Evaluation of colour space effect on estimation accuracy of hyperspectral image by dimension extension based on RGB image

Ryoji Sato\textsuperscript{a}, Yuri Hamada\textsuperscript{a}, Takashi Kaburagi\textsuperscript{b} and Yosuke Kurihara\textsuperscript{a}

\textsuperscript{a}Department of Industrial and Systems Engineering, Aoyama Gakuin University, Sagamihara, Japan; \textsuperscript{b}Department of Natural Sciences, International Christian University, Tokyo, Japan

\textbf{ABSTRACT}

Recently, the utilization of hyperspectral images containing several hundred wavelength information has been increasing in various fields. If a hyperspectral image can be estimated from a low-cost RGB image that has only R, G, and B wavelength information without using a hyperspectral camera, it would be useful in various fields. Herein, we propose a hyperspectral image estimation method based on RGB images, wherein RGB components and YUV colour space information calculated from the RGB are applied to a neural network for tuning, and the hyperspectral image is estimated by inputting the output from the tuning neural network to a decoding function of the trained autoencoder. To evaluate the estimation accuracy of hyperspectral images based on differences in the combination of RGB and colour space models, we conducted validity experiments for the estimation of hyperspectral images in three scenarios with different colour spaces: RGB and YUV, RGB and HSV, only RGB. The results showed that the scenario with RGB and YUV colour space exhibited the highest estimation accuracy of 0.913 by averaging all similarities for wavelength among the three scenarios; thus, the validity of the proposed method as an estimation method for hyperspectral images was verified.

\textbf{ARTICLE HISTORY}

Received 21 October 2021
Revised 24 January 2022
Accepted 21 February 2022

\textbf{KEYWORDS}

Hyperspectral image; neural network; autoencoder; YUV colour space; HSV colour space

\section{1. Introduction}

Recently, hyperspectral images have been receiving increased attention as it contains several hundred wavelength information. Compared with RGB images, which have three-wavelength information, hyperspectral images have a wide range of wavelengths; thus, it is possible to recognize and classify some specific phenomena using hyperspectral images with high accuracy. Various studies utilizing hyperspectral images instead of RGB images have been conducted in various fields. In [1–3], certain diseases or stress were diagnosed noninvasively using hyperspectral images in the medical field. In [4–7], certain objects or targets that cannot be recognized by RGB images were recognized and tracked using hyperspectral images in the security control field. In [8–10], analysis systems of soil or plants for environmental protection or application for agriculture were designed using hyperspectral images of land, plants, or crops. It should be noted that hyperspectral cameras are utilized for capturing hyperspectral images, and they cost approximately 5–10 million yen. Hence, it is difficult to install hyperspectral cameras for extended fields. Therefore, if a hyperspectral image can be estimated from a low-cost RGB image that has only R, G, and B wavelength information without using a hyperspectral camera, it would be useful in various fields.

To extend low-dimensional information of the images to high-dimensional information, several studies have been conducted on colourization, in which RGB images are estimated by extending one dimension of monochrome images to three dimensions of RGB images. In [11–17], high-quality colourization methods were proposed using a neural network. In [18,19], colourization methods using generative adversarial networks were proposed and it showed high accuracy. It is extremely difficult to apply the methods in these studies directly to the estimation of hyperspectral images with wavelength information in several hundred dimensions from RGB images with three wavelengths. Additionally, the computational cost for training and number of training datasets would be very expensive if the colourization method, which is a dimension expansion from one dimension to three dimensions, is adopted to perform dimension expansion from three dimensions to several hundred dimensions because the rate of dimensional expansion is more than a hundred times greater. To accelerate learning, the compressed values of high-dimensional data in hyperspectral images can be used as feature values. Because the closed-form equations for compressing high-dimensional hyperspectral images are unknown, we utilize machine learning methods (i.e. neural network and autoencoder) to generate models that will...
transform RGB images into the desired compressed values. To ensure the effective training of the machine learning models, we add additional information to supplement RGB components, which are insufficient as input values for estimating the high-dimensional wavelengths of hyperspectral images. As additional information, we focus on two colour spaces with different perspectives and independent information relative to the RGB space, namely a colour space with luminance and colour difference signals, and a colours space with hue, saturation, and lightness. YUV colour space information is utilized for image encoding strategies as an alternative to common image compression schemes [20] and an effective novel content-aware chroma reconstruction method for screen content images [21]. HSV colour space information is utilized for skin colour enhancement to make skin colours displayed on large-screen flat-panel TVs consistent [22] and an image defogging algorithm with colour correction for video processing [23]. Furthermore, we proposed an estimation method for hyperspectral images with 462-dimensional wavelength information using neural networks from the three-dimensional wavelength information of RGB images [24]. In [24], a high-dimensional wavelength image was estimated by applying not only RGB information but also HSV colour space information calculated from RGB to the neural network. In addition, the estimation accuracy due to differences in the characteristics of the light source in the image capturing environment was compared.

