Pansharpening by Complementing Compressed Sensing with Spectral Correction

Naoko Tsukamoto *, Yoshihiro Sugaya and Shinichiro Omachi

Graduate School of Engineering, Tohoku University, Sendai, Miyagi 980-8579, Japan; sugaya@iic.ecei.tohoku.ac.jp (Y.S.); machi@ecei.tohoku.ac.jp (S.O.)
* Correspondence: natsuka@iic.ecei.tohoku.ac.jp

Received: 30 June 2020; Accepted: 18 August 2020; Published: 21 August 2020

Featured Application: Satellite image processing for change detection, target recognition, classification, map application, visual image analysis, etc.

Abstract: Pansharpening (PS) is a process used to generate high-resolution multispectral (MS) images from high-spatial-resolution panchromatic (PAN) and high-spectral-resolution multispectral images. In this paper, we propose a method for pansharpening by focusing on a compressed sensing (CS) technique. The spectral reproducibility of the CS technique is high due to its image reproducibility, but the reproduced image is blurry. Although methods of complementing this incomplete reproduction have been proposed, it is known that the existing method may cause ringing artifacts. On the other hand, component substitution is another technique used for pansharpening. It is expected that the spatial resolution of the images generated by this technique will be as high as that of the high-resolution PAN image, because the technique uses the corrected intensity calculated from the PAN image. Based on these facts, the proposed method fuses the intensity obtained by the component substitution method and the intensity obtained by the CS technique to move the spatial resolution of the reproduced image close to that of the PAN image while reducing the spectral distortion. Experimental results showed that the proposed method can reduce spectral distortion and maintain spatial resolution better than the existing methods.

Keywords: pansharpening; spectrum correction; intensity correction; compressed sensing; tradeoff process; IKONOS

1. Introduction

The optical sensors installed in satellites acquire panchromatic (PAN) and multispectral (MS) region images. PAN sensors observe a wide range of visible and near-infrared (NIR) regions as one band with a high spatial resolution, and the MS sensor observes multiple bands. Pansharpening (PS) is a method used to generate a high-resolution MS images from these two types of data. Due to physical constraints [1], MS sensors are not designed to acquire high-resolution images. In general, high-resolution MS images are obtained via the pansharpening process. PS techniques are used for change detection, target recognition, classification, backgrounds for map application, visual image analysis, etc.

PS has been studied for decades [2,3], and its methodologies can roughly be classified into four types. The first is methods used to generate a pansharpened image by substituting the intensity component of the MS images for that of the PAN image via component substitution. The intensity–hue–saturation (IHS) transform [4], generalized IHS method (GIHS) [5], Brovey transform [6], principal component analysis [6], and Gram–Schmidt transform [7] belong to this group. The second group contains methods used to extract high-frequency components of PAN images via multiresolution analysis.
(MRA) and then add them to the MS images. To extract high-frequency components, high-pass filter, decimated wavelet transforms [8], a “trous” wavelet transform [9], Laplacian pyramid [10,11], and non-subsampled contourlet transform [12] methods are used. The third group contains methods that use machine-learning techniques such as compressed sensing [13] and deep learning [14,15]. The fourth group is the hybrid methods that combine multiple methods described above. The methods proposed by Vivone et al. [16], Fei et al. [17], and Yin [18] are the ones that combined the component substitution method and the MRA. The combination of the MRA, convolutional neural network (CNN), and sparse modeling was proposed by Wang et al. [19], and the combination of the MRA and CNN was proposed by He et al. [20].

Recently, proposals for hybrid methods using machine-learning techniques have been increasing. Many methods based on compressed sensing (CS) have been proposed based on the 2006 theory [21]. Since Yang et al. [13] proposed a method for super-resolution, it has been frequently applied to PS. Li et al. proposed a method using a learning-free dictionary and CS [22], followed by the proposal of a method using dictionary learning and CS [23]. Although they showed the possibility of using a dictionary without learning, the feasibility of these methods was low because high-resolution MS images that were not realistically available were necessary for dictionary construction. Jiang et al. proposed a method using a high-resolution dictionary generated using a set of pairs of low-resolution MS images and high-resolution PAN images [24]. They then proposed a method for reconstruction by calculating sparse representations in two stages using a learning-free dictionary [25]. Zhu et al. constructed a pair of high-resolution and low-resolution learning-free dictionaries from PAN images [26,27]. Guo et al. proposed a method called online coupled dictionary learning (OCDL) [28], which iteratively performs dictionary learning and reconstruction processes until the reconstructed image becomes stable, based on the sparse representation (SR) theory described by Elad [29]. SR theory shows that better reconstruction results are obtained when the dictionary’s atoms are highly related to the reconstructed image. Vicinanze et al. proposed the generation of a learning-free dictionary using the high-resolution and low-resolution dictionaries as the detailed image information [30]. Ghahremani et al. proposed a learning-free dictionary with low-resolution PAN images and high-resolution detailed information extracted from the ripplet transform of the edges and textures of high-resolution PAN images [31]. Zhang et al. introduced non-negative matrix factorization and proposed estimation of a high-resolution matrix by solving an optimization problem of decomposing the matrix into basis and coefficient matrices [32]. Ayas et al. created a dictionary of high-resolution MS image features by incorporating the tradeoff parameter [24,26] and back-projection [13,33]. It was shown that the spectral distortion was reduced by incorporating back-projection. Yin proposed the cross-resolution projection and the offset [18]. The cross-resolution projection generates high-resolution MS images by assuming that the position estimated by CS is the same for high-resolution and low-resolution images. The offset is used to adjust the reconstructed image.

In these studies, the results of PS depended on the model selection and dictionary selection. Various studies on the structure and construction of dictionaries, model construction, and optimization processes have been conducted. In addition, it was pointed out that CS-based reconstruction does not guarantee the reproduction of the original image [13,29,34], which means that the spatial characteristics may not be incorporated accurately. The fact that it is not an exact reconstruction should be considered when the method is used for spectral analysis. The process of back-projection was introduced by Yang et al. [13] to improve the reconstructed image. This process was also incorporated by Ayas et al. [34] to reduce the spectral distortion. On the other hand, it is also known that ringing artefacts occur when back-projection is performed. In the PS process, low spectral distortion is important as well as high resolution. It is important to consider how to achieve the fidelity of the reproduced image to the ideal image in terms of the spectral and spatial resolution.

In this paper, we focused on reducing the spectral distortion more effectively than the back-projection for resolution enhancement of a visible light image. To this end, we propose a method for pansharpening by combining the CS technique and a component substitution method
that calculates the intensity with high spatial resolution and low spectral distortion to enhance the reproducibility. As the component substitution, spectrum correction using modeled panchromatic image (SCMP) [35] is introduced. The observation band of the PAN sensor of IKONOS, an earth observation satellite, was 526–929 nm. The red, green, blue, and NIR bands of the MS sensor of IKONOS were 632–698 nm, 505–595 nm, 445–516 nm, and 757–853 nm, respectively. Therefore, the PAN sensor covered the observation band of the MS sensor including NIR. The NIR information needs to be included when generating PS images using component substitution in order to avoid spectral distortion. The SCMP is a model that can correct this distortion. On the other hand, processing using the CS is expected to have high data reproducibility and it reflects the characteristics of the input image, while the resolution is lower than that generated by the SCMP. In order to improve the fidelity of the reproduced image to the original image, a tradeoff process was applied on the high-resolution intensity images obtained by the CS and the SCMP. It was found that a spatial resolution equivalent to that of PAN image was obtained and spectral distortion was reduced by the proposed method.

