

Removing Reflection From a Single Image With Ghosting Effect

Yan Huang, Yuhui Quan*, Yong Xu, Ruotao Xu, Hui Ji

Abstract—How to remove undesired reflections of images taken through glass is an important problem in digital photography and other vision applications. When taking images through thick or insulating glass windows, the so-called ghosting effect, *i.e.* pattern repetitiveness in reflection, is exploited in existing reflection removal techniques. By formulating reflection removal as a two-layer separation problem, the existing methods take a two-stage approach that first estimates the parameters of ghosting effect and then separates the reflection and latent image layer. This paper aims at addressing one main challenge, *i.e.* how to accurately distinguish repetitive patterns on the two layers. Motivated by the observation on the difference of the number of patterns repeated in two layers, a wavelet transform based regularization with a novel weighting scheme is proposed for separating the two layers. The experiments showed that the proposed method is capable of accurately separating the latent image layer and the reflection layer, and outperforms the existing ones on both synthetic data set and real data set.

Index Terms—Reflection removal, Image separation, Ghosting effects, Image decomposition

I. INTRODUCTION

It often happens that one has to take pictures through glass windows, *e.g.* taking pictures of the paintings in the museum [1], [2], [3], [4], [5], the dresses in the showcase [6], [7], [8], the landscape outside when in a train or plane [9], [10]. Unfortunately, it is known that the resulting picture taken through glass often contains the reflection of the scene behind the camera. The avoidance of the inclusion of such reflection requires either specific hardware (*e.g.* polarizers) or manual control of lighting conditions around cameras. There is certainly the need to develop effective techniques to remove the reflection from an input image taken through a glass window.

In general, it is not possible to separate the reflection caused by glass and the content of the scene of interest, without additional priors or assumptions on the reflection. One popular cue for identifying reflection in the image is the so-called *ghosting effect*. See Fig. 1 for an illustration. The ghosting effects of an image caused by the reflection of glass refer to the two or multiple reflections of the reflected scene off the different glass surfaces. For example, for a two-panel window, both panels will reflect the scene behind and the two reflections differ from each other by their positions and

Yan Huang, Yuhui Quan, Ruotao Xu and Yong Xu are with School of Computer Science and Engineering at South China University of Technology, Guangzhou 510006, China. (email: huangkaiyan@scut.edu.cn; ruotao.xu@mail.scut.edu.cn; yxu@scut.edu.cn; csyhquan@scut.edu.cn)

Hui Ji is with Department of Mathematics at National University of Singapore, Singapore 119076. (email: matjh@nus.edu.sg)

Asterisk indicates the corresponding author.

brightness strengths. For a single-panel window with thick glass, both sides of the glass will yield similar effects, but with weakened brightness. See Fig. 2 for an illustration of such two types of reflections. In practice, since the brightness strength of the third reflection is much less than the primary reflection, the double reflection instead of the multiple one is usually considered; see e.g. [11], [12].



Fig. 1: Illustration of reflection removal from a single image with ghosting effect. (a) The image with ghosting effect. (b) The result after reflection removal by the proposed method.

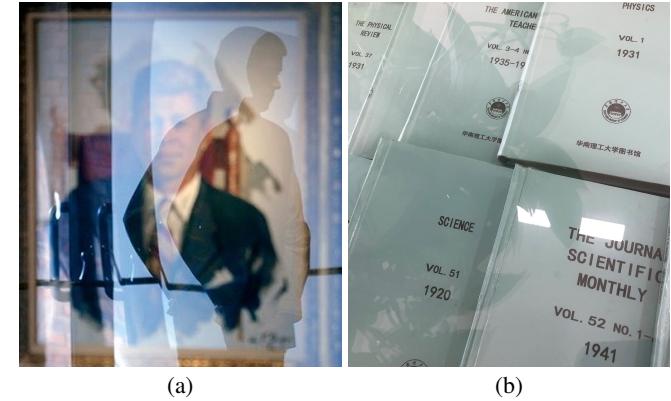


Fig. 2: Illustration of ghosting effect. (a) Ghosting reflection caused by two sides of glass. (b) Ghosting reflection caused by thick glass.

The image with ghost effect is modeled in [11], [12] by the following composite model:

$$\mathbf{Y} = \mathbf{T} + \mathbf{R} \otimes \mathbf{k} + \mathbf{N}, \quad (1)$$

where \otimes denotes the convolution operation, \mathbf{Y} denotes the input image, \mathbf{N} denotes the noise, \mathbf{T} denotes the transmission layer which is the captured frame of the scene behind the window, and $\mathbf{R} \otimes \mathbf{k}$ denotes the reflection layer which is the captured frame of the scene in front of the window. The

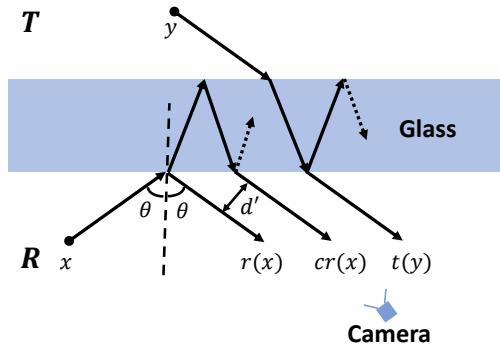


Fig. 3: Diagram of ghosting effects. Each object point is sensed simultaneously into different pixels due to the dissipation of its energy among the different reflection orders, and a spatial effect is created by the secondary reflection. With unpolarized light and flat glass, both t and r are only dependent on the incident angle. The distance between the primary reflection and the successive secondary reflection is d' for constant incident angle. More details can be found in [11].

reflection layer is modelled by an image pattern R convolved with a two-impulse kernel of the form:

$$k = \delta(x) + c\delta(x - d), \quad (2)$$

where δ denotes the Dirac impulse function, and c, d are two parameters. The convolution model of the reflection layer is illustrated in Fig. 3. It can be seen that the effect can be modeled by two replicates of the same object with the displacement d' and with different average brightness, which then can be formulated as a convolution with a specific kernel function parameterized by the ghosting shift d which is proportional to the physical distance d' and the attenuation factor c which is the ratio between the brightness strength of the first reflection and the successive second reflection. In other words, The displacement and the difference of average brightness is parameterized by d and c in the kernel k . These two parameters depend on the reflectance properties and the configurations of glass surfaces.

Existing methods, *e.g.* [12], take a two-stage approach for reflection removal, which first estimates the two parameters of the kernel k and then estimates the reflected image region R using the estimated k . The task in the second stage requires to solve a challenging ill-posed inverse problem, *i.e.*, simultaneously estimating the transmission layer T and the reflection layer $R \otimes k$ from Equation (1). Motivated by the fact that the reflection layer $R \otimes k$ contains the replicates of the same object, the method proposed in [12] imposes the periodicity prior on $R \otimes k$ for the separation of two layers. Unfortunately, such a prior also often holds true for the transmission layer as well, especially for the scenes in urban regions. As a result, the periodicity prior on the reflection layer used in [12] is not sufficient for distinguishing the repeating patterns in two layers. It can be seen that the outputs from the method proposed in [12] often show noticeable artifacts on the transmission layer with local periodic patterns. See Fig. 4 for such a demonstration, where the artifacts appear on the windows as well as the long white roof.

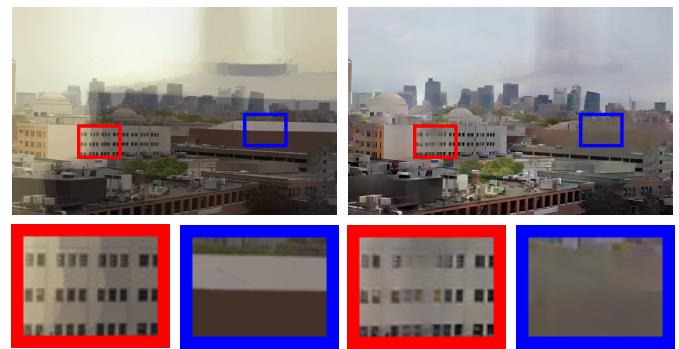


Image with local periodic patterns Results by Shih *et al.* [12]

Fig. 4: Demonstration of the results of the method proposed by Shih *et al.* [12] on the image which contain local periodic patterns.

Motivated by the weakness of the existing methods that use the ghosting effect for reflection removal, this paper aims at developing a better method for removing reflection from the input image, with a particular focus on better regularization model for separating the two layers: reflection layer and transmission layer. Built on the same two-stage framework as the previous method, *e.g.* [12], we proposed a new prior on the repeating patterns of the two layers to distinguish them, which is based on the observation that the number of repeating patterns in the reflection layer is two and that in the transmission layer is usually larger than two. Thus, by checking the number of repeating patterns in the direction of the estimated ghosting shift, one can accurately tell which layer the periodic patterns (*e.g.* edges) belong to. Such a prior is then implemented in a wavelet transform based regularization model with a novel re-weighting scheme for the separation of the two layers. In short, the contributions of the proposed method are summarized as follows:

- A new periodicity-based prior for accurately distinguishing repeating patterns in the two layers
- A wavelet transform based regularization model with a novel re-weighting scheme for effectively extracting the two layers.

In addition, the ghosting kernel estimation scheme used in the first stage of [12] is also refined with a better implementation technique. The experiments on both the synthetic data set and real data set showed that the proposed method outperformed the existing ones. In addition to the performance gain, the proposed one has other merits over existing ones. For example, Shih *et al.*'s method [12] requires a large data set for GMM training and requires solving a complex optimization problem. In contrast, the proposed one does not need any data for training, and has a much simpler optimization problem to solve.

The paper is organized as follows. In section II we give a brief discussion on the related work on reflection removal. In section III, we present the proposed method, as well as the implementation details. In section IV the experiments are done to evaluate the proposed work, with the comparison to other related methods. In Section V we conclude the paper

and discuss the future work along this line.

