End-to-end Projector Photometric Compensation
- Supplementary Materials -

Bingyao Huang[1,2,*] Haibin Ling[1,†]
[1]Department of Computer and Information Sciences, Temple University, Philadelphia, PA USA
[2]Meitu HiScene Lab, HiScene Information Technologies, Shanghai, China
{bingyao.huang, hbling}temple.edu

1. Introduction

In this supplementary material, firstly, we report a comprehensive evaluation of the proposed CompenNet trained with different hyperparameters in §2. Then we discuss the configurations of the benchmark in §3. Finally, the camera perceived compensation results are shown in §4.

2. CompenNet Hyperparameters

We train and evaluate the proposed CompenNet with different number of iterations, batch sizes and number of training images using the proposed benchmark dataset. As shown in Table 1, CompenNet achieves higher PSNR and SSIM and lower RMSE when number of iterations and number of training images increase. In addition, when the number of iterations is 1000 and the number of training images is 125, a batch size of 32 outperforms a batch size of 64 on training time, PSNR, RMSE and SSIM. In this paper, we use a default setting of iterations = 1000, batch size = 64 and #train = 500 to balance training/prediction time and sampling data size. However, if an application prefers accuracy to speed, it can increase the number of iterations and capture more training data accordingly.

Table 1: Qualitative results of the proposed CompenNet trained using different number of iterations, batch sizes and number of training images. The training time in minutes is shown in the last column. Note the default CompenNet's hyperparameters are: iterations = 1000, batch size = 64 and #train = 500.

Iterations	**Batch size**	**#Train**	**PSNR**	**RMSE**	**SSIM**	**Time**
1000	32	125	21.3595	0.1503	0.7381	~6m
		250	21.2740	0.1541	0.7396	~6m
		500	21.7348	0.1435	0.7490	~6m
	64	125	21.0542	0.1574	0.7314	~10m
		250	21.2991	0.1536	0.7420	~10m
		500	**21.7998**	**0.1425**	**0.7523**	~10m
2000	32	125	21.5335	0.1474	0.7448	~12m
		250	21.4579	0.1509	0.7479	~12m
		500	21.9369	0.1404	0.7571	~12m
	64	125	21.5505	0.1471	0.7441	~20m
		250	21.4891	0.1505	0.7484	~20m
		500	**22.0005**	**0.1393**	**0.7596**	~20m
Uncompensated	-	-	12.1673	0.4342	0.4875	-

*Work partly done during internship with HiScene.
†Corresponding author.

3. Benchmark Configuration

The benchmark dataset consists of 24 different setups. The following settings are varied: the projection surface texture, the pose between the camera, the projector and the projection surface, lighting conditions, projector brightness and contrast, camera shutter speed, f-number and focal length. It is worth noting that the camera-projector parameters are coadjusted such that the brightest projected input image (plain white) slightly overexposes the camera captured image. Similarly, the darkest projected input image (plain black) slightly underexposes the camera captured image. This allows the projector dynamic range to cover the full camera dynamic range.

It takes about 15 minutes to capture one setup, which includes 125 plain color images for TPS [1], and 500 training images and 200 validation/testing images for TPS textured, pix2pix [2] and CompenNet. To compare the final camera captured compensated projected results, it takes extra 40 minutes to train the four models. Then the 200×4=800 compensated images by each model are projected and captured by the camera, as shown in Fig. 1. During this process the setup must stay unchanged. As we can see, this process is time consuming and requires a large amount of manual efforts. In this work, we provide a surrogate evaluation protocol that requests no actual projection of the algorithm output. As a result, this surrogate allows us to construct, for the first time, a sharable setup-independent compensation benchmark, which is expected to facilitate future works in this direction.