In this study, we propose a method for the estimation of hyperspectral images based on RGB image and YUV colour space information calculated from the RGB by utilizing a neural network and an autoencoder’s decoding function. In the learning phase, images captured by a hyperspectral camera are learned. Next, a hyperspectral image is estimated from an RGB image utilizing a neural network and autoencoder decoding function. At this time, a hyperspectral camera is not required for the estimation phase. We then evaluate the effect of the colour space on the estimation accuracy of the hyperspectral image due to differences in the combination of RGB and colour space models.

2. Proposed method

Figure 1 shows the architecture of the proposed method for estimating a hyperspectral image from an RGB image. The proposed method mainly consists of a tuning part and dimension extension part. The tuning and dimension extension parts are implemented utilizing a neural network model and the decoding function of the autoencoder, and both parts are trained in the learning phase in advance as shown in Figure 1(a). The autoencoder in the dimension extension part is trained by a hyperspectral image; then, the neural network in the tuning part is trained by applying the middle layer of the trained autoencoder and feature vector containing the RGB components and YUV colour space information. In the estimation phase shown in Figure 1(b), a hyperspectral image is estimated by applying a feature vector with an RGB components and YUV information from an RGB image to the trained tuning neural network and decoder of the autoencoder for dimension extension.

2.1. Defined variables

Herein, we define variables related to image information, hyperspectral image, autoencoder and neural network, YUV colour space, and HSV colour space as shown in Table 1.

2.2. Learning phase

2.2.1. Dimension extension part based on the autoencoder

The dimension extension part was implemented using the decoding function of the autoencoder as shown in the dimension extension part of Figure 1(a). To obtain the decoding function, the autoencoder was trained using the Ny hyperspectral images with an image size of \( N_x \times N_y \). In the nth hyperspectral image for training the autoencoder, an arbitrary pixel \( i \) and \( j \) has brightness value of \( l_{\lambda,n}(i,j) \) with 256 levels for each index \( \lambda \) (1, 2, \ldots, \( K \)) corresponding to the wavelength. In the autoencoder, both the input and output layers have \( K \) units corresponding to the brightness values \( l_1^m(i,j), l_2^m(i,j), \ldots, l_K^m(i,j) ; M (< K) \) of units \( Q_m \) \((m = 1, 2, \ldots, M) \) were arranged in the middle layer of the autoencoder, as shown in Figure 1. Because \( N_y \) images with an image size of \( N_x \times N_y \) were used for training, the autoencoder was trained by \( N_y \times N_x \times N_y \) pixels in total. From the result of the training, an encoder which compresses the dimension \( K \) of brightness value \( l_\lambda^m(i,j) \) for each pixel to the dimension \( M \) of units \( Q_m \), and a decoder which extends dimension \( M \) of units \( Q_m \) to dimension \( K \) of brightness value \( l_\lambda^m(i,j) \) were obtained. Let \( f_{\text{decode}}(Q_m) \) be the decoding function of the decoder which outputs the brightness value \( l_\lambda^m(i,j) \) in dimension \( K \) from unit \( Q_m \) in dimension \( M \).

2.2.2. Tuning part based on the neural network

To estimate hyperspectral image \( l_\lambda^m(i,j) \) by an RGB image, a tuning neural network that creates \( Q_m \) from the RGB components was trained as shown in the tuning part of Figure 1(a). For an arbitrary pixel \( l_\lambda^m(i,j) \) of the training hyperspectral image, the brightness values \( l_{R,G,B}^m(i,j) \), \( l_{R,G}^m(i,j) \), \( l_{R,B}^m(i,j) \) represent the brightness of each R, G, and B components, respectively. Hence, the \( l_{R,G}^m(i,j) \), \( l_{R,B}^m(i,j) \), and \( l_{R,B}^m(i,j) \) were used to train the tuning neural network as RGB components. In the tuning neural network, the feature vector \( Z^m(i,j) \) for the input layer consisted of RGB components and YUV colour space as \( Z^m(i,j) = [l_{R,G}^m(i,j), l_{R,G}^m(i,j), l_{R,B}^m(i,j)] \).
The YUV colour space for pixel \( (i, j) \) were calculated from RGB components \( I_{\lambda}^{R}(i, j) \), \( I_{\lambda}^{G}(i, j) \), \( I_{\lambda}^{B}(i, j) \) as follows, as defined by the International Telecommunication Union:

\[
F_{y}(i, j) = 0.299 \times \left( \frac{I_{\lambda}^{R}(i, j)}{255} \right)^{1/\gamma} + 0.587 \times \left( \frac{I_{\lambda}^{G}(i, j)}{255} \right)^{1/\gamma} + 0.114 \times \left( \frac{I_{\lambda}^{B}(i, j)}{255} \right)^{1/\gamma}
\]

\[
F_{u}(i, j) = 0.492 \times \left( \frac{I_{\lambda}^{B}(i, j)}{255} \right)^{1/\gamma} - F_{y}(i, j)
\]

\[
F_{v}(i, j) = 0.877 \times \left( \frac{I_{\lambda}^{R}(i, j)}{255} \right)^{1/\gamma} - F_{y}(i, j)
\]

where \( \gamma = 2.2 \)

The gamma values (\( \gamma = 2.2 \)) for gamma correction were taken from [25].

