognized metrics: Peak Signal-to-Noise Ratio (PSNR), Root Mean Square Error (RMSE), Structural Similarity Index Measure (SSIM) [68], Color Difference (ΔE, CIEDE2000) [55], Learned Perceptual Image Patch Similarity (LPIPS) [74], and Fréchet Inception Distance (FID) [18].

We train SIComp using an $\ell_1 + $ SSIM loss combination, which provides the strongest performance on the primary pixel-based metrics (PSNR and SSIM). The comparison of different loss functions is shown in Supplementary Material.

4.2 Evaluation on surrogate compensation

We first evaluate performance using the surrogate compensation pipeline [20], which enables a direct comparison between the predicted compensation $\hat{x}$ and the ground-truth input x in a controlled digital environment. Quantitative results for Set A (8 unseen setups) are detailed in Tab. 3.

As expected, the setup-dependent CompenNeSt++ achieves the

Table 2: Quantitative comparison of **real** projector compensation methods on **Set A** (Same ProCams devices, unseen setups, averaged over 8 setups) and **Set B** (Novel ProCams devices, unseen setups, averaged over 5 setups). SI indicates Setup-Independent methods (✔). Time and GRAM report the operational overhead for the training+inference cycle (or inference only for SI methods). The best results within each category are **bolded**.

Method	SI	Set A: Same ProCams devices, unseen setups						Set B: Novel ProCams devices, unseen setups						Overhead	
		PSNR↑	RMSE↓	SSIM↑	ΔE↓	LPIPS↓	FID↓	PSNR↑	RMSE↓	SSIM↑	ΔE↓	LPIPS↓	FID↓	Time↓	GRAM↓
CompenNeSt++ [22]	✗	**24.64**	**0.063**	**0.883**	**4.48**	**0.110**	**72.77**	**23.75**	0.067	**0.802**	4.98	**0.156**	**96.11**	∼9min	25.3GB
CompenHR [66]	✗	23.69	0.069	0.851	5.33	0.153	91.52	23.74	**0.067**	0.797	5.55	0.178	99.56	∼2min	**1.40GB**
DeProCams [21]	✗	22.32	0.080	0.807	6.25	0.202	119.39	21.96	0.082	0.740	6.65	0.226	128.29	∼8min	9.28GB
DPCS [39]	✗	22.48	0.078	0.819	6.06	0.165	92.70	20.53	0.097	0.715	8.14	0.250	140.55	∼30min	7.10GB
FF+CompenNeSt++	✔	20.76	0.097	0.826	6.95	0.156	119.63	21.48	0.087	0.779	7.26	0.165	123.00	0.49s	0.38GB
FF+CompenNeSt++ (Origin)	✔	19.17	0.115	0.704	7.50	0.182	120.34	19.43	0.110	0.641	8.20	0.208	127.61	0.49s	0.38GB
FF+PANet	✔	21.78	0.085	0.823	6.70	0.170	117.32	21.20	0.090	0.759	7.95	0.188	131.09	**0.45s**	**0.31GB**
FF+PANet (Origin)	✔	21.13	0.091	0.799	8.21	0.206	135.04	20.95	0.091	0.735	8.25	0.217	134.52	**0.45s**	**0.31GB**
SIComp (#surf=1)	✔	18.23	0.134	0.752	10.02	0.191	140.45	22.07	0.083	0.788	7.53	0.153	98.57	0.57s	0.58GB
SIComp (#surf=3)	✔	22.79	0.076	0.840	6.24	**0.144**	95.07	23.11	0.071	0.793	6.06	0.157	100.34	0.58s	0.60GB
SIComp (#surf=5)	✔	**23.51**	**0.070**	**0.846**	**5.76**	0.146	**90.95**	**23.46**	**0.069**	**0.802**	**5.87**	**0.148**	**89.60**	0.60s	0.62GB

Table 3: Quantitative comparison of **surrogate** projector compensation methods on **Set A** (Same ProCams devices, unseen setups, averaged over 8 setups). SI indicates Setup-Independent methods (✔). The best results within each category are **bolded**. See qualitative comparisons in Supplementary Material).

