# Robust RGB-Guided Super-Resolution of Hyperspectral Images via TV$^3$ Minimization

Weixiao Wan, Bowen Zhang, Marija Vella, João F.C. Mota, *Member, IEEE*, Wei Chen, *Senior Member, IEEE*

*Abstract*—**We consider the problem of increasing the resolution of a hyperspectral image (HSI) with the aid of a high-resolution RGB image of the same scene. The current state-of-the-art algorithms for this task are based on convolutional neural networks (CNNs) and generally assume that the relation between the RGB image and the HSIs remains constant during training and testing. In particular, their performance quickly degrades if we use different color spaces, e.g., CIEXYZ or CIERGB during these stages. In this paper, we propose a method that addresses this problem. Specifically, our method requires no RGB images during training, but still can leverage an RGB image during testing to improve the performance of super-resolution. Furthermore, the method works even if the relation between the RGB and HSI images, captured by the camera spectral response (CSR), is not known precisely. Our experiments demonstrate that the proposed method not only outperforms state-of-the-art methods for joint RGB-HSI super-resolution, but also works for various types of color images.**

*Index Terms*—**RGB-guided hyperspectral image super-resolution, TV$^3$ minimization**

## I. Introduction

HYPERSPECTRAL images (HSIs) are composed of images in several spectral bands, ranging from infrared to ultraviolet. As different materials have different spectral signatures, HSIs enable the identification of different types of materials in a given scene, a feature important in applications such as remote sensing [1], object detection [2], or tracking [3]. However, hardware constraints impose limits on the spatial resolution of each band. RGB cameras, on the other hand, produce images of much higher-resolution (HR), but integrate information across several bands, and thus have low spectral resolution.

The task of leveraging a HR RGB image to increase the spatial resolution of a HSI is known as *RGB-guided hyperspectral image super-resolution (HSI-SR)*. Algorithms for this task can be classified as model-driven or data-driven. Model-driven methods include matrix factorization [4]–[7], Bayesian inference [8], [9], and tensor factorization methods [10]–[13]. In all these methods, the camera spectral response (CSR), which maps the HSI onto the RGB image, is encoded explicitly via a physical model that is often assumed known, fixed, and linear. Specifically, the RGB image is obtained by multiplying the HSI cube with a matrix obtained from the CSR. Data-driven methods learn this map by exploiting a training dataset containing HSIs and their corresponding RGB images. The current state-of-the-art methods use convolution neural networks (CNNs), which results in boosted spatial resolution of HSIs [14]–[16].

There also exist single-image HSI-SR methods [17]–[19] that exploit the spatial and spectral correlation characteristics in hyperspectral data with hand-crafted or learned priors from some HSI dataset. These methods do not employ an RGB image for guidance, and thus have relatively poor performance compared to RGB-guided methods.

Although current state-of-the-art CNN methods have achieved promising results, they have a major shortcoming: if a method is designed on images acquired from a camera with a specific CSR, its performance quickly degrades when applied to images acquired from a camera with a different CSR function [16]. In particular, there exist several types of color spaces, e.g., CIEXYZ, CIERGB [20], and different (RGB) cameras construct images under different color spaces [21]–[23]. Different camera manufacturers often optimize the color spaces of their own cameras during manufacturing, and the final color spaces are often undisclosed. The proposed method addresses this problem by requiring no knowledge of the transformation matrix between the different color spaces. In our experiments, for example, we use different color spaces during training and testing, namely, the standard CIEXYZ and CIERGB spaces, to assess performance. They show that if a method (model-driven or data-driven) for RGB-guided HSI-SR is designed assuming RGB images in CIEXYZ color space, its performance degrades when applied to RGB images in CIERGB color space. Although this problem can be addressed by considering different CSR functions or providing training data with various color spaces, it is usually difficult to obtain accurate CSR functions or enough training samples for all the relevant color spaces. Furthermore, registering the different RGB images to the HSI cube is not only challenging, but also cumbersome.