II. RELATED WORK

General reflection removal considers the following composite model for images with undesired reflection (see *e.g.* [13], [2], [6], [14], [7], [15]):

$$\mathbf{Y} = \mathbf{T} + \tilde{\mathbf{R}}, \quad (3)$$

where $\mathbf{Y} \in \mathbb{R}^{M \times N}$ is the input which is the superposition of the transmission layer $\mathbf{T} \in \mathbb{R}^{M \times N}$ and the reflection layer $\tilde{\mathbf{R}} \in \mathbb{R}^{M \times N}$. The goal of reflection removal is then to estimate the transmission layer \mathbf{T} from \mathbf{Y} . Clearly, the problem of reflection removal is an ill-posed problem with the ambiguities between \mathbf{T} and $\tilde{\mathbf{R}}$. Additional priors on \mathbf{T} and $\tilde{\mathbf{R}}$ are needed to separate these two layers.

Based on the number of input sources, reflection removal can be classified into two categories, *i.e.* multiple-image based approach and single-image based approach.

A. Multiple-image based general reflection removal

The early reflection removal algorithms treated the problem as a blind source separation problem, which separates the mixed signal into two independent signals. Farid *et al.* [16] collected two images taken with different polarization angles and estimated the mixing coefficients using independent component analysis. Richard *et al.* [17] proposed to use constrained least squares to recover the background layers based on an arbitrary number of composite images with known motion estimates. Bernard and Michal [18] presented an approach for separating reflections under various real-world scenarios, including images obtained under different polarizations or videos with non-rigid transparent motions. Ni *et al.* [19] aimed at processing images captured by light field cameras which could provide extra directional information of incoming rays. Such methods directly model the reflection generation process and solve it ignoring the intrinsic properties of each layer.

In recent years, most proposed approaches [20], [6], [4], [9], [14], [7], [21], [8] focused on discovering the intrinsic properties of each layer or exploiting the different motion behaviors between layers. The most often seen natural image prior is the sparsity of image gradients [20], [6], [4], [9], [14], [21]. The Laplacian approximation is used in [6], [9] for its quick convergence and simplicity. The Hyperlaplacian approximation used in [14], [21] is regarded as a better distribution to model the long-tail distribution of the image gradients. In order to describe the distribution of image gradients more accurately, Gai *et al.* [20] proposed to learn the distribution of gradients from the given data.

In addition to the image prior based approaches, there are also a class of approaches which utilizes motion cues from different layers for reflection removal. Based on the fact that the reflection layer moves slower than the background since it is closer to the camera than the transmission layer, Xue *et al.* [9] proposed to separate the two layers using the difference of motion speed between layers. Built upon the observation that the background layer as the major part of

an image is rather static, a number of approaches used the SIFT flow [22] to first warp the images to the reference one, and then to estimate the edge map using different techniques, *e.g.* the sparsity of the edge occurrence [6], the low-rank matrix completion [8], the motion and intensity score [7]. A main challenge to such flow-based methods is that the flow estimation may be not stable under reflections. To have better robustness on optical flow estimated from videos with reflections, Yang *et al.* [23] considered a generalized version of the brightness constancy constraint. Ajay *et al.* [24] constructed a spatio-temporal optimization model which jointly solves the flow estimation and reflection separation.

B. Single-image based general reflection removal

Since the single-image based reflection removal problem is severely ill-posed, extra information is required for the separation. Levin *et al.* [13] searched for a decomposition of two layers with the minimal amount of edges and corners. Wan *et al.* [15] proposed to retrieve external patches based on a natural image database and combine the sparsity prior and non-local similarity prior of images for the removal. In addition to the help of external databases, Levin *et al.* [25] also requires user interactions for marking a number of edges within background or reflection labels, and then refined the separation framework based on the sparsity of derivative filters with an iterative re-weighted least squares scheme.

Another consideration for the single-image based reflection removal problem is the same as the multiple-image based one: to discover the different properties of the two mixed layers. In [26], [27], [10], an assumption is made that the background is in focus of the camera and the reflection is out of focus, and thus the reflection layer is relatively smoother than the background layer. Based on the assumption, Li *et al.* [26] separated the two layers by enforcing different sparsity properties on two layers respectively: first order derivative filters on the transmission layer and second order derivative filter on the reflection layer. Wan *et al.* [27] used the multi-scale technique to calculate the confidence map for background edge selection. Nikolaos *et al.* [10] directly removed the reflections based on a Laplacian data fidelity and a ℓ_0 gradient sparsity term. Shih *et al.* [12] exploited the ghosting cues appeared in most of the double-panel window shooting images to remove the reflections from the input image. The work of Shih *et al.* is the most related one to the proposed method.

Recently, taking the advantages of deep learning, several approaches [28], [29], [30], [31] have been proposed for reflection removal using convolutional neural networks. Fan *et al.* [28] used a large number of synthetic images to train a deep neural network. This method performed well on synthetic data sets. However, it did not show noticeable performance gain on real images when compared to the traditional model-driven methods. Such a phenomena comes from the big gap between synthetic images and real images in terms of complex optical characteristics of reflection. To address such a weakness, Wan *et al.* [32], [29] established a dataset of real images with reflections, and combined it with synthetic images to train a multi-scale edge-guided concurrent network for reflection

removal. Using both real data sets and synthetic data sets, Zhang *et al.* [31] proposed a convolutional network that is optimized in terms of perceptual loss.

C. Reflection removal for a single image with ghosting effect

The previous results on removing reflection from a single image with the ghosting effect are the most related ones to this paper. Indeed, there are few results on single-image based reflection removal problem that are specifically designed for images with ghosting effects. The seminal work is done in [11] which formulated multiple reflections (*i.e.* the ghosting reflection) by the convolutional model in (1) with a multi-impulse convolution kernel. The clear image without reflection is then recovered by a deconvolution process. The advantage of convolution model for ghosting effect in (1), *i.e.*, $\tilde{\mathbf{R}} = \mathbf{R} \otimes \mathbf{k}$, is that it introduces the different characters of the transmission layer \mathbf{T} and the reflection layer $\tilde{\mathbf{R}}$ in the model of (3), owing to the convolution process existing in $\tilde{\mathbf{R}}$. The resulting difference helps resolving the ambiguities between \mathbf{T} and $\tilde{\mathbf{R}}$, making the separation plausible.

Following the model of [11], Shih *et al.* [12] proposed to solve the reflection removal problem by a two-stage approach. The proposed method first estimates the two parameters of kernel \mathbf{k} , *i.e.* the spatial shift \mathbf{d} and the attenuation c in (2), and then impose the mixture of Gaussian prior on each layer for layer separation. Fairly good results on both the real and synthetic images are obtained using such an approach. However, the method [12] is built on the patch-based Gaussian Mixture Model [33] learned from a large set of natural images.

Moreover, the ghosting effect of the reflection layer is implemented by imposing periodic prior on reflection layer, which also holds true for the repeating patterns in the transmission layer. As a result, the method often yields poor results when the transmission layer also contain repeating patterns, which is quite often in urban scenes. See Fig. 4 for an example of such a case. It can be seen that the method [12] wrongly treats the background objects with the period similar to the shift distance as the reflections and tries to remove such objects from the background, which results in poor separation of the two layers.

It is noted that the ghosting effect is also studied in the merging of multi-exposure images; see *e.g.* [34], [35], [36]. In such a case, the pixels of the different images are to be aligned. The motion of the camera or moving objects in the scene will cause ghosting artifacts if the motion is large. Then, a de-ghosting is needed for recovering images from such a degradation. Such problems have a different setting from ours in terms of available sources, and they are usually not formulated as a layer separation problem as ours.

III. PROPOSED METHOD

Throughout the paper, if there is no special illustration, bold upper letters are used for matrices, *e.g.* \mathbf{Y}, \mathbf{T} , bold lower letters for column vectors, *e.g.* \mathbf{y}, \mathbf{t} , light lower letters for scalars, *e.g.* y, t , hollow letters for sets *e.g.* \mathbb{R}, \mathbb{Z} , and calligraphic letters for operators, *e.g.* \mathcal{S}, \mathcal{T} . Let $\mathbf{0}$ and \mathbf{I} denote the zero matrix and the identity matrix with appropriate sizes

respectively. Let \mathbb{N} denote the set of natural numbers, \mathbb{Z}_M denote the set $1, \dots, M$, and \mathbb{R} denote the set of real numbers. Given a sequence $\{\mathbf{y}^{(t)}\}_{t \in \mathbb{N}}$, $\mathbf{y}^{(t)}$ denotes the t -th element in the sequence. For a vector $\mathbf{x} \in \mathbb{R}^N$, let $\mathbf{x}(i)$ denote the i -th element in \mathbf{x} , and define $\|\mathbf{x}\|_2 = \sqrt{\sum \mathbf{x}(i)^2}$. For a matrix \mathbf{X} , let $\mathbf{X}(i, j)$ denote the element of \mathbf{X} at the i -th row and j -th column, and define $\|\mathbf{X}\|_{\text{F}} = \sqrt{\sum_{i,j} (\mathbf{X}(i, j))^2}$ and $\|\mathbf{X}\|_1 = \sum_{i,j} |\mathbf{X}(i, j)|$. For matrix concatenation, semicolons are used for adding elements in columns and commas are used for adding elements in rows.

