6-2018

# Region-aware reflection removal with unified content and gradient priors

Renjie WAN

Boxin SHI

Ling-Yu DUAN

Ah-hwee TAN
*Singapore Management University*, ahtan@smu.edu.sg

Wen GAO

*See next page for additional authors*



## Citation

WAN, Renjie; SHI, Boxin; DUAN, Ling-Yu; TAN, Ah-hwee; GAO, Wen; and KOT, Alex C.. Region-aware reflection removal with unified content and gradient priors. (2018). *IEEE Transactions on Image Processing*. 27, (6), 2927-2941.
**Available at:** https://ink.library.smu.edu.sg/sis_research/5194

# Region-Aware Reflection Removal With Unified Content and Gradient Priors

Renjie Wan, *Student Member, IEEE*, Boxin Shi, *Member, IEEE*, Ling-Yu Duan, *Member, IEEE*, Ah-Hwee Tan, *Senior Member, IEEE*, Wen Gao, *Fellow, IEEE*, and Alex C. Kot, *Fellow, IEEE*

***Abstract*— Removing the undesired reflections in images taken through the glass is of broad application to various image processing and computer vision tasks. Existing single image-based solutions heavily rely on scene priors such as separable sparse gradients caused by different levels of blur, and they are fragile when such priors are not observed. In this paper, we notice that strong reflections usually dominant a limited region in the whole image, and propose a region-aware reflection removal approach by automatically detecting and heterogeneously processing regions with and without reflections. We integrate content and gradient priors to jointly achieve missing contents restoration, as well as background and reflection separation, in a unified optimization framework. Extensive validation using 50 sets of real data shows that the proposed method outperforms state-of-the-art on both quantitative metrics and visual qualities.**

***Index Terms*— Reflection removal, internal patch recurrence, content prior, sparse representation.**

## I. Introduction

CAPTURING images through a transparent glass is unavoidable in many daily scenarios such as looking through a window or in front of a glass show case at the museum. Images taken under such circumstances usually have the objects of interests overlaid by the undesired reflections of the scene behind the camera. Feeding such images as input into general image processing and computer vision algorithms may result in degenerated output, since most algorithms only process images captured under reflection-free conditions with the desired background scene being clean and clear. Given such a mixture image, reflection removal aims at enhancing the clearity of the desired background scene while removing the undesired reflections. It has been shown to be an important pre-processing step for many image processing (*e.g.*, image stitching [1]) and computer vision (*e.g.*, image classification [2]) tasks to largely increase their accuracy. The reflection removal problem can be represented by the following equation:

$$\mathbf{I} = \mathbf{B} + \mathbf{R} + \mathbf{n}, \tag{1}$$

where $\mathbf{I}$ is the input mixture image, $\mathbf{B}$ is the background image to be clearly recovered, $\mathbf{R}$ is the reflection to be removed, and $\mathbf{n}$ is the additive noise. We show an example of $\mathbf{I}$, $\mathbf{B}$ and $\mathbf{R}$ in Figure 1.

This problem is challenging due to its obviously ill-posed nature — the number of unknowns is twice the number of equations. Besides, the structures and properties of reflections can be similar with that of background, which makes it difficult to simultaneously remove the reflections and restore the contents in the background. To reduce ill-posedness of the problem, reflection removal can be solved by using a sequence of images (*e.g.*, from different viewpoints [1], [3] or using images under various polarization status [4]), but such requests for special data capture setup or device largely limit their wide applicability, for instance to images downloaded from the Internet.

With image statistics and priors being taken into consideration, reflection removal can also be solved using only a single image. Motivated by the fact that natural image gradients have the heavy-tailed distribution, the gradient priors are widely used to fit the heavy-tailed distribution in solving different problems, *e.g.*, image deblurring [6] and single image reflection removal [5], [7], [8]. To apply such gradient priors, special properties have to be observed on one of $\mathbf{I}$, $\mathbf{B}$, and $\mathbf{R}$, or two or three of them simultaneously. For example, the edges from $\mathbf{B}$ and $\mathbf{R}$ need to be clearly distinguishable and could be manually annotated, then Laplacian mixtures are applied to model the distributions of $\mathbf{B}$ and $\mathbf{R}$ [7]; the image is assumed to be captured with $\mathbf{B}$ in focus and $\mathbf{R}$ out of focus, *i.e.*, the blur levels from $\mathbf{B}$ and $\mathbf{R}$ show obvious difference, then narrow Gaussian distributions can model their differences [5]. However, when the reflection dominates the contents, gradient priors can hardly be used and content priors are introduced to solve the problem through restoration of the missing contents instead of separating two images. For example, GMM patch prior can be adopted to learn the image prior from the external

Manuscript received July 4, 2017; revised November 23, 2017; accepted February 5, 2018. Date of publication February 22, 2018; date of current version March 26, 2018. This research was carried out at the Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore. The ROSE Lab is supported by the National Research Foundation, Singapore, and the Infocomm Media Development Authority, Singapore. This work was supported in part by the National Natural Science Foundation of China under Grant 61661146005, Grant U1611461, and Grant 61390515, in part by the National Key Research and Development Program of China under Grant 2016YFB1001501, and in part by the NTU-PKU Joint Research Institute through the Ng Teng Fong Charitable Foundation. The work of B. Shi was supported by the Recruitment Program of Global Experts (Youth Program) in China (a.k.a. 1000 Youth Talents). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ying Wu. *(Corresponding authors: Renjie Wan; Ling-Yu Duan.)*

R. Wan is with the Interdisciplinary Graduate School and the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: rwan001@e.ntu.edu.sg).

B. Shi, L.-Y. Duan, and W. Gao are with the Institute of Digital Media, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: lingyu@pku.edu.cn).

A.-H. Tan is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798.

A. C. Kot is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2808768

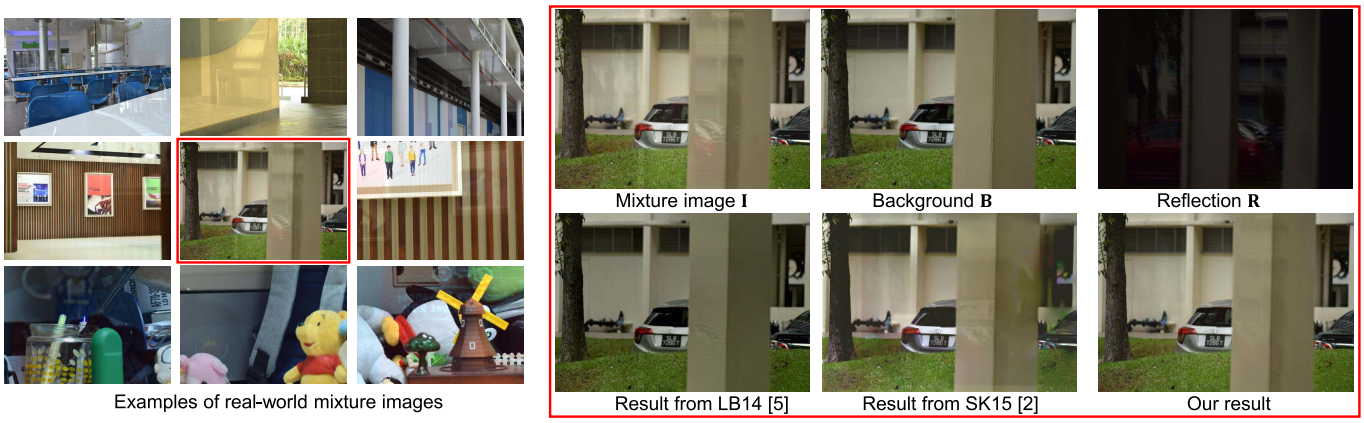



Fig. 1. Examples of real-world mixture images and reflection removal results using LB14 [5], SK15 [2], and our method.

sources [2], and such external patch recurrence can be applied to repair the mixture patch with several clean patches [9]. However, these methods are either based on the observations that reflections show ghosting effects observed through a piece of thick glass [2], or require reflection-free patches similar to the background image from an external database [9], which limit their practicability for a general scenario taken in the wild.

In many real-world mixture images taken in the real world, the reflections only occupy a part of the whole image plane like regional 'noise', as the examples shown in Figure 1. However, existing single image methods treat the whole mixture image in a global manner. Either the gradient prior based separation (*e.g.*, LB14 [5]) or the content prior based restoration (*e.g.*, SK15 [2]) shows artifacts in regions with weak reflections, as shown in the example of Figure 1. The result of LB14 [5] becomes globally darker than the ground truth background **B** and the result of SK15 [2] suffers from the patchy effect where the color becomes non-uniform; both methods are not able to effectively handle locally strong reflections, which results in residue edges on the pillar next to the car.

In this paper, we propose a Region-aware Reflection Removal (R3) approach to address these limitations. Given regions with and without strong reflections automatically detected, we apply customized strategies to handle them, so that the regional part focuses on removing the reflections with fewer artifacts and the global part keeps the consistency of the color and gradient information. We integrate both the content and gradient priors into a unified framework, with the content priors restoring the missing contents caused by the reflections (regional) and the gradient priors separating the two images (global). As an example, the result of our method shown in Figure 1 shows less reflection residues and more complete image content than previous methods.

The framework of our method is illustrated in Figure 2. Given the input mixture image **I**, we consider the reflection removal as image restoration with complementary priors to restore the missing contents, which is similar to [9] in the patch matching stage, but we utilize the internal patch recurrence from the input mixture image itself instead of relying on external database like [9], which extends the practicability of our method to more diverse scenes. In the removal stage, we model the gradient distributions of **B** and **R** with long- and short-tail distributions respectively, to avoid the direct dependency on commonly assumed image properties (*e.g.*, blur levels [5] or ghosting effect [2]) and hence better suppress artifacts by residual reflections. Our major contributions are summarized as follows:

- We build a R3 framework by automatically detecting regions with and without strong reflections and applying customized processing on different regions for more thorough reflection removal and more complete image content restoration;
- We develop a new content prior based on the internal patch recurrence to effectively restore missing contents covered by reflections;
- We integrate the content prior with newly designed gradient priors that distinctively model the distributions of reflection **R** and background **B** to achieve robust separation in a jointly optimized manner.

Our method is evaluated on a real world dataset of 50 scenes with the mixture images and ground truth background and shows superior performance both quantitatively and visually.

The remainder of this paper is organized as follows. Section II introduces relevant prior works. Section III and Section IV describe the proposed reflection removal method and its corresponding optimization solution, respectively. Experimental results and discussions are presented in Section V. Finally, we conclude the paper in Section VI.

## II. Related Work

In addition to reflection removal, our method is also related to patch based image restoration and sparse representation. We will briefly review relevant prior works about these three topics in the following paragraphs.

### A. Reflection Removal

Previous work on reflection removal can be roughly categorized as single image and multiple image based methods.