In this letter, we propose a novel RGB-guided HSI SR framework that requires no RGB images during training, nor an accurate CSR function during testing. These features make the proposed method robust to variations of the CSR function during training and testing, making it applicable to scenarios in which training and testing images are acquired by different

This work is supported by the National Key R&D Program of China (2018YFE0207600, 2020YFB1807201); the Natural Science Foundation of China (62122012, 61911530216, U1834210, 61725101); the State Key Laboratory of Rail Traffic Control and Safety (RCS2021ZZ004); the Beijing Natural Science Foundation (L211012, L202019); the Fundamental Research Funds for the Central Universities (2020JBZD005); the Royal Society and NSFC (IEC/NSFC/181255); the EPSRC New Investigator Award (IEC/NSFC/181255).

Weixiao Wan, Bowen Zhang and Wei Chen are with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China (email: 20120132, 18120171@bjtu.edu.cn, weich@bjtu.edu.cn). Marija Vella and João F.C. Mota are with the School of Engineering and Physical Sciences, Heriot-Watt University, UK (email: mv37@hw.ac.uk, j.mota@hw.ac.uk). Weixiao Wan and Bowen Zhang contributed equally to this work. Corresponding author: Wei Chen.

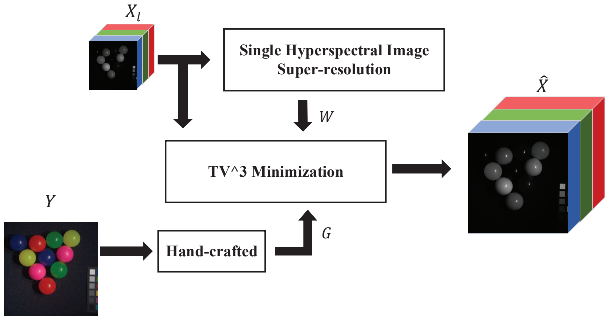


Fig. 1: Our framework for RGB-guided HSI SR.

cameras. Fig. 1 shows the block diagram of our method, which is explained in more detail in Section II. Essentially, we start by super-resolving the low-resolution (LR) HSI cube $\mathbf{X}_l$ using a single HSI-SR method, i.e., a method that requires no RGB images during training. Then, using the resulting HR HSI cube $\mathbf{W}$, and a HR RGB image $\mathbf{Y}$ of the scene , created by an arbitrary CSR function and processed into $\mathbf{G}$ by extracting hand-crafted features, we further improve the resolution of the HR HSI cube by solving a problem which we call TV$^3$ minimization. This method is inspired by the TV-TV minimization in [24], [25], which post-processes the output of CNNs for single-image super-resolution. The TV$^3$ regularization term in the formulated optimization problem renders our algorithm robust to mismatches in the CSR function. Experimental results demonstrate that the proposed method outperforms the current state-of-the-art (SOTA) for RGB-guided HSI super-resolution [4], [8], [12], [14]. To the best of our knowledge, this work is the first to study the effect of CSR mismatch in hybrid hyperspectral imaging systems.

## II. RGB-Guided HSI Super-resolution with an Arbitrary CSR Function

### A. Problem Formulation

We aim to restore a HR HSI $\mathbf{X} \in \mathbb{R}^{MN\times B}$ from a LR HSI $\mathbf{X}_l \in \mathbb{R}^{mn\times B}$ with the guidance of a HR RGB image $\mathbf{Y} \in \mathbb{R}^{MN\times b}$. $M$, $N$ and $B$ are the height, width and number of bands of the HR HSI $\mathbf{X}$, respectively. Correspondingly, $m$ and $n$ denote the height and width of the LR HSI $\mathbf{X}_l$, and $b$ denotes the number of bands of the HR RGB image $\mathbf{Y}$. Obviously, $b = 3$. Each column of $\mathbf{X}$, $\mathbf{X}_l$, and $\mathbf{Y}$ contains vectorized images. We assume that $\mathbf{X}_l$ is obtained from $\mathbf{X}$ by downsampling it along the spatial dimensions, and $\mathbf{Y}$ is obtained from $\mathbf{X}$ by downsampling it along the spectral dimension with some CSR function. That is,

$$\mathbf{X}_l = \mathbf{B}\mathbf{X}, \mathbf{Y} = \mathbf{X}\mathbf{C}, \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{mn\times MN}$ denotes the spatial downsampling matrix, and $\mathbf{C} \in \mathbb{R}^{B\times b}$ denotes the CSR function, which integrates the spectra into R, G and B channels. We assume that the downsampling matrix $\mathbf{B}$ is known, while the CSR matrix $\mathbf{C}$ is unkown.