Method	SI	Set A (surrogate)					
		PSNR↑	RMSE↓	SSIM↑	ΔE↓	LPIPS↓	FID↓
CompenNeSt++ [22]	✗	**23.77**	**0.066**	**0.767**	**5.31**	**0.281**	**42.55**
CompenHR [66]	✗	22.36	0.077	0.703	6.48	0.353	73.68
DeProCams [21]	✗	17.86	0.129	0.508	10.92	0.471	121.24
DPCS [39]	✗	16.20	0.161	0.467	12.90	0.467	123.09
FF+CompenNeSt++	✔	20.25	0.098	0.698	8.20	0.319	65.70
FF+CompenNeSt++ (Origin)	✔	18.57	0.119	0.571	9.28	0.350	69.76
FF+PANet	✔	19.77	0.104	0.681	8.77	0.340	76.38
FF+PANet (Origin)	✔	19.05	0.112	0.635	10.00	0.380	95.38
SIComp (#surf=1)	✔	21.54	0.085	0.709	7.29	0.313	57.63
SIComp (#surf=3)	✔	22.24	0.078	0.721	6.84	0.295	50.07
SIComp (#surf=5)	✔	**22.33**	**0.077**	**0.722**	**6.51**	**0.291**	**48.41**

highest fidelity, as it is optimized for fixed environments. However, among Setup-Independent (SI) methods, our SIComp (#surf=5) significantly outperforms existing baselines, including FF+PANet and FF+CompenNeSt++, nearly matching the performance of setup-dependent CompenHR. The performance improvement from #surf=1 to 5 shows the importance of diverse surface priors in modeling complex photometric conditions.

4.3 Evaluation on real compensation

We further validate our approach through physical projection and capture, comparing the camera-captured compensated image against the desired target x'. Results for Set A (same devices, unseen setups) and Set B (novel devices, unseen setups) are summarized in Tab. 2 and Fig. 6. We also show SIComp results on sharp and wavy surfaces in Fig. 7.

4.3.1 Generalization to unseen setups (Set A)

While setup-dependent methods like CompenNeSt++ [22] and CompenHR [66] obtain high fidelity, they require a time-consuming retraining phase (∼2-30 minutes) for every new setup. In contrast, SIComp generalizes to unseen setups.

Specifically, SIComp (#surf=5) outperforms other SI baselines and significantly narrows the performance margin to setup-dependent models. The metric improvement as the number of surface priors increases confirms that richer intensity-varying surface priors effectively enhance photometric robustness. As shown in Tab. 2, SIComp achieves inference in ∼0.6s, which is orders of magnitude faster than the training+inference cycle of setup-dependent methods, making it ideal for dynamic, real-world deployment.

Qualitatively, Fig. 6 (rows 1-2) shows that artifacts (e.g., "heart-shaped" distortions) gradually diminish as the number of surface priors increases.

4.3.2 Generalization to novel devices (Set B)

To evaluate generalization capability of SIComp across different devices, we test on **Set B**, which features entirely novel ProCams devices and unseen setups. As shown in Tab. 2, SIComp (#surf=5) achieves competitive results without any fine-tuning. Remarkably, its performance is on par with CompenNeSt++ on several metrics, despite the latter being specifically retrained for these new devices. This confirms that SIComp successfully decouples device-specific characteristics from the compensation logic.

Visual comparisons in Fig. 6 (rows 3-4) also validate these findings. While other SI baselines exhibit edge artifacts or color shifts, SIComp produces high-fidelity results with accurate color restoration across diverse poses and lighting. SIComp provides a "plug-and-play" solution, drastically reducing operational overhead (Time and GRAM) while maintaining high-quality compensation that is visually indistinguishable from per-setup retrained models.

4.4 Ablation studies

To evaluate the contribution of each key component in SIComp, we conduct ablation studies focusing on the number of intensity-varying surface priors and loss function configurations. Unless otherwise specified, all ablation experiments are performed on the eight setups of Set A to ensure controlled and reproducible evaluation.