Therefore, the number of units in the input layer was six to receive each element of the feature vector \( Z_{\lambda}^{h}(i, j) \). \( N_{f} \times N_{x} \times N_{y} \) of feature vectors \( Z_{\lambda}^{h}(i, j) \) were obtained from the hyperspectral images for training. The tuning neural network was trained using the set of \( Z_{\lambda}^{h}(i, j) \).
and \( Q_m \). From the results of the training, a tuning function \( F_{\text{tune}}(Z^n(i,j)) \), which outputs \( Q_m \) by applying the feature vector \( Z^n(i,j) \) was obtained.

### 2.3. Estimation phase

In the estimation phase, a hyperspectral image was estimated per pixel as shown in Figure 1(b). Hence, for arbitrary pixel \( i, j \), a pixel \( I^n_{\lambda}(i,j) \) including \( K \) of brightness corresponding to each \( \lambda \) (\( \lambda = 1, 2, \ldots, K \)) was estimated by a pixel \( I^n_{\lambda}(i,j) \) of the RGB image including brightness values of RGB components \( I^n_{\lambda,R}(i,j), I^n_{\lambda,G}(i,j), I^n_{\lambda,B}(i,j) \) as follows.

The YUV colour space information \( F^y_{\lambda}(i,j), F^y_{\lambda}(i,j), F^y_{\lambda}(i,j) \) of the pixel \( i, j \) were calculated by substituting \( I^n_{\lambda,R}(i,j), I^n_{\lambda,G}(i,j), I^n_{\lambda,B}(i,j) \) with \( I^n_{\lambda}(i,j), I^n_{\lambda}(i,j), I^n_{\lambda}(i,j) \) in Equations (1)–(3), respectively, and the feature vector \( Z^n(i,j) = [I^n_{\lambda,R}(i,j), I^n_{\lambda,G}(i,j), I^n_{\lambda,B}(i,j), F^y_{\lambda}(i,j), F^y_{\lambda}(i,j), F^y_{\lambda}(i,j)] \) were obtained. Then, the feature vector \( Z^n(i,j) \) was input to the tuning function as \( F_{\text{tune}}(Z^n(i,j)) \) to obtain \( Q_m \) (\( m = 1, 2, 3, \ldots, M \)). The \( Q_m \) was applied to the decoding function as \( F_{\text{decode}}(Q_m) \); then, the hyperspectral image \( I^n_{\lambda}(i,j) \) at the pixel \( i, j \) including \( K \) of brightness corresponding to each \( \lambda \) (\( \lambda = 1, 2, \ldots, K \)) was estimated. Finally, the entire hyperspectral image was estimated by repeating the above-mentioned procedures for all pixels \( I^n_{\lambda}(i,j) \) (\( i = 1, 2, \ldots, N_x, j = 1, 2, \ldots, N_y \)).

### 3. Validity experiment

In the validity experiment, we verified how accurately hyperspectral images are estimated by the proposed method. We also compared the estimation accuracies of the hyperspectral image among three scenarios with different combinations of RGB components and two models of colour spaces.

#### 3.1. Experimental systems

Figure 2 shows the entire experimental system used in the validity experiment. The experimental system comprises a hyperspectral camera (RESONON, Pika RC2) to capture hyperspectral images, and a PC to control the hyperspectral camera. The number of wavelengths of the hyperspectral images was 462-level (\( K = 462 \)) wavelengths and corresponded to the range from 398.67 nm (\( \lambda = 1 \)) to 1016.78 nm (\( \lambda = 462 \)) with wavelength resolution of \( \Delta \lambda = 1.34 \) nm. The number of pixels in each image corresponding to each wavelength...
level of a hyperspectral image was 1,600,000 pixels \((N_x \times N_y = 1,000 \times 1,600)\). An object that would be captured by the hyperspectral camera was set at a distance of 0.15 m from the camera lens. To irradiate a light source containing widespread wavelengths to the object, the experiment was conducted under sunlight, as shown in Figure 2. An RGB image is obtained by extracting three components from the 462-level wavelengths in the captured hyperspectral images because it may not be possible to evaluate the results of estimation accurately by changing the perspective between an RGB camera and hyperspectral camera if RGB images and hyperspectral images are captured from different cameras. The number of pixels in each image for the R, G, and B channels is 1,600,000.