### B. The Proposed RGB-guided HSI SR Framework

The proposed framework is shown in Fig 1. First, the LR HSI $\mathbf{X}_l$ is super-resolved into $\mathbf{W} \in \mathbb{R}^{MN\times B}$ via an arbitrary single HSI super-resolution method. We adopt CNN-based methods owing to their current outstanding performance. These networks are trained by using datasets with HR HSI and LR HSI pairs, and RGB images are not required.

At the same time, we construct a HR guidance image $\mathbf{G} \in \mathbb{R}^{MN\times B}$ from the HR RGB image $\mathbf{Y} \in \mathbb{R}^{MN\times b}$ heuristically. Namely, we assume that the R, G, B bands of the RGB images represent the high-resolution spatial information around the 700nm, 540nm and 430nm wavelengths. The remaining bands are constructed according to the following steps:

**Step 1**. We choose a HR reference image $y_{\text{ref}}$ for different band intervals. Specifically, bands from $y_{\text{ref}}$ in 400nm $\sim$ 480nm are assigned the B channel of $\mathbf{Y}$, in 480nm $\sim$ 580nm the G channel, and 580nm $\sim$ 700nm the R channel.

**Step 2**. The band $i = 1, \ldots, B$ of the guidance image $\mathbf{G} \in \mathbb{R}^{MN\times B}$ is determined by

$$\mathbf{G}_i = y_{\text{ref}}\,, \tag{2}$$

where $\mathbf{G}_i$ denotes band $i$ of $\mathbf{G}$ and $y_{\text{ref}}$ denotes the reference image chosen for band $i$.

This simple procedure leads to a coarse approximation of the HR HSI (produced only from the RGB data), and does not require knowledge of the CSR function.

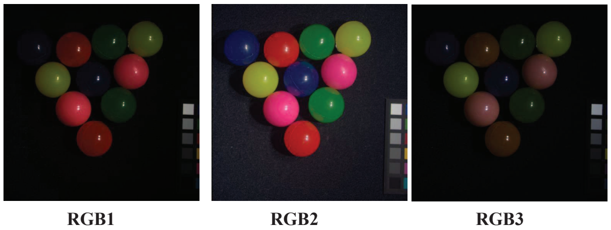
**TV$^3$ minimization.** We use the super-resolved HR HSI $\mathbf{W}$, the hand-crafted HR guidance $\mathbf{G}$, and the LR HSI image $\mathbf{X}_l$ to formulate a new HSI HR problem called TV$^3$ minimization. This method creates an estimate $\widehat{\mathbf{X}}$ of the desired HR HSI, and operates under the following assumptions on the ground truth $\mathbf{X}$: i) $\mathbf{X}$ has a small number of edges, i.e. a small TV-norm; ii) $\mathbf{X}$ is close to $\mathbf{W}$, the HR output generated by CNN-based methods, and the distance is measured by TV-norm; iii) $\mathbf{X}$ is also close to the HR guidance $\mathbf{G}$, and the distance is measured by TV-norm. We explicitly use measurement $\mathbf{X}_l = \mathbf{B}\mathbf{X}$ as a constraint. The TV$^3$ block implements a solver for the following optimization problem:

$$\begin{array}{ll} \underset{\mathbf{X}}{\min} & \|\mathbf{X}\|_{\text{TV}} + \beta\|\mathbf{X}-\mathbf{W}\|_{\text{TV}} + \gamma\|\mathbf{X}-\mathbf{G}\|_{\text{TV}} \\ \text{s.t.} & \mathbf{X}_l = \mathbf{B}\mathbf{X}\,, \end{array} \tag{3}$$