We used the following regularization-based optimization model for estimating both layers: \mathbf{R} and \mathbf{T} :

$$\min_{\mathbf{T}, \mathbf{R}} \frac{1}{2} \|\mathbf{Y} - \mathbf{T} - \mathbf{R} \otimes \mathbf{k}\|_{\text{F}}^2 + \lambda \Gamma_1(\mathbf{T}) + \beta \Gamma_2(\mathbf{R}), \quad (4)$$

where the first term denotes the fidelity term when one assumes *i.i.d.* Gaussian noise, the second term and the third term represent the regularizations on the two layers respectively. Such a model has the same form as that of [12] where the GMM prior is used. It is noted that the ghosting kernel \mathbf{k} is also unknown. Following the work [12], we take a two-stage approach to solve the above problem:

- First stage: estimate the ghosting kernel \mathbf{k} . In this stage, our method is mainly based on the techniques proposed in [11], [12], with some refinements.
- Second stage: estimate two layers \mathbf{R} and \mathbf{T} via solving (4). Regarding this stage, we mainly focus on how to construct the regularizations Γ_1, Γ_2 so as to well distinguish \mathbf{T} and \mathbf{R} . The key issue we aim to address is how to distinguish the repeating patterns on the transmission layer from those on the reflection layer, which is done based on the observation that the repeating times of such patterns are different between the two layers. We also exploit the difference of the sharpness between the two layers. With such cues, a weighted wavelet transform based regularization was developed for effectively separating the two layers. Compared to Shih *et al.*'s method [12], ours use a different way to exploit the prior of pattern recurrence to identify reflection layer and transmission layer. Besides, compared to the GMM prior used in [12], our used wavelet sparsity prior is less data-dependent and leads to an easier optimization.

A. Estimation of ghosting kernel

Recall that the ghosting kernel \mathbf{k} defined by (2) is a two-impulse kernel with two parameters: the amplitude c and displacement \mathbf{d} of the displaced impulse. We mainly use the techniques proposed in [12] to estimate such two parameters, with some refinements. In [12], the displacement \mathbf{d} is first estimated and then the amplitude c is evaluated. Followings are the details.

The displacement \mathbf{d} of the kernel \mathbf{k} is actually the displacement of the reflected contents, which can be estimated by checking the position difference of the edges of reflected contents. Given an input image $\mathbf{Y} \in \mathbb{R}^{M \times N}$, its edge map

is first obtained, which is denoted by $\mathbf{S} \in \mathbb{R}^{M \times N}$, using the Sobel edge detector as follows:

$$\mathbf{S}(i, j) = \sqrt{|(\mathbf{g}_x \otimes \mathbf{Y})(i, j)|^2 + |(\mathbf{g}_y \otimes \mathbf{Y})(i, j)|^2}, \quad (5)$$

where $\mathbf{g}_x = [-1, 0, 1; -2, 0, 2; -1, 0, 1]$ and $\mathbf{g}_y = [1, 2, 1; 0, 0, 0; -1, -2, -1]$. Then the autocorrelation map, denoted by $\mathbf{C} \in \mathbb{R}^{M \times N}$, is computed on the edge map \mathbf{S} to measure the self-similarity of \mathbf{S} with different displacements. The autocorrelation map is defined by

$$\mathbf{C} = \mathbf{S} \otimes \mathbf{S}(\cdot\cdot), \quad (6)$$

where $\mathbf{S}(\cdot\cdot)$ denotes a matrix created by reversing the indexes of \mathbf{S} .

Since \mathbf{C} tells how well an image correlates with itself under conditions where the image is displaced with respect to itself in all possible directions, a local maximum of \mathbf{C} would correspond to some repetitive patterns in \mathbf{S} , implying that it is likely to associate with the shifted copies of the reflection layer \mathbf{R} in the context of ghosting effect. Inspired by this property, we apply the max pooling to \mathbf{C} , *i.e.* to search on \mathbf{C} for all local maxima using a neighbor of size $r \times r$. Then for robust estimation, the local maxima that meet any of the following conditions are filtered out:

- The local maximum in \mathbf{C} corresponds to a strong edge in \mathbf{S} whose magnitude is larger than a threshold σ which is set to 0.2 in the implementation, since edges with magnitudes over 0.2 are often stable and statistically not dense in many images. This is for cancelling out the strong edges of repetitive patterns in the transmission layer.
- The local maximum is within t pixels from the origin, where $t = 5$ is used in the implementation for a moderate neighbourhood size. This is an often-used trick to remove the influence of wide edges, since a wide edge itself may contain several repetitive patterns.

Finally, among all the remaining local maxima, the largest one is chosen and its displacement to the origin is used as the ghosting displacement \mathbf{d} . There is still an ambiguity on the sign of \mathbf{d} , and we resolve this by choosing \mathbf{d} such that $c < 1$, using the property that the second reflection has lower energy than the first one.

Based on the estimated displacement parameter \mathbf{d} , the amplitude parameter c is evaluated as follows. Firstly, a set of interest points are detected from the input using a Harris corner detector. Secondly, a 5×5 contrast normalized patch is extracted around each corner feature. Lastly, the variances of each pair of matched patches, denoted by $\text{var}[\mathbf{p}_i]$ and $\text{var}[\mathbf{p}_j]$ respectively, are calculated, and the amplitude parameter c is computed as the weighted summation over the ratios of ($\text{var}[\mathbf{p}_i]$ and $\text{var}[\mathbf{p}_j]$) as follows:

$$c = \frac{1}{Z} \sum_{ij} w_{ij} \sqrt{\frac{\text{var}[\mathbf{p}_i]}{\text{var}[\mathbf{p}_j]}}, \quad (7)$$

where Z is the normalization factor, and w_{ij} is the weight for evaluating the contribution of each pair of patches to estimate the attenuation c . In the implementation, w_{ij} is experientially set by $w_{ij} = \exp(-\|\mathbf{p}_i - \mathbf{p}_j\|_2^2 / 0.1)$.

It is noted that the used kernel estimation scheme above can be viewed as a refined version of that of [11], [12] in terms of simplicity and performance gain. The main refinements over that of [11], [12] are listed as follows.

- The Sobel edge operator instead of Laplacian operator is used to get the edge map. Compared to the Laplacian operator, the Sobel operator can detect edges with stronger strength and is more robust to noise than the Laplacian operator.
- A thresholding strategy is implemented to filter out the strong edges that are likely to belong to background objects.
- Those local maximum points, *i.e.* the points that the first and second maxima in their neighbourhood are closer than a pre-defined threshold, are not discarded as the existing ones. As it complicates the process while no performance gain is observed by doing it.

B. Design of regularization model

In the first stage, we estimate the ghosting kernel \mathbf{k} . Then, in the second stage, we estimate two layers via solving the optimization model (4). The design of the model (4) is about how to design effective regularizations for the two layers such that they can be accurately separated. There are two components in the regularization: one is for wavelet transform based regularization for image recovery, and the other is for the separation of two layers by indicating the likeliness of each edge belongs to which layer. More specifically, let $\{\mathbf{f}_i\}_{i=1}^L$ denote the filter bank of a wavelet transform, and let \mathbf{A} denote the weighting matrix which is very related to the ownership of the edges to which layer. Then, we propose to separate the transmission layer \mathbf{T} and reflection layer \mathbf{R} from the input \mathbf{Y} by considering the following model:

$$\min_{\mathbf{T}, \mathbf{R}} \frac{1}{2} \|\mathbf{Y} - \mathbf{T} - \mathbf{R} \otimes \mathbf{k}\|_F^2 + \lambda \sum_i \|\mathbf{A} \odot (\mathbf{f}_i \otimes \mathbf{T})\|_1 + \beta \sum_i \|(1 - \mathbf{A}) \odot (\mathbf{f}_i \otimes \mathbf{R})\|_1, \quad (8)$$

where \odot denotes the element-wise multiplication. In the following, we will give a detailed discussion on the model above.

C. Wavelet transform based regularization

It can be seen that there are two operations involved in the ℓ_1 -norm related regularization for each layer. The coefficients $\{\mathbf{f}_i \otimes \mathbf{T}\}$ represent the high-pass wavelet coefficient of the layer \mathbf{T} , and the coefficients $\{\mathbf{f}_i \otimes \mathbf{R}\}$ represent the high-pass wavelet coefficient of the layer \mathbf{R} . Wavelet transform [37] based regularizations have been widely used in image recovery for suppressing undesired artifacts of the results; see *e.g.* [38], [39]. The regularization in wavelet transform is based on the fact that wavelet coefficient of a natural image is sparse, which can be exploited by regularizing the estimation of images using the ℓ_1 norm of its high-pass wavelet coefficients.

A wavelet transform is implemented by convolving input image with a set of wavelet filters. More specifically, for a

filter $\mathbf{f} \in \ell_2(\mathbb{Z})$, the convolution $S_{\mathbf{f}} : \ell_2(\mathbb{Z}) \rightarrow \ell_2(\mathbb{Z})$ can be defined by $[S_{\mathbf{f}}\mathbf{v}](n) = [\mathbf{f} * \mathbf{v}](n) = \sum_{k \in \mathbb{Z}} \mathbf{f}(n-k)\mathbf{v}(k), \forall \mathbf{v} \in \ell_2(\mathbb{Z})$. For a set of wavelet filters $\{\mathbf{f}_i\}_{i=1}^L \subset \ell_2(\mathbb{Z})$, the corresponding analysis operator of wavelet transform can be expressed in matrix form:

$$\mathbf{W} = [S_{\mathbf{f}_1}^\top(-\cdot), S_{\mathbf{f}_2}^\top(-\cdot), \dots, S_{\mathbf{f}_L}^\top(-\cdot)]^\top, \quad (9)$$

and the corresponding synthesis operator is \mathbf{W}^\top . A wavelet transform has the so-called perfect reconstruction property: $\mathbf{W}^\top \mathbf{W} = \mathbf{I}$, which avoids calculating $\mathbf{W}^\top \mathbf{W}$ in our numerical scheme. Also, the filter bank associated with wavelet transform contains the filters that correspond to the difference operators with different orders at different orientations, which is very helpful to suppressing artifacts. In the implementation, the linear B-spline wavelet transform [40] is implemented, which contains totally nine two-dimensional filters composed by the tensor product of the following three one-dimensional filters:

$$\mathbf{f}_1 = \frac{1}{4}(1, 2, 1)^\top; \mathbf{f}_2 = \frac{\sqrt{2}}{4}(1, 0, -1)^\top; \mathbf{f}_3 = \frac{1}{4}(-1, 2, -1)^\top.$$

The wavelet transform based regularization is effective on suppressing the undesired artifacts along image edges when recovering two layers. Such artifacts are often local variations or oscillations which are not sparse in the wavelet domain. However, only using such a regularization does not provide sufficient discrimination capability to accurately separate two layers. The discrimination information of two layers are indeed encoded in the weighting matrix \mathbf{A} . Recall that high-pass wavelet coefficients are very related to the image gradients with different orientations and different orders. The role of the weighting matrix \mathbf{A} is then for the assignment of image gradients to which layer, with a particular focus on distinguishing the difference in terms of edge periodicity.