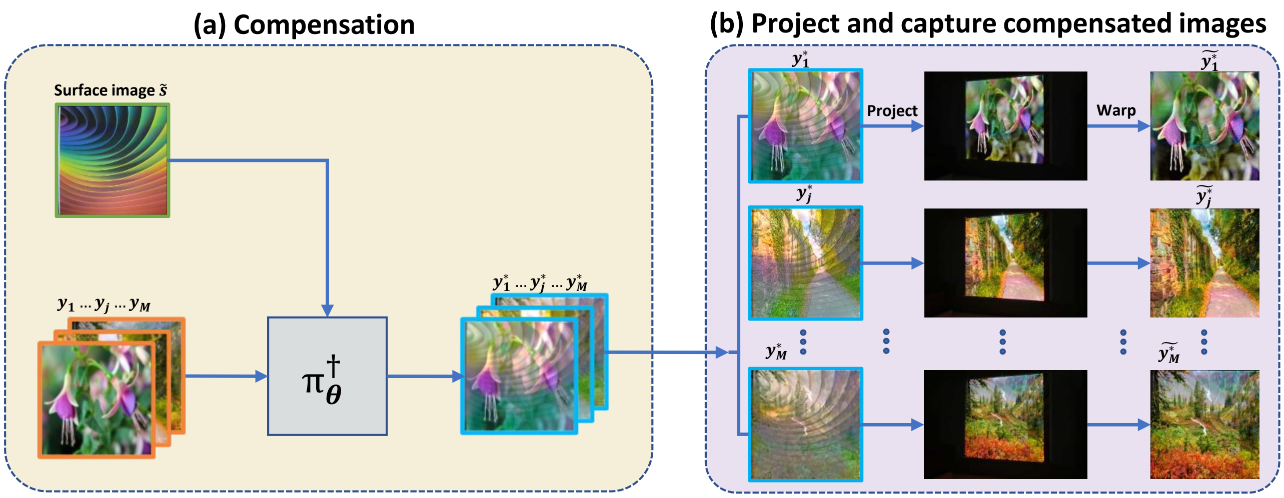


Figure 1: Flowchart of obtaining the final camera perceived compensation results. **(a)** With the trained CompenNet $\pi^{\dagger}_{\boldsymbol{\theta}}$, input images $\boldsymbol{y} \dots \boldsymbol{y}_M$ are compensated, then **(b)** projected to the surface and captured by the camera.