2. Materials and Methods

2.1. Image Datasets

Table 1 shows the two image datasets used for the experiments. The first was collected in May 2008 and covers the city of Nihonmatsu, Japan. The second was collected in May 2006 and covers the city of Yokohama, Japan. The two IKONOS images datasets were provided by the Japan Space Imaging Corporation, Japan. The spatial resolutions of the PAN and the MS images in these datasets were 1 m and 4 m, respectively. The original dataset contained PAN images with 1024 × 1024 pixels and MS images with 256 × 256 pixels for Nihonmatsu, and PAN images with 1792 × 1792 pixels and MS images with 448 × 448 pixels for Yokohama.

| Image          | Nihonmatsu | Yokohama |
|----------------|------------|----------|
| Original image | PAN image  | 1024 × 1024 | 1792 × 1792 |
|                | MS image   | 256 × 256  | 448 × 448  |
| Training image | PAN image  | 256 × 256  | 448 × 448  |
| (for high-resolution dictionary) |            |          |
| Test image     | PAN image  | 256 × 256  | 448 × 448  |
|                | MS image   | 64 × 64    | 112 × 112  |

The training image datasets had high-resolution PAN images with 256 × 256 pixels for Nihonmatsu, and high-resolution PAN images with 448 × 448 pixels for the Yokohama.

To evaluate the quality of the PS images, we experimented with the test images and original images according to the Wald protocol [36]. The test images were used to evaluate the numerical image quality, and the original images were used as reference images for numerical and visual evaluation. We regarded the original images as ground truth images. The spatial resolution of the test PAN images was reduced from 1 to 4 m and that of the test MS images was reduced from 4 to 16 m. Hence, the test image datasets had PAN images with 256 × 256 pixels and MS images with 64 × 64 pixels for the Nihonmatsu, and PAN images with 448 × 448 pixels and MS images with 112 × 112 pixels for the Yokohama. The training and test images were downsampled images of the original image with bicubic spline interpolation.

2.2. Compressed Sensing

Compressed sensing (CS) is a technique used to reconstruct unknown data from a small number of observed data. In theory, the original data can be estimated when the data is sparse [28].
We considered the problem of reconstructing an image $z^h \in \mathbb{R}^{m_0}$ with a higher resolution than the observed low-resolution image $z^l \in \mathbb{R}^{m_0}$ ($m_0 < n_0$). The relationship between the high-resolution image and the low-resolution image can be expressed by Equation (1).

$$z^l = SHz^h = Lz^h,$$

where $S$ is a downsampling operator and $H$ is a filter that lowers the resolution. At this time, since the dimensionality of $z^l$ is smaller than that of $z^h$, the solution cannot be uniquely determined.

Based on the compressed sensing theory, the high-resolution image $z^h$ is estimated by Equation (2) from the image element $D_h$ and sparse representation $a$. The element for reproducing the image is called atom $d_i \in \mathbb{R}^{n_0}$, and the set of atoms is called the dictionary $D_h \in \mathbb{R}^{n_0 \times N_d}$.

$$z^h = D_h a \text{ s.t.} \|a\|_0 \leq m_0,$$

Using Equation (2), Equation (1) can be expressed as

$$z^l = Lz^h = D_h a,$$

$$D_l = LD_h.$$  \hspace{1cm} (3)

$z^h$ can be reproduced by $z^l$ using the sparse representation $a$ obtained from Equation (4), in which the sparsity constraint is added to Equation (3).

$$\min_{a} \|a\|_0 \text{ s.t.} \|z^l - LD_h a\|_2^2 \leq \epsilon.$$  \hspace{1cm} (4)

Equation (4) can be solved using optimization methods.

2.3. Proposed Method

Our proposed method combines the advantages of super-resolution based on the theory of compressed sensing and component substitution. In this method, the high-resolution intensity obtained by the SCMP and the high-resolution intensity obtained by the CS-based method are linearly combined using the tradeoff process, and the obtained high-resolution intensity images are fused via the GIHS method to generate PS images. The flow of the proposed method is shown in Figure 1.

![Figure 1. Flow of the proposed method.](image-url)
2.4. Notation

In the proposed method, four features are extracted from low-resolution images, and the set of features is called the feature map. We used the gradient map proposed by Yang et al. [13] as the feature map. Let $I_{\text{high}} \in \mathbb{R}^{p \times 1}$ be a patch of size $p \times p$ extracted from a high-resolution image, and $I_{\text{low}} \in \mathbb{R}^{4p \times 1}$ be a set of four patches of size $p \times p$ extracted from the feature map. The $i$th training data patch $x_i \in \mathbb{R}^{4p^2 \times 1}$ is defined as $x_i = \begin{pmatrix} I_{\text{high}}^i \\ I_{\text{low}}^i \end{pmatrix}$, $X = \{x_1, \ldots, x_N\} \in \mathbb{R}^{5p^2 \times N}$, $X_{\text{high}} = \{I_{\text{high}}^1, \ldots, I_{\text{high}}^N\} \in \mathbb{R}^{5p^2 \times N}$, and $X_{\text{low}} = \{I_{\text{low}}^1, \ldots, I_{\text{low}}^N\} \in \mathbb{R}^{4p^2 \times N}$ represent the set of patches of the training data, high-resolution training data, and low-resolution training data, respectively, where $N$ is the number of training data. $x_i^\text{high}$ indicates the mean value of the intensity values of the $i$th training data patch $x_i^\text{high}$.

D = \begin{bmatrix} D_{\text{high}} \\ D_{\text{low}} \end{bmatrix} = \{d_1, \ldots, d_N\}, D \in \mathbb{R}^{5p^2 \times N}$ is called a dictionary and $d_i = \begin{pmatrix} d_i^\text{high} \\ d_i^\text{low} \end{pmatrix}$ is the $i$th atom of the dictionary where $N_d$ is the number of atoms. $d_i^\text{high} \in \mathbb{R}^{4p \times 1}$ is a high-resolution dictionary atom of size $p \times p$, and $d_i^\text{low} \in \mathbb{R}^{4p^2 \times 1}$ is a low-resolution dictionary atom of size $p \times p$. All the atoms are arranged in raster scan order. The high-resolution dictionary $D_{\text{high}} \in \mathbb{R}^{2p \times N_d}$ and the low-resolution dictionary $D_{\text{low}} \in \mathbb{R}^{4p^2 \times N_d}$ are defined as $D_{\text{high}} = \{d_1^\text{high}, \ldots, d_N^\text{high}\}$ and $D_{\text{low}} = \{d_1^\text{low}, \ldots, d_N^\text{low}\}$, respectively. $Y_{\text{low}} \in \mathbb{R}^{4p \times N_{\text{patch}}}$ indicates the feature map of the low-resolution input image to be reconstructed, where $N_{\text{patch}}$ is the number of patches of the input image. $I_{\text{sr}} \in \mathbb{R}^{2p \times N_{\text{patch}}}$ represents the reconstructed high-resolution image. $I_{\text{sr}} \in \mathbb{R}^{2p \times N_{\text{patch}}}$ represents the mean value of the intensity values of the reconstructed high-resolution image patch. The sparse representations are denoted as $\alpha \in \mathbb{R}^{N_d \times N_p}$ and $\beta \in \mathbb{R}^{N \times N_p}$, and $\lambda$ represents the sparsity regularization parameter.