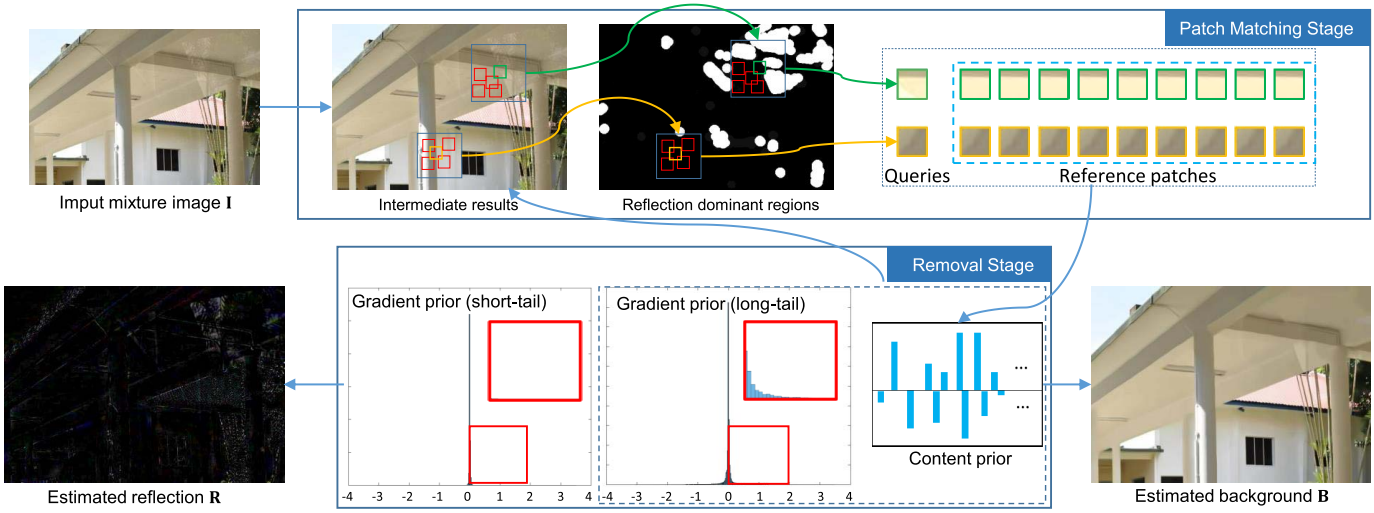



Fig. 2. The framework of our method. In the patch matching stage, we obtain reference patches from intermediate results of background in the detected reflection dominant regions using internal patch recurrence; then in the removal stage, the information from reference patches are used to refine the sparse codes of the query patches to generate the content prior. With the content prior and long-tail gradient prior, the background image **B** is recovered; based on the short-tail gradient prior, the reflection **R** is also estimated.

The single image approaches exploit the special properties of the background and reflection layers to deal with the highly ill-posed nature of this problem. Different priors are employed to make this problem more tractable: For example, Levin and Weiss [7] adopted the sparsity priors to decompose the input image. However, their method relied on the users to label the background and reflection edges, which is quite labor-intensive and may fail in textured regions. Li and Brown [5] made use of the different blur levels of the background and reflection layers. Recently, Shih *et al.* [2] adopted the GMM patch prior to model the ghosting effects of the reflections. Nikolaos *et al.* [10] adopted the Laplacian data fidelity term to solve this problem. However, the requirements for the properties of the background and reflection layers (different blur levels [5], [8], visible ghosting effects [2]) make them difficult to handle many general scenes when these properties are weakly observed.

Multiple image based methods remove the reflection by using a set of images taken from different viewpoints [1], [11]. By exploiting the motion cues between the background and reflection from multiview captures and assuming the glass is closer to the camera, the projected motion of the two layers is different due to the visual parallax. The motion of each layer can be represented by using parametric model, such as the translative motion [12], the affine transformation [11], and the homography [11]. In contrast to the fixed parametric motion, dense motion fields provide a more general modeling of layer motions represented by per-pixel motion vectors. Existing reflection removal methods estimate the dense motion fields for each layer using the optical flow [13], SIFT flow [3], [14], [15], the pixel-wise flow field [1]. Special capture conditions and camera settings, such as flash and no-flash image pair [16], different focuses [17], and light field cameras [18] also provide useful cues for reflection removal. Although the multiple image methods show more reliable results when input data are appropriately prepared, the requirement for special facilities or capturing limits such methods for practical use, especially for mobile devices or image downloaded from the Internet.

### *B. Patch Based Image Restoration*

Patch based image restoration has been used as a strong prior to solve a variety of ill-posed vision problems. It makes use of the patch recurrence property and can be roughly divided into two categories: internal methods and external methods. For internal methods, references patches are found from the noisy image itself; while for external methods reference patches are extracted from an external database of patches from clean images. Popular internal methods include the non-local means algorithm [19], BM3D [20], LPG-PCA [21] *etc.*, and external methods include the EPLL [22], external-BM3D [23], dictionary-based methods [24], and so on.

These two categories have their own strengths and limitations. Internal methods suffer from the rare-patch effect. Though this issue can be ameliorated in external methods by using a large database of patches, the external approaches are computationally expensive. Under high noise levels, internal approaches are generally more effective than the external approaches unless specific databased are employed [25].

### *C. Sparse Representation*

By assuming the signals or images only have a small number of non-zero entries w.r.t. some representation bases, the sparse representation has shown its effectiveness in many low-level image processing and high-level computer vision tasks, *e.g.*, image denoising [26], image deblurring [27], and anomaly detection [28]. The sparse representation is also widely used in the signal and image separation problems. Jafari *et al.* [29], [30] adopted the sparse dictionary for the

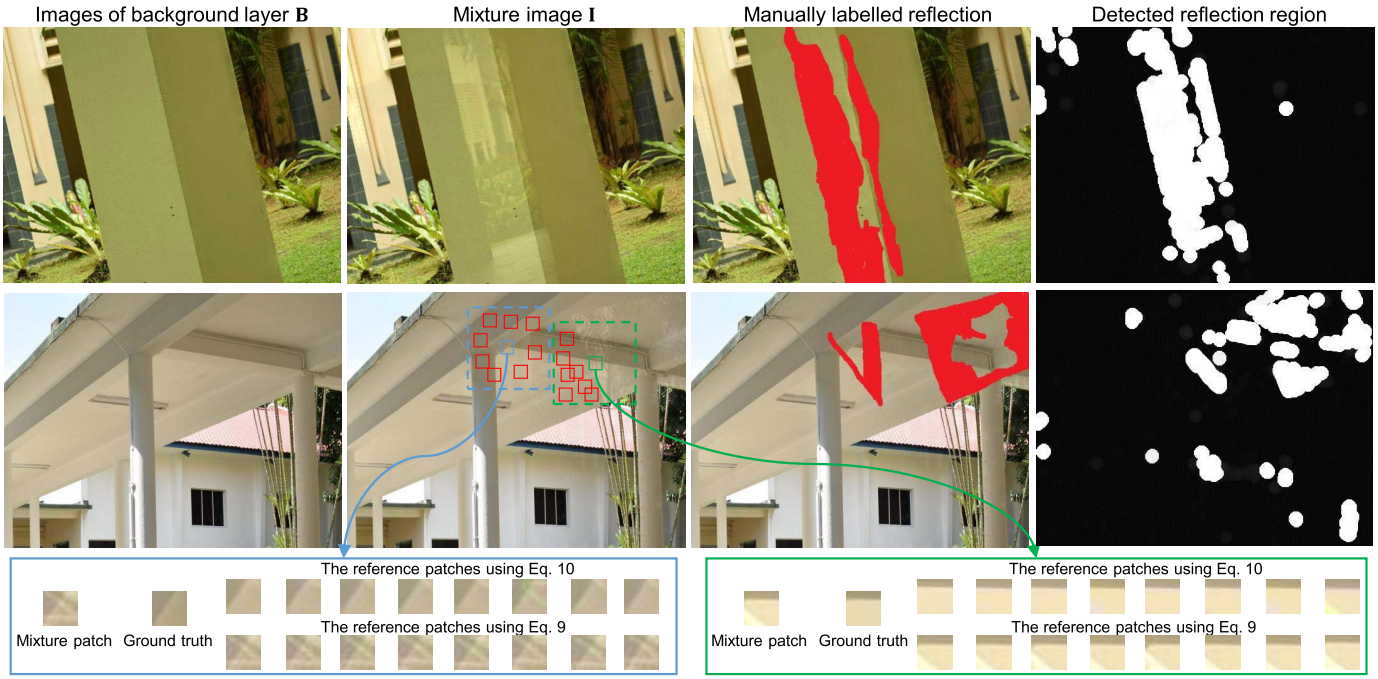



Fig. 3. Two examples of the detected reflection dominant regions (white pixels in the rightmost column) with their corresponding images of background, mixture images, and reference of reflections identified by humans (red pixels in the third column). At the bottom row, we show two examples of the patch matching results (the patch brightness $\times 2$ for better visualization). Given the mixture patch with reflections, we show its corresponding ground truth without reflections (extracted from $\mathbf{B}$) and the reference patches found using Equation (9) and Equation (10) (extracted from $\mathbf{I}$, with dashed box as the searching window), respectively. The reference patches found using Equation (10) are more similar to ground truth than the patches found using Equation (9).

separation of speech mixtures. Bobin *et al.* proposed the multichannel morphological component analysis (MMCA) [31] and generalized morphological component analysis (GMCA) [32] to separate the mixture images, by assuming that each source is sparse in its corresponding domain and can be modeled as the linear combination of a number of sparse morphological components. Such methods only perform well when prior knowledge about the sparse domain of each source is available. To address this limitation, Abolghasemi *et al.* [33] proposed a method to adaptively obtain the sparse domain from the mixture image using the K-SVD. We recommend the readers to check the survey in [34] for more methods and details.