### 3.2. Experimental procedures

As objects to be captured for the images to verify the proposed method, we used tomato, banana, and eggplant to cover a wide range of wavelengths. Figure 3 shows the RGB images with three RGB components \((\lambda_R = 183 \text{ (640.01 nm)}, \lambda_G = 115 \text{ (549.43 nm)}, \lambda_B = 48 \text{ (460.66 nm)})\) from the 462 wavelengths in the captured hyperspectral images for training and test image. In this experiment, three images were used in the learning phase \((N_l = 3)\), and an image was applied as a test image for the estimation phase. The images in Figure 3 are RGB, but hyperspectral images corresponding to each RGB image were also captured simultaneously.

The number of units in the middle layer \(M\) for the autoencoder was set to six. To compare the estimation accuracies of the hyperspectral image among different combinations of RGB components and two models of colour spaces that are input to the neural network in the tuning part, the following three scenarios with different feature vectors \(Z^{\text{rn}}(i,j)\), and \(Z^s(i,j)\) were applied to estimate the hyperspectral image.

#### 3.2.1. 1st scenario

In this scenario, a hyperspectral image was estimated via the proposed method utilizing the RGB and YUV colour space. Hence, the feature vectors \(Z^{\text{rn}}(i,j) = [I^{\text{rn}}_{\lambda_R}(i,j), I^{\text{rn}}_{\lambda_G}(i,j), I^{\text{rn}}_{\lambda_B}(i,j), F^r(i,j), U^s(i,j), F^s(i,j)]\) and \(Z^s(i,j) = [I^s_{\lambda_R}(i,j), I^s_{\lambda_G}(i,j), I^s_{\lambda_B}(i,j), F^r(i,j), U^s(i,j), F^s(i,j)]\) were used in the learning and estimation phases, as mentioned in the proposed method.

#### 3.2.2. 2nd scenario

In this scenario, RGB and HSV colour space were applied to estimate the hyperspectral image. Hence, in the learning phase, the feature vector \(Z^{\text{rs}}(i,j)\) consisted of RGB components and HSV colour space as \(Z^{\text{rs}}(i,j) = [I^{\text{rs}}_{\lambda_R}(i,j), I^{\text{rs}}_{\lambda_G}(i,j), I^{\text{rs}}_{\lambda_B}(i,j), U^s_{\lambda_R}(i,j), U^s_{\lambda_G}(i,j), U^s_{\lambda_B}(i,j)]\) and it was applied to the input layer in the tuning neural network. The \(U^s_{\lambda_R}(i,j), U^s_{\lambda_G}(i,j), U^s_{\lambda_B}(i,j)\) of the HSV colour space information at pixels \(i\) and \(j\) are calculated as follows:

\[
U^s_{\lambda_R}(i,j) = \begin{cases} 
60 \times \frac{I^{\text{rn}}_{\lambda_R}(i,j) - I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j)}{I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j) - I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j)} + 120 \quad \text{if } I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j) - I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j) > 0 \\
60 \times \frac{I^{\text{rn}}_{\lambda_B}(i,j) - I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j)}{I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j) - I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j)} + 240 \quad \text{if } I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j) - I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j) < 0 \\
\text{Undefined} \quad \text{if } I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j) = I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j) \end{cases}
\]

\[
U^s_{\lambda_G}(i,j) = \frac{I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j) - I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j)}{I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j)}
\]

\[
U^s_{\lambda_B}(i,j) = I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j)
\]

where \(I^{\text{rn}}_{\lambda_{\text{MAX}}}(i,j) = \max[I^{\text{rn}}_{\lambda_R}(i,j), I^{\text{rn}}_{\lambda_G}(i,j), I^{\text{rn}}_{\lambda_B}(i,j)]\) and \(I^{\text{rn}}_{\lambda_{\text{MIN}}}(i,j) = \min[I^{\text{rn}}_{\lambda_R}(i,j), I^{\text{rn}}_{\lambda_G}(i,j), I^{\text{rn}}_{\lambda_B}(i,j)]\)

Furthermore, in the estimation phase, the HSV colour space information \(U^s_{\lambda_R}(i,j), U^s_{\lambda_G}(i,j), U^s_{\lambda_B}(i,j)\) were calculated by substituting \(I^s_{\lambda_R}(i,j), I^s_{\lambda_G}(i,j), I^s_{\lambda_B}(i,j)\) in Equations (4–6), respectively. The feature vector \(Z^s(i,j) = [I^s_{\lambda_R}(i,j), I^s_{\lambda_G}(i,j), I^s_{\lambda_B}(i,j), U^s_{\lambda_R}(i,j), U^s_{\lambda_G}(i,j), U^s_{\lambda_B}(i,j)]\) was used to estimate the hyperspectral image \(I^s_{\lambda}(i,j)\).
3.2.3. 3rd scenario
In this scenario, only RGB components were applied to the input layer in the tuning neural network. Hence, the feature vector $Z^{(i,j)} = [I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}]$ and $Z^{(i,j)} = [I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}]$ were used applied in the learning and estimation phases to estimate the hyperspectral image $\hat{I}_{\lambda}^{(i,j)}$.