2.5. Estimation of Coefficients for Intensity Correction

The coefficient estimated by SCMP is used for the intensity correction in the proposed method. It is possible to obtain an intensity correction value with little spectral distortion with SCMP. This coefficient is calculated by

$$\arg\min_{c} \|Ac - d\|_2^2 \text{ s.t. } c \geq 0,$$

where

$$A = \begin{bmatrix} -\text{MS}_{\text{air}}^\text{low} (1) & \text{MS}_{\text{blu}}^\text{low} (1) & \text{MS}_{\text{grn}}^\text{low} (1) & \text{MS}_{\text{red}}^\text{low} (1) \\ \vdots & \vdots & \vdots & \vdots \\ -\text{MS}_{\text{air}}^\text{low} (k) & \text{MS}_{\text{blu}}^\text{low} (k) & \text{MS}_{\text{grn}}^\text{low} (k) & \text{MS}_{\text{red}}^\text{low} (k) \\ \vdots & \vdots & \vdots & \vdots \\ -\text{MS}_{\text{air}}^\text{low} (N) & \text{MS}_{\text{blu}}^\text{low} (N) & \text{MS}_{\text{grn}}^\text{low} (N) & \text{MS}_{\text{red}}^\text{low} (N) \end{bmatrix} \in \mathbb{R}^{N \times 4},$$

$$c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} \in \mathbb{R}^{4 \times 1},$$

$$d = \begin{bmatrix} I_{\text{low}}^1 - \text{PAN}_{\text{high}}^\text{low} (1) \\ \vdots \\ I_{\text{low}}^k - \text{PAN}_{\text{high}}^\text{low} (k) \\ \vdots \\ I_{\text{low}}^N - \text{PAN}_{\text{high}}^\text{low} (N) \end{bmatrix} \in \mathbb{R}^{N \times 1},$$

where $k$ indicates the pixel position. $N$ is the number of pixels and $\text{PAN}_{\text{high}}^\text{low} \in \mathbb{R}^N$ is the downsampled PAN test image, of which the size is the same as that of $I_{\text{low}}$ obtained via the bicubic interpolation.
The suffixes nir, blu, grn, and red represent the NIR, blue, green, and red color components of the MS image, respectively. For $MS_{\text{nir}}$, $MS_{\text{blu}}$, $MS_{\text{grn}}$, and $MS_{\text{red}}$, test MS images are used. $I_{\text{low}}$ is calculated by

$$I_{\text{low}} = \frac{MS_{\text{low red}} + MS_{\text{low grn}} + MS_{\text{low blu}}}{3}$$  \hspace{1cm} (6)

Note that Equation (6) does not include NIR because it is the intensity of the RGB image.

2.6. Dictionary Learning

The high-resolution dictionary and the low-resolution dictionary were constructed via Equation (7) using the corresponding pair of training images. These are shown as Equation (8).

$$X_{\text{high}} = D_{\text{high}} \beta,$$
$$X_{\text{low}} = D_{\text{low}} \beta,$$

$$X = D \beta,$$  \hspace{1cm} (7, 8)

where $\beta \in \mathbb{R}^{N_d \times N_t}$ represents the sparse representation and $X = \begin{bmatrix} X_{\text{high}} \\ X_{\text{low}} \end{bmatrix}$, $D = \begin{bmatrix} D_{\text{high}} \\ D_{\text{low}} \end{bmatrix}$. The dictionary was obtained by solving the optimization problem of Equation (9), where the regularization conditions and constraints are added to Equation (8).

$$\min_{\beta, D} \frac{1}{2} \|X - D \beta \|^2_2 + \lambda \|\beta\|_1 \text{ s.t. } \|d_i\|_2 \leq 1, \ i = 1, 2, \cdots, N_t,$$  \hspace{1cm} (9)

where $\lambda$ is the normalization parameter.

The training data used for dictionary learning were training PAN images for the high-resolution dictionary and the feature map obtained from its corresponding low-resolution training PAN image for the low-resolution dictionary as shown in Table 1. $X_{\text{high}}$ and $X_{\text{low}}$ were obtained from a high-resolution training PAN image and its corresponding low-resolution training PAN image. Given a high-resolution training PAN image, it was divided into regions of size $p \times p$. The high-resolution patch $x_{\text{high}}^i$ was then obtained for each region by

$$x_{\text{high}}^i = x_{\text{high raw}}^i - x_{\text{high raw}}^i,$$

where $x_{\text{high raw}}^i$ is the $p \times p$ image and $x_{\text{high raw}}^i$ is the mean intensity value of $x_{\text{high raw}}^{i,Nt}$. $X_{\text{high}} = \{x_{\text{high}}^1, \cdots, x_{\text{high}}^{N_t}\}$ is obtained. Given the low-resolution training PAN image, the feature map was calculated with four filters of the first derivative and the second derivative defined by

$$F_1 = [-1, 0, 1], \ F_2 = F_1^T, \ F_3 = [1, 0, -2, 0, 1], \ F_4 = F_3^T,$$

where $T$ indicates transposition. The feature map was divided into patches of $p \times p$, and each patch was normalized. Since the feature map was calculated from the entire image, each patch contained the information of its adjacent patch. By arranging the normalized feature map in the raster scan order for each patch, $X_{\text{low}} = \{x_{\text{low}}^1, \cdots, x_{\text{low}}^{N_t}\}$, $x_{\text{low}}^i = \begin{bmatrix} F_1(i) \\ F_2(i) \\ F_3(i) \\ F_4(i) \end{bmatrix}$, $i = 1, 2, \cdots, N_t$ was obtained.

The algorithm of the dictionary learning is described as follows.

1. Obtain $X_{\text{high}}$ and $X_{\text{low}}$ from the training data. Each column of $X = \begin{bmatrix} X_{\text{high}} \\ X_{\text{low}} \end{bmatrix}$ is then normalized.

2. Set the initial value of the dictionary $D$. Random numbers that follow the Gaussian distribution with mean 0 and variance 1 are normalized for each patch region.
(3) Estimate the sparse representation $\beta$ by solving the optimization problem of Equation (10) by fixing the dictionary $D$.

$$
\beta = \arg\min_{\beta} \frac{1}{2} \|X - D\beta\|_2^2 + \lambda \|\beta\|_1
$$

(10)

(4) Estimate the dictionary $D$ by solving the optimization problem of Equation (11) by fixing the sparse representation $\beta$.

$$
D = \arg\min_{D} \|X - D\beta\|_2^2 \text{ s.t. } \|d_i\|_2 \leq 1, \ i = 1, 2, \cdots, N_t
$$

(11)

(5) Steps (3) and (4) are repeated. (In the experiment, we repeated 40 times.)

(6) The obtained dictionary $D$ is normalized for each patch and used as the final trained dictionary $D$.

2.7. Reconstruction Process

Assuming that the low-resolution image and the high-resolution image have the same sparse representation, the sparse representation of the low-resolution image can be obtained. The sparse representation $\alpha$ was estimated by solving the optimization problem of Equation (12).

$$
\alpha = \arg\min_{\alpha} \frac{1}{2} \|D_{\text{low}} \alpha - Y_{\text{low}}\|_2^2 + \lambda \|\alpha\|_1
$$

(12)

The reconstruction process is performed as follows. In this study, the resolution of RGB image was increased.