## III. Proposed Method

We formulate the reflection removal as the maximum a posteriori (MAP) estimation problem, which is expressed using the Bayes' theorem as

$$\begin{aligned}\{\hat{\mathbf{B}}, \hat{\mathbf{R}}\} &= \underset{\mathbf{B},\mathbf{R}}{\operatorname{argmax}}\, f(\mathbf{B}, \mathbf{R}, \sigma^2|\mathbf{I}) \\ &= \underset{\mathbf{B},\mathbf{R}}{\operatorname{argmax}}\, f(\mathbf{I}|\mathbf{B}, \mathbf{R}, \sigma^2) f(\mathbf{B}) f(\mathbf{R}) \\ &= \underset{\mathbf{B},\mathbf{R}}{\operatorname{argmin}}\, L(\mathbf{I}|\mathbf{B}, \mathbf{R}, \sigma^2) + L(\mathbf{B}) + L(\mathbf{R}), \end{aligned} \tag{2}$$

where $f(\cdot)$ is the prior distribution and $L(\cdot) = -\log(f(\cdot))$. As commonly adopted by many reflection removal methods [5], [7], we assume the background and reflection distributions are independent, so we have $f(\mathbf{B}, \mathbf{R}) = f(\mathbf{B}) f(\mathbf{R})$. The noise term $\mathbf{n}$ in Equation (1) is assumed to follow $i.i.d.$ Gaussian distribution with the standard deviation as $\sigma$, then the likelihood model is represented as

$$L(\mathbf{I}|\mathbf{B}, \mathbf{R}, \sigma^2) = \frac{1}{2\sigma^2} \|\mathbf{I} - \mathbf{B} - \mathbf{R}\|_2^2 . \tag{3}$$

$L(\mathbf{B})$ is our unified prior which is formulated as

$$L(\mathbf{B}) = L_c(\mathbf{B}) + L_g(\nabla \mathbf{B}), \tag{4}$$

where $L_c(\mathbf{B})$ is the content prior and $L_g(\nabla \mathbf{B})$ is the gradient prior.

In the following, we will first introduce how we determine the regions with and without strong reflections, then we introduce the detailed formulation of content prior based on the region labels, and finally we introduce our gradient priors for background and reflection, respectively.

### A. Detecting Regions With and Without Reflections

As shown in Figure 1, in many real world scenarios, visually obvious reflections only dominate a part of the whole image plane, which we call reflection dominant region. Analogously, for other small regions showing less obvious or no visual artifacts caused by reflections, we call them reflection non-dominant regions. The reflection (non-)dominant regions can be automatically detected by checking the difference between the input mixture image and the results from single-image reflection removal algorithms [2], [5], [7], [8].

We borrow the idea in [8] which makes use of slightly different blur levels between the background and reflection due to the depth of field to differentiate the two types of regions. Similar to [8], we first calculate the KL divergence between the input mixture image and its blurred version to get a background map denoted as $E_{\mathbf{B}}$, which indicates the pixels

belonging to the background. Then, based on the fact that the reflections are generally with small image gradients [35], the initial reflection map $E'_{\mathbf{R}}$ is obtained by choosing the image gradients below a threshold (set as 0.3). Combining $E_{\mathbf{B}}$ obtained before, the refined reflection map $E_{\mathbf{R}}$ is obtained as

$$E_{\mathbf{R}} = \overline{E}_{\mathbf{B}} \odot E'_{\mathbf{R}}, \tag{5}$$

where $\overline{E}_{\mathbf{B}}$ denotes *not* operation over $E_{\mathbf{B}}$ and $\odot$ is the element-wise multiplication. Such an operation enhances $E_{\mathbf{R}}$ with many misclassified pixels in $E'_{\mathbf{R}}$ removed. Finally, we apply a dilation operation $S$ over $E_{\mathbf{R}}$ to further merge isolated pixels and regions in $E_{\mathbf{R}}$ as

$$\mathcal{D} = S(E_{\mathbf{R}}). \tag{6}$$

The dialtion operator $S(\cdot)$ we use is a non-flat ball-shaped structuring element with neighborhood and height values all set as 5. $\mathcal{D}(\cdot)$ is a binary matrix, whose element as 1 indicates reflection dominant regions and 0 indicates non-dominant regions.

We show examples of the reflection detection results calculated from Equation (5) and Equation (6) in the rightmost column of Figure 3. Comparing with the manually labelled reference and the mixture image, we observe that pixels with strong reflections and covering large areas are correctly detected as reflection dominant regions. Misclassified pixels covering some sparse regions show little influence to the next stage of operations. The detected reflection dominant regions will be used in two parts of the following processings: 1) the patch matching step for content prior which will be introduced in the next subsection and 2) the optimization stage which will be introduced later in Section IV.

### *B. Content Prior*

The proposed R3 solution utilizes the patch recurrence property within the input mixture image itself. Given $\mathbf{q}_i$, an image patch overlaid with reflections and centered at position $i$, the patch recurrence property aims at using the estimation of $\mathbf{q}_i$ with the $L$ nearest patches $\{\mathbf{p}_{i,l}\}_{l=1}^{L}$ from its surroundings to restore it. We assume that we have already obtained a set of reference patches $\{\mathbf{p}_{i,l}\}_{l=1}^{L}$ for now. Then the estimation of $\mathbf{q}_i$, denoted as $\mathbf{u}_i$, can be obtained as the weighted average of $\{\mathbf{p}_{i,l}\}_{l=1}^{L}$ as follows:

$$\mathbf{u}_i = \sum_{l=1}^{L} v_{i,l}\mathbf{p}_{i,l}. \tag{7}$$

Here, $v_{i,l}$ is the similarity weight expressed as $v_{i,l} = \exp(-\|\mathbf{p}_{i,l} - \mathbf{q}_i\|_2^2/2\sigma^2)/c$; $c$ is the normalization constant to guarantee $\sum_l v_{i,l} = 1$ and the parameter $\sigma$ controls the tolerance to noise due to illumination changes, compression, and so on.

We adopt the NCSR model [36] as the content prior and it can be formulated as follows:

$$L_c(\mathbf{B}) = \sum_i \|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\|_1, \quad s.t.\ \mathbf{M}_i\mathbf{B} = \mathbf{D}\boldsymbol{\alpha}_i, \tag{8}$$

where $\mathbf{M}_i$ is the matrix extracting an image patch of size $N \times N$ from the background image $\mathbf{B}$; $\mathbf{D}$ denotes the dictionary built from the mixture image $\mathbf{I}$ and $\boldsymbol{\alpha}_i$ is the sparse coefficients corresponding to $\mathbf{q}_i$. Then $\boldsymbol{\beta}_i$ is the nonlocal estimation of $\boldsymbol{\alpha}_i$ in the sparse domain. Equation (8) minimizes the difference between $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$, which means that the missing contents in the mixture patch $\mathbf{q}_i$ can be restored by its similar patch $\mathbf{u}_i$. Without losing generality, we choose the K-PCA dictionaries [36], [37] as $\mathbf{D}$. To be specific, the image patches are extracted from the input mixture image $\mathbf{I}$, and clustered into $K$ clusters using K-means. For each cluster, a dictionary of PCA bases is learned to encode the patches in this cluster. Due to the orthogonal property of the PCA bases, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ can be easily computed as $\boldsymbol{\alpha}_i = \mathbf{D}^\top\mathbf{q}_i$ and $\boldsymbol{\beta}_i = \mathbf{D}^\top\mathbf{u}_i$. Please refer to [36] and [37] for more details.

*Patch Matching:* Here, we explain how to obtain the reference patches $\{\mathbf{p}_{i,l}\}_{l=1}^{L}$ in Equation (7). If external images with similar contents to the ground truth of background $\mathbf{B}$ are available, patch matching can be accurately performed by searching the whole image and measuring the $l_2$ distance [9]. For each $\mathbf{q}_i$, its reference patches $\{\mathbf{p}_{i,l}\}_{l=1}^{L}$ are searched within $W_H(i)$, a window with size $H \times H$, using $l_2$ distance:

$$d(\mathbf{q}_i, \mathbf{p}_{i,l}) = \|\mathbf{q}_i - \mathbf{p}_{i,l}\|_2, \quad \forall l \in W_H(i). \tag{9}$$

Such a process is illustrated in Figure 2 and Figure 3.

Note that such an approach can provide quite clean patches only when the input mixture image contains some landmarks or objects that can be retrieved from an external database. To provide a more broaderly applicable solution, the patch matching should be performed within the input mixture image itself. However, we cannot directly apply the simple matching strategy in Equation (9), due to that 1) the mixture images include regions with strong reflections (while external patches are all clean) and 2) these strong reflections make the simply $l_2$ distance measuring rather unreliable. To address these two problems, we develop our patch matching solution as guided by the reflection (non-)dominant regions detected in Section III-A with a robust distant function:

$$d(\mathbf{q}_i, \mathbf{p}_{i,l}) = \rho_s(\mathbf{q}_i, \mathbf{p}_{i,l}) + \lambda\rho_r(\overline{\mathbf{q}}_i, \mathbf{p}_{i,l}), \quad \forall l \in W_H(i), \quad \sum \mathcal{D}(\mathbf{p}_{i,l}) < N/2 \tag{10}$$

Some reference patches found using Equation (10) may still contain reflections, which affect the accuracy of the patch matching and the subsequent reflection removal. To eliminate these negative effects and make sure that enough reference patches can be found, we add the constraint $\sum \mathcal{D}(\mathbf{p}_{i,l}) < N/2$ in Equation (10) to require that fewer than half of all pixels in a patch (note $N$ is total number of pixels of a patch) are labelled as reflection dominant, *i.e.*, we limit the searching of reference patches only within reflection non-dominant regions. $\overline{\mathbf{q}}_i$ denotes as a patch being processed by TV-decomposition and $\lambda$ is a balancing weight.

Taking the intrinsic image structure into consideration, we define the first robust distance term $\rho_s$ by making use of the image gradient information as a structure-aware criterion:

$$\rho_s(\mathbf{q}_i, \mathbf{p}_{i,l}) = \|\mathbf{q}_i - \mathbf{p}_{i,l}\|_2 + \eta\|\nabla\mathbf{q}_i - \nabla\mathbf{p}_{i,l}\|_2. \tag{11}$$

We then define the second robust distance term $\rho_r$ to specifically handle the patches in the reflection dominant regions.

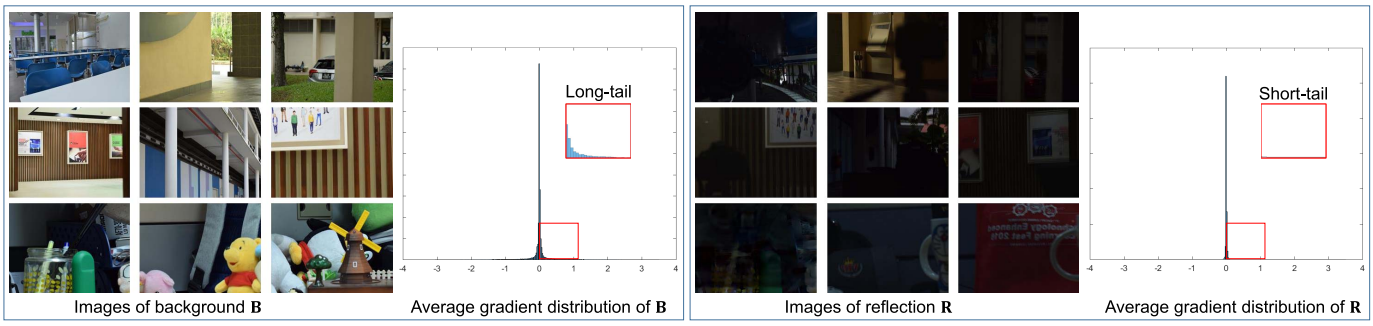



Fig. 4. Some sample images of the background **B** and reflection **R** and their corresponding long-tail and short-tail gradient distributions.