4. Camera Perceived Compensated Images

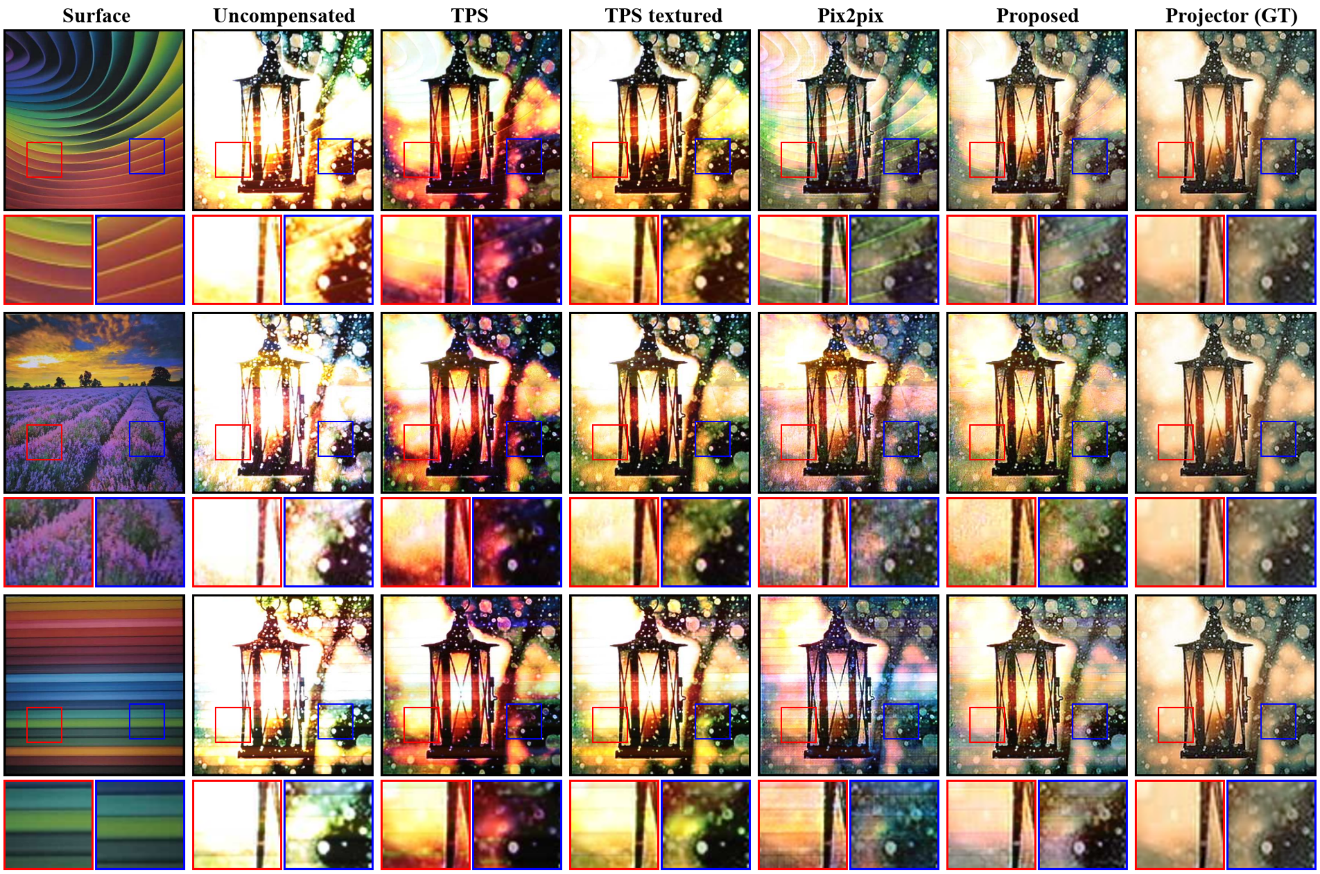
As we mentioned in the paper §4 **Benchmark**, despite the proposed benchmark is setup-independent and does not require future works to replicate our setup for comparison, we demonstrate the effectiveness of the proposed method on actual camera perceived compensation results below.

In the following figures, we show the comparisons of TPS [1], TPS textured, pix2pix [2] and CompenNet on three different surfaces. The 1st column is the camera-captured projection surface. The 2nd column is the camera-captured uncompensated projected image. The 3rd to 6th columns are the camera-captured compensated projected images using different methods. The last column is the ground truth input image. Each image is provided with two zoomed-in patches for detailed comparison. We list six comparison images due to the 20MB limit of the supplementary material. The source code, benchmark and experimental results are available at `https://github.com/BingyaoHuang/CompenNet`.