(1) The low-resolution RGB images are upsampled via bicubic interpolation to the size of the PAN image. The upsampled low-resolution intensity $I_{\text{low}}^{\text{up}}$ is calculated using Equation (6) with the upsampled RGB image.

(2) The feature map $Y_{\text{low}}$ is obtained from the upsampled low-resolution intensity $I_{\text{low}}^{\text{up}}$. After applying the four filters shown in Section 2.6 to $I_{\text{low}}^{\text{up}}$, the feature map $Y_{\text{low}} \in \mathbb{R}^{4p^2 \times N_{\text{patch}}}$ is obtained, where the overlap of adjacent patches are $p \times 1$ and $1 \times p$ for horizontal and vertical directions.

$I_{\text{low}}^{\text{patch}} \in \mathbb{R}^{p^2 \times N_{\text{patch}}}$ is then obtained from $I_{\text{low}}^{\text{up}}$ where the overlap of adjacent patches are $p \times 1$ and $1 \times p$ for horizontal and vertical directions.

(3) The sparse representation $\hat{\alpha}$ is calculated using Equation (12).

(4) The high-resolution intensity image $I_{sr}$ is obtained from the sparse representation $\hat{\alpha}$ and the high-resolution dictionary $D_{\text{high}}^{\text{patch}}$ by Equation (13), and it is normalized for each patch. The average value of the $j$th patch, $I_{sr}(j)$, is calculated with the upsampled low-resolution intensity $I_{\text{low}}^{\text{patch}}(j)$, and it is added to the patch of the high-resolution intensity $I_{sr}(j)$ using Equation (14).

$$
I_{sr}(j) = D_{\text{high}}^{\text{patch}} \hat{\alpha}(j), \quad \hat{\alpha} \in \mathbb{R}^{Nd \times N_{\text{patch}}}
$$

(13)

$$
I_{sr}(j) = I_{sr}(j) + I_{sr}(j)
$$

(14)

(5) Using the patches of the obtained high-resolution intensity $I_{sr}(j)$, the image is reconstructed. The mean value of the overlapped pixels is used as the value of the pixel in the adjacent overlapping patches.

2.8. Tradeoff Process

The intensity of SCMP is calculated using Equation (15) after obtaining the intensity $I_{\text{low}}^{\text{up}}$ via Equation (6).
were combined linearly using the tradeoff parameter $\tau$.

Equation (17). 

$$I_{\text{scmp}}(k) = \frac{I_{\text{up}}^\text{low}(k)}{PAN_{\text{low}}(k)}PAN_{\text{high}}^{\text{high}}(k),$$

where $k$ indicates the pixel position. The low-resolution PAN image, $PAN_{\text{low}}$, is obtained using Equation (16).

$$PAN_{\text{low}} = I_{\text{up}}^\text{low} + c_1 MS_{\text{up},\text{nir}} + c_2 MS_{\text{up},\text{blu}} + c_3 MS_{\text{up},\text{grn}} + c_4 MS_{\text{up},\text{red}},$$

where $I_{\text{up}}^\text{low}$, $MS_{\text{up},\text{nir}}$, $MS_{\text{up},\text{blu}}$, $MS_{\text{up},\text{grn}}$, and $MS_{\text{up},\text{red}}$ are the intensities of the low-resolution RGB image, NIR, blue, green, and red, respectively, and these are upsampled to the same sizes as those of the PAN image.

If the PAN image is corrected using the intensity $I_{\text{scmp}}$ obtained via SCMP, there will be little loss of spatial information. Since the intensity correction is performed appropriately on the image, the spectral distortion when using SCMP is smaller than that of the other component substitution methods. Figure 2 shows the intensity of the original RGB image, the intensity image generated by SCMP, and the intensity image obtained by CS of Nihonmatsu and Yokohama images. From this figure, it can be seen that the intensity obtained by CS had less spatial information than SCMP. In order to increase the quality of the intensity reproduced by CS using SCMP, the intensities obtained by the CS and the SCMP were combined linearly using the tradeoff parameter $\tau$. In the tradeoff process, the high-resolution intensity images were obtained by SCMP and CS, and these were linearly combined by Equation (17).

$$I_{\text{high}}^{\text{high}} = \tau I_{\text{up}} + (1 - \tau)I_{\text{scmp}} \text{ s.t. } |\tau| \leq 1$$

**Figure 2.** Intensity of RGB images of (a) Nihonmatsu, (b) Yokohama. (i) Intensity of the original image, (ii) intensity of SCMP, (iii) intensity obtained via compressed sensing.
2.9. Generalized IHS Method

For the fusion process, the generalized IHS method (GIHS) [5] proposed by Tu et al. was used. The intensity was calculated using Equation (6), and Equation (18) was applied to red-, green-, and blue-band RGB images.

\[
\begin{bmatrix}
MS_{\text{red}}^{\text{high}} \\
MS_{\text{grn}}^{\text{high}} \\
MS_{\text{blu}}^{\text{high}}
\end{bmatrix} = \begin{bmatrix}
MS_{\text{red}}^{\text{low}} + I_{\text{high}} - I_{\text{low}} \\
MS_{\text{grn}}^{\text{low}} + I_{\text{high}} - I_{\text{low}} \\
MS_{\text{blu}}^{\text{low}} + I_{\text{high}} - I_{\text{low}}
\end{bmatrix},
\]

where \(MS_{\text{red}}^{\text{high}}, MS_{\text{grn}}^{\text{high}}, MS_{\text{blu}}^{\text{high}}, MS_{\text{red}}^{\text{low}}, MS_{\text{grn}}^{\text{low}},\) and \(MS_{\text{blu}}^{\text{low}}\) are the intensities of the high-resolution and low-resolution RGB images.

3. Results

3.1. Experimental Setup

3.1.1. Quality Evaluation

Visual and numerical evaluations were performed according to the Wald protocol [36]. The original MS image was used as the reference image. Correlation coefficient (CC), universal image quality index (UIQI) [37], erreur relative global adimensionnelle de synthese (ERGAS) [38], and spectral angle mapper (SAM) [39] were used for numerical evaluation. These are major evaluation criteria and used in almost all PS-related research [2]. The CC is given by

\[
CC = \frac{1}{|B|} \times \sum_{b \in B} CC_b,
\]

\[
CC_b = \frac{\sum_{i=1}^{N} (O_b(i) - \overline{O}_b) \times (PS_b(i) - \overline{PS}_b)}{\sqrt{\sum_{i=1}^{N} (O_b(i) - \overline{O}_b)^2} \times \sqrt{\sum_{i=1}^{N} (PS_b(i) - \overline{PS}_b)^2}},
\]

where a value closer to 1.0 implies a smaller loss of the intensity correlation and a better result. \(N\) and \(|B|\) are the total number of pixels in the entire image for each band and the number of bands in the PS image, respectively. \(O_b(i)\) and \(\overline{O}_b\) denote the \(i\)th pixel value of the \(b\)-band reference image and its mean value, respectively, and \(PS_b(i)\) and \(\overline{PS}_b\) denote the \(i\)th pixel value of the \(b\)-band PS image and its mean value, respectively. UIQI is an index for measuring the loss of intensity correlation, intensity distortion, and contrast distortion and is given by

\[
UIQI = \frac{1}{|B|} \times \sum_{b \in B} UIQI_b,
\]

\[
UIQI_b = \frac{O_b - \overline{O}_b}{\sigma_{O_b}} \times \frac{2 \sigma_{PS_b}}{\overline{O}_b ^2 + \sigma_{PS_b} ^2} \times \frac{2 \sigma_{O_b} \sigma_{PS_b}}{O_b ^2 + \sigma_{PS_b} ^2},
\]

where \(\sigma_{O_b}\) and \(\sigma_{PS_b}\) are the standard deviation of the reference and PS images in the \(b\)-band, respectively, and \(\sigma_{O_b,PS_b}\) denotes the covariance of the reference and PS images in the \(b\)-band. A value closer to 1.0 implies that these losses are small. The size of the UIQI sliding window was \(8 \times 8\).