where $\beta$ and $\gamma$ are regularization parameters. We chose the TV-norm since it is a widely used prior for image processing tasks, and our framework can be simply revised to incorporate other priors. The three terms in the objective function of (3) correspond to the assumption i), ii) and iii), respectively. As the objective function is convex and the constraint is linear, (3) is a convex optimization problem. The TV-norm $\|\mathbf{X}\|_{\text{TV}}$ in (3) is defined for a spatial-vectorized HSI $\mathbf{X} \in \mathbb{R}^{MN\times B}$ as the sum of the 2D TV-norms along the spectra, i.e., $\|\mathbf{X}\|_{\text{TV}} = \sum_{i=1}^{B} ||\mathbf{X}_i||_{\text{TV}}$, where $\mathbf{X}_i \in \mathbb{R}^{M\times N}$ denotes the matrix folded from the band $i$ of $\mathbf{X}$, and the TV-norm on the right-hand side is the 2D TV-norm. Specifically, for an image $\mathbf{P} \in \mathbb{R}^{M\times N}$ with a spatial-vectorized version $\boldsymbol{p} \in \mathbb{R}^{MN}$, the 2D TV-norm is

$$\begin{aligned} \|\boldsymbol{p}\|_{\text{TV}} &= \sum_{i=1}^{M}\sum_{j=1}^{N} |\boldsymbol{v}_{ij}^T\boldsymbol{p}| + |\boldsymbol{h}_{ij}^T\boldsymbol{p}| \\ &= \left\| \begin{bmatrix} \mathbf{V} \\ \mathbf{H} \end{bmatrix} \boldsymbol{p} \right\|_1 = \|\mathbf{D}\boldsymbol{p}\|_1\,, \end{aligned} \tag{4}$$

where $\boldsymbol{v}_{ij} \in \mathbb{R}^{MN}$ and $\boldsymbol{h}_{ij} \in \mathbb{R}^{MN}$ extract the vertical and horizontal differences at pixel $(i, j)$ of $\mathbf{P}$. $\mathbf{V} \in \mathbb{R}^{MN\times MN}$ and



RGB1 RGB2 RGB3

Fig. 2: Example of HR color images generated according to different schemes.

$\mathbf{H} \in \mathbb{R}^{MN \times MN}$ are matrices that concatenate $\boldsymbol{v}_{ij}$ and $\boldsymbol{h}_{ij}$ for all pixels, respectively, and $\mathbf{D} = \left[\mathbf{V}^T, \mathbf{H}^T\right]^T \in \mathbb{R}^{2MN \times MN}$. $\|\cdot\|_1$ denotes the $\ell_1$-norm. As the TV-norm in (3) decomposes across bands, problem (3) also decomposes into $B$ independent problems that can be solved in parallel. Namely, representing $\boldsymbol{w}_k$ as the $k$th band of $\mathbf{W}$, $\boldsymbol{g}_k$ the $k$th band of $\mathbf{G}$, and $\boldsymbol{x}_{lk}$ the $k$th band of $\mathbf{X}_l$, the $k$th problem in (3) can be written as

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \|\boldsymbol{x}\|_{\text{TV}} + \beta\|\boldsymbol{x} - \boldsymbol{w}_k\|_{\text{TV}} + \gamma\|\boldsymbol{x} - \boldsymbol{g}_k\|_{\text{TV}} \\ \text{s.t.} \quad & \boldsymbol{x}_{lk} = \mathbf{B}\boldsymbol{x}\,, \end{aligned} \tag{5}$$

whose optimal solution is the spatial-vectorized result of the band $k$th of the RGB-guided HR hyperspectral image $\widehat{\mathbf{X}}$.