D. Construction of weighting matrix

One main challenge when attempting to separate the transmission layer and the reflection layer is that the contents in the transmission layer with strong globally repetitive patterns can also be well modeled by the convolution process that is used in the reflection layer. As a result, the repetitive patterns in the transmission layer might be wrongly assigned to the reflection layer, and vice versa. The role of the weighting matrix \mathbf{A} in the model (8) is to resolve such ambiguities. The value of each entry of the matrix \mathbf{A} falls into $[0, 1]$, which can be viewed as the likeliness that a point contains the content of the reflection layer. In other words, a larger value is assigned to $\mathbf{A}(i, j)$ if an edge of the reflection layer is more likely on the point (i, j) . It corresponds to a larger penalty on the image gradients of T at the point (i, j) such that there is no image gradients along this point for T , which thus make the image gradient to be biased to the reflection layer.

We propose the following scheme to define \mathbf{A} :

$$\mathbf{A} = \frac{\mathbf{A}_1 + \mathbf{A}_2}{2}, \quad (10)$$

where $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{M \times N}$ are constructed with the following two assumptions respectively:

- *Difference in sharpness.* The background objects in the transmission layer are often sharper and much more clear than that in the reflection layer. This is also the assumption used in kernel estimation for filtering strong edges in the transmission layer. It is noted that it plays different roles in different stages.

- *Difference in periods.* The repeating patterns of the objects in the transmission layer often repeat several times within a local region (*i.e.* the periods of such patterns are larger than one), while the objects in the reflection layer only repeat once in the image for ghosting effect. Such an observation is particularly useful for the weak edges in the image which cannot be correctly identified by the first assumption.

Regarding the assumption of sharpness, we consider the magnitude of the edge map \mathbf{S} defined in (5) to select the background edges. First, we use K -means to cluster on the values of \mathbf{S} into three clusters. Such a single-dimensional clustering form four intervals: $\{[a_0, a_1], [a_1, a_2], [a_2, a_3], [a_3, a_4]\}$ where a_0, \dots, a_4 are in the decreasing order. The first interval $[a_0, a_1]$ correspond to the sharpest edges, while the last interval $[a_3, a_4]$ correspond to the flat or nearly-flat regions. In Fig. 7 (a) we show an example on the clustering results. It can be seen that the first interval basically contains all the strong edges in transmission layer. Thus, we binarize the edge map according to the largest cluster and define $\mathbf{A}_1(i, j)$ as follows:

$$\mathbf{A}_1(i, j) = \begin{cases} 1, & \mathbf{S}(i, j) < a_1 \\ 0, & \mathbf{S}(i, j) \geq a_1 \end{cases}. \quad (11)$$

In other words, \mathbf{A}_1 assigns an edge point to the reflection layer if its associated edge is relatively weak. Note that since the values of \mathbf{S} may be distributed in several clusters, using two clusters is not optimal. We experiencedly found that 4 intervals with 3 clusters in our scheme can lead to satisfactory results.

Regarding the assumption of periodicity difference, we show some images in Fig. 5 to demonstrate its rationality. It can be seen that the repetitive patterns in transmission layer often appear more than twice while the repetitive patterns occur in reflection layer usually only twice. Based on this assumption, we propose to utilize the periods of the repetitive patterns in the image for calculating \mathbf{A}_2 as follows. Firstly, the edge map $\mathbf{M} \in \mathbb{R}^{M \times N}$ of the input image $\mathbf{Y} \in \mathbb{R}^{M \times N}$ is calculated by

$$\mathbf{M} = \mathbf{g} \otimes (\mathcal{C} \circ \mathbf{Y}), \quad (12)$$

where \mathcal{C} is the Canny edge detector¹ and \mathbf{g} is a Gaussian smoothing kernel with the standard deviation 0.5 which is a default value in many image processing methods. Secondly, for each position $\mathbf{x} \in \mathbb{Z}_M \times \mathbb{Z}_N$ in \mathbf{M} , we extract the patch centered at \mathbf{x} with size $p \times p$, which is denoted by $\mathbf{P}(\mathbf{x})$. The correlation coefficients between $\mathbf{P}(\mathbf{x})$ and $\mathbf{P}(\mathbf{x} + \mathbf{d}), \mathbf{P}(\mathbf{x} - \mathbf{d}), \mathbf{P}(\mathbf{x} + 2\mathbf{d}), \mathbf{P}(\mathbf{x} - 2\mathbf{d})$ are calculated respectively where \mathbf{d} is the displacement parameter of the ghosting kernel that has been estimated, and the first two maxima among these four coefficients are stored as $\mathbf{Q}_1(\mathbf{x}), \mathbf{Q}_2(\mathbf{x})$ respectively. As

¹We use the Canny detector instead of the previously-used Sobel detector for obtaining additional information.

a result, we obtain two maximal correlation maps $\mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{R}^{M \times N}$. We denote $\mathbf{Q} = |\mathbf{Q}_1 - \mathbf{Q}_2|$, then we can define

$$\mathbf{A}_2(i, j) = \begin{cases} 1, & \mathbf{Q}(i, j) > \alpha \text{ \& } M(i, j) \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where the constant α is a predefined threshold (0.5 in the implementation). In other words, \mathbf{A}_2 tells whether an edge point on a weak edge belongs to the reflection layer based on the periodicity of its associated repetitive patterns. If an edge point on a weak edge has repetitive patterns that repeat only once around its neighborhood, then it is likely to be classified into the reflection layer.

See Fig. 6 for an example of the correlation coefficients. It can be seen that for an edge pixel on the reflection layer (e.g. for the center pixel in the third blue rectangle), its four correlation coefficients have one large value and three small values, and thus the difference between \mathbf{Q}_1 and \mathbf{Q}_2 would be large. In contrast, for an edge pixel on the transmission layer (e.g. the center pixels in the third yellow, red, magenta rectangles), there are at least 2 large values or all 4 small values in the four correlation coefficients, and thus the difference between \mathbf{Q}_1 and \mathbf{Q}_2 would be small. Thus, the reflections from background can be recognized according to the value of the difference of matrix \mathbf{Q}_1 and \mathbf{Q}_2 . See Fig. 7 (b) for the visualization of the difference map \mathbf{Q} .



Fig. 5: Periodic patterns of real objects and reflected contents.

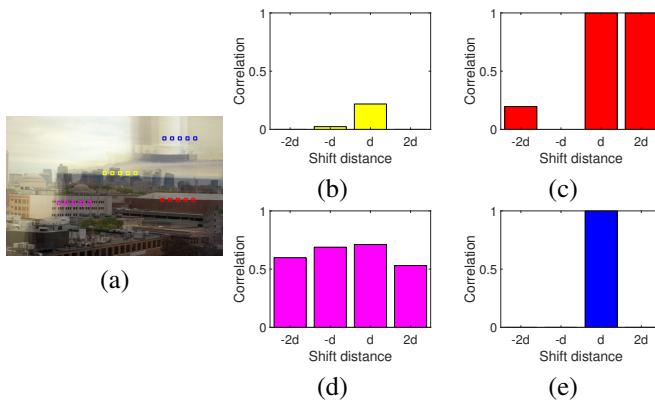


Fig. 6: Local correlation coefficients of a synthetic image. (a) Image with ghosting reflections. (b)-(e) Correlation coefficients between the central patch and its neighboring four patches with displacement $-2d, -d, d, 2d$ respectively. The bar in each sub-figure corresponds to the patches with the same color in (a).

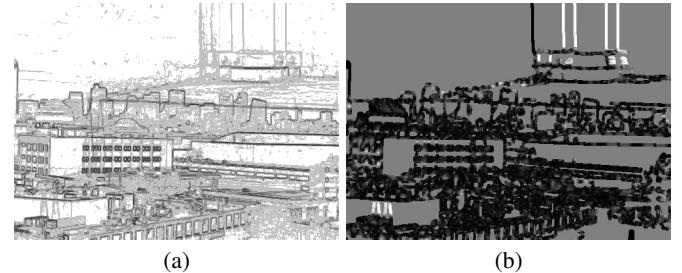


Fig. 7: Illustration of intermediate results for calculating \mathbf{A}_1 and \mathbf{A}_2 . (a) Clustered intervals on the Sobel edge map for generating \mathbf{A}_1 . (b) The map \mathbf{Q} computed on the sorted maximal correlation coefficients for generating \mathbf{A}_2 .

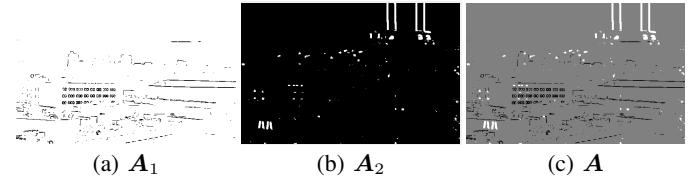


Fig. 8: Illustration of matrices \mathbf{A}_1 , \mathbf{A}_2 and \mathbf{A} . (a) Matrix \mathbf{A}_1 . (b) Matrix \mathbf{A}_2 . (c) Spatial weight matrix \mathbf{A} .