3.3. Evaluation method
The test image had an RGB image corresponding to $I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}, I_{\lambda}^{(i,j)}$ and hyperspectral image $I_{\lambda}^{(i,j)}$ which can be a reference for the estimated hyperspectral image $\hat{I}_{\lambda}^{(i,j)}$. Therefore, in this experiment, we defined the similarity $S_{\lambda}$ to express cosine similarity. Cosine similarity has been widely used in machine learning applications such as natural language processing and image processing. Mathematically, cosine similarity is very similar to Euclidean distance in that both methods aim to define the similarity between two vectors. The main difference between the two methods is that Euclidean distance considers both the sizes and directions of two vectors, whereas cosine similarity considers only the directions. In the context of this study, the size of a vector corresponds to the brightness of an image. Because we are more interested in the relative brightness between pixels, we adopted cosine similarity as an evaluation index. The similarity $S_{\lambda}$ is calculated for every wavelength $\lambda$ between the estimated hyperspectral image $\hat{I}_{\lambda}^{(i,j)}$ and reference hyperspectral image $I_{\lambda}^{(i,j)}$ as follows:

$$S_{\lambda} = \frac{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} I_{\lambda}^{(i,j)} \cdot \hat{I}_{\lambda}^{(i,j)}}{\sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} I_{\lambda}^{(i,j)}^2} \cdot \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \hat{I}_{\lambda}^{(i,j)}^2}}$$

where $\hat{c} = \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \hat{I}_{\lambda}^{(i,j)}^2}$, $c = \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} I_{\lambda}^{(i,j)}^2}$.

Because the similarity $S_{\lambda}$ is calculated for each wavelength, we obtained 462 similarities ($S_1, S_2, \ldots, S_{462}$) in total.

The accuracy of the method was evaluated for eight wavelength bands (B1–B8) and full wavelength band. For each band, we calculated an evaluation index $\bar{S}(\lambda_{low}, \lambda_{high})$, which was defined as the average value of the similarities from $S_{\lambda_{low}}$ to $S_{\lambda_{high}}$ as follows:

$$\bar{S}(\lambda_{low}, \lambda_{high}) = \frac{1}{N_b} \sum_{\lambda=\lambda_{low}}^{\lambda_{high}} S_{\lambda}$$

where $N_b = \lambda_{high} - \lambda_{low} + 1$.

Table 2 lists the $\lambda_{low}$ and $\lambda_{high}$ values for each band (B1–B8) and full wavelength band, respectively.
Table 2. \( \lambda_{\text{low}}, \lambda_{\text{high}} \), and corresponding wavelength for each wavelength band and full wavelength band.

| Wavelength band | \( \lambda_{\text{low}} \) | \( \lambda_{\text{high}} \) | Wavelength of \( \lambda_{\text{low}} \) | Wavelength of \( \lambda_{\text{high}} \) |
|-----------------|----------------|----------------|----------------|----------------|
| (B1): Purple    | 1              | 24             | 398.67 nm      | 428.97 nm      |
| (B2): Indigo    | 25             | 47             | 430.29 nm      | 459.34 nm      |
| (B3): Blue      | 48             | 76             | 460.66 nm      | 497.70 nm      |
| (B4): Green     | 77             | 129            | 499.02 nm      | 568.04 nm      |
| (B5): Yellow    | 130            | 144            | 569.37 nm      | 588.00 nm      |
| (B6): Orange    | 145            | 159            | 589.33 nm      | 607.98 nm      |
| (B7): Red       | 160            | 286            | 609.32 nm      | 778.15 nm      |
| (B8): Infrared  | 287            | 462            | 779.49 nm      | 1016.78 nm     |
| Full wavelength | 1              | 462            | 398.67 nm      | 1016.78 nm     |

Figure 5. Reference image and estimated hyperspectral images at wavelength of 959.54 nm (\( \lambda = 420 \)).

respectively, were estimated to be darker than those of 2nd scenario but brighter than the reference image \( I_{420}(i,j) \), as shown in Figure 5(a). Similarities, \( S_{420} \), in Figure 5(b–d) with Figure 5(a) were 0.803, 0.810, and 0.796, respectively. The similarity \( S \) is cosine similarity, which considers the directions of two vectors corresponding to the relative brightness of the images used in this study. This is why the similarity \( S_{420} \) in Figure 5(c) is higher than that in Figure 5(b), even though the estimated image in Figure 5(c) differs more significantly from the reference image in Figure 5(a) compared to that in Figure 5(b).

Because the wavelength at 959.54 nm is in the infrared band, objects in the image \( I_{420}(i,j) \) of Figure 5(a) were darker and the edge between objects and background was not clear. In contrast, the objects in the estimated image presented in Figure 5(b–d) were clearer. In addition, eggplant in the correct image in Figure 5(a) was whitish, but in Figure 5(b–d), the eggplant in each \( \hat{I}_{420}(i,j) \) was estimated to be black, as in the visible light band image. Hence, comparing the similarities \( S_{420} \) and \( S_{100} \) for each scenario, it was found that all \( S_{120} \) values were lower than \( S_{100} \).