4.4.1 Effect of intensity-varying surface priors

To explore the effectiveness of incorporating intensity-varying surface priors, we experimented with different configurations of sampled surface images (Fig. 5). Specifically, SIComp (#surf=1) utilized one surface image (corresponding to a projection grayscale image level of 64). Configuration #surf=3 employed three surface images obtained at projection grayscale image levels 0, 128, and 255. Configuration #surf=5 used five surface images corresponding to projection grayscale image levels 0, 64, 128, 191, and 255.

As detailed in Tab. 2 and Fig. 6, increasing the number of surface priors from SIComp (#surf=1) to SIComp (#surf=3) yields a substantial improvement across all key metrics, significantly enhancing compensation performance and robustness. While SIComp (#surf=5) further improves these results, the gains are relatively marginal compared to the sharp increase observed when moving from #surf=1 to #surf=3.

5 LIMITATIONS AND FUTURE WORK

While SIComp achieves robust setup-independent performance, it presents several constraints that offer directions for future research.
View-dependency. As shown in Fig. 8, SIComp assumes a static canonical viewpoint. Significant observer movement introduces degradation due to non-Lambertian surface reflections and projector-camera parallax. Integrating real-time gaze tracking or view-aware appearance modeling remains a promising future direction.
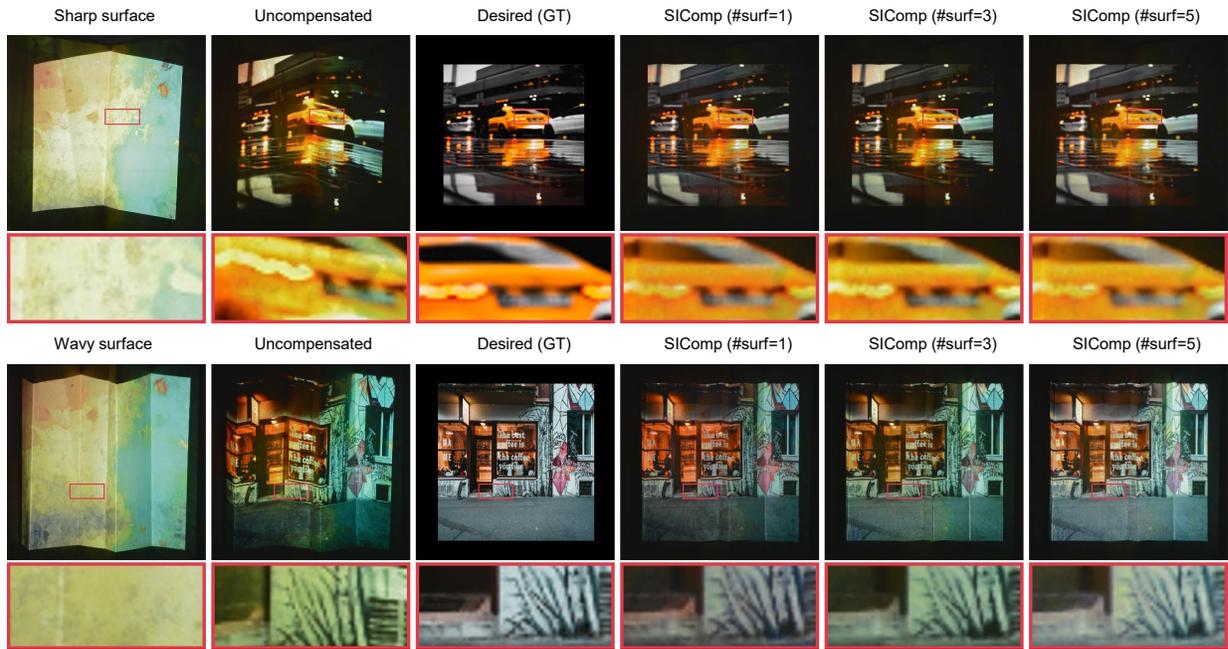
Fig. 7: Real compensation of SIComp on **sharp (top)** and **wavy (bottom)** surfaces, demonstrating improved quality as #surf increases from 1 to 5.
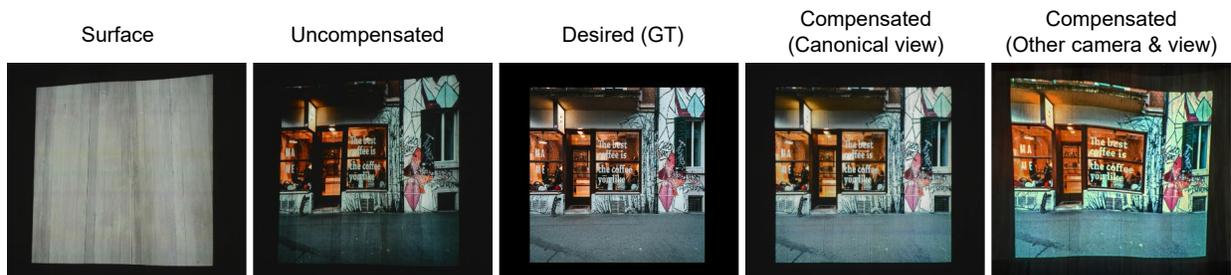