Once we calculate \mathbf{A}_1 and \mathbf{A}_2 , we obtain the weighting matrix \mathbf{A} using (10). An example of the calculated matrices $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}$ is shown in Fig. 8. It can be seen that from \mathbf{A}_1 the reflection can be located by the relatively large values, and \mathbf{A}_2 tells the locations of edges in the reflection layer. Then the weight matrix \mathbf{A} combines $\mathbf{A}_1, \mathbf{A}_2$ for better separating the reflection and transmission layers.

Now, we have all details of the proposed method. See Algorithm 1 for the outline of the method. The only remains is how to solve the minimization of (4).

E. Numerical solver

In this section, we present a numerical algorithm for solving model (8). For the convenience of presentation, we represent the related variables by the vectorization form. Let $\mathbf{y}, \mathbf{t}, \mathbf{r}, \mathbf{a} \in \mathbb{R}^{MN \times 1}$ denote the vectorization of the input image \mathbf{Y} , the transmission layer \mathbf{T} , the reflection layer \mathbf{R} , and the weighting matrix \mathbf{A} respectively. The vectorization is done by sequentially concatenating the columns of the matrix. Let $\mathbf{K} \in \mathbb{R}^{MN \times MN}$ denote the convolution matrix for the two-dimensional kernel \mathbf{k} , that is, the $\mathbf{K}\mathbf{r}$ is equivalent to the vectorization of $\mathbf{k} \otimes \mathbf{R}$. Let $\mathcal{D}(\mathbf{x})$ denote the diagonal matrix with \mathbf{x} as its diagonal elements. Let \mathbf{W} denote the analysis operator of the two-dimensional B-spline wavelet frame. Using these notations, the model (4) can be rewritten as

$$\min_{\mathbf{t}, \mathbf{r}} \left\| \begin{bmatrix} \lambda \cdot \mathcal{D}(\mathbf{a})\mathbf{W} & \mathbf{0} \\ \mathbf{0} & \beta \cdot (\mathbf{I} - \mathcal{D}(\mathbf{a}))\mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{r} \end{bmatrix} \right\|_1 + \frac{1}{2} \|\mathbf{y} - [\mathbf{I} \quad \mathbf{K}] \begin{bmatrix} \mathbf{t} \\ \mathbf{r} \end{bmatrix}\|_F^2. \quad (14)$$

The model above can be further simplified as:

$$\min_{\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{u}\|_F^2 + \|\mathbf{D}\mathbf{u}\|_1, \quad (15)$$

where $\mathbf{u} = [\mathbf{t}^\top, \mathbf{r}^\top]^\top$, $\mathbf{D} = \begin{bmatrix} \mathcal{D}(\lambda\mathbf{a})\mathbf{W} & \mathbf{0} \\ \mathbf{0} & (\beta - \mathcal{D}(\beta\mathbf{a}))\mathbf{W} \end{bmatrix}$, and $\mathbf{H} = [\mathbf{I} \quad \mathbf{K}]$. The problem (15) is a standard ℓ_1 -minimization problem which can be efficiently solved by the split-Bregman method [41]. Please refer to Appendix A for the details of the numerical algorithm. A simple initialization of the algorithm is to set $\mathbf{u}^{(0)} = [\mathbf{t}^{(0)}, \mathbf{r}^{(0)}]$ with $\mathbf{t}^{(0)} = \mathbf{y}$ and $\mathbf{r}^{(0)} = \mathbf{0}$. To obtain a better initialization, we first set $\mathbf{a} = [0.5, \dots, 0.5]$ and run the algorithm for 5 iterations. Then the outputs are used as the initialization to restart the algorithm. The final output \mathbf{t} is reshaped into an image, with values projected to $[0, 255]$.

Algorithm 1 Removing reflection from single image with ghosting effect

INPUT: Image \mathbf{Y} with ghosting reflections

OUTPUT: Transmission layer \mathbf{T}

Main procedure:

- 1) Estimate the ghosting kernel k ;
- 2) Calculate the weighting matrix \mathbf{A} ;
- 3) Solve \mathbf{T} from (4).

IV. EXPERIMENTS

To demonstrate the effectiveness of the proposed method, we conducted several experiments on both the synthetic data sets and real images. For comparison, in addition to the method of Shih *et al.* [12] which is the most related work to ours, we also compared our method against the other state-of-the-art methods of Li *et al.* [26] and Nikolaos *et al.* [10]. The results of these methods for comparison are either directly cited from published paper or generated using the codes from the authors with suggested parameter setting.

Through all the experiments, the parameters of the proposed method are set as follows. The two parameters in the model (4) are set as $\lambda = 0.01$, $\beta = 0.01$, according to the simple parameter tuning on several images with the candidate values $\{0.001, 0.01, 0.1, 0.5, 1\}$. In kernel estimation, the neighbourhood size used in searching local maximums is set to $r = 5$. The patch size for computing local correlation coefficients in the calculation of weighting matrix \mathbf{A} is set to $p = 8$. The parameters in the split-Bregman are simply set as $\gamma_1 = 1$, $\gamma_2 = 1$.

A. Results on synthetic images

To verify the effectiveness of the proposed method, in Fig. 9 we show the results by the proposed method and the method [12] on a synthetic image. The synthetic image is synthesized as follows. The transmission layer is composed of a long rectangle and five short rectangles, which are the simulation of real objects, and the reflection layer contains the ghosting version of the circle, which is generated using model (1) with the convolution kernel parameters $\mathbf{d} = [25, 0]$, $c = 0.5$. Moreover, the Gaussian noise with standard deviation 0.001 is added. The periods of the short rectangles are the same as the shifted distance of ghosting circle, and the period of the long rectangle is less than half the length. The decomposition results by [12] are shown in Fig. 9 (b-c) and

the results of our method are shown in Fig. 9 (e)-(f). It can be seen that the results by [12] are not good, which mix up the reflections and the period patterns in transmission layer. In contrast, by utilizing the difference between reflections and period patterns in terms of periodicity, our method can identify the background objects which show similar shifting properties and thus achieved better results.

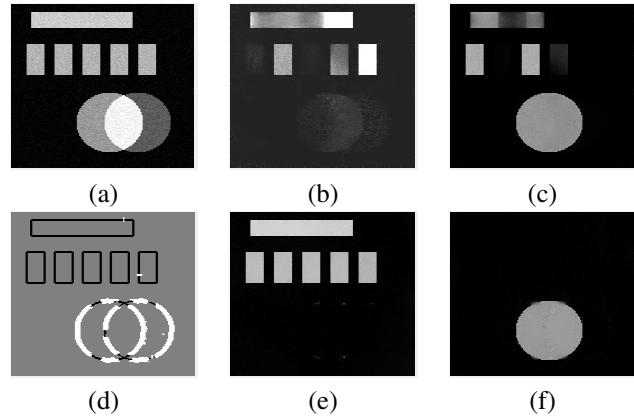


Fig. 9: Comparison of reflection removal results of [12] and the proposed method on a synthesized image. (a) Synthetic image which contains a long rectangle and five short ones in the transmission layer, as well as ghosting reflections of a circle in the reflection layer. (b) Transmission layer \mathbf{T} computed by [12]. (c) Reflection layer \mathbf{R} computed by [12]. (d) Spatial weight matrix \mathbf{A} computed by the proposed scheme. (e) Transmission layer \mathbf{T} computed by our method. (f) Reflection layer \mathbf{R} computed by our method.

In order to quantitatively evaluate the performance of the proposed method, we synthesized 143 images with ghosting reflections using the model (1). In the synthesis, the brightness parameter c is randomly sampled from $[0.5, 1]$, and the displacement parameter \mathbf{d} is a random vector with $1 < \|(\mathbf{d})\|_2 \leq 20$. The transmission layer \mathbf{T} and reflection layer \mathbf{R} are randomly sampled from the BSDS-500 dataset [42]. See Fig. 10 for some examples of the synthetic images samples.

We applied the proposed reflection removal method to the synthesized images and compared the output transmission layers with the ground-truths in the synthesis in terms of both PSNR and SSIM. The results are listed in Table I. On average, our method achieved the PSNR of 20.43dB and SSIM [43] of 0.81, while Li *et al.*'s method [26] only achieved a PSNR of 14.44dB and SSIM of 0.57. The other two methods for comparison are better than Li *et al.*'s method [26] but still worse than ours. In Fig. 11, we show some recovered results. It can be seen that our method achieved the best results among all the compared methods in terms of both visual quality and quantity metrics. The images recovered by Li *et al.*'s method [26] are generally darker and thus have weakness in color fidelity, *e.g.* the airplane and the old man. Nikolaos *et al.*'s method [10] can preserve more details and achieve better color quality, but it failed in removing the ghosting reflections in the images, *e.g.* the tall building and the island. In contrast, both Shih *et al.*'s method [12] and the proposed one obtained



Fig. 10: Samples of synthetic images used for evaluation.

better results on the island, tall building, and shoes. Compared with Shih *et al.*'s method [12], the proposed method shows better performance in dealing with images with repetitive transmission objects, and thus obtains better results on the boundaries of objects. Note that the results of Shih *et al.*'s method show severe boundary condition artifacts in Fig. 11. However, even eliminating the image boundaries, our method still outperformed than Shih *et al.*'s method. To verify this, we cut 25 pixels off around the boundaries and recalculate the numerical results, in which the PSNR and SSIM of our method are 20.55dB and 0.8146, while those of Shih *et al.*'s method are 18.29dB and 0.7993.

TABLE I: PSNR and SSIM values by different methods on synthetic images.