4.2. Results of evaluation indices for each scenario

Table 3 shows evaluation indices \( \hat{S}(\lambda_{\text{low}}, \lambda_{\text{high}}) \) for each wavelength band (B1)–(B8) and all 462 wavelengths \( \hat{S}(\lambda_{1}, \lambda_{462}) \) in the 1st, 2nd, and 3rd scenarios, respectively.

The results for 1st scenario exhibited the highest accuracy. In the visible light band from the (B1) purple band to the (B7) red band, the evaluation indices \( \hat{S}(\lambda_{\text{low}}, \lambda_{\text{high}}) \) of 1st scenario were the highest among the three scenarios. The evaluation index \( \hat{S}(1,462) \) for all 462 wavelengths was also the highest at 0.913 in 1st scenario. Regarding the (B8) infrared band, \( \hat{S}(287,462) \) of the three scenarios were 0.838, 0.852, and 0.831, respectively, and there were no significant differences.

From Table 3, it was confirmed that by applying the RGB components and YUV colour space information to the input layer of the tuning neural network, the estimation accuracy of the hyperspectral image in the visible light band is higher than that in the case of applying RGB components and HSV colour space information or only RGB components to the input layer.

5. Discussions

The results of the validity experiments showed that the 1st scenario, where RGB and YUV colour space information are applied, had the highest accuracy from a comprehensive perspective and the evaluation index on full wavelength bands \( \hat{S}(1,462) \) were 0.913, 0.812, and 0.811 in 1st, 2nd, and 3rd scenarios, respectively. This is listed in Table 3. In each wavelength band, there was a large difference in the index of the visible light bands, that is, from the purple band (B1) to the red band (B7). In the visible light bands, evaluation indices in 1st scenario were above 0.9, whereas all evaluation indices of (B1)–(B7) in 2nd and 3rd scenarios were less than 0.9. The YUV colour space was utilized in telecast for high-quality moving images. By using YUV colour space information including a wider luminance signal \( Y \), differences occurred between the components on the YUV even for objects that were displayed in the same way between the components on RGB. Thus, the input of the YUV colour space as additional information, which has characteristics different from RGB, played a major role in training the tuning neural network and is considered to be a primary factor for the improved accuracy. In contrast, the HSV colour space information, which is devised based on the intuitive human sense of colour recognition, was not suitable for training, and did not improve the accuracy in the 2nd scenario.
Table 3. Evaluation indices $\tilde{S}(\lambda_{\text{low}}, \lambda_{\text{high}})$ for each wavelength band (B1)–(B8) in the 1st, 2nd, and 3rd scenarios.

| Wavelength band | $\tilde{S}(\lambda_{\text{low}}, \lambda_{\text{high}})$ | (1st scenario) RGB + YUV | (2nd scenario) RGB + HSV | (3rd scenario) RGB |
|-----------------|-------------------------------------------|---------------------|-------------------|-----------------|
| (B1): Purple    | $\tilde{S}(1, 24)$                      | 0.919               | 0.804             | 0.723           |
| (B2): Indigo    | $\tilde{S}(25, 47)$                     | 0.956               | 0.801             | 0.785           |
| (B3): Blue      | $\tilde{S}(48, 76)$                     | 0.970               | 0.738             | 0.832           |
| (B4): Green     | $\tilde{S}(77, 129)$                    | 0.985               | 0.719             | 0.787           |
| (B5): Yellow    | $\tilde{S}(130, 144)$                   | 0.982               | 0.715             | 0.767           |
| (B6): Orange    | $\tilde{S}(145, 159)$                   | 0.980               | 0.726             | 0.737           |
| (B7): Red       | $\tilde{S}(160, 286)$                   | 0.949               | 0.832             | 0.822           |
| (B8): Infrared  | $\tilde{S}(287, 462)$                   | 0.838               | 0.852             | 0.831           |
| Full wavelength | $\tilde{S}(1, 462)$                     | 0.913               | 0.812             | 0.811           |

Table 4. Evaluation indices $\tilde{S}(\lambda_{\text{low}}, \lambda_{\text{high}})$ for each wavelength band in (B1) to (B8) for the image of the colour checker used as testing data.

| Wavelength band | $\tilde{S}(\lambda_{\text{low}}, \lambda_{\text{high}})$ | (1st scenario) RGB + YUV |
|-----------------|-------------------------------------------|---------------------|
| (B1): Purple    | $\tilde{S}(1, 24)$                      | 0.784               |
| (B2): Indigo    | $\tilde{S}(25, 47)$                     | 0.769               |
| (B3): Blue      | $\tilde{S}(48, 76)$                     | 0.785               |
| (B4): Green     | $\tilde{S}(77, 129)$                    | 0.899               |
| (B5): Yellow    | $\tilde{S}(130, 144)$                   | 0.886               |
| (B6): Orange    | $\tilde{S}(145, 159)$                   | 0.878               |
| (B7): Red       | $\tilde{S}(160, 286)$                   | 0.838               |
| (B8): Infrared  | $\tilde{S}(287, 462)$                   | 0.653               |
| Full wavelength | $\tilde{S}(1, 462)$                     | 0.768               |

Figure 6. RGB image of a colour checker derived from the 462 wavelengths in the captured hyperspectral images.