Fig. 8: View-dependent limitation of SIComp. Our proposed SIComp method is restricted to a single viewpoint. The last two columns display the projected compensations on the surface, captured from two different perspectives: the canonical view (using the same DSLR camera as in columns 1-3) and a side view (using a cell phone camera). The significant degradation in fidelity from the side view clearly demonstrates that view changes compromise compensation quality.

Fig. 9: Failure cases under extreme lighting conditions. We evaluate SIComp in a challenging surrogate scenario with highly reflective surface and occlusions. The surrogate compensation in the 3rd column shows clearly limited texture/color recovery in overexposed regions, and poor geometry correction in occluded areas.

Extreme lighting conditions. As illustrated in Fig. 9, SIComp faces challenges in extreme scenarios like specularities and occlusions (shadows). Highly reflective surfaces can saturate camera sensors, causing irreversible information loss. While our intensity-varying surface priors enhance robustness, they cannot recover signals in clipped, overexposed regions. The geometric correction module's dependence on optical flow is also affected by self-occlusions, while the photometric compensation module does not explicitly account for hard shadows. These violations of brightness constancy and flow continuity lead to local artifacts. Extending this framework by incorporating HDR imaging

and multi-projectors is an interesting direction to explore in the future.

6 CONCLUSION

In this paper, we introduced SIComp, to our knowledge, the first setup-independent framework for projector compensation. Our framework effectively addresses both geometric and photometric distortions across a wide array of diverse projection setups, encompassing varying surface properties, ambient lighting conditions, and projector-camera poses. A key innovation lies in formulating projector compensation as a generalizable problem, leveraging attention-based modules and incorporating intensity-varying surface priors. This design enables SIComp to adaptively handle unseen configurations without the need for time-consuming retraining or fine-tuning, significantly enhancing the practical applicability of ProCams. Furthermore, we established a large-scale, diverse dataset to facilitate effective pre-training and promote strong generalization capabilities.

REFERENCES

[1] H. Aoki, T. Tochimoto, Y. Hiroi, and Y. Itoh. Towards co-operative beaming displays: Dual steering projectors for extended projection volume and head orientation range. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2309–2318, 2024. 1

[2] S. Audet and M. Okutomi. A user-friendly method to geometrically calibrate projector-camera systems. In *2009 IEEE computer society con-*

ference on computer vision and pattern recognition workshops, pp. 47–54. IEEE, 2009. 2

[3] A. H. Bermano, M. Billeter, D. Iwai, and A. Grundhöfer. Makeup lamps: Live augmentation of human faces via projection. In *Computer Graphics Forum*, vol. 36, pp. 311–323. Wiley Online Library, 2017. 1

[4] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer. The visual computing of projector-camera systems. In *ACM SIGGRAPH 2008 classes*, pp. 1–25. ACM, 2008. 2