Metric	Li <i>et al.</i> [26]	Nikolaos <i>et al.</i> [10]	Shih <i>et al.</i> [12]	Ours
PSNR (dB)	14.48	17.77	17.48	20.43
SSIM	0.5694	0.7862	0.7820	0.8098

B. Results on real images

The proposed method is also evaluated on real images. We cite the three images from [12] for the test. The visual results are shown in Fig. 12, 13, 14.

It can be seen that our method outperforms the other three compared methods in terms of the effectiveness of removing reflections and the visual quality of recovered images. In Fig. 12, the chimney at the top right is hardly removed by the methods of Li *et al.* [26] and Nikolaos *et al.* [10], while our method can well remove the chimney and yields consistent color at the upper right corner of the recovered image. Shih *et al.*'s method [12] can deal with the chimney well but not good at dealing with the color consistency. Moreover, it cannot distinguish the periodic patterns of reflections and real objects, and thus leave the long white roof and repetitive windows in the image. In contrast, benefiting from the proposed weighting strategy which can correctly identify such patterns as the contents of transmission layer, our method achieved better visual results on the long white roof and repetitive windows. A similar phenomenon can also be observed in Fig. 13. The repetitive windows recovered by the method of Shih *et al.* [12] are bad both on color consistency and structure, while our method can obtain better visual quality.

In Fig. 14, since there are a very few repetitive patterns, our method performed comparably with Shih *et al.*'s approach [12]

and performed much better than those of Li *et al.* [26] as well as Nikolaos *et al.* [10]. Li *et al.*'s method [26] generated darker images with noticeable artifacts. Nikolaos *et al.*'s method [10] can suppress weak reflections, but fails to tackle strong reflections.

C. Results on SIR2-Postcard dataset

We also tested our method on the SIR2-Postcard dataset [32] whose creators use postcards to compose 20 different controlled scenes. The images on this dataset are not synthetic but real images. From the dataset we used 20 groups of images which are taken under three controlled thickness of glass (3mm, 5mm, 10mm). Each group of images is composed of a triplet, which contains the mixture image, as well as the ground truth of background and reflection. The mixture image is captured through the thick glass, the ground truth of the reflection is captured with a sheet of black paper behind the glass, and the ground truth of the background is obtained by removing the glass. The numerical results are listed in Table II, where we show the average PSNR values on the dataset as well as on each glass thickness. It can be seen that on average our method obtained 0.11dB PSNR improvement over Shih *et al.*'s method. Some visual results are shown in Fig. 15-Fig. 17, where our method yielded visually better results. In particular, the test images in Fig. 15-Fig. 17 contain obvious repeating patterns in the reflected scene and the transmitted scenes. We can see that our method works well on distinguishing such repeating patterns on the two layers.

TABLE II: PSNR(dB) under different thickness of glass on the SIR2-Postcard dataset.

	3mm	5mm	10mm	Average
Shih <i>et al.</i> [12]	19.60	19.29	19.50	19.46
Ours	19.72	19.35	19.63	19.57

V. SUMMARY

In this paper, based on the convolutional composite model for ghosting reflection, we proposed a regularization model for separating the transmission layer and reflection layer. The key issue we addressed is how to distinguish the repeating patterns on the transmission layer from those on the reflection layer, which is done based on the observation that the repeating times of such patterns are different between the two layers. With this

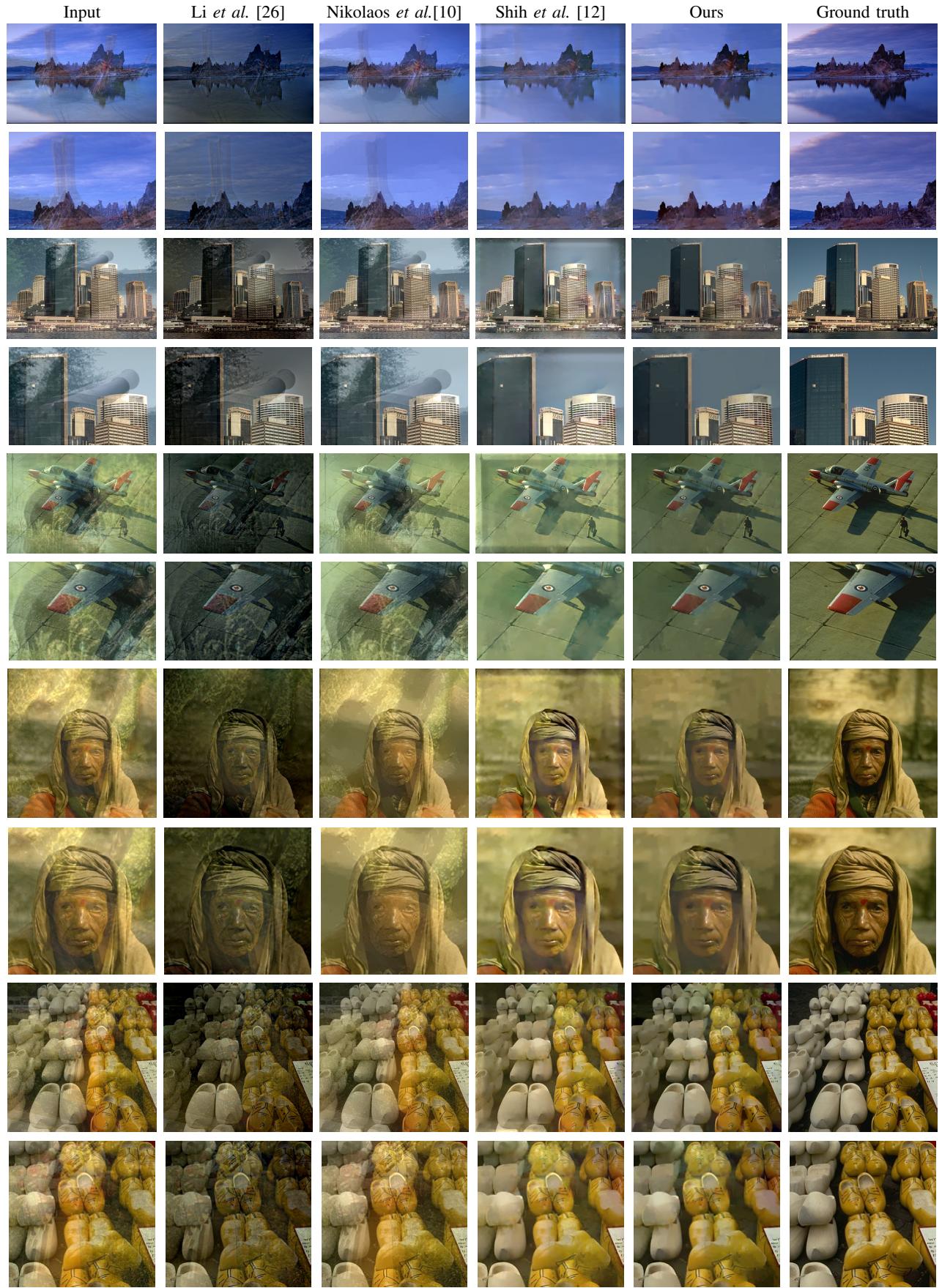


Fig. 11: Reflection removal results on synthetic images. The odd/even rows show the whole/zoomed-in results.

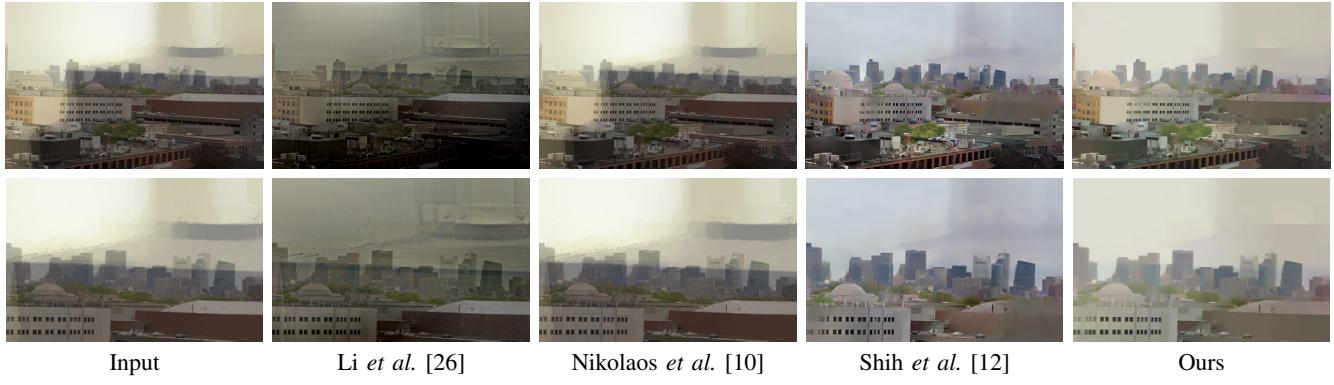


Fig. 12: Reflection results on image "Factory". The upper/bottom row shows the whole/zoomed-in results.