The ERGAS is given by

\[
ERGAS = 100 \times \frac{b}{7} \times \sqrt{\frac{1}{|B|} \times \sum_{b \in B} \left(\frac{\text{RMSE}_b}{PS_b} \right)^2},
\]

\[
\text{RMSE}_b = \frac{1}{N} \times \sum_{i=1}^{N} (O_b(i) - PS_b(i))^2,
\]

where \(\text{RMSE}_b\) is the root mean square error of the \(b\)-band PS image, \(\overline{O}_b\) is the mean value of the \(b\)-band reference image, \(PS_b(i)\) is the \(i\)th pixel value of the \(b\)-band PS image, and \(|B|\) is the number of bands in the PS image.
where \( h \) and \( l \) denote the spatial resolution of the PAN and MS images, respectively. The smaller the ERGAS value, the better the image quality. The SAM is an index for measuring spectral distortion and is given by

\[
\text{SAM} = \frac{1}{N} \sum_{i=1}^{N} \text{SAM}(i),
\]

\[
\text{SAM}(i) = \cos^{-1}\left( \frac{\sum_{b \in B} O_b(i) \times P_b(i)}{\sqrt{\sum_{b \in B} (O_b(i))^2} \times \sqrt{\sum_{b \in B} (P_b(i))^2}} \right)
\]  

(22)

If the value is closer to 0.0, the spectrum ratio of each band is closer to the reference image.

3.1.2. Dictionary Learning

We used PAN images from the Nihonmatsu and Yokohama datasets as training images for dictionary learning. For training image \( X \) and dictionary, the number of training times was 40, the number of atoms of the dictionary \( D \) was 1024, and the size of each atom was \( p = 4 \). The number of patches was 6433 (Nihonmatsu) and 10,000 (Yokohama). The training dataset was a set of 1000 patches randomly selected from these patches. As the training image \( X \), a training PAN image for the high-resolution dictionary was used as the high-resolution data \( X^{\text{high}} \), and a training PAN image for the low-resolution dictionary was used as the low-resolution data \( X^{\text{low}} \). The low-resolution image was generated by downsampling and upsampling via bicubic interpolation. The sparsity regularization parameter was \( \lambda = 0.1 \). The sparse representation \( \beta \) was calculated using Equation (10) by solving the L1 norm regularized least-squares problem, and the dictionary \( D \) was obtained using Equation (11) by solving the least-squares problem with quadratic constraints. We used the code provided by Lee et al. [40]. The learned high-resolution dictionaries are shown in Figure 3. For Nihonmatsu images, the dictionary created using Nihonmatsu images was used. For Yokohama images, the dictionary created using Yokohama images was used.

![Trained high-resolution dictionaries: (a) Nihonmatsu, (b) Yokohama.](image)

Figure 3. Trained high-resolution dictionaries: (a) Nihonmatsu, (b) Yokohama.

3.1.3. Reconstruction Process

The bicubic interpolation was used to upsample the intensity \( I^{\text{low}} \) of the low-resolution RGB image. The size of the patch of the input low-resolution data \( Y^{\text{low}} \) was \( p = 4 \), and the overlapping region of adjacent patches was \( p \times 1 \) for the horizontal direction and \( 1 \times p \) for the vertical direction. The filter
size used for the back-projection process was $5 \times 5$. The sparse representation $\alpha$ was calculated using Equation (12) by solving the L1 norm regularized least-squares problem. We used the code provided by Lee et al. [40].

3.1.4. Coefficients for Intensity Correction

The correction coefficients estimated by SCMP are shown in Table 2. These values were used for the fusion process.

| Coefficient | Nihonmatsu | Yokohama |
|-------------|------------|----------|
| $c_1$ (NIR) | 0.3857     | 0.3789   |
| $c_2$ (Blue)| 0.2199     | 0.2549   |
| $c_3$ (Green)| 0.1980   | 0.1123   |
| $c_4$ (Red) | 0.0486     | 0.1099   |

3.1.5. Tradeoff Parameter

In the tradeoff process, the intensity images obtained by SCMP and CS were combined by Equation (17) using the tradeoff parameter $\tau$ ($0 \leq \tau \leq 1$). The tradeoff parameter is a hyperparameter that should be determined in advance. In general, experimental validation is carried out via cross-validation when there are hyperparameters. Since we used two kinds of satellite images (Ninohmatsu and Yokohama), the value of $\tau$ was determined using one of these datasets, and the other dataset was processed using the determined value. In our experiment, the tradeoff parameter was determined using the correlation coefficient (CC) and ERGAS [38]. The results of applying the tradeoff parameter with a step size of 0.1 for image quality evaluation are shown in Table 3. From this result, it was found that the resolution and numerical evaluation were improved when the sum of squares of 1-CC and ERGAS was the smallest. This relationship is shown in Equations (23) and (24).

$$
\arg\min_{\tau} S(\tau) \text{ s.t. } 0 \leq S(\tau) \leq 2
$$

$$
S(\tau) = \left( \frac{1 - \text{CC}(\tau)}{\max(1 - \text{CC})} \right)^2 + \left( \frac{\text{ERGAS}(\tau)}{\max\text{ERGAS}} \right)^2
$$

The tradeoff parameter was determined as the value that minimizes Equation (24). Figure 4 displays the results of evaluation of the intensity of RGB images with various tradeoff parameters. The red line shows Ninohmatsu, the blue line shows Yokohama, the solid line is the result of CC, and the dotted line is the result of ERGAS. Figure 5 shows three images with some tradeoff parameter values. The lower the evaluation index $S$, the better the visual appearance. From these results, the value that minimized $S(\tau)$ was $\tau = 0.4$ for Ninohmatsu and $\tau = 0.3$ for Yokohama. Therefore, in the following experiments, $\tau = 0.3$ was used for Ninohmatsu and $\tau = 0.4$ was used for Yokohama.
Table 3. CC, ERGAS, and the evaluation index S that determines the tradeoff parameter. The best values are given in bold.