[5] J. Bluteau, I. Kitahara, Y. Kameda, H. Noma, K. Kogure, and Y. Ohta. Visual support for medical communication by using projector-based augmented reality and thermal markers. In *Proceedings of the 2005 international conference on Augmented tele-existence*, pp. 98–105, 2005. 1

[6] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In A. Fitzgibbon et al. (Eds.), ed., *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pp. 611–625. Springer-Verlag, Oct. 2012. 4, 5

[7] J.-C. Chien, J.-D. Lee, C.-W. Chang, and C.-T. Wu. A projection-based augmented reality system for medical applications. *Applied Sciences*, 12(23):12027, 2022. 1

[8] C. Cruz-Neira, D. Sandin, and T. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave (1993). *ResearchGate, Chicago*, 2023. 1

[9] Q. Deng, J. Li, H. Ling, and B. Huang. Gs-procams: Gaussian splatting-based projector-camera systems, 2024. 2, 3

[10] P. Edgcumbe, R. Singla, P. Pratt, C. Schneider, C. Nguan, and R. Rohling. Follow the light: projector-based augmented reality intracorporeal system for laparoscopic surgery. *Journal of Medical Imaging*, 5(2):021216–021216, 2018. 1

[11] Y. Erel, D. Iwai, and A. H. Bermano. Neural projection mapping using reflectance fields. *IEEE Transactions on Visualization and Computer Graphics*, 29(11):4339–4349, 2023. 2, 3

[12] Y. Erel, O. Kozlovsky-Mordenfeld, D. Iwai, K. Sato, and A. H. Bermano. Casper dpm: Cascaded perceptual dynamic projection mapping onto hands. In *SIGGRAPH Asia 2024 Conference Papers*, pp. 1–10, 2024. 2

[13] 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, pp. 814–821. IEEE, 2005. 2

[14] A. Grundhofer. Practical non-linear photometric projector compensation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 924–929, 2013. 2

[15] A. Grundhöfer and D. Iwai. Robust, error-tolerant photometric projector compensation. *IEEE Transactions on Image Processing*, 24(12):5086–5099, 2015. 1, 2

[16] A. Grundhöfer and D. Iwai. Recent advances in projection mapping algorithms, hardware and applications. In *Computer graphics forum*, vol. 37, pp. 653–675. Wiley Online Library, 2018. 1

[17] M. Harville, B. Culbertson, I. Sobel, D. Gelb, A. Fitzhugh, and D. Tanguay. Practical methods for geometric and photometric correction of tiled projector. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pp. 5–5. IEEE, 2006. 1

[18] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 7

[19] B. Huang and H. Ling. Compennet++: End-to-end full projector compensation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7165–7174, 2019. 1, 2

[20] B. Huang and H. Ling. End-to-end projector photometric compensation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6810–6819, 2019. 1, 3, 4, 6, 7

[21] B. Huang and H. Ling. Deprocams: Simultaneous relighting, compensation and shape reconstruction for projector-camera systems. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2725–2735, 2021. 1, 3, 6, 8

[22] B. Huang, T. Sun, and H. Ling. End-to-end full projector compensation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):2953–2967, 2021. 1, 2, 3, 4, 5, 6, 8

[23] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pp. 1–11, 2024. 3

[24] Z. Huang, X. Shi, C. Zhang, Q. Wang, K. C. Cheung, H. Qin, J. Dai, and H. Li. Flowformer: A transformer architecture for optical flow. In

[25] M. T. Ibrahim, M. Gopi, and A. Majumder. Projector-camera calibration on dynamic, deformable surfaces. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 905–906. IEEE, 2023. 2

[26] M. T. Ibrahim, M. Gopi, and A. Majumder. Self-calibrating dynamic projection mapping system for dynamic, deformable surfaces with jitter correction and occlusion handling. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 293–302. IEEE, 2023. 2

[27] M. T. Ibrahim, M. Gopi, and A. Majumder. Real-time seamless multi-projector displays on deformable surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2527–2537, 2024. 2