We solve a reformulation of (5) obtained by introducing two auxiliary variables $\boldsymbol{u} \in \mathbb{R}^{2MN}$ and $\boldsymbol{v} \in \mathbb{R}^{MN}$, and defining $\overline{\boldsymbol{w}_k} = \mathbf{D}\boldsymbol{w}_k$, and $\overline{\boldsymbol{g}_k} = \mathbf{D}\boldsymbol{g}_k$. Problem (5) is then equivalent to

$$\begin{aligned} \min_{\boldsymbol{u},\boldsymbol{x},\boldsymbol{v}} \quad & \|\boldsymbol{u}\|_1 + \beta\|\boldsymbol{u} - \overline{\boldsymbol{w}_k}\|_1 + \gamma\|\boldsymbol{u} - \overline{\boldsymbol{g}_k}\|_1 \\ \text{s.t.} \quad & \boldsymbol{x}_{lk} = \mathbf{B}\boldsymbol{x} \\ & \boldsymbol{u} = \mathbf{D}\boldsymbol{v} \\ & \boldsymbol{x} = \boldsymbol{v} \end{aligned} \tag{6}$$

The above optimization problem can be solved by applying the alternating direction method of multipliers (ADMM) [26], where the iteration steps are similar to the ones in [24]. We omit the detailed derivations owing to the limited space.

## III. Experiments

To assess the effectiveness of the proposed method, we carried out experiments on the CAVE dataset [21]. The CAVE dataset consists of 32 HR HSI images, each of which with dimensions $512 \times 512$ and 31 spectral bands. These spectral images are taken within the wavelength range 400nm $\sim$ 700nm with an interval of 10 nm. The CAVE dataset also contains a representative sRGB image for each HSI image, which is rendered under a neutral daylight illuminant (D65).

The LR HSI images $\mathbf{X}_l$ were obtained by applying 1/4 bicubic interpolation to the HR HSI images $\mathbf{X}$ via Matlab's imresize function, following the settings in [17], which is also considered in many relevant works [6], [17], [27]–[29]. We used three different methods for generating the HR RGB images Y from the ground truth HR HSI images:

**RGB1**: The HR RGB images $\mathbf{Y}$ were generated by multiplying the HR HSI images $\mathbf{X}$ with the given spectral response matrix $\mathbf{F} \in \mathbb{R}^{B \times b}$ of Nikon D700, i.e., $\mathbf{Y} = \mathbf{XF}$, whose spectral response is defined in CIEXYZ color space.

**RGB2**: We adopted the sRGB renderings in the CAVE dataset as the guided HR RGB images $\mathbf{Y}$. Color rendering is needed to produce picture on the device to make it appear like the original scene. In this case, $\mathbf{Y} = \mathbf{XFT}$, where $\mathbf{T}$ is a $b \times b$ transformation matrix from CIEXYZ to CIERGB color space.

**RGB3**: We created the HR RGB images $\mathbf{Y}$ by applying Matlab's rgb2xyz to the sRGB renderings. In this case, $\mathbf{Y} = \mathbf{XFTL}$, where $\mathbf{L} \in \mathbb{R}^{b \times b}$ is a reverse transformation from CIERGB to CIEXYZ color space, which to some extent reduces the influence of $\mathbf{T}$.

The CSR matrices $\mathbf{C}$ of the schemes in RGB1, RGB2, and RGB3 are $\mathbf{F}$, $\mathbf{FT}$, and $\mathbf{FTL}$, respectively. We show an example of images generated according to these schemes in Fig 2. Clearly, RGB2 resembles a natural color image. These different methods of generating HR RGB images are useful to illustrate how different color schemes affect the performance of RGB-guided HSI-SR methods.

### A. Experimental Settings

In our experiments, we normalized all images of the CAVE dataset to the interval $[0, 1]$. To generate $\mathbf{W}$, we selected SSPSR [17] as the single HSI SR network. We used 22 images in the CAVE dataset for training SSPSR, and the remaining 10 for testing. For the parameters in the proposed TV$^3$ algorithm, we found by grid search that $\beta = 1$ and $\gamma = 2$ lead to the best performance.

The proposed method was compared against four SOTA methods including Bayesian sparse representation method BSR [8], matrix factorization method CNMF [4], tensor factorization method NLSTF [12] and deep learning method uSDN [14]. We used the code provided by the authors, but generated $\mathbf{X}_l$ via 1/4 bicubic downsampling, and the CSR function remained unchanged for different color schemes, which coincided with the RGB1 procedure. We also compared the proposed method against SSPSR to evaluate the performance gain achieved by our TV$^3$ block.