| Tradeoff Parameter | Nihonmatsu | Yokohama |
|--------------------|------------|----------|
|                    | CC | ERGAS | S | CC | ERGAS | S |
| 0.0                | 0.899 | 2.415 | 1.186 | 0.944 | 2.393 | 0.405 |
| 0.1                | 0.908 | 2.250 | 1.016 | 0.949 | 2.148 | 0.327 |
| 0.2                | 0.915 | 2.123 | 0.892 | 0.952 | 1.996 | 0.283 |
| 0.3                | 0.920 | 2.041 | 0.815 | 0.953 | 1.961 | **0.274** |
| 0.4                | **0.922** | **2.008** | **0.783** | 0.949 | 2.050 | 0.302 |
| 0.5                | 0.921 | 2.027 | 0.800 | 0.941 | 2.248 | 0.371 |
| 0.6                | 0.916 | 2.097 | 0.873 | 0.926 | 2.529 | 0.491 |
| 0.7                | 0.906 | 2.213 | 1.011 | 0.903 | 2.670 | 0.679 |
| 0.8                | 0.891 | 2.369 | 1.228 | 0.871 | 3.251 | 0.963 |
| 0.9                | 0.871 | 2.557 | 1.550 | 0.828 | 3.660 | 1.385 |
| 1.0                | 0.846 | 2.769 | 2.000 | 0.775 | 4.088 | 2.000 |

Figure 4. The values of CC and ERGAS obtained in the tradeoff process. CC: solid line, ERGAS: dotted line; Nihonmatsu: blue line, Yokohama: red line.
\[ \tau = 0.4, S = 0.783 \]
\[ \tau = 0.5, S = 0.8 \]
\[ \tau = 0.4, S = 0.302 \]
\[ \tau = 0.3, S = 0.815 \]
\[ \tau = 0.3, S = 0.274 \]

**Figure 5.** Comparison of intensity of RGB images against the tradeoff parameter. \( S \) represents the index that determines the tradeoff parameter. (a) Nihonmatsu, (b) Yokohama.

### 3.2. Experimental Results

Table 4 shows the effect of the back-projection (BP). The results of the intensity \( I_p \) obtained using the sparse representation (SR) and the intensity obtained by repeating BP ten times with a size 5 filter are shown. These were compared by CC, UIQI, and ERGAS, and applying BP was better in all cases. Table 5 shows the effect of the tradeoff process (TP). The intensity images generated by SCMP, SR, and TP are shown. TP was the best in all evaluations. Table 6 shows the comparison of BP and TP. TP was better in every case.

Tables 7 and 8 show numerical evaluations of the existing methods and the proposed method. The existing methods include fast IHS [5], Gram–Schmidt method (GS) [7], band-dependent spatial detail (BDSD) [41], weighted least-squares (WLS)-filter-based method (WLS) [42], multiband images with adaptive spectral-intensity modulation (MDSIm) [43], spectrum correction using modeled panchromatic image (SCMP) [35], image super-resolution via sparse representation (ISSR) [13] using natural images (Dict-natural) and corresponding images (Dict-self), the sparse representation of injected details (SR-D) [30], and the method of sparse representation described by Ayas et al. (SRayas) [34].

**Table 4.** Numerical evaluation of image intensities with or without the back-projection process of the intensity obtained via sparse representation. The best values are given in bold.

| Method  | CC Nihonmatsu | CC Yokohama | UIQI Nihonmatsu | UIQI Yokohama | ERGAS Nihonmatsu | ERGAS Yokohama |
|---------|----------------|-------------|-----------------|---------------|------------------|----------------|
| Ideal   | 1.0            | 1.0         | 1.0             | 0.0           |                  |                |
| SR      | 0.846          | 0.775       | 0.807           | 0.709         | 2.769            | 4.088          |
| SR+BP   | 0.853          | 0.785       | 0.822           | 0.742         | 2.709            | 4.009          |
Table 5. Numerical evaluation of image intensities obtained by SCMP, SR, and the tradeoff process. The best values are given in bold.

| Method  | CC Nihonmatsu | CC Yokohama | UIQI Nihonmatsu | UIQI Yokohama | ERGAS Nihonmatsu | ERGAS Yokohama |
|---------|---------------|-------------|-----------------|---------------|------------------|---------------|
| Ideal   | 1.0           | 1.0         |                 |               |                  |               |
| SCMP    | 0.899         | 0.944       | 0.779           | 0.915         | 2.415            | 2.393         |
| SR      | 0.846         | 0.775       | 0.807           | 0.709         | 2.769            | 4.088         |
| TP      | 0.920         | 0.949       | 0.825           | 0.926         | 2.041            | 2.050         |

Table 6. Comparison of the back-projection and the tradeoff process. The best values are given in bold.

| Method | CC Nihonmatsu | CC Yokohama | UIQI Nihonmatsu | UIQI Yokohama | ERGAS Nihonmatsu | ERGAS Yokohama |
|--------|---------------|-------------|-----------------|---------------|------------------|---------------|
| Ideal  | 1.0           | 1.0         |                 |               |                  |               |
| BP     | 0.853         | 0.785       | 0.822           | 0.742         | 2.709            | 4.009         |
| TP     | 0.920         | 0.949       | 0.825           | 0.926         | 2.041            | 2.050         |

Table 7. Numerical evaluation of the existing methods and the proposed method by CC and UIQI. The highest scores are printed in bold, and the second highest scores are underlined.

| Method          | CC Nihonmatsu | CC Yokohama | UIQI Nihonmatsu | UIQI Yokohama |
|-----------------|---------------|-------------|-----------------|---------------|
| Ideal           | 1.0           | 1.0         |                 |               |
| fast IHS [5]    | 0.783         | 0.914       | 0.717           | 0.901         |
| GS [7]          | 0.508         | 0.860       | 0.373           | 0.838         |
| BDSD [41]       | 0.860         | 0.887       | 0.864           | 0.851         |
| WLS [42]        | 0.866         | 0.884       | 0.870           | 0.759         |
| MDSIm [43]      | 0.791         | 0.865       | 0.736           | 0.823         |
| SCMP [35]       | 0.883         | 0.928       | 0.864           | 0.910         |
| ISSR (Dict-natural) [13] | 0.831   | 0.765       | **0.918**       | 0.704         |
| ISSR (Dict-self) [13] | 0.831  | 0.759       | **0.917**       | 0.701         |
| SR-D            | 0.845         | 0.786       | 0.914           | 0.710         |
| SRayas [34]     | 0.787         | 0.758       | 0.851           | 0.703         |
| Proposed        | **0.906**     | **0.937**   | **0.903**       | **0.918**     |

The size of the local estimation of the distinct block of BDSD was 256 × 256 for Nihonmatsu and 448 × 448 for Yokohama. For the ISSR settings, the training images for the dictionary included training PAN images and natural images, the number of training times was 40, the number of atoms in the dictionary was 1024, the atom size of the dictionary was 4 × 4 × 5, the sparsity regularization parameter was 0.1, randomly selected 1000 training image patches were used, the upscale was 4 (ratio of resolution of PAN images and MS images of IKONOS), overlap pixel of patches in the reconstruction process was \( \mathbb{R}^{5 \times 5} \) in horizontal direction and \( \mathbb{R}^{p \times 1} \) in vertical direction, the size of the back-projection filter was 5 × 5, and the number of iterations was 10. For SR-D, high-resolution and low-resolution dictionaries were constructed from the original PAN images without training. The atom size of the high-resolution dictionary was \( \mathbb{R}^{28 \times 28 \times 4} \) and the overlapping areas of the adjacent atoms were \( \mathbb{R}^{16 \times 16} \) and \( \mathbb{R}^{p \times 16} \); the atom size of the low-resolution dictionary was \( \mathbb{R}^{7 \times 7 \times 4} \) and the overlapping area of the adjacent atoms were \( \mathbb{R}^{4 \times 4} \) and \( \mathbb{R}^{p \times 4} \). For the SRayas setting, the original IKONOS MS images were used as the training images for the dictionary, the number of training times was 20, the number of atoms in the dictionary...
was 4096, the size of the dictionary atom was $8 \times 8 \times 4$, the sparsity regularization parameter was $\lambda = 0.15$, the number of training image patches was 2000, the upscale was 4 (ratio of resolution of PAN images and MS images of IKONOS), $\beta = 0.25$, the weight of each spectral band of IKONOS was $w = [0.1071, 0.2646, 0.2696, 0.3587]$, the overlap pixel in the patch of reconstruction process was 0, the back-projection filter size was $3 \times 3$, and the number of repetitions was 20. These settings of the existing methods followed those described in the original papers except the distinct block size of BDSD. The code of Vivone et al. [2] was used for GS and BDSD, and the code of Yang et al. [13] was used for ISSR.