In this experiment, three images of a tomato, banana, and eggplant were used as objects for training data because the accuracy of estimation was verified by utilizing images of these vegetables as testing data. Additionally, another object was inputted into the learning model used in the 1st scenario as testing data for this verification. RGB and YUV colour space information regarding the image of the colour checker used as a different object in the testing data for this experiment (see Figure 6) were inputted into the same models used in the 1st scenario (4th scenario). Table 4 lists the evaluation indices $\tilde{S}(\lambda_{\text{low}}, \lambda_{\text{high}})$ for each wavelength band from (B1) to (B8) and all 462 wavelengths $\tilde{S}(\lambda_1, \lambda_{462})$ in the 1st and 4th scenarios, respectively.

All similarities in the eight wavelength bands and full wavelength spectrum are lower than those in the 1st scenario when inputting the image shown in Figure 3 as testing data, as shown in Table 3. Therefore, it is possible to obtain more accurate estimation results by learning images of objects that are closer in colour and material to the objects used as testing data.

6. Conclusions

In this study, we proposed a hyperspectral image estimation method based on RGB images. In the proposed method, the RGB component and YUV colour space information calculated from the RGB were applied to the tuning function provided by the tuning neural network. The hyperspectral image was estimated by inputting the output from the tuning function to the decoding function of the trained autoencoder, which functioned as a dimension extension. To evaluate estimation accuracy for different combinations of RGB and colour space models, we conducted validity experiments for the estimation of hyperspectral images in three scenarios with different colour spaces: RGB and YUV, RGB and HSV, and only RGB. The result showed that the scenario with RGB and YUV colour space information had the highest estimation accuracy of 0.913 by averaging all similarities for wavelength among the three scenarios; thus, the validity of the proposed method as an estimation method for hyperspectral images was verified. Using YUV colour space information, which corresponds to visible light wavelengths, the similarity in the visible light bands were improved compared to our previous work [24].

In our future study, we will conduct validity experiments to increase the estimation accuracy of the
infrared band by utilizing other colour spaces which are suitable for the characteristics of images at infrared wavelengths instead of using YUV colour space information. Furthermore, we will validate a set of appropriate parameters for the proposed method. Additionally, to construct a model with greater generalizability, we must conduct verification on the choice of learning data for generating such a model.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Parts of this work were supported by JSPS KAKENHI [grant number 19K20386], [grant number 20K12769], [grant number 16K16392].

Notes on contributors

Ryoji Sato received the B.E. degree from Aoyama Gakuin University, in 2021. He is currently pursuing the master’s degree from the Graduate School of Science and Engineering, Aoyama Gakuin University. His research interests include image estimation and machine learning.

Yuri Hamada received B.E., M.E. and Ph.D. degrees in Engineering from Chuo University, Japan, in 2008, 2010, and 2017, respectively. From 2018 to 2020, she served as an Assistant Professor at Chuo University. Since 2020, she is currently an Assistant Professor in Industrial and Systems Engineering at Aoyama Gakuin University. Her research interests include modeling of communication processes and analysis of decision making processes. She is a member of Japan Society of Kansei Engineering, Japanese Cognitive Science Society.

Takashi Kaburagi received his B.E., M.E., and Ph.D. degrees from Waseda University in 2003, 2005, and 2009, respectively. He has been a research associate at Waseda University from 2008 to 2010, Assistant Professor at Gakushuin University from 2010 to 2013, Assistant Professor at Aoyama Gakuin University from 2013 to 2019. He is an Instructor at the College of Liberal Arts, International Christian University, where he has been a faculty member since 2019. His research interests are machine learning and time-series data analysis.

Yosuke Karihara received his ME and PhD degrees from Hosei University, Tokyo, in 2003 and 2009, respectively. He joined Hitachi Software Engineering, Ltd., in 2003. From 2009 to 2012, he served as an Assistant Professor at Seikei University. From 2013 to 2018, he was an Associate Professor at Aoyama Gakuin University; and from 2019, he has been serving as a Professor. He is the author of five books, more than 70 journal papers, and more than 100 international conference proceedings, and he holds four patents. His research interests include system engineering, sensing methods, biosensing, and system information engineering. He is a member of the Japanese Society for Medical and Biological Engineering, the Society of Instrument and Control Engineers, the Electrical Engineers of Japan, the Japan Society for Fuzzy Theory and Intelligent Informatics, and the IEEE.