[28] M. T. Ibrahim and A. Majumder. Multi-projector dynamic spatially augmented reality on deformable surfaces. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 1156–1157. IEEE, 2024. 2

[29] M. T. Ibrahim, G. Meenakshisundaram, and A. Majumder. Dynamic projection mapping of deformable stretchable materials. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–5, 2020. 2

[30] D. Iwai. Projection mapping technologies: A review of current trends and future directions. *Proceedings of the Japan Academy, Series B*, 100(3):234–251, 2024. 2

[31] S. Kagami and K. Hashimoto. Animated stickies: Fast video projection mapping onto a markerless plane through a direct closed-loop alignment. *IEEE transactions on visualization and computer graphics*, 25(11):3094–3104, 2019. 2

[32] S. Kagami and K. Hashimoto. Interactive stickies: Low-latency projection mapping for dynamic interaction with projected images on a movable surface. In *ACM SIGGRAPH 2020 Emerging Technologies*, pp. 1–2. 2020. 2

[33] Y. Kageyama, D. Iwai, and K. Sato. Online projector deblurring using a convolutional neural network. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2223–2233, 2022. 1

[34] Y. Kageyama, D. Iwai, and K. Sato. Efficient distortion-free neural projector deblurring in dynamic projection mapping. *IEEE Transactions on Visualization and Computer Graphics*, 30(12):7544–7557, 2024. 1

[35] O. V. Kanivets, I. M. Kanivets, N. V. Kononets, T. M. Gorda, and E. O. Shmeltser. Development of mobile applications of augmented reality for projects with projection drawings. In *Augmented Reality in Education: Proceedings of the 2nd International Workshop (AREdu 2019), Kryvyi Rih, Ukraine, March 22, 2019*, pp. 262–273. CEUR Workshop Proceedings, 2020. 1

[36] D. P. Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[37] P. Kurth, V. Lange, C. Siegl, M. Stamminger, and F. Bauer. Auto-calibration for dynamic multi-projection mapping on arbitrary surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2886–2894, 2018. 1

[38] P. Kurth, M. Leuschner, M. Stamminger, and F. Bauer. Content-aware brightness solving and error mitigation in large-scale multi-projection mapping. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3607–3617, 2022. 1

[39] J. Li, Q. Deng, H. Ling, and B. Huang. Dpcs: Path tracing-based differentiable projector-camera systems. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 2, 3, 6, 8

[40] 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. 2, 3, 4

[41] L.-Y. Liu and M.-T. Cheng. Interactive projector as an interactive teaching tool in the classroom: Evaluating teaching efficiency and interactivity. *Turkish Online Journal of Educational Technology-TOJET*, 14(2):110–115, 2015. 1

[42] S. Liu, Q. Ruan, and X. Li. The color calibration across multi-projector display. *journal of signal and information processing*, 2(2):53, 2011. 1, 2

[43] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021. 2, 3, 6

[44] 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, pp. 674–679, 1981. 1

[45] D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pp. 464–471. IEEE, 2012. 2

[46] G. Narita, Y. Watanabe, and M. Ishikawa. Dynamic projection mapping onto deforming non-rigid surface using deformable dot cluster marker. *IEEE transactions on visualization and computer graphics*, 23(3):1235–1248, 2016. 2

[47] S. K. Nayar, H. Peri, M. D. Grossberg, and P. N. Belhumeur. A projection system with radiometric compensation for screen imperfections. In *ICCV workshop on projector-camera systems (PROCAMS)*, vol. 3, 2003. 1, 2, 5

[48] J. Park, D. Jung, and B. Moon. Projector compensation framework using differentiable rendering. *IEEE Access*, 10:44461–44470, 2022. 2, 3

[49] H.-L. Peng, K. Sato, S. Nakagawa, and Y. Watanabe. Perceptually-aligned dynamic facial projection mapping by high-speed face-tracking method and lens-shift co-axial setup. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 1