Table 8. Numerical evaluation of the existing methods and the proposed method by ERGAS and SAM. The highest scores are printed in bold, and the second highest scores are underlined.

| Method                  | ERGAS (Nihonmatsu) | SAM (Nihonmatsu) | ERGAS (Yokohama) | SAM (Yokohama) |
|-------------------------|--------------------|------------------|------------------|----------------|
|                         | Ideal              |                  | Ideal            |                |
| fast IHS [5]            | 3.471              | 0.0              | 1.898            | 2.367          |
| GS [7]                  | 5.046              | 3.563            | 2.748            | 1.950          |
| BDSD [41]               | 3.026              | 3.361            | 1.925            | 2.326          |
| WLS [42]                | 2.815              | 3.816            | 1.699            | 2.123          |
| MDSIm [43]              | 3.321              | 3.515            | **1.658**        | 2.189          |
| SCMP [35]               | 2.673              | 2.697            | 1.753            | 2.176          |
| ISSR (Dict-natural) [13]| 3.118              | 4.419            | 1.749            | 2.324          |
| ISSR (Dict-self) [13]   | 3.124              | 4.474            | 1.750            | 2.331          |
| SR-D                    | 3.007              | 4.226            | 1.897            | 2.649          |
| SRayas [34]             | 3.497              | 4.476            | 1.765            | 2.299          |
| Proposed                | **2.336**          | **2.424**        | **1.688**        | **2.159**      |

Figure 6. Scores of quality metrics.

In Tables 7 and 8, the highest scores are printed in bold, and the second highest scores are underlined. GS was generally not good except for the SAM of Yokohama. The results of BDSD were unremarkable but stable. SCMP was stable and gave good results. In ISSR, differences in training images had little effect on results. SRayas did not perform as well overall as ISSR and SRayas. Although some results, such as the CC of SR-D of Table 7, MDSIm of Table 8, and SAM of GS were better in part
than the proposed method, in many other cases they were less accurate than the proposed method. The results of the proposed method were generally good, although there are some differences due to the tradeoff parameter.

Figure 6 shows the ranking of the quality metric of the numerical evaluations of Tables 7 and 8, except for the proposed method. For each test, the best result was worth three points, the second-best result was worth two points, and the third-best result was worth one point. The highest score was 24 points. This figure shows that only three methods, WLS, SCMP, and the proposed method, were good for both of the images, and the proposed method got the highest score.

Figure 7. Reference and pansharpened Nihonmatsu images: (a) original PAN image (reference image), (b) original RGB image, (c) ground truth RGB image, (d) RGB image upsampled by bicubic interpolation, (e) fast IHS, (f) GS, (g) BDSD, (h) WLS, (i) MDSIm, (j) SCMP, (k) ISSR (natural images were used for training), (l) ISSR (Nihonmatsu images were used for training), (m) SR-D, (n) SRayas method, (o) proposed method ($\tau = 0.3$).
From the results in Table 4, it was found that the back-projection (BP) was effective in improving spectral distortion. Furthermore, since the TP was better than the existing methods, it was clarified that these methods complementarily improve the spectral distortion. From the results in Table 5, it was found that the tradeoff process (TP) is effective in improving spectral distortion.

![Figure 8](image_url)

**Figure 8.** Reference and pansharpened Yokohama images: (a) original PAN image (reference image), (b) original RGB image, (c) ground truth RGB image, (d) RGB image upsampled by bicubic interpolation, (e) fast IHS, (f) GS, (g) BDSD, (h) WLS, (i) MDSIm, (j) SCMP, (k) ISSR (natural images were used for training), (l) ISSR (Yokohama images were used for training), (m) SR-D, (n) SRayas method, (o) proposed method ($\tau = 0.4$).

Figures 7 and 8 show the reference and PS images from the Nihonmatsu and Yokohama datasets, respectively. Since the images were small, the enlarged image surrounded by the yellow frame of the original RGB image (b) is shown in (c)–(o). In Figure 7, in GS (f), the color of green was darker in the rice field area (indicated by the green arrow), while the forest area was whitish. BDSD (g) was more blurred than other images. In MDSIm (i), the color of the forest area was also whitish (indicated by the yellow arrow). In Figure 8, the vegetation area was whitish in GS (f) and MDSIm (i) (indicated by the red arrow). WLS (h) looked hazy. SR-D (m) had lower resolution than the other methods using sparse representation with back-projection. In both Figures 7 and 8, the results of ISSR (k) (l) and SRayas (n) had ringing artefacts. Other images appeared to be reproduced without problems.

4. Discussion

From the results in Table 4, it was found that the back-projection (BP) was effective from the viewpoint of improving spectral distortion. From the results in Table 5, it was found that the tradeoff process (TP) is effective in improving spectral distortion. Furthermore, since the TP was better than the individual methods of sparse representation (SR) and SCMP, it was clarified that these methods complementarily improve the spectral distortion. From the results in Table 6, it was found that the...
TP improved spectral distortion more than BP. In the results shown in Tables 7 and 8, the method using GIHS (WLS and SCMP) was better than the existing methods using SR. In addition, it was found that the proposed method, the linear combination of SR and SCMP, gave better results than SCMP alone. One of the problems to be solved in PS processing is the independence of the processed image. In other words, it is important to obtain stable and good results rather than obtaining good results only on a specific image. Although there are some methods shown in Tables 7 and 8 that gave better results than the proposed method, comparing the other evaluation results shows that the results were inconsistent. The reason why the evaluation results were so different could be that these processing methods depend on the processed image.

As shown in Figures 7 and 8, it was found that the reproduction of vegetation area by the GS was unstable. Since the image quality differs depending on the size of the local estimation on the distinct blocks used in BDSD, we evaluated the visibly good images with good numerical values, but the resolution of the image was low. The WLS gave good numerical results with two images, but the images were blurred. Both ISSR and SRayas using BP generated ringing artifacts. Other images seemed to be reproduced without problems in resolution and color. Among them, the proposed method gave the best results in the numerical evaluation.

In this method, resolution enhancement was achieved by using the visible and NIR regions. On the other hand, it can be applied only to the resolution enhancement of RGB images, and not NIR images.