### B. Ablation Experiments

The proposed method contains three TV-based minimizations including $\|\mathbf{X}\|_{\text{TV}}$, $\|\mathbf{X} - \mathbf{W}\|_{\text{TV}}$, $\|\mathbf{X} - \mathbf{G}\|_{\text{TV}}$. In order to validate the effectiveness of these terms, we assess their contribution by considering the average performance over 10 testing images from CAVE dataset in RGB1. It depicts five different performance metrics: peak signal-to-noise ratio (PSNR), structural similarity (SSIM), root mean squared error (RMSE), spectral angle mapper (SAM) and dimensionless global relative error of synthesis (ERGAS). The higher (resp. lower) PSNR and SSIM (resp. RMSE, SAM, and ERGAS), the better the reconstruction quality.

Terms 1, 2, and 3 in Table I represent $\|\mathbf{X}\|_{\text{TV}}$, $\|\mathbf{X} - \mathbf{W}\|_{\text{TV}}$, $\|\mathbf{X} - \mathbf{G}\|_{\text{TV}}$, respectively. Different TV-based minimizations lead to a considerable performance improvement. Specifically,

TABLE I: Average performance among three different combinations of TV-based minimization over 10 testing images from CAVE dataset.

| Term 1 | Term 2 | Term 3 | Performance Metrics | | | | |
|---|---|---|---|---|---|---|---|
| | | | PSNR | SSIM | RMSE | SAM | ERGAS |
| ✓ | ✗ | ✗ | 39.12 | 0.963 | 3.445 | 4.158 | 3.154 |
| ✓ | ✓ | ✗ | 43.15 | 0.977 | 2.214 | 3.991 | 2.857 |
| ✓ | ✗ | ✓ | 43.09 | 0.976 | 2.345 | 4.016 | 2.943 |
| ✓ | ✓ | ✓ | **45.15** | **0.988** | **1.560** | **3.902** | **1.487** |

each term captures complementary information, as translated by the observed gains: Terms 2 and 3, when considered individually, lead to 4.03 dB and 3.97 dB gains in PSNR. But jointly they lead to an impressive 6.03 dB gain. As for other performance metrics, the gains are also considerable. Note that $\|\mathbf{X}\|_{\text{TV}}$ encodes the assumption that $\mathbf{X}$ has a small number of edges, so we treat it as a regular term in our experiments.

TABLE II: Average performance of all methods in different color schemes, i.e. RGB1, RGB2, and RGB3.

| **RGB1** | Method | | | | | |
|---|---|---|---|---|---|---|
| | SSPSR | BSR | uSDN | CNMF | NLSTF | **Ours** |
| PSNR | 38.96 | 37.16 | 38.70 | 43.37 | 44.48 | **45.15** |
| SSIM | 0.962 | 0.958 | 0.960 | 0.984 | 0.987 | **0.988** |
| RMSE | 3.475 | 5.565 | 3.615 | 2.380 | 1.774 | **1.560** |
| SAM | 4.158 | 10.833 | 11.276 | 5.735 | 4.068 | **3.902** |
| ERGAS | 3.155 | 4.697 | 3.233 | 1.812 | 1.610 | **1.487** |
| **RGB2** | Method | | | | | |
| | SSPSR | BSR | uSDN | CNMF | NLSTF | **Ours** |
| PSNR | 38.96 | 15.00 | 19.68 | 32.19 | 34.98 | **42.85** |
| SSIM | 0.962 | 0.450 | 0.587 | 0.910 | 0.908 | **0.966** |
| RMSE | 3.475 | 46.583 | 29.491 | 7.358 | 4.903 | **3.234** |
| SAM | 4.158 | 15.750 | 22.059 | 11.088 | 11.416 | **4.066** |
| ERGAS | 3.155 | 47.284 | 27.880 | 5.782 | 5.019 | **2.884** |
| **RGB3** | Method | | | | | |
| | SSPSR | BSR | uSDN | CNMF | NLSTF | **Ours** |
| PSNR | 38.96 | 26.29 | 29.78 | 42.56 | 40.49 | **44.42** |
| SSIM | 0.962 | 0.839 | 0.865 | 0.980 | 0.978 | **0.986** |
| RMSE | 3.475 | 16.194 | 12.575 | 2.712 | 2.738 | **1.708** |
| SAM | 4.158 | 19.625 | 19.054 | 6.084 | 4.750 | **4.032** |
| ERGAS | 3.155 | 14.187 | 19.054 | 1.988 | 2.520 | **1.558** |