References

[1] Pani S, Saifuddin SC, Ferreira FIM, et al. High energy resolution hyperspectral X-ray imaging for low-dose contrast-enhanced digital mammography. IEEE Trans Med Imag. 2017;36(9):1784–1795.

[2] Pike R, Lu G, Wang D, et al. A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging. IEEE Trans Bio Med Eng. 2016;63(3):653–663.

[3] Fabelo H, Ortega S, Szolna A, et al. In-vivo hyperspectral human brain image database for brain cancer detection. IEEE Access. 2019;7:39098–39116.

[4] Eslami M, Mohammazadeh A. Developing a spectral-based strategy for urban object detection from airborne hyperspectral TIR and visible data. IEEE J Sel Top Appl Earth Obs Remote Sens. 2016;9(5):1808–1816.

[5] Wang T, Zhu Z, Blasch E. Bio-inspired adaptive hyperspectral imaging for real-time target tracking. IEEE Sens J. 2010;10(3):647–654.

[6] Koz A. Ground-based hyperspectral image surveillance systems for explosive detection: part II—radiance to reflectance conversions. IEEE J Sel Top Appl Earth Obs Remote Sens. 2019;12(12):4754–4765.

[7] Xiong F, Zhou J, Qian Y. Material based object tracking in hyperspectral videos. IEEE Trans Image Process. 2020;29:3719–3733.

[8] Li X, Wang H, Li X, et al. Identifying degraded grass species in inner Mongolia based on measured hyperspectral data. IEEE J Sel Top Appl Earth Obs Remote Sens. 2019;12(12):1–15.

[9] Yang W, Yang C, Hao Z, et al. Diagnosis of plant cold damage based on hyperspectral imaging and convolutional neural network. IEEE Access. 2019;7:118239–118248.

[10] Schmid T, Rodriguez-Rastrero M, Escrivan P, et al. Characterization of soil erosion indicators using hyperspectral data from a Mediterranean rainfed cultivated region. IEEE J Sel Top Appl Earth Obs Remote Sens. 2016;9(2):845–860.

[11] Karlik B, Sariöz M. Coloring gray-scale image using artificial neural networks. 2009 2nd International conference on adaptive science & technology (ICAST). 2009, p. 366–371.

[12] Cheng Z, Yang Q, Sheng B. Colorization using neural network ensemble. IEEE Trans Image Process. 2017;26(11):3541–3555.

[13] Dong X, Li W, Hu X, et al. A colorization framework for monochrome-color dual-lens systems using a deep convolutional network. IEEE Trans Vis Comput Graph. 2020;28(3):1469–1485.

[14] Manolakis D, Golowich S, DiPietro RS. Long-Wave infrared hyperspectral remote sensing of chemical clouds: a focus on signal processing approaches. IEEE Signal Process Mag. 2019;31(4):120–141.
[15] Song Q, Xu F, Jin Y. Radar image colorization: converting single-polarization to fully polarimetric using deep neural networks. IEEE Access. 2018;6:1647–1661.

[16] Ye W, Chen H, Zhang Z, et al. Hybrid scheme of image’s regional colorization using mask R-CNN and poisson editing. IEEE Access. 2019;7:115901–115913.

[17] Wan S, Xia Y, Qi L, et al. Automated colorization of a grayscale image with seed points propagation. IEEE Trans Multimedia. 2020;22(7):1756–1768.

[18] Ren H, Li J, Gao N. Two-stage sketch colorization with color parsing. IEEE Access. 2020;8:44599–44610.

[19] Ozcelik F, et al. Rethinking CNN-based pansharpening: guided colorization of panchromatic images via GANs. IEEE Trans Geosci Remote Sens. 2021;59(4):3486–3501.

[20] Mancuso R, Smorfa S, Olivieri M. A novel high-quality YUV-based image coding technique for efficient image storage in portable electronic appliances. IEEE Trans Consum Electron. 2008;54(2):695–702.

[21] Chung K, Liang Y, Wang C. Effective content-aware chroma reconstruction method for screen content images. IEEE Trans Image Process. 2019;28(3):1108–1117. DOI: 10.1109/TIP.2018.2875340.

[22] Zhang X, Jiang J, Liang Z, et al. Skin color enhancement based on favorite skin color in HSV color space. IEEE Trans Consum Electron. 2010;56(3):1789–1793. DOI:10.1109/TCE.2010.5606327.

[23] Yoon I, Kim S, Kim D, et al. Adaptive defogging with color correction in the HSV color space for consumer surveillance system. IEEE Trans Consum Electron. 2012;58(1):111–116. DOI: 10.1109/TCE.2012.6170062.

[24] Sato R, Hamada Y, Kaburagi T, et al. Evaluation of the effect of light source spectrum on the estimation accuracy of high dimensional spectral images using neural networks based on RGB images. 2021 SICE annual conference. 2021.

[25] Shimizu M. Digital image processing revised. 2nd ed. CG-ARTS. 2020, p. 77–79.