5. Conclusions

In this paper, we proposed a method for pansharpening based on CS theory. In the proposed method, the intensity obtained from the component substitution method and the intensity obtained via the method based on CS theory are fused to reproduce the intensity close to the original. We introduced SCMP as the intensity substitution method and used the tradeoff process for image fusion. Experimental results showed that the proposed method outperformed existing methods in terms of numerical and visual evaluation. The proposed method was also effective for satellites with panchromatic sensors (observed areas are visible and NIR regions) and multispectral sensors (observed areas are red, blue, green, and NIR bands) like IKONOS.

Generally, the intensity image generated by a CS-based method is blurrier than the intensity image generated by the component substitution method, because component substitution captures the intensity of the PAN image. On the other hand, it is expected that the spectral distortion of the intensity image generated by the CS-based method will be lower than that of the image generated by the component substitution method. Since complete restoration is not guaranteed, in order to get the image close to a complete reproduction, back-projection methods can be used. However, they may cause ringing artifacts. Based on these considerations, our proposed method combines the intensities generated by the CS-based method and the component-substitution-based method via the tradeoff process instead of the back-projection to achieve both an improvement of spatial resolution and a reduction of spectral distortion. Experimental results show that the tradeoff process was more effective than the back-projection in generating a pansharpened image of which the spatial resolution was equivalent to that of the PAN image and reducing spectral distortion. Improvement of the accuracy by parameter tuning is important future work.

Author Contributions: Methodology, N.T. and Y.S.; Software, N.T.; Validation, Y.S.; Writing—Original Draft Preparation, N.T.; Writing—Review and Editing, Y.S. and S.O.; Supervision, S.O.; Project Administration, S.O.; Funding Acquisition, S.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the JSPS KAKENHI Grant Number 18K19772 and the Yotta Informatics Project by MEXT, Japan.

Acknowledgments: The authors thank the Japan Space Imaging Corporation and Space Imaging, LLC for providing the images.

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Amro, I.; Mateos, J.; Vega, M.; Molina, R.; Katsaggelos, A.K. A survey of classical methods and new trends in pansharpening of multispectral images. *EURASIP J. Adv. Signal Process.* 2011, 2011, 79. [CrossRef]

2. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 2565–2586. [CrossRef]

3. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; et al. Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geosci. Remote Sens. Mag.* 2019, 7, 6–39. [CrossRef]

4. Pohl, C.; Van Genderen, J.L. Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* 1998, 19, 743–757. [CrossRef]

5. Zhou, J.; Civco, D.L.; Silander, J.A. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *Int. J. Remote Sens.* 1998, 19, 743–757. [CrossRef]

6. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. U.S. Patent 6,011,875, 4 January 2000.

7. Mallat, S.G. A theory for multiResolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 1989, 17, 674–693. [CrossRef]

8. Shensa, M.J. The Discrete Wavelet Transform: Wedding the À Trous and Mallat Algorithms. *IEEE Trans. Signal Process.* 1992, 40, 2464–2482. [CrossRef]

9. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super—Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2016, 38, 295–307. [CrossRef] [PubMed]

10. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by convolutional neural networks. *Remote Sens.* 2016, 8, 594. [CrossRef]

11. Vivone, G.; Addesso, P.; Restaino, R.; Dalla Mura, M.; Chanussot, J. Pansharpening Based on Deconvolution for Multiband Filter Estimation. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 540–553. [CrossRef]

12. Wei, S.; Yin, H.; Zhang, J.; Liu, D.; Hu, J. Convolutional Sparse Representation of Injected Details for Pansharpening. *IEEE Geosci. Remote Sens. Lett.* 2019, 16, 1595–1599. [CrossRef]
25. Jiang, C.; Zhang, H.; Shen, H.; Zhang, L. Two-step sparse coding for the pan-sharpening of remote sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2014, 7, 1792–1805. [CrossRef]
26. Zhu, X.X.; Bamler, R. A sparse image fusion algorithm with application to pan-sharpening. IEEE Trans. Geosci. Remote Sens. 2013, 51, 2827–2836. [CrossRef]
27. Zhu, X.X.; Grohnfeldt, C.; Bamler, R. Exploiting Joint Sparsity for PanSharpening: The J-SparseFI Algorithm. IEEE Trans. Geosci. Remote Sens. 2016, 54, 2664–2681. [CrossRef]
28. Guo, M.; Zhang, H.; Li, J.; Zhang, L.; Shen, H. An online coupled dictionary learning approach for remote sensing image fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2014, 7, 1284–1294. [CrossRef]
29. Elad, M. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing; Springer: New York, NY, USA, 2010; ISBN 978-1441970107.
30. Vicinanza, M.R.; Restaino, R.; Vivone, G.; Dalla Mura, M.; Chanussot, J. A pansharpening method based on the sparse representation of injected details. IEEE Geosci. Remote Sens. Lett. 2015, 12, 180–184. [CrossRef]
31. Ghahremani, M.; Ghassemian, H. Remote Sensing Image Fusion Using Ripplet Transform and Compressed Sensing. IEEE Geosci. Remote Sens. Lett. 2015, 12, 502–506. [CrossRef]
32. Zhang, K.; Wang, M.; Yang, S.; Xing, Y.; Qu, R. Fusion of Panchromatic and Multispectral Images via Coupled Sparse Non-Negative Matrix Factorization. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2016, 9, 5740–5747. [CrossRef]
33. Irani, M.; Peleg, S. Motion analysis for image enhancement: Resolution, occlusion, and transparency. J. Vis. Commun. Image Represent. 1993, 4, 324–335. [CrossRef]
34. Ayas, S.; Gormus, E.T.; Ekinci, M. An efficient pan sharpening via texture based dictionary learning and sparse representation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2018, 11, 2448–2460. [CrossRef]
35. Tsukamoto, N.; Sugaya, Y.; Omachi, S. Spectrum Correction Using Modeled Panchromatic Image for Pansharpening. J. Imaging 2020, 6, 20. [CrossRef]
36. Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. Photogramm. Eng. Remote Sens. 1997, 63, 691–699.
37. Wang, Z.; Bovik, A.C. A Universal Image Quality Index. IEEE Signal Process. Lett. 2002, 9, 81–84. [CrossRef]
38. Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In Proceedings of the Third Conference Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images; Ranchin, T., Wald, L., Eds.; SEE/URISCA: Nice, France, 2000; pp. 99–103.
39. Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data. Remote Sens. Environ. 1993, 44, 145–163. [CrossRef]
40. Lee, H.; Battle, A.; Raina, R.; Ng, A.Y. Efficient sparse coding algorithms. In Proceedings of the 19th International Conference on Neural Information Processing Systems; Schölkopf, B., Platt, J.C., Hoffman, T., Eds.; MIT Press: Cambridge, MA, USA, 2006; pp. 801–808.
41. Garzelli, A.; Nencini, F.; Capobianco, L. Optimal MMSE pan sharpening of very high resolution multispectral images. IEEE Trans. Geosci. Remote Sens. 2008, 46, 228–236. [CrossRef]
42. Song, Y.; Wu, W.; Liu, Z.; Yang, X.; Liu, K.; Lu, W. An adaptive pansharpening method by using weighted least squares filter. IEEE Geosci. Remote Sens. Lett. 2016, 13, 18–22. [CrossRef]
43. Yang, Y.; Wu, L.; Huang, S.; Tang, Y.; Wan, W. Pansharpening for Multiband Images with Adaptive Spectral-Intensity Modulation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2018, 11, 3196–3208. [CrossRef]