### C. Experimental Results

We provide the averaged performance over the 10 test images in Table II. Table II shows that the proposed method outperforms all the competing methods in all quality indicators. In addition, it is more robust to the change of color images. Specifically when the RGB generator changes from RGB1 to RGB2, the PSNR of our method drops 2.3 dB (5%), of BSR drops 22.16 dB (60%), of uSDN drops 19.02 dB (49%), of CNMF drops 11.18 dB (26%), of NLSTF drops 9.8 dB (22%) respectively. The reason for the reduction in performance of all these methods is that CSR function in the CIERGB color space differs significantly from that in the CIEXYZ color space. Thus, methods that explicitly use the CSR function, i.e., BSR, uSDN and NLSTF, suffer severe performance degradation. Although CNMF does not require specifying a particular CSR matrix, it assumes it is a non-negative matrix. This assumption fails to hold when the transformation matrix $\mathbf{T}$ has negative values. Thus the performance of CNMF also degrades. In the case of the RGB3 transformation, the RGB renderings are in CIEXYZ color space, which results in better performance, while is still inferior to the case of RGB1. Specifically the PSNR of our method degrades 0.73 dB (1.6%) while BSR degrades 10.87 dB (29%), uSDN degrades 8.92 dB (23%), CNMF 0.81 dB (1.8%) and NLSTF 3.99 dB (8.9%) respectively. Table II also shows that the proposed method prominently improved the performance of W by SSPSR, the backbone method used by our algorithm in Fig. 1. Specifically, the PSNR and SSIM values of different color schemes (RGB1/RGB2/RGB3) are 6.19/3.89/5.46 dB and 0.026/0.004/0.024 higher than SSPSR, respectively. It also has notable gain in other quality indicators like RMSE, SAM and ERGAS.

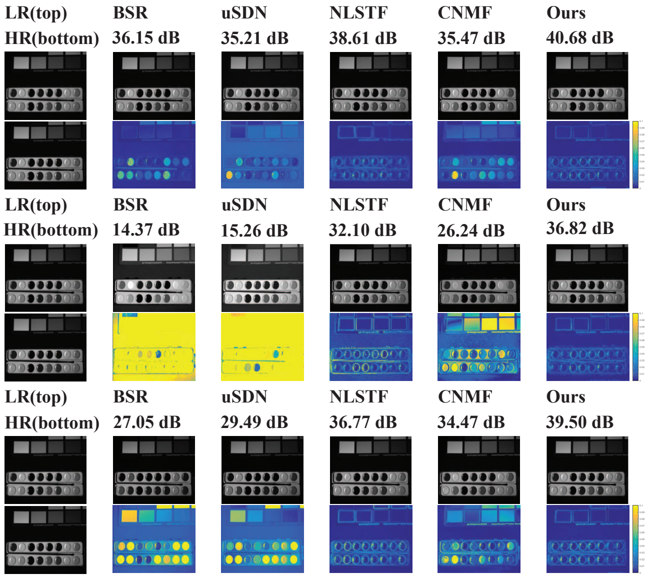


Fig. 3: Reconstructed images in RGB1 (top two rows), RGB2 (middle two rows) and RGB3 (bottom two rows). PSNR values indicated.

Fig. 3 displays visual results of all the algorithms on the ”paint_ms” image in the CAVE dataset. It shows the reconstructed results (top) and the corresponding error images (bottom) for better comparison. The superiority of the proposed method can be observed from the visual results. Specifically, the proposed method is closer to the ground truth than BSR, uSDN, NLSTF and CNMF for all the RGB1, RGB2 and RGB3 cases. Furthermore, it is noticed that the performance of the proposed method is stable in all the cases, while the performance of BSR and uSDN degrade significantly in the RGB2 and RGB3 cases.