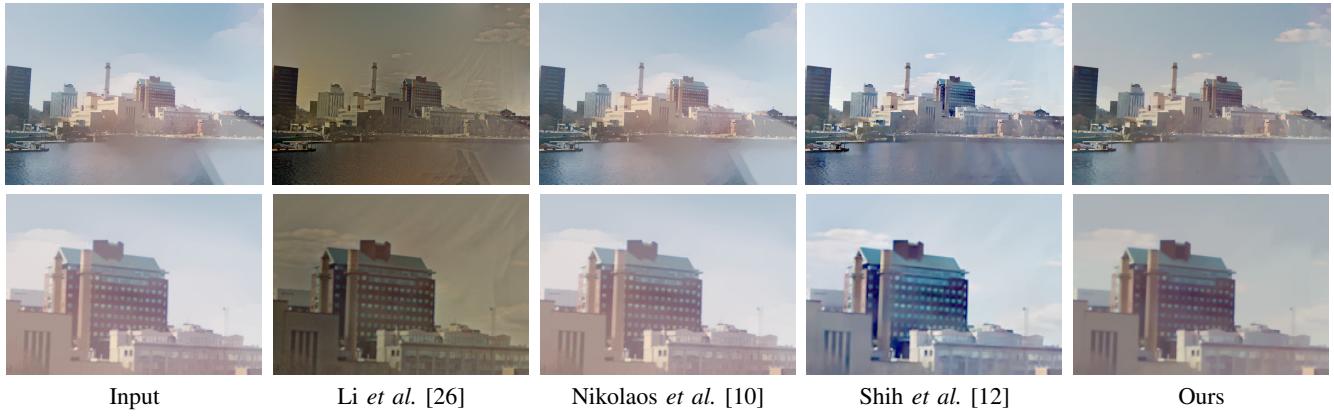


Fig. 13: Reflection results on image "Lake". The upper/bottom row shows the whole/zoomed-in results.

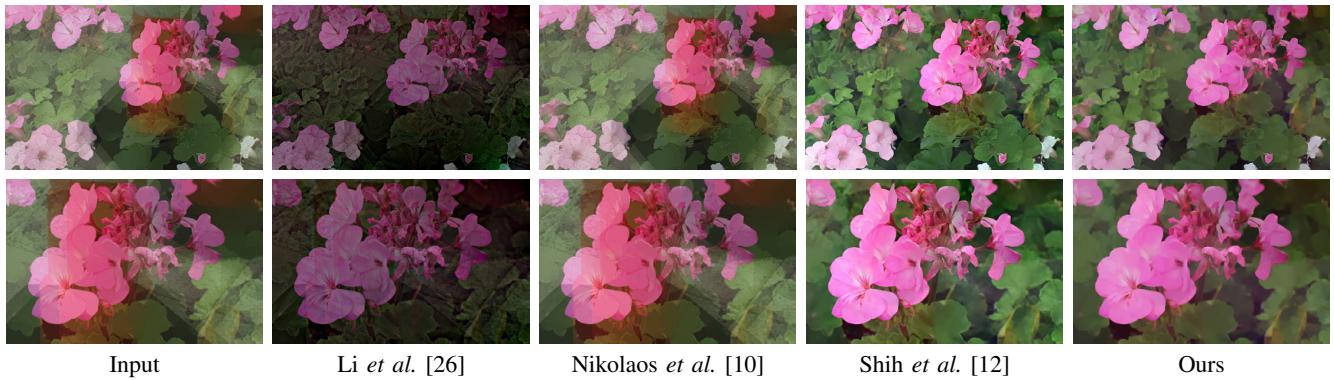


Fig. 14: Reflection results on image "Flower". The upper/bottom row shows the whole/zoomed-in results.

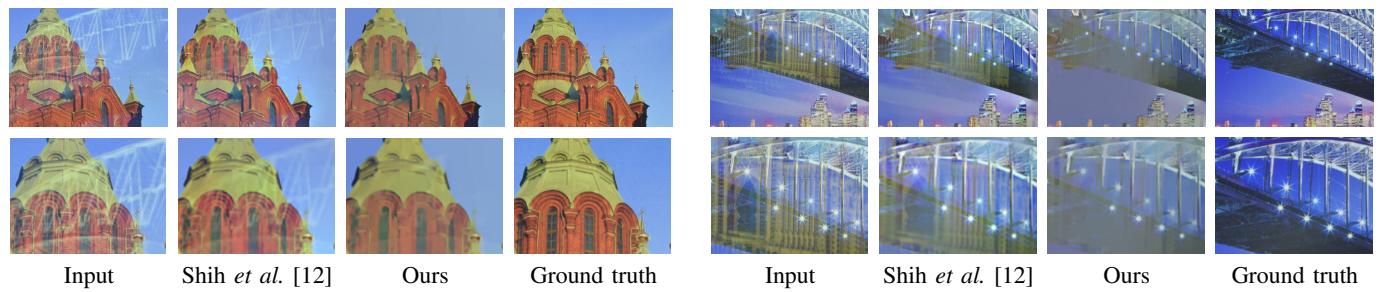


Fig. 15: Reflection results on image "Church". The upper/bottom row shows the whole/zoomed-in results.

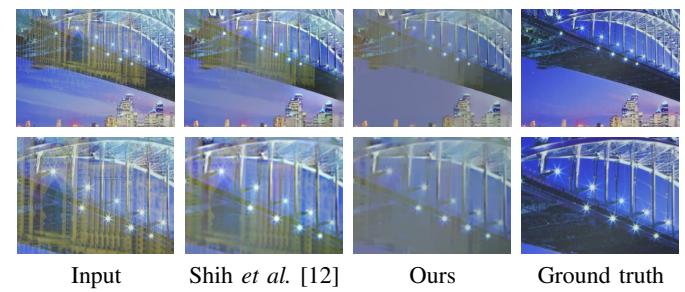


Fig. 16: Reflection results on image "Bridge". The upper/bottom row shows the whole/zoomed-in results.

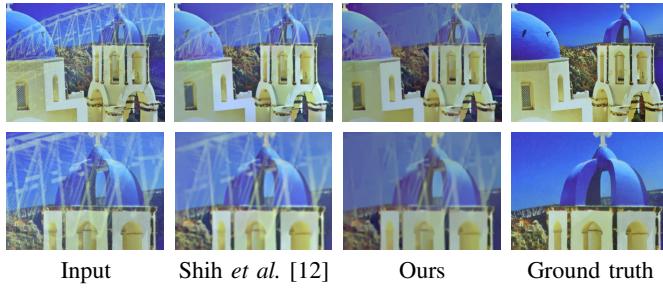


Fig. 17: Reflection results on image "House". The upper/bottom row shows the whole/zoomed-in results.

observation, a weighted wavelet transform based regularization was developed for effectively separating the two layers. Combining the regularization model with the refinement of existing ghosting kernel estimation techniques, we proposed an effective approach for removing reflections from a single image. In the experiments on both the synthetic data and real images, the proposed method showed noticeable improvement over the existing ones. In future, we would like to investigate iterative schemes for further improving the estimation of ghosting kernel and weighting matrix. Moreover, we would like to extend the proposed method to video reflection removal, by using additional motion cues to refine the model.

APPENDIX A

THE DETAIL OF NUMERICAL OPTIMIZATION SCHEME

The split-Bregman algorithm for solving (15) is as follows. First, the problem of (15) is rewritten as the constrained problem:

$$\min_{\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{u}\|_F^2 + \|\mathbf{d}\|_1, \quad \text{s.t. } \mathbf{d} = \mathbf{D}\mathbf{u}, \quad (16)$$

which is the "splitting" step to separate \mathbf{u} from the ℓ_1 penalty. Next, the Bregman iteration is used to solve (16) as follows:

$$\begin{cases} (\mathbf{u}^{(k+1)}, \mathbf{d}^{(k+1)}) = \operatorname{argmin}_{\mathbf{u}, \mathbf{d}} \frac{1}{2} \|\mathbf{H}\mathbf{u} - \mathbf{y}\|_2^2 + \|\mathbf{d}\|_1 + \frac{\gamma_1}{2} \|\mathbf{D}\mathbf{u} - \mathbf{d} + \mathbf{b}^{(k)}\|_2^2 \\ \mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \gamma_2(\mathbf{D}\mathbf{u}^{(k+1)} - \mathbf{d}^{(k+1)}) \end{cases} \quad (17)$$

for $k = 0, 1, \dots$, where $\gamma_1, \gamma_2 > 0$ are two parameters arising from the Bregman iteration. By decomposing its first subproblem into two subproblems, we further rewrite (17) as follows:

$$\begin{cases} \mathbf{u}^{(k+1)} = \operatorname{argmin}_{\mathbf{u}} \frac{1}{2} \|\mathbf{H}\mathbf{u} - \mathbf{y}\|_2^2 + \frac{\gamma_1}{2} \|\mathbf{D}\mathbf{u} - \mathbf{d}^{(k)} + \mathbf{b}^{(k)}\|_2^2 \\ \mathbf{d}^{(k+1)} = \operatorname{argmin}_{\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\gamma_1}{2} \|\mathbf{D}\mathbf{u}^{(k+1)} - \mathbf{d} + \mathbf{b}^{(k)}\|_2^2 \\ \mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \gamma_2(\mathbf{D}\mathbf{u}^{(k+1)} - \mathbf{d}^{(k+1)}) \end{cases} \quad (18)$$

In (18), the first subproblem is a quadratic problem with the analytic solution given by the least square:

$$\begin{aligned} \mathbf{u}^{(k+1)} &= (\mathbf{H}^\top \mathbf{H} + \gamma_1 \mathbf{D}^\top \mathbf{D})^{-1} \\ &\quad (\mathbf{D}^\top \mathbf{y} + \gamma_2 \mathbf{D}^\top (\mathbf{d}^{(k)} - \mathbf{b}^{(k)})), \end{aligned} \quad (19)$$

which is calculated by the conjugate gradient method in our implementation to avoid computing the inverse. The second

subproblem is separable to each dimension of \mathbf{d} and thus has the analytic solution given by

$$\mathbf{d}^{(k+1)} = \mathcal{S}_{\frac{1}{\gamma_1}}(\mathbf{D}\mathbf{u}^{(k+1)} + \mathbf{b}^{(k)}), \quad (20)$$

where $\mathcal{S}_\beta(\cdot)$ is the soft-thresholding operation defined by

$$\mathcal{S}_\beta(\mathbf{x}) = \operatorname{sgn}(\mathbf{x}) \max(|\mathbf{x}| - \beta, 0). \quad (21)$$

Combining (18), (19) and (20), the problem (15) is solved by the following iteration:

$$\begin{cases} \mathbf{u}^{(k+1)} = (\mathbf{H}^\top \mathbf{H} + \gamma_1 \mathbf{D}^\top \mathbf{D})^{-1} \\ \quad (\mathbf{H}^\top \mathbf{y} + \gamma_1 \mathbf{D}^\top (\mathbf{d}^{(k)} - \mathbf{b}^{(k)})) \\ \mathbf{d}^{(k+1)} = \mathcal{S}_{\frac{1}{\gamma_1}}(\mathbf{D}\mathbf{u}^{(k+1)} + \mathbf{b}^{(k)}) \\ \mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \gamma_2(\mathbf{D}\mathbf{u}^{(k+1)} - \mathbf{d}^{(k+1)}) \end{cases}.$$

ACKNOWLEDGEMENT

Yuhui Quan would like to thank the support by National Natural Science Foundation of China (Grant No. 61602184, 61872151), Natural Science Foundation of Guangdong Province (Grant No. 2017A030313376), Science and Technology Program of Guangzhou (Grant No. 201707010147), and Fundamental Research Funds for the Central Universities (x2js-D2181690). Yong Xu would like to thank the support by National Natural Science Foundation of China (Grant No. U1611461, 61672241, 61602184, 61528204) and Natural Science Foundation of Guangdong Province (Grant No. 2016A030308013). Hui Ji would acknowledge the support from Singapore MOE AcRF Research Grant R146000229114 and MOE2017-T2-2-156.