Due to the interference of the reflections, the candidate patches may not be truly relevant to the mixture patch. Considering the fact that the reflections are more related with the low-frequency component of images [38], we apply the TV-decomposition [39] to pre-process the input mixture image **I**, so that structures with large gradient values are retained and the low-frequency components are filtered out. $\rho_r$ is defined as

$$\rho_r(\overline{\mathbf{q}}_i, \mathbf{p}_{i,l}) = \|\overline{\mathbf{q}}_i - \mathbf{p}_{i,l}\|_2 + \eta\|\nabla\overline{\mathbf{q}}_{i,l} - \nabla\mathbf{p}_{i,l}\|_2. \quad (12)$$

Equation (10) is simply a linear combination of $\rho_s$ and $\rho_r$, which shows a balance between the original mixture patch $\mathbf{q}_i$ and the TV-decomposed $\overline{\mathbf{q}}_i$. In the reflection non-dominant regions, $\rho_s$ can easily find sufficient numbers of patches, thus $\lambda$ is given a smaller value to decrease the influence of $\rho_r$; in contrast, we need a larger $\lambda$ in the reflection dominant regions. Since the searching of reference patches is limited only within reflection non-dominant regions, we need a larger $H$ for patches from the reflection dominant regions to increase the searching window size for matching sufficient numbers of reference patches. Comparing with the vanilla solution using Equation (9), our region-aware robust strategy in Equation (10) could find reflection-free patches with much closer appearances to the ground truth, as shown by the two examples in the bottom row of Figure 3.

### C. Gradient Prior

The gradient priors play important roles in the reflection removal stage, as shown in Figure 2. A popular choice is fitting the heavy tailed gradient distribution such as the Laplacian mixtures [7] to both background and reflection. We find such a homogeneous processing cannot take advantages of our R3 framework. Since the regional reflections only cover a part of the whole image, its corresponding gradient distributions should be different from the distributions of the background image, due to its sparser property. Therefore, we design the gradient prior of **B** and **R** in a heterogeneous manner using different types of distributions.

*Assumption Verification:* To verify the above assumptions, we capture mixture images (through the glass) with ground truth background (by removing the glass) and reflection images (by putting a black sheet of paper behind the glass) to analyze the gradient properties. Following such a three-step procedure, we capture image triplet for 50 different scenes in the wild (150 images in total), with a DSLR camera under fully manual control.[1] All images are not corrected to the linear response. These scenes include substantial real-wrold objects of complex reflectance (car, tree leaves, glass windows, *etc.*), various distances and scales (residential halls, gardens, and lecture rooms, *etc.*), and different illuminations (direct sunlight, cloudy sky light, and twilight, *etc.*).

Nine (out of 50) sample scenes used in our analysis are shown in Figure 4 and the corresponding average gradient distributions (over 50 scenes) are plotted next to them. The plotted distributions clearly show that the background and reflection images belong to the long-tail and short-tail distribution, respectively. Similar heterogeneous distributions are reported in [5], but their observations are only applicable to images where the background is in focus and reflection is out of focus. Our analysis here shows such heterogeneous distributions also apply to images with the reflection being in focus. We adopt the prior proposed in [40], which regularizes the high frequency part by manually manipulating the image gradients, to fit our gradient distribution for background as

$$L_g(\nabla\mathbf{B}) = \sum_x \phi(\nabla\mathbf{B}(x)), \quad (13)$$

where

$$\phi(\nabla\mathbf{B}(x)) = \begin{cases} \frac{1}{\epsilon^2}|\nabla\mathbf{B}(x)|^2, & \text{if } |\nabla\mathbf{B}(x)| < \epsilon, \\ 1, & \text{otherwise}, \end{cases} \quad (14)$$

where $x$ is pixel locations. $L_g(\cdot)$ approximates $L_0$ norm by thresholding a quadratic penalty function parameterized by $\epsilon$ to avoid the distribution dropping too fast. Such a prior restores sharper edges belonging to the background image with less noise. Based on the proof in [40], Equation (14) is equivalent to

$$\phi(\nabla\mathbf{B}(x)) = \min_{\mathbf{l}_{mx}}\{|\mathbf{l}_{mx}|^0 + \frac{1}{\epsilon}(\nabla_m\mathbf{B}_x - \mathbf{l}_{mx})^2\}, \quad (15)$$

where $m \in \{h, v\}$ corresponding to the horizontal and vertical directions, respectively; $\mathbf{l}$ is an auxiliary variable and $x$ is the pixel position.

The gradient distribution of **R** belongs to short-tail distribution partly due to the higher blur levels of **R** [5]. However, as we show in Figure 4, the majority of regions in **R** have brightness values closing to zero, *i.e.*, its gradient distribution should also have the sparse property when compared with the background. Therefore, we model it using a $L_0$-regularized

[1] Similar data capture procedure has been conducted in [1] for a smaller scale quantitative evaluation.

prior as

$$L(\mathbf{R}) = \|\nabla \mathbf{R}\|_0, \tag{16}$$

where $\|\cdot\|_0$ counts the number of non-zero values in $\nabla\mathbf{R}$. Such a prior enforces the sparsity property of $\mathbf{R}$ in its gradient domain.

By substituting Equation (3), Equation (15), Equation (8), and Equation (16) into Equation (2), our complete energy function is represented as

$$\begin{aligned}\{\hat{\mathbf{B}}, \hat{\mathbf{R}}\} = \underset{\mathbf{B},\mathbf{R},\boldsymbol{\alpha}_i}{\operatorname{argmin}} & \|\mathbf{I}-\mathbf{B}-\mathbf{R}\|_2^2 + \omega \sum_i \|\mathbf{M}_i\mathbf{B} - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 \\ & + \xi \sum_i \|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\|_1 + \delta \sum_{m\in\{h,v\}} \|\nabla_m \mathbf{R}\|_0 \\ & + \gamma \sum_{m\in\{h,v\}} \sum_x \{|\mathbf{l}_{mx}|^0 + \frac{1}{\epsilon}(\nabla_m \mathbf{B}_x - \mathbf{l}_{mx})^2\}, \end{aligned} \tag{17}$$

where $i$ denotes the $i$-th patch or atoms, $x$ is the pixel position, and $\hat{\mathbf{B}}$, $\hat{\mathbf{R}}$ are the intermediate results of $\mathbf{B}$, $\mathbf{R}$ generated at each iteration. It will be optimized in the next subsection.

## IV. Optimization

The direct minimization of Equation (17) is difficult due to the multiple variables involved in different terms. Thus, we divide the original problem into several subproblems by following the half-quadratic splitting technique [41] advocated by the previous methods in image deblurring and denoising [26]. The proposed algorithm iteratively updates the variables, reduces the objective function values in each iteration, and finally converges to a local minima. We summarize each step of our method as Algorithm 1, and the details are described in the following paragraphs.

### A. Solving for $\boldsymbol{\alpha}_i$

Given fixed $\mathbf{B}$ and $\mathbf{R}$, Equation (17) becomes a $l_1$ minimization problem:

$$\hat{\boldsymbol{\alpha}_i} = \underset{\boldsymbol{\alpha}_i}{\operatorname{argmin}}\, \omega\|\mathbf{M}_i\mathbf{B} - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \xi \sum_i \|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\|_1. \tag{18}$$

With fixed $\boldsymbol{\beta}_i$, Equation (18) can be solved iteratively by the surrogate based algorithm [42]:

$$\boldsymbol{\alpha}_i{}^{(t+1)} = \mathcal{S}_\tau(v_i^{(t)} - \boldsymbol{\beta}_i) + \boldsymbol{\beta}_i, \tag{19}$$

where $v_i^{(t)} = \mathbf{D}^\top(\mathbf{M}_i\mathbf{B} - \mathbf{D}\boldsymbol{\alpha}_i{}^{(t)})/c + \boldsymbol{\alpha}_i{}^{(t)}$, $\mathcal{S}_\tau(\cdot)$ represents the soft-thresholding operator with threshold $\tau = \xi/\omega c$, and $c$ is a constant to guarantee the convexity. Equation (19) balances the influence of $\boldsymbol{\beta}_i$ to $\boldsymbol{\alpha}_i$, and a larger $\tau$ generally allows a quicker convergence. Due to the orthogonal properties of the local PCA dictionaries $\mathbf{D}$, the sparse coding problem of Equation (18) can be solved in just one step [43].

### B. Solving for $\mathbf{B}$

When $\mathbf{R}$ and $\boldsymbol{\alpha}_i$ are fixed, $\mathbf{B}$ can be estimated by solving the following optimization problem:

$$\begin{aligned}\hat{\mathbf{B}} = \underset{\mathbf{B}}{\operatorname{argmin}} & \|\mathbf{I}-\mathbf{B}-\mathbf{R}\|_2^2 + \omega \sum_i \|\mathbf{M}_i\mathbf{B} - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 \\ & + \gamma \sum_{m\in\{h,v\}} \sum_x \{|\mathbf{l}_{mx}|^0 + \frac{1}{\epsilon}(\nabla_m \mathbf{B}_x - \mathbf{l}_{mx})^2\}, \end{aligned} \tag{20}$$

**Algorithm 1** Region-Aware Reflection Removal Algorithm

**Input:** Mixture image $\mathbf{I}$ and patch size $N$.

```
Estimate the reflection dominant regions D(·) using Equation (5) and Equation (6);
Compute the dictionaries D by K-means and PCA;
for m = 1 to M do
    for j = 1 to J do
        Find the reference patches {p_{i,l}}_{l=1}^L corresponding to each patch in I using Equation (10);
        Calculate the weighted average for each mixture patch using Equation (7);
        Calculate the sparse codes α_i and β_i;
        Update sparse codes α_i^{j+1} by solving Equation (18);
        Update B^{j+1} by solving Equation (20);
        Update R^{j+1} by solving Equation (16);
        if j reaches maximum number of iteration then
            Set B^{m+1} = B^{j+1} and R^{m+1} = R^{j+1};
        end if
    end for
    if mod(m, 5) = 0 then
        Update D and the region labelling D(·);
    end if
end for
return B^{m+1} and R^{m+1}.
```

**Output:** Estimated background $\mathbf{B}^*$ and reflection $\mathbf{R}^*$.

whose closed-form solution can be easily obtained by alternating between updating $\mathbf{l}$ and computing $\mathbf{B}$. Updating $\mathbf{l}$ is calculated as

$$\mathbf{l} = \begin{cases} \nabla\mathbf{B}, & \text{if } |\nabla\mathbf{B}| > \epsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

With $\mathbf{l}$ being fixed, the closed-form solution for Equation (20) is obtained similar to the strategy adopted by previous method [44]:

$$\hat{\mathbf{B}} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{I}) - \mathcal{F}(\mathbf{R}) + \gamma\,\mathcal{F}(\sum_i \mathbf{M}_i{}^\top \mathbf{D}\boldsymbol{\alpha}_i) + \frac{1}{\epsilon^2}\mathcal{F}_L}{\mathbf{E} + \gamma\,\mathcal{F}(\sum_i \mathbf{M}_i{}^\top \mathbf{M}_i) + \frac{1}{\epsilon^2}\mathcal{F}_D^2}\right). \tag{22}$$

$\mathbf{E}$ is a matrix with all elements being equal to one; $\mathcal{F}(\cdot)$ and $\mathcal{F}(\cdot)^{-1}$ denotes the Fourier transform and its inverse transform, respectively; $\mathcal{F}(\cdot)^*$ is the corresponding complex conjugate operator; and $\mathcal{F}_L = \sum_{m\in\{h,v\}} \mathcal{F}(\nabla_m)^* \mathcal{F}(\mathbf{l}_m)$ and $\mathcal{F}_D = \sum_{m\in\{h,v\}} \mathcal{F}(\nabla_m)^* \mathcal{F}(\nabla_m)$, where $\nabla_h$ and $\nabla_v$ are the horizontal and vertical differential operators, respectively.