On average, to upsample a $128 \times 128 \times 31$ LR HSI image with a $4\times$ factor and a $512 \times 512 \times 3$ RGB image, BSR requires 1425.71 seconds, CNMF requires 5.47 seconds, NLSTF requires 12.65 seconds, uSDN requires 231.37 seconds while our method requires 312.24 seconds. It should be pointed out that our method needs to obtain super-resolved HR HSI W by SSPSR, which requires 2.44 seconds. Although the proposed method needs more computation time than some other methods, it outperforms state-of-the-art methods for joint RGB-HSI super-resolution, and is also robust for various types of color images.

## IV. Conclusions

We presented a new framework for RGB-guided HSI SR. The proposed method uses a TV$^3$ optimization problem to merge the HR color information with the results of single HSI SR methods. The method requires no explicit modeling of the relation between HSI and RGB images, rendering it robust to the choice of the color spaces. Experimental results show the superior performance of the proposed method for RGB-guided HSI SR against state-of-the-art methods. Although the method is efficient, there is margin to improve its computing time, e.g., by unrolling the proposed method with neural network. We will follow this in future work.

## References

[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral remote sensing: principles and applications*. CRC press, 2007.

[2] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, "Hyperspectral image classification with markov random fields and a convolutional neural network," *IEEE Transactions on image processing*, vol. 27, no. 5, pp. 2354–2367, 2018.

[3] H. Van Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 44–51.

[4] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2011.

[5] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.

[6] K. Zhang, M. Wang, and S. Yang, "Multispectral and hyperspectral image fusion based on group spectral embedding and low-rank factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 3, pp. 1363–1371, 2016.

[7] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *CVPR 2011*. IEEE, 2011, pp. 2329–2336.

[8] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3631–3640.

[9] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multiband images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1117–1127, 2015.

[10] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3034–3047, 2019.

[11] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118–4130, 2018.

[12] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5344–5353.

[13] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2672–2683, 2019.

[14] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2511–2520.

[15] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1585–1594.

[16] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Hyperspectral image super-resolution with optimized rgb guidance," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 661–11 670.

[17] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *arXiv preprint arXiv:2005.08752*, 2020.

[18] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sensing*, vol. 9, no. 11, p. 1139, 2017.

[19] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2018, pp. 1–4.

[20] B. Funt, R. Ghaffari, and B. Bastani, "Optimal linear rgb-to-xyz mapping for color display calibration," in *Color and Imaging Conference*, vol. 2004, no. 1. Society for Imaging Science and Technology, 2004, pp. 223–227.

[21] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *IEEE transactions on image processing*, vol. 19, no. 9, pp. 2241–2253, 2010.

[22] M. Kretkowski, R. Jablonski, and Y. Shimodaira, "Development of an xyz digital camera with embedded color calibration system for accurate color acquisition," *IEICE transactions on information and systems*, vol. 93, no. 3, pp. 651–653, 2010.

[23] J. Jiang, D. Liu, J. Gu, and S. Süsstrunk, "What is the space of spectral sensitivity functions for digital color cameras?" in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 2013, pp. 168–179.

[24] M. Vella and J. F. C. Mota, "Robust Single-Image Super-Resolution via CNNs and TV-TV Minimization," *IEEE Transactions on Image Processing*, vol. 30, pp. 7830–7841, 2021.

[25] M. Vella and J. Mota, "Single image super-resolution via CNN architectures and TV-TV minimization," 2019.

[26] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[27] H. Irmak, G. B. Akar, and S. E. Yüksel, "A map-based approach to resolution enhancement of hyperspectral images," in *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2015, pp. 1–4.

[28] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model-based fusion of multi-and hyperspectral images using pca and wavelets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2652–2663, 2014.

[29] H. Kwon and Y.-W. Tai, "Rgb-guided hyperspectral image upsampling," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 307–315.