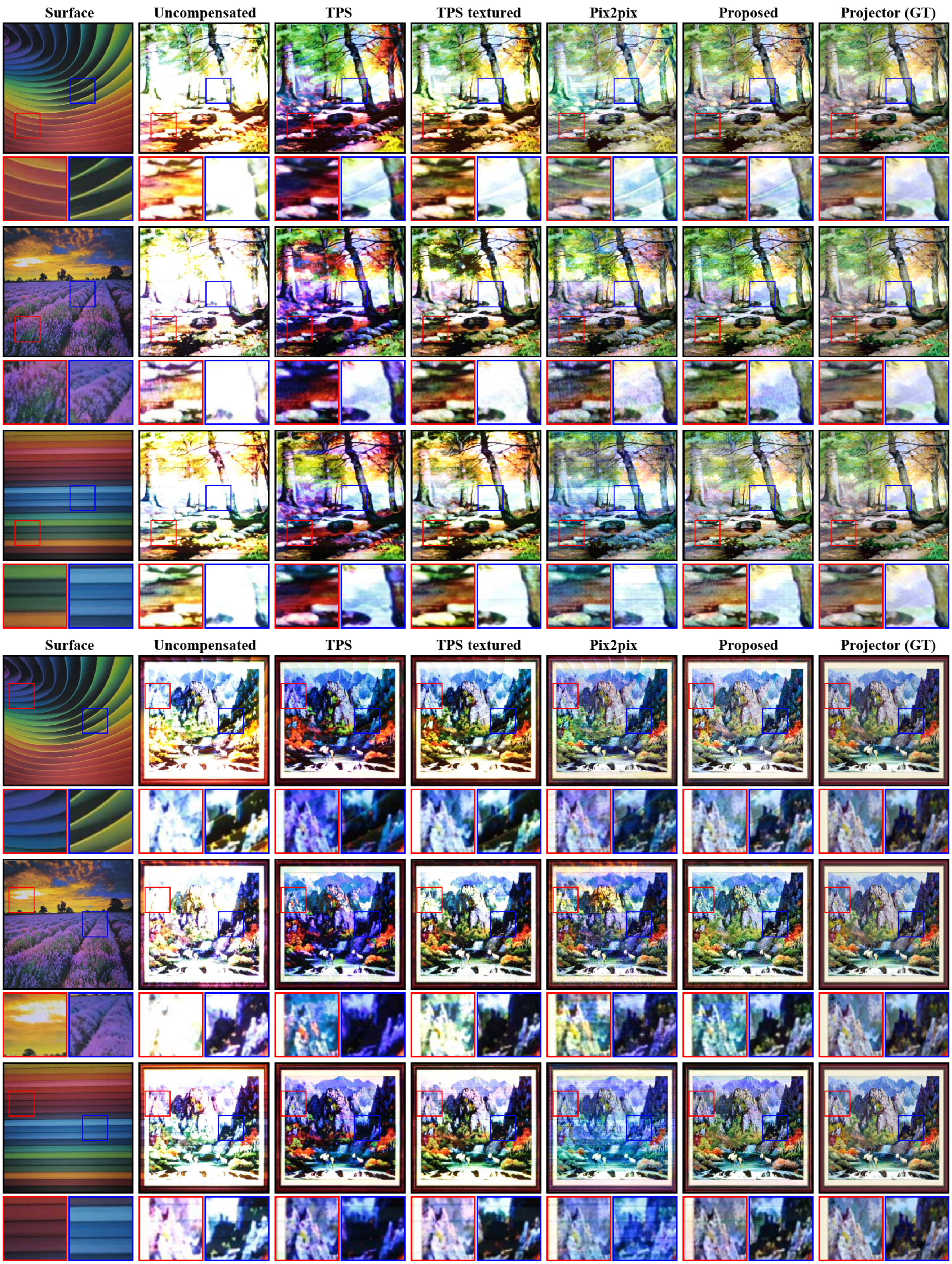
References

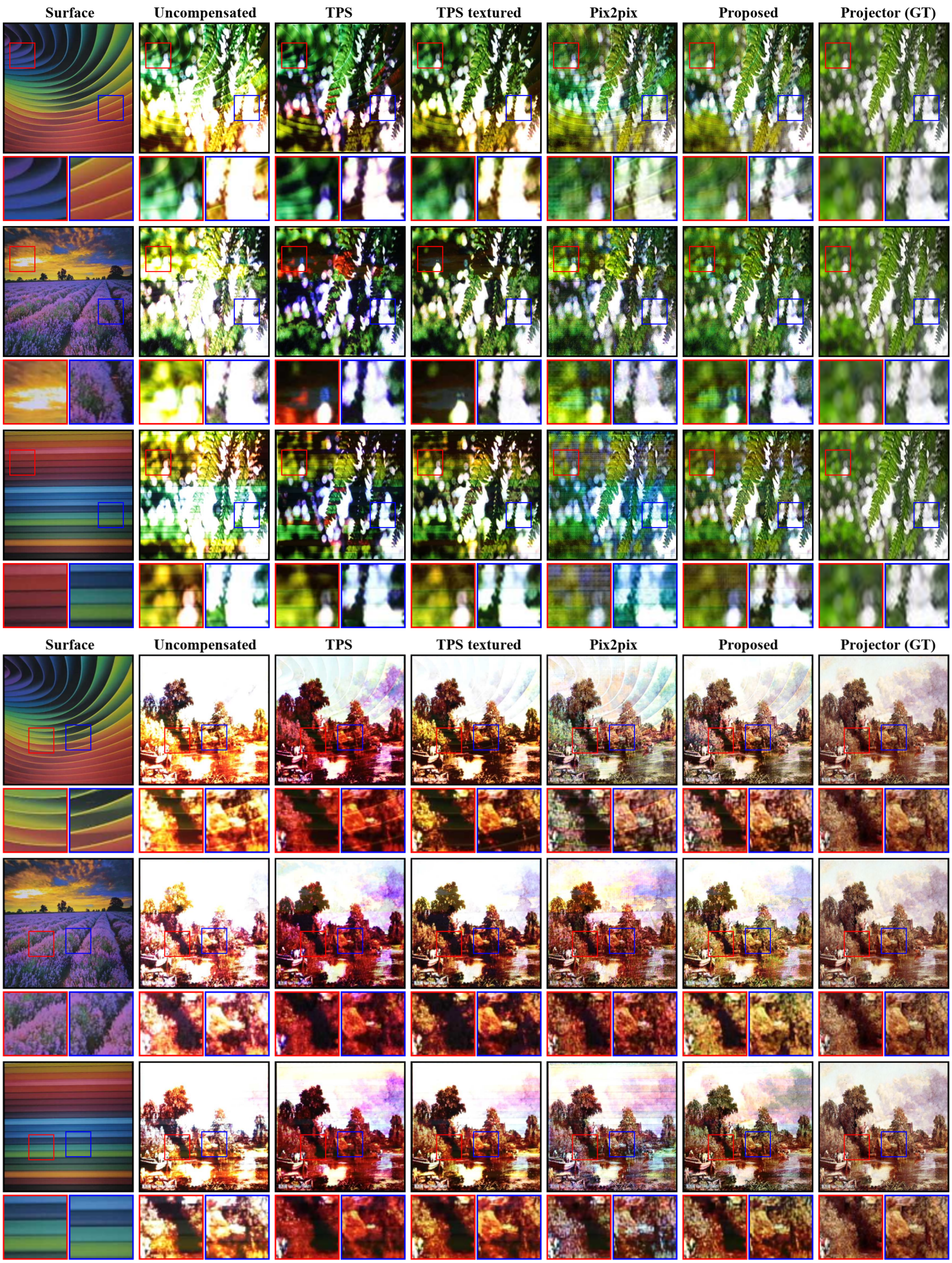
[1] A. Grundhöfer and D. Iwai. Robust, error-tolerant photometric projector compensation. *IEEE TIP*, 24(12):5086–5099, 2015. 2

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017. 2


Surface
Uncompensated
TPS
TPS textured
Pix2pix
Proposed
Projector (GT)


Surface
Uncompensated
TPS
TPS textured
Pix2pix
Proposed
Projector (GT)


Surface
Uncompensated
TPS
TPS textured
Pix2pix
Proposed
Projector (GT)
Surface
Uncompensated
TPS
TPS textured
Pix2pix
Proposed
Projector (GT)


Surface
Uncompensated
TPS
TPS textured
Pix2pix
Proposed
Projector (GT)
Surface
Uncompensated
TPS
TPS textured
Pix2pix
Proposed
Projector (GT)