### C. Solving for $\mathbf{R}$

With all variables unrelated to $\mathbf{R}$ being fixed, the optimization problem for $\mathbf{R}$ becomes

$$\hat{\mathbf{R}} = \underset{\mathbf{R}}{\operatorname{argmin}} \|\mathbf{I}-\mathbf{B}-\mathbf{R}\|_2^2 + \delta\|\nabla\mathbf{R}\|_0. \tag{23}$$

Equation (23) can be solved by introducing the auxiliary variables $\mathbf{g} = (\mathbf{g}_h, \mathbf{g}_v)$ w.r.t. the image gradients of $\nabla\mathbf{R}$ in horizontal and vertical directions, which is also adopted

by [45]. Equation (23) can be expressed as

$$\hat{\mathbf{R}} = \underset{\mathbf{R}}{\operatorname{argmin}} \|\mathbf{I} - \mathbf{B} - \mathbf{R}\|_2^2 + \mu\|\nabla\mathbf{R} - \mathbf{g}\|_2^2 + \delta\|\mathbf{g}\|_0. \quad (24)$$

The values of $\mathbf{g}$ are initialized to be zeros. In each iteration, the solution of $\mathbf{R}$ is obtained by solving

$$\min_{\mathbf{R}} \|\mathbf{I} - \mathbf{B} - \mathbf{R}\|_2^2 + \mu\|\nabla\mathbf{R} - \mathbf{g}\|_2^2 \quad (25)$$

The closed-form solution for the least squares problem above can be easily obtained as

$$\mathbf{R} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{I}) - \mathcal{F}(\mathbf{B}) + \mu\mathcal{F}_G}{1 + \mu\sum_{m\in\{h,v\}}\mathcal{F}(\nabla_m)^*\mathcal{F}(\nabla_m)}\right), \quad (26)$$

where $\mathcal{F}_G = \mathcal{F}(\nabla_h)^* F(\mathbf{g}_h) + \mathcal{F}(\nabla_v)^* F(\mathbf{g}_v)$. Finally, given $\mathbf{R}$, we compute $\mathbf{g}$ by

$$\min_{\mathbf{g}} \mu\|\nabla\mathbf{R} - \mathbf{g}\|_2^2 + \delta\|\mathbf{g}\|_0. \quad (27)$$

Equation (27) is a pixel-wise minimization problem, whose solution is calculated as

$$\mathbf{g} = \begin{cases} \nabla\mathbf{R}, & |\nabla\mathbf{R}|^2 > \dfrac{\delta}{\mu}, \\ 0, & \text{otherwise.} \end{cases} \quad (28)$$

## V. Experiment Results

To evaluate the performance of reflection removal, the majority of existing methods compare the visual quality of the estimated background images on 3 to 5 sets of real data [7], [8], or perform the quantitative evaluations using the synthetic images [5], [10]. Due to the lack of real-world dataset with ground truth, quantitative comparison using real data has seldom been done. Thanks to the dataset introduced in Section III-C, we compare our R3 method with state-of-the-art methods for both quantitative accuracies (w.r.t. its corresponding ground truth) and visual quality based on the 50 sets of real data. Though the images in our dataset are taken by DSLR camera with high resolution, considering the computation time and to make the image size compatible to all evaluated algorithms, all images are resized to $400 \times 500$. Since the computations in our methods all belong to the per-pixel computation, such kinds of operation does not influence the final results, which are also adopted by previous methods [2], [5].

The main parameters used in our method are set as follows: $\delta$, $\omega$, and $\gamma$ in Equation (17) are set to 0.004, 1.5, and 1, respectively. Empirically, for the patches from the reflection non-dominant regions, $\xi$ in Equation (17) and Equation (18) is set to 10.5 (with $\tau = 7$); $\lambda$ and the initial value of $H$ in Equation (10) are set to 1 and 30, respectively. For the patches from the reflection dominant regions, $\xi$ is set to 22.5 (with $\tau = 15$); $\lambda$ and the initial value of $H$ are set to 0.01 and 10, respectively. $\mu$ in Equation (27) is set to 0.008. The patch size is set to $7 \times 7$. $L$ in Equation (7) is set to 8. The initial value of $\epsilon$ in Equation (14) is set to 0.05 and is divided by 2 in each iteration. $H$ is added by 10 if the number of reference patches found within current window is less than $L$.

TABLE I

Quantitative Evaluation Results Using Five Different Error Metrics and Compared With AY07 [7], LB14 [5], SK15 [2], WS16 [8], and NR17 [10]

| | Baseline | Ours | AY07 | LB14 | SK15 | WS16 | NR17 |
|---|---|---|---|---|---|---|---|
| sLMSE | 0.969 | **0.980** | 0.927 | 0.939 | 0.830 | 0.963 | 0.969 |
| SSIM | 0.940 | **0.944** | 0.906 | 0.862 | 0.870 | 0.937 | 0.930 |
| SI | **0.965** | 0.958 | 0.943 | 0.958 | 0.913 | 0.955 | 0.950 |
| $\text{SSIM}_r$ | 0.857 | **0.936** | 0.880 | 0.847 | 0.858 | 0.921 | 0.908 |
| $\text{SI}_r$ | 0.886 | **0.942** | 0.905 | 0.906 | 0.886 | 0.921 | 0.925 |

### A. Error Metrics

We adopt the structural similarity index (SSIM) and local mean square error (LMSE), which are widely used by previous methods [5], [10], [46], as error metrics for quantitative evaluation. To make the value of LMSE consistent with SSIM, we convert it to a similarity measure as follows:

$$\text{sLMSE}(\mathbf{B}, \mathbf{B}^*) = 1 - \text{LMSE}(\mathbf{B}, \mathbf{B}^*), \quad (29)$$

where $\mathbf{B}$ is the ground truth and $\mathbf{B}^*$ is the estimated background image.

The luminance and contrast similarity in the original SSIM definition are sensitive to the intensity variance, so we define the structure index (SI) to focus only on the structural similarity between $\mathbf{B}$ and $\mathbf{B}^*$. SI shares similar format as the error metric proposed in [47], but it omits the luminance and contrast part in its original form as

$$\text{SI} = \frac{2\sigma_{\mathbf{B}}\sigma_{\mathbf{B}^*} + c}{\sigma_{\mathbf{B}}^2 + \sigma_{\mathbf{B}^*}^2 + c}, \quad (30)$$

where $\sigma_{\mathbf{B}}$ and $\sigma_{\mathbf{B}}^*$ are the variances of $\mathbf{B}$ and $\mathbf{B}^*$, respectively, and $\sigma_{\mathbf{B}}\sigma_{\mathbf{B}}^*$ is the corresponding covariance.

SSIM, sLMSE, and SI are error metrics evaluating the global similarity between $\mathbf{B}$ and $\mathbf{B}^*$. In our region-aware context, the reflections only dominate a part of the whole image. Based on our observations, though some methods [5], [8] downgrade the quality of the whole images, they can remove the local reflections quite effectively. We define the regional SSIM and SI, denoted as $\text{SSIM}_\text{r}$ and $\text{SI}_\text{r}$, to complement the limitations of global error metrics. We manually label the reflection dominant regions (*e.g.*, like the third column of Figure 3) and evaluate the SSIM and SI values at these regions similar to the evaluation method proposed in [48].

### B. Comparison With the State-of-the-Arts

We compare our method with state-of-the-art single image reflection removal methods, including AY07 [7], LB14 [5], SK15 [2], WS16 [8], and NR17 [10]. We use the codes provided by their authors and set the parameters as suggested in their original papers. Except that for SK15 [2] we adjust its pre-defined threshold (set as 70 in their code) that chooses some local maxima values, since we find the default value shows degenerated results on our data and we manually adjust it for different images to make sure that a similar number of local maxima values to their original demo are generated. AY07 [7] requires the user annotations of background and reflection edges, and we follow their guidance to do the annotation manually.