[50] P. Pjanic, S. Willi, D. Iwai, and A. Grundhöfer. Seamless multi-projection revisited. *IEEE transactions on visualization and computer graphics*, 24(11):2963–2973, 2018. 2

[51] A. G. Porreca. *A Comparison of the Effectiveness of an Overhead Projector Presentation and a Programmed Textbook Presentation in Economic Education*. Boston University School of Education, 1971. 1

[52] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8):2666–2680, 2010. 2

[53] Y. Sato, D. Iwai, and K. Sato. Responsive-extendedhand: Adaptive visuo-haptic feedback recognizing object property with rgb-d camera for projected extended hand. *IEEE Access*, 12:38247–38257, 2024. 1

[54] M. Shahpaski, L. Ricardo Sapaico, G. Chevassus, and S. Susstrunk. Simultaneous geometric and radiometric calibration of a projector-camera pair. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4885–4893, 2017. 3

[55] G. Sharma, W. Wu, and E. N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 30(1):21–30, 2005. 7

[56] 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 Transactions on Graphics (TOG)*, 34(6):1–11, 2015. 1, 3

[57] T. Takahashi, T. Kawano, K. Ito, T. Aoki, and S. Kondo. Performance evaluation of a geometric correction method for multi-projector display using sift and phase-only correlation. In *2010 IEEE International Conference on Image Processing*, pp. 1189–1192. IEEE, 2010. 2

[58] M. Takeuchi, H. Kusuyama, D. Iwai, and K. Sato. Projection mapping under environmental lighting by replacing room lights with heterogeneous projectors. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2151–2161, 2024. 1

[59] J. Tardif. Multi-projectors for arbitrary surfaces without explicit calibration and reconstruction. *DIM*, pp. 1–14, 2003. 2

[60] Z. Teed and J. Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pp. 402–419. Springer, 2020. 3

[61] M. A. Tehrani, M. Gopi, and A. Majumder. Automated geometric registration for multi-projector displays on arbitrary 3d shapes using uncalibrated devices. *IEEE transactions on visualization and computer graphics*, 27(4):2265–2279, 2019. 2

[62] M. A. Tehrani, M. T. Ibrahim, A. Majumder, and M. Gopi. 3d gamut morphing for non-rectangular multi-projector displays. *IEEE Transactions on Visualization and Computer Graphics*, 30(8):4724–4738, 2023. 2

[63] T. Ueda, D. Iwai, T. Hiraki, and K. Sato. Illuminated focus: Vision augmentation using spatial defocusing via focal sweep eyeglasses and high-speed projector. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2051–2061, 2020. 1

[64] D. Wang, I. Sato, T. Okabe, and Y. Sato. Radiometric compensation in a projector-camera system based properties of human vision system. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pp. 100–100. IEEE, 2005. 2

[65] J. Wang, Z. Zhang, W. Lu, and X. J. Jiang. High-accuracy calibration of high-speed fringe projection profilometry using a checkerboard. *IEEE/ASME Transactions on Mechatronics*, 27(5):4199–4204, 2022. 2

[66] 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)*, pp. 135–145. IEEE, 2023. 1, 2, 3, 4, 5, 6, 8

[67] Y. Wang, H. Ling, and B. Huang. Vicomp: Video compensation for projector-camera systems. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 3, 4

[68] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 7

[69] G. Wetzstein, O. Bimber, et al. Radiometric compensation through inverse light transport. *PG*, 7:391–399, 2007. 2, 3

[70] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018. 2, 3, 6

[71] K. Yamamoto, D. Iwai, I. Tani, and K. Sato. A monocular projector-camera system using modular architecture. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5586–5592, 2022. 1

[72] M. Yasui, R. Iwataki, M. Ishikawa, and Y. Watanabe. Projection mapping with a brightly lit surrounding using a mixed light field approach. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2217–2227, 2024. 1

[73] T. Yoshida, C. Horii, and K. Sato. A virtual color reconstruction system for real heritage with light projection. In *Proceedings of VSMM*, vol. 3, pp. 1–7. Citeseer, 2003. 1, 2

[74] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018. 7