REFERENCES

- [1] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *IEEE Trans. Graphics*, vol. 24, no. 3, pp. 828–835, 2005.
- [2] A. Levin and Y. Weiss, "User assisted separation of reflections from a single image using a sparsity prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, 2007.
- [3] S. N. Sinha, J. Kopf, M. Goesele, D. Scharstein, and R. Szeliski, "Image-based rendering for scenes with reflections," *IEEE Trans. Graphics*, vol. 31, no. 4, pp. 100–1, 2012.
- [4] X. Guo, X. Cao, and Y. Ma, "Robust separation of reflection from multiple images," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2014, pp. 2187–2194.
- [5] L. Bedini, P. Savino, and A. Tonazzini, "Removing achromatic reflections from color images with application to artwork imaging," in *Image and Signal Processing and Analysis*. IEEE, 2015, pp. 126–130.
- [6] Y. Li and M. S. Brown, "Exploiting reflection change for automatic reflection removal," in *Proc. Int. Conf. Comput. Vision*. IEEE, 2013, pp. 2432–2439.
- [7] C. Sun, S. Liu, T. Yang, B. Zeng, Z. Wang, and G. Liu, "Automatic reflection removal using gradient intensity and motion cues," in *Proc. Multimedia Conf.*, 2016, pp. 466–470.
- [8] B.-J. Han and J.-Y. Sim, "Reflection removal using low-rank matrix completion," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2017.
- [9] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman, "A computational approach for obstruction-free photography," *IEEE Trans. Graphics*, vol. 34, no. 4, p. 79, 2015.
- [10] N. Arvanitopoulos, R. Achanta, and S. Süstrunk, "Single image reflection suppression," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2017, pp. 1752–1760.
- [11] Y. Diamant and Y. Y. Schechner, "Overcoming visual reverberations," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2008.

- [12] Y. Shih, D. Krishnan, F. Durand, and W. T. Freeman, "Reflection removal using ghosting cues," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2015, pp. 3193–3201.
- [13] A. Levin, A. Zomet, and Y. Weiss, "Separating reflections from a single image using local features," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, vol. 1. IEEE, 2004, pp. I–I.
- [14] C. Simon and I. Kyu Park, "Reflection removal for in-vehicle black box videos," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2015, pp. 4231–4239.
- [15] R. Wan, B. Shi, A.-H. Tan, and A. C. Kot, "Sparsity based reflection removal using external patch search," in *Proc. Int. Conf. Multimedia Expo*. IEEE, 2017, pp. 1500–1505.
- [16] H. Farid and E. H. Adelson, "Separating reflections and lighting using independent components analysis," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, vol. 1. IEEE, 1999, pp. 262–267.
- [17] R. Szeliski, S. Avidan, and P. Anandan, "Layer extraction from multiple images containing reflections and transparency," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, vol. 1. IEEE, 2000, pp. 246–253.
- [18] B. Sarel and M. Irani, "Separating transparent layers through layer information exchange," *Proc. European Conf. Comput. Vision*, pp. 328–341, 2004.
- [19] Y. Ni, J. Chen, and L.-P. Chau, "Reflection removal based on single light field capture," in *Proc. Int. Symp. on Circuits and Systems*. IEEE, 2017, pp. 1–4.
- [20] K. Gai, Z. Shi, and C. Zhang, "Blind separation of superimposed moving images using image statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 19–32, 2012.
- [21] J. Y. Cheong, C. Simon, C.-S. Kim, and I. K. Park, "Reflection removal under fast forward camera motion," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6061–6073, 2017.
- [22] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, 2011.
- [23] J. Yang, H. Li, Y. Dai, and R. T. Tan, "Robust optical flow estimation of double-layer images under transparency or reflection," in *Proc. Int. Conf. Comput. Vision*. IEEE, 2016, pp. 1410–1419.
- [24] A. Nandoriya, M. Elgharib, C. Kim, M. Hefeeda, and W. Matusik, "Video reflection removal through spatio-temporal optimization," in *Proc. Int. Conf. Comput. Vision*. IEEE, 2017, pp. 2430–2438.
- [25] A. Levin and Y. Weiss, "User assisted separation of reflections from a single image using a sparsity prior," in *Proc. European Conf. Comput. Vision*. Springer, 2004, pp. 602–613.
- [26] Y. Li and M. S. Brown, "Single image layer separation using relative smoothness," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2014, pp. 2752–2759.
- [27] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot, "Depth of field guided reflection removal," in *Proc. Int. Conf. Image Process*. IEEE, 2016, pp. 21–25.
- [28] Q. Fan, J. Yang, G. Hua, B. Chen, and D. P. Wipf, "A generic deep architecture for single image reflection removal and image smoothing," in *Proc. Int. Conf. Comput. Vision*. IEEE, 2017, pp. 3258–3267.
- [29] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, "Crnn: Multi-scale guided concurrent reflection removal network," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2018, pp. 4777–4785.
- [30] Z. Chi, X. Wu, X. Shu, and J. Gu, "Single image reflection removal using deep encoder-decoder network," *arXiv preprint arXiv:1802.00094*, 2018.
- [31] X. Zhang, R. Ng, and Q. Chen, "Single image reflection separation with perceptual losses," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, June 2018.
- [32] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, "Benchmarking single-image reflection removal algorithms," in *Proc. Int. Conf. Comput. Vision*. IEEE, 2017.
- [33] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vision*. IEEE, 2011, pp. 479–486.
- [34] K. K. Hadziabdic, J. H. Telalovic, and R. Mantiuk, "Comparison of deghosting algorithms for multi-exposure high dynamic range imaging," in *Proc. 29th Spring Conf. on Comput. Graphics*. ACM, 2013, pp. 21–28.
- [35] K. Karadzovic-Hadziabdic, J. H. Telalovic, and R. Mantiuk, "Subjective and objective evaluation of multi-exposure high dynamic range image deghosting methods," 2016.
- [36] W. Zhang and W.-K. Cham, "Gradient-directed composition of multi-exposure images," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2010, pp. 530–536.
- [37] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1999.
- [38] H. Ji, J. Li, Z. Shen, and K. Wang, "Image deconvolution using a characterization of sharp images in wavelet domain," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 295–304, 2012.
- [39] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," *IEEE Trans. Image Process*, vol. 21, no. 2, pp. 562–572, 2012.
- [40] I. Daubechies, B. Han, A. Ron, and Z. Shen, "Framelets: Mra-based constructions of wavelet frames," *Applied and computational harmonic analysis*, vol. 14, no. 1, pp. 1–46, 2003.
- [41] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM journal on imaging sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [42] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int. Conf. Comput. Vision*, vol. 2. IEEE, 2001, pp. 416–423.
- [43] 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, no. 4, pp. 600–612, 2004.

Yan Huang received the Ph.D. degree in Computer Science from South China University of Technology in 2018. She is currently the postdoctoral research fellow in Computer Science from South China University of Technology. Her research interests include computer vision, image processing and sparse representation.

Yuhui Quan received the Ph.D. degree in Computer Science from South China University of Technology in 2013. He worked as the postdoctoral research fellow in Mathematics at National University of Singapore from 2013 to 2016. He is currently the associate professor at School of Computer Science and Engineering in South China University of Technology. His research interests include computer vision, image processing and sparse representation.

Yong Xu received the B.S., M.S., and Ph.D. degrees in mathematics from Nanjing University, Nanjing, China, in 1993, 1996, and 1999, respectively. He was a Post-Doctoral Research Fellow of computer science with South China University of Technology, Guangzhou, China, from 1999 to 2001, where he became a Faculty Member and where he is currently a Professor with the School of Computer Science and Engineering. His current research interests include image analysis, video recognition, and image quality assessment. Dr. Xu is a member of the IEEE Computer Society and the ACM.

Ruotao Xu received the B.Eng degree in Computer Science from South China University of Technology in 2015. He is currently a PH.D candidate in South China University of Technology. His research interests include computer vision, image processing, and sparse coding.

Hui Ji received the B.Sc. degree in Mathematics from Nanjing University in China, the M.Sc. degree in Mathematics from National University of Singapore and the Ph.D. degree in Computer Science from the University of Maryland, College Park. In 2006, he joined National University of Singapore as an assistant professor in Mathematics. Currently, he is an associate professor in mathematics at National University of Singapore. His research interests include computational harmonic analysis, optimization, computational vision, image processing and biological imaging.