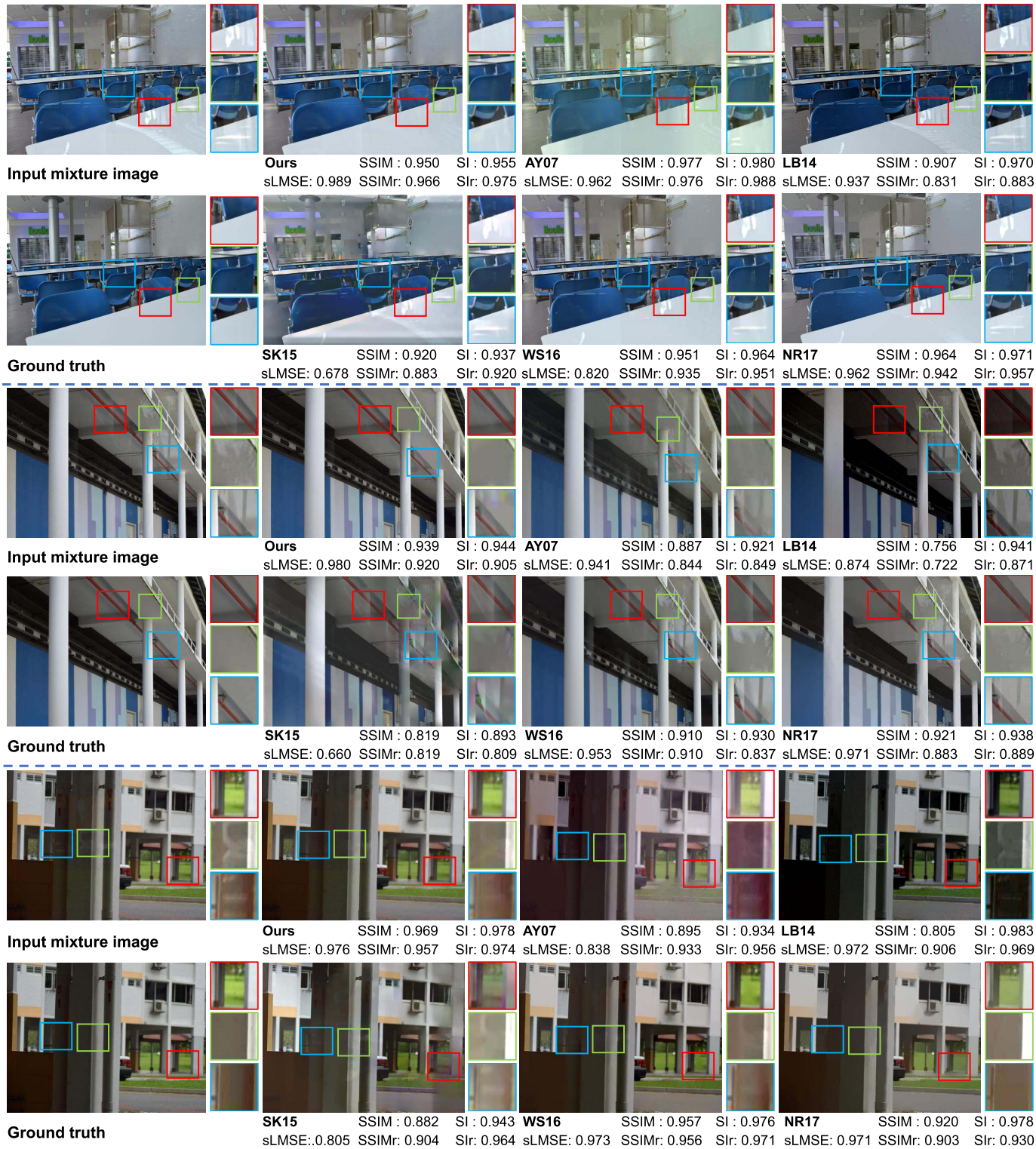



Fig. 5. Reflection removal results on three natural images under weak reflections, compared with AY07 [7], LB14 [5], SK15 [2], WS16 [8], and NR17 [10]. Corresponding close-up views are shown next to the images (the patch brightness $\times 2$ for better visualization), and SSIM and sLMSE values are displayed below the images.

*1) Quantitative Evaluations:* The quantitative evaluation results using five different error metrics and compared with five state-of-the-art methods are summarized in Table I, where the errors between the input mixture images and the corresponding ground truth are used as the baseline comparison. The numbers displayed are the mean values over all 50 images in our dataset. As shown in Table I, the proposed algorithm consistently outperforms other methods for all five error metrics. The higher SSIM and sLMSE values indicate that our method recovers the whole background image with better quality, whose global appearance is closer to the ground truth. For SI values, all methods are lower than the baseline, which

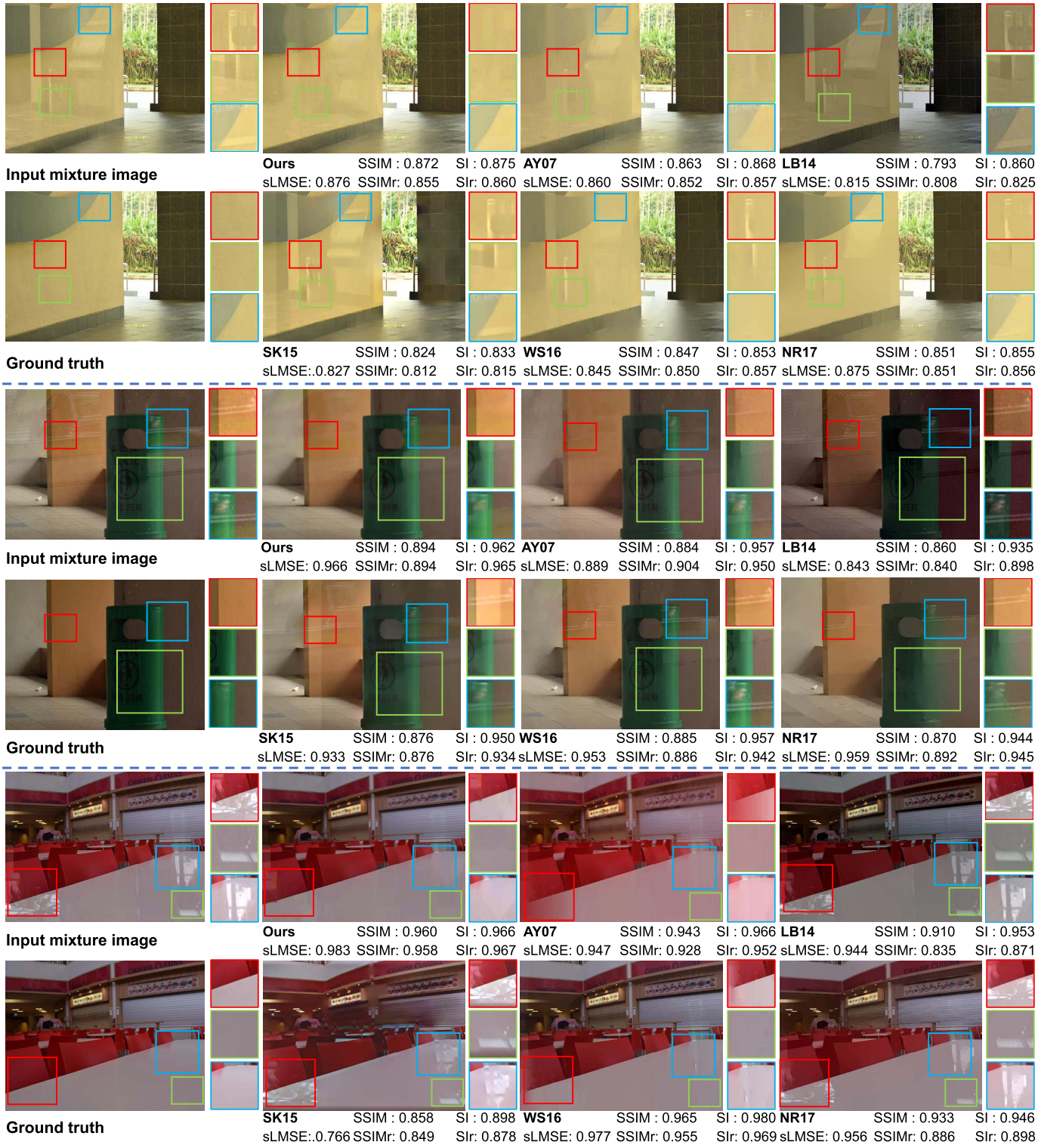



Fig. 6. Reflection removal results on three natural images under strong reflections, compared with AY07 [7], LB14 [5], SK15 [2], WS16 [8], and NR17 [10]. Corresponding close-up views are shown next to the images (the patch brightness $\times 2$ for better visualization), and SSIM and sLMSE values are displayed below the images.

is partly because all methods impair the global structures of the input images. However, due to the the regional strategy in our R3 method, it still beats other five methods and achieve the second best result. The higher SI values tell that our method preserves the structural information more accurately. The higher $\text{SSIM}_\text{r}$ and $\text{SI}_\text{r}$ values mean that our method can remove strong reflections more efficiently in the reflection dominated regions than other methods. LB14 [5] shows the second best result on SI; the most recent method NR17 [10] shows the second best results on SSIM, sLMSE, $\text{SSIM}_\text{r}$ and $\text{SI}_\text{r}$.

*2) Visual Quality Comparison:* We then show examples of estimated background images by our method and five other methods in Figure 5 (three examples with weak reflections)


Fig. 7. Results with and without reflection dominant region (the patch brightness $\times 1.3$ for better visualization).

and Figure 6 (three examples with strong reflections) to check their visual quality. In these examples, our method removes the reflections more effectively and recovers the details of the background image more clearly. NR17 [10] and LB14 [5] remove the reflections to some extent, but from the results shown in the third example of Figure 5 and Figure 6, some residue edges remain visible for the reflections that are not out of focus. LB14 [5] also causes a color change of the input mixture image, where the results are much darker than the ground truth. Both LB14 [5] and WS16 [8] show some over-smooth artifacts, when they are not able to differentiate the background and reflection clearly. When the edges can be correctly labelled, AY07 [7] shows acceptable results in some examples (*e.g.*, the third example in Figure 5), but the performance is poor when the edges cannot be clearly differentiated by human labelling (*e.g.*, the first example in Figure 6). The performance of SK15 [2] is a bit degenerated with these examples, and it shows some pacthy artifacts. When the reflection is strong (*e.g.*, the first example in Figure 6), our method not only removes the undesired reflections, but also restores the missing contents of the background caused by the reflection, thanks to the region-aware content prior.

### C. The Effect of the Reflection Dominant Region

Comparing with existing methods, region-aware processing is unique in the proposed whole framework. To evaluate whether it effectively recovers the details in reflection dominant regions and avoid artifacts in reflection non-dominant regions, we show recovered background image with and without the reflection (non-)dominant region labelling. Two examples are shown in Figure 7. In both examples, the methods without reflection dominant regions only attenuates the reflections but fails to remove them, but the region-aware approach successfully removes the reflections; in the top example, the image details (*e.g.*, the patch in the red box) of the method without the reflection dominant regions are rather blurred.

### D. The Effect of the Gradient Prior

We conduct another experiment to show the effectiveness of the gradient priors in Figure 8. Although for the image patches in the red and green boxes, both reflections are removed regardless of whether gradient priors are considered, the image patches in the blue boxes clearly show that the gradient prior helps keeping the sharpness of the edges so that the structural information is better recovered in the background image.

### E. Comparison With WS17

The recent approach WS17 [9] also makes use of the patch recurrence from several similar images and content priors, by assuming that reflection-free images with similar content are available from an external database. To make their assumptions satisfied, we use images containing objects which can be easily retrieved from an external database, and provide both the input mixture image and external database to WS17 [9]. The comparison between our method and WS17 [9] are illustrated in Figure 9. With the help of an external database, WS17 [9] shows superior performance in some parts (the blue box in Figure 9). But our method still provides comparable results to WS17 [9] with only internal image recurrence, thanks to the robust patch matching in reflection dominant regions. Note our method can be applied to much broader categories of images.

### F. Convergence Analysis

The last experiment shows the convergence of our algorithm. As we have claimed in Section IV, a larger $\tau$ in

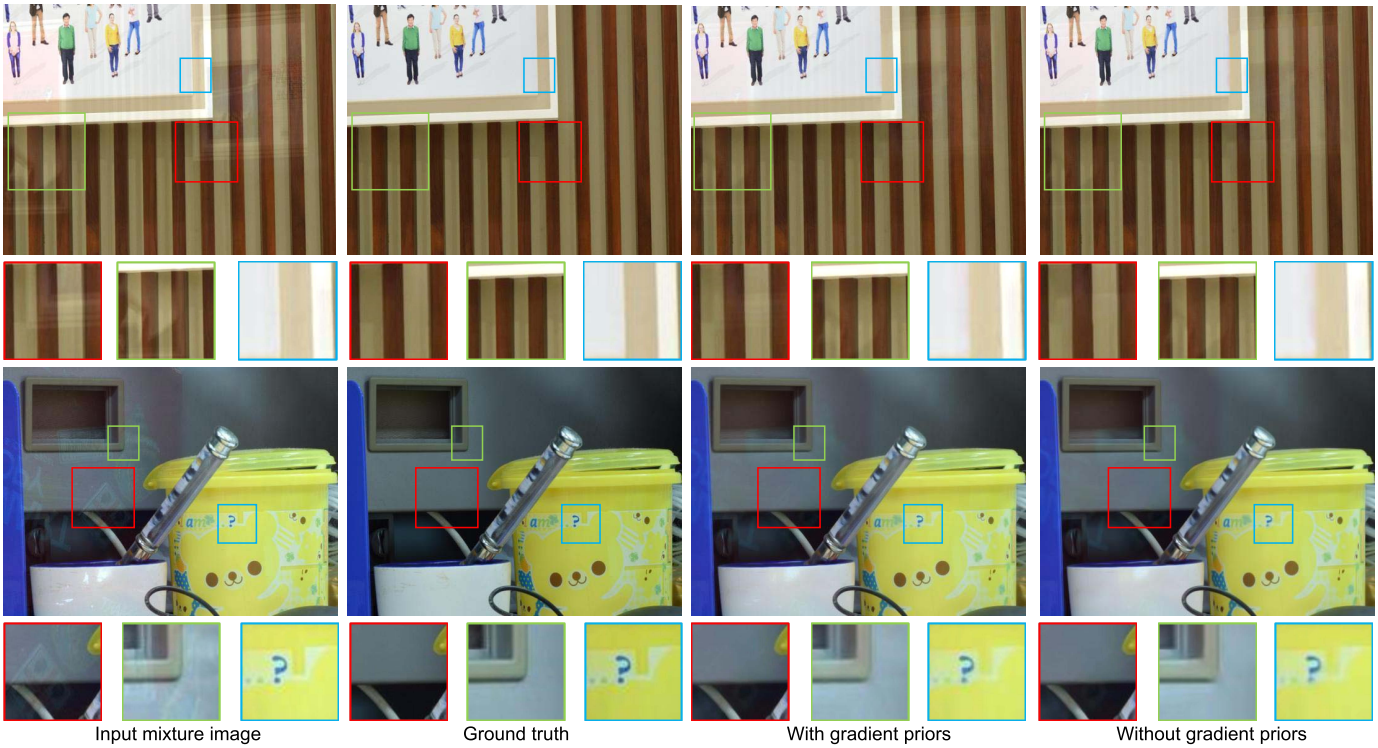



Fig. 8. Results with and without the gradient priors (the patch brightness $\times 1.3$ for better visualization).

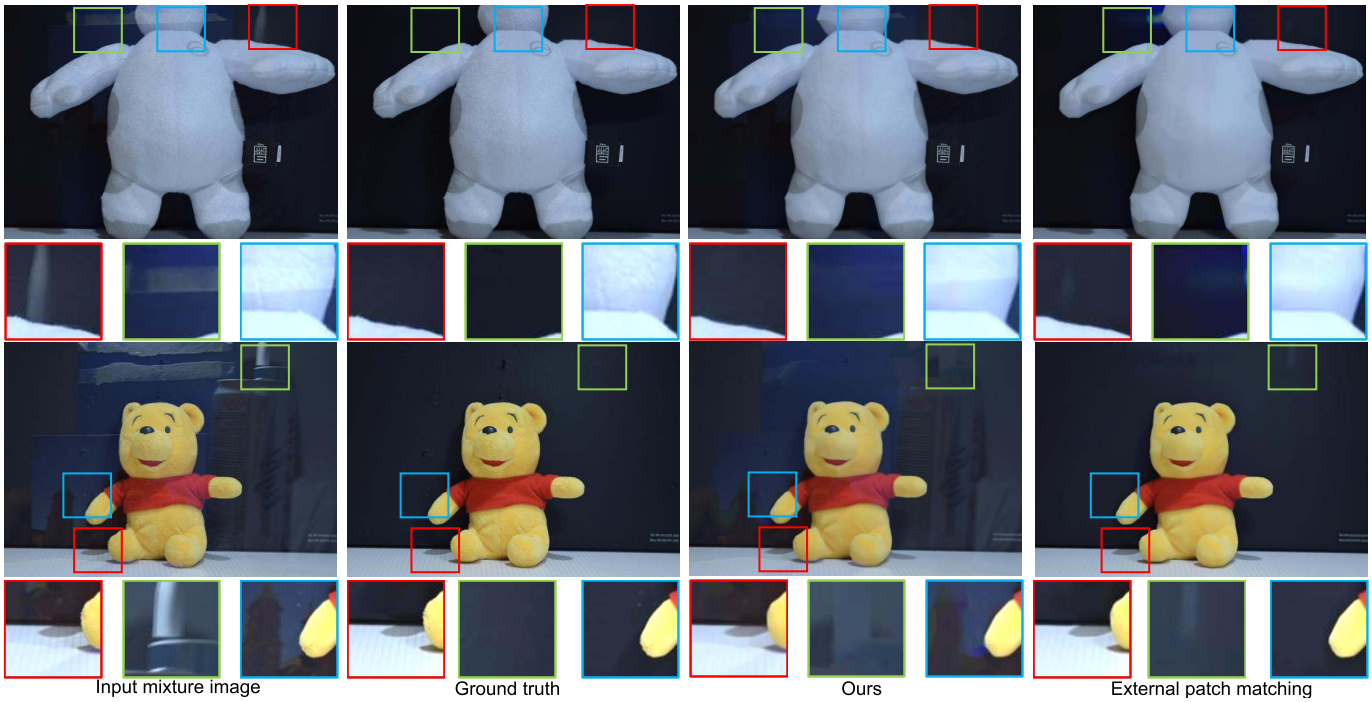



Fig. 9. Comparison between our proposed method and WS17 [9] (the patch brightness $\times 1.3$ for better visualization).

Equation (19) generally allows a quicker convergence of our R3 method. In our settings, the patches from the reflection dominant regions are given a larger $\tau$ values defined as $\tau_1$ here and the patches from the reflection non-dominant regions are assigned a smaller $\tau$ defined as $\tau_2$ here. We set $\tau_1 = 15$ and $\tau_2 = 7$ in our experiments. To validate the settings, we test different values by fixing one and changing another one. The performances with different values are illustrated in Figure 10. By fixing $\tau_2 = 7$, $\tau_1$ is set to 10, 15 (the values used in our experiments), 100 and 200. A larger $\tau_1$ can achieve better results in the first iteration and converge faster under a larger value. By fixing $\tau_1 = 15$, $\tau_2$ is set to 5.5, 7

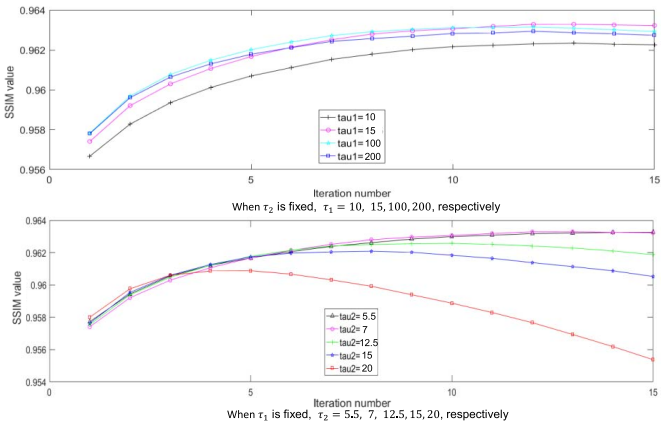



Fig. 10. The convergence analysis of our proposed method under different $\tau$ values.

(the values used in our experiments), 12.5, 15 and 20. A larger $\tau_2$ decrease SSIM values after approximate six iterations, which indicates that the image structure are impaired. It is partly due to the over-smooth effect of the non-local image prior we adopt, which is explained in [49]. A smaller $\tau_2$ achieves similar performances when compared with the value used in our experiment. Considering the performance variation with different $\tau$, the parameters in our experiments ($\tau_1 = 15$ and $\tau_2 = 7$) can achieve good results and keep stable after six iterations.

## VI. Conclusion

We introduce reflection dominant regions to single image reflection removal problem to efficiently remove reflections and avoid artifacts caused by incompletely removed reflections in an adaptive manner. We integrate the content prior and gradient prior into a unified R3 framework to take account of both content restoration and reflection suppression. By refining the sparse coefficients learned from the mixture images with the reference patches to generate a more accurate sparse regularization term. We show better performances than state-of-the-art methods for both the quantitative and visual qualities.

*Limitations:* In spite of the effectiveness of our R3 method, it also has several limitations:

- The patch selection step is computationally expensive. Its complexity increases linearly with the window size. However, our current implementation is an unoptimized Matlab implementation, which takes three minutes for the patch matching and fewer than 30 seconds for other steps on a modern PC. Based on the experience in denoising [23] with similar formulation, the computation can be sped up by using more efficient programming language (*e.g.*, C++) and parallel implementations;
- Our method adopts the non-local image prior as the content prior. As mentioned in [49], it is prone for the non-local image priors to over-smooth highly textured regions, especially in the case of strong artifacts. The performance drops when the background is textured;
- Our method is based on the observation that reflection only dominates a part of an image. However, in real scenes, it is possible that the whole image is overlaid with strong reflections; in such a case our method may fail due to the 'rare patch effect' [50].
- Since our method utilizes the reference patches around the mixture patch to remove the reflections, the information of the background must be kept more or less. If very strong reflections exist in a scene, the reference patches cannot be found since very fews details of the background are kept. In this situation, the reflection removal problems degrades to an image inpainting problem;
- Though our method does not explicitly rely on image priors (*e.g.*, the blur levels [5] or ghosting effects [2]), the reflection dominant region detection is based on the depth of field of the input mixture image. When the depth of field is not uniform, the detection may be less accurate. In such a situation, our performance is similar to that in Figure 7 where sharp edge information cannot be clearly recovered.

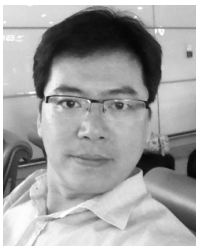
## References

[1] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman, "A computational approach for obstruction-free photography," *ACM Trans. Graph.*, vol. 34, no. 4, p. 79, 2015.
[2] Y. C. Shih, D. Krishnan, F. Durand, and W. T. Freeman, "Reflection removal using ghosting cues," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3193–3201.
[3] Y. Li and M. S. Brown, "Exploiting reflection change for automatic reflection removal," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2432–2439.
[4] N. Kong, Y.-W. Tai, and J. S. Shin, "A physically-based approach to reflection separation: From physical modeling to constrained optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 209–221, Feb. 2014.
[5] Y. Li and M. S. Brown, "Single image layer separation using relative smoothness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2752–2759.
[6] T. A. Javaran, H. Hassanpour, and V. Abolghasemi, "Local motion deblurring using an effective image prior based on both the first- and second-order gradients," *Mach. Vis. Appl.*, vol. 28, nos. 3–4, pp. 431–444, 2017.
[7] 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, pp. 1647–1654, Sep. 2007.
[8] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot, "Depth of field guided reflection removal," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 21–25.
[9] R. Wan, B. Shi, A.-H. Tan, and A. C. Kot, "Sparsity based reflection removal using external patch search," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2017, pp. 1500–1505.
[10] A. Nikolaos, A. Radhakrishna, and S. Sabine, "Single image reflection suppression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 1752–1760.
[11] 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, Jan. 2012.
[12] E. Be'Ery and A. Yeredor, "Blind separation of superimposed shifted images using parameterized joint diagonalization," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 340–353, Mar. 2008.
[13] R. Szeliski, S. Avidan, and P. Anandan, "Layer extraction from multiple images containing reflections and transparency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2000, pp. 246–253.
[14] C. Sun, S. Liu, T. Yang, B. Zeng, Z. Wang, and G. Liu, "Automatic reflection removal using gradient intensity and motion cues," in *Proc. ACM Multimedia Conf.*, 2016, pp. 466–470.
[15] T. Sirinukulwattana, G. Choe, and I. S. Kweon, "Reflection removal using disparity and gradient-sparsity via smoothing algorithm," in *Proc. Int. Conf. Image Process. (ICIP)*, 2015, pp. 1940–1944.
[16] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 828–835, 2005.

[17] Y. Y. Schechner, N. Kiryati, and R. Basri, "Separation of transparent layers using focus," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 1998, pp. 25–39.
[18] Q. Wang, H. Lin, Y. Ma, S. Kang, and J. Yu. (2015). "Automatic layer separation using light field imaging." [Online]. Available: https://arxiv.org/abs/1506.04721
[19] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 60–65.
[20] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[21] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognit.*, vol. 43, no. 4, pp. 1531–1549, 2010.
[22] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 479–486.
[23] E. Luo, S. H. Chan, and T. Q. Nguyen, "Adaptive image denoising by targeted databases," *IEEE Trans. Image Process.*, vol. 24, no. 7, pp. 2167–2181, Jul. 2015.
[24] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[25] M. Zontak and M. Irani, "Internal statistics of a single natural image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 977–984.
[26] H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 770–777.
[27] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Sparse representation based blind image deblurring," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2011, pp. 1–6.
[28] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3449–3456.
[29] M. G. Jafari and M. D. Plumbley, "Separation of stereo speech signals based on a sparse dictionary algorithm," in *Proc. IEEE Eur. Signal Process. Conf.*, Aug. 2008, pp. 1–5.
[30] M. G. Jafari, M. D. Plumbley, and M. E. Davies, "Speech separation using an adaptive sparse dictionary algorithm," in *Proc. IEEE Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, May 2008, pp. 25–28.
[31] J. Bobin, Y. Moudden, J. L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 409–412, Jul. 2006.
[32] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2662–2674, Nov. 2007.
[33] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Blind separation of image sources via adaptive dictionary learning," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 2921–2930, Jun. 2012.
[34] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, no. 1, pp. 490–530, May 2015.
[35] Q. Yan, Y. Xu, X. Yang, and T. Nguyen, "Separation of weak reflection from a single superimposed image," *IEEE Signal Process. Lett.*, vol. 21, no. 10, pp. 1173–1176, Oct. 2014.
[36] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
[37] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
[38] Y.-C. Chung, S.-L. Chang, J.-M. Wang, and S.-W. Chen, "Interference reflection separation from a single image," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Dec. 2009, pp. 1–6.
[39] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1519–1524, Sep. 2006.
[40] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1107–1114.
[41] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Signal Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
[42] X. Zhang, *Matrix Analysis and Applications*. Beijing, China: Springer, 2004.
[43] Y. Li, W. Dong, G. Shi, and X. Xie, "Learning parametric distributions for image super-resolution: Where patch matching meets sparse coding," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 450–458.
[44] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "$L_0$-regularized intensity and gradient prior for deblurring text images and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 342–355, Feb. 2017.
[45] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring text images via L0-regularized intensity and gradient prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2901–2908.
[46] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2009, pp. 2335–2342.
[47] S.-H. Sun, S.-P. Fan, and Y.-C. F. Wang, "Exploiting image structural similarity for single image rain removal," in *Proc. Int. Conf. Image Process. (ICIP)*, 2014, pp. 4482–4486.
[48] L. Qu, J. Tian, S. He, Y. Tang, and R. Lau, "Deshadownet: A multi-context embedding deep network for shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 2308–2316.
[49] P. Qiao, Y. Dou, W. Feng, and Y. Chen. (2017). "Learning non-local image diffusion for image denoising." [Online]. Available: https://arxiv.org/abs/1702.07472
[50] C. A. Deledalle, V. Duval, and J. Salmon, "Non-local methods with shape-adaptive patches (NLM-SAP)," *J. Math. Imag. Vis.*, vol. 43, no. 2, pp. 103–120, Jun. 2012.

**Renjie Wan** (S'14) received the B.S. degree from the University of Electronic Science and Technology of China in 2012. He is currently pursuing the Ph.D. degree with the Nanyang Technological University, Singapore. He feels interested in the computation photography and the history of ancient China.

**Boxin Shi** (M'14) received the B.E. degree from the Beijing University of Posts and Telecommunications in 2007, the M.E. degree from Peking University (PKU) in 2010, and the Ph.D. degree from The University of Tokyo in 2013. He was a Post-Doctoral Researcher with the MIT Media Lab, the Singapore University of Technology and Design, and Nanyang Technological University from 2013 to 2016. He was a Researcher with the National Institute of Advanced Industrial Science and Technology from 2016 to 2017. He is currently an Assistant Professor (1000 Youth Talents Professorship) with PKU, where he leads the Camera Intelligence Group. He received the Best Paper Runner-up Award from the International Conference on Computational Photography in 2015. He served/is serving as an Area Chair for MVA 2017 and ACCV 2018.



**Ling-Yu Duan** (M'06) received the M.Sc. degree in automation from the University of Science and Technology of China, Hefei, China, in 1999, the M.Sc. degree in computer science from the National University of Singapore, Singapore, in 2002, and the Ph.D. degree in information technology from The University of Newcastle, Callaghan, Australia, in 2008. He has been serving as an Associate Director of the Rapid-Rich Object Search Laboratory, a joint lab between Nanyang Technological University, Singapore, and Peking University (PKU), China, since 2012. He is currently a Full Professor with the National Engineering Laboratory of Video Technology, School of Electronics Engineering and Computer Science, PKU. His research interests include multimedia indexing, search, and retrieval, mobile visual search, visual feature coding, and video analytics. He received the *EURASIP Journal on Image and Video Processing* Best Paper Award in 2015, the Ministry of Education Technology Invention Award (first prize) in 2016, the National Technology Invention Award (second prize) in 2017, the China Patent Award for Excellence in 2017, and the National Information Technology Standardization Technical Committee Standardization Work Outstanding Person Award in 2015. He is serving as a Co-Chair for the MPEG Compact Descriptor for Video Analytics. He was a Co-Editor of MPEG Compact Descriptor for Visual Search Standard (ISO/IEC 15938-13). He is currently an Associate Editor of *ACM Transactions on Intelligent Systems and Technology* and *ACM Transactions on Multimedia Computing, Communications, and Applications*.

**Ah-Hwee Tan** (SM'04) received the B.Sc. (Hons.) and M.Sc. degrees in computer and information science from the National University of Singapore, Singapore, in 1989 and 1991, respectively, and the Ph.D. degree in cognitive and neural systems from Boston University, Boston, MA, USA, in 1994. He was a Research Manager with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore, spearheading the Text Mining and Intelligent Agents research programs. He is currently a Professor of computer science and an Associate Chair (research) with the School of Computer Science and Engineering, Nanyang Technological University. His current research interests include cognitive and neural systems, brain inspired intelligent agents, machine learning, knowledge discovery, and text mining. He is an Editorial Board Member of the IEEE ACCESS and an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.

**Wen Gao** (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from The University of Tokyo, Tokyo, Japan, in 1991. He was a Professor of computer science with the Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He is currently a Professor of computer science with the School of Electronic Engineering and Computer Science, Institute of Digital Media, Peking University, Beijing. He has authored extensively including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interfaces, and bioinformatics. He is a member of the China Engineering Academy. He has been the Chair of a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE International Conference on Multimedia and Expo and ACM Multimedia. He has served on the advisory and technical committees of numerous professional organizations. He served/serves on the Editorial Board of several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, and the *Journal of Visual Communication and Image Representation*.

**Alex C. Kot** (S'85–M'89–SM'98–F'06) has been with Nanyang Technological University, Singapore, since 1991. He headed the Division of Information Engineering, School of Electrical and Electronic Engineering, for eight years and served as an Associate Chair/Research. He was a Vice Dean Research with the School of Electrical and Electronic Engineering and an Associate Dean of the College of Engineering for eight years. He is currently a Professor with the School of Electrical and Electronic Engineering and the Director of the Rapid-Rich Object Search Lab. He has published extensively in the areas of signal processing for communication, biometrics, image forensics, information security, and computer vision. He is a fellow of IES and the Academy of Engineering, Singapore. He was recognized as the IEEE Distinguished Lecturer of the Signal Processing Society. He was a recipient of the Best Teacher of the Year Award and co-recipient of several best paper awards, including for ICPR, IEEE WIFS, and IWDW. He has served for the IEEE Signal Processing Society in various capacities, such as a General Co-Chair at the 2004 IEEE International Conference on Image Processing and the Vice President of the IEEE Signal Processing Society. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, *IEEE Signal Processing Magazine*, the IEEE JOURNAL OF SPECIAL TOPICS IN SIGNAL PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: FUNDAMENTAL THEORY AND APPLICATIONS, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: ANALOG AND DIGITAL SIGNAL PROCESSING