

# Journal of Electronic Imaging

SPIEDigitalLibrary.org/jei

## **Stereo one-shot six-band camera system for accurate color reproduction**

Masaru Tsuchida  
Shuji Sakai  
Mamoru Miura  
Koichi Ito  
Takahito Kawanishi  
Kashino Kunio  
Junji Yamato  
Takafumi Aoki



# Stereo one-shot six-band camera system for accurate color reproduction

**Masaru Tsuchida**

NTT Communication Science Laboratories  
Atsugi-shi, Kanagawa 243-0198, Japan  
E-mail: [tsuchida.masaru@lab.ntt.co.jp](mailto:tsuchida.masaru@lab.ntt.co.jp)

**Shuji Sakai**

**Mamoru Miura**

**Koichi Ito**

Tohoku University  
Graduate School of Information Sciences  
Sendai-shi 980-8579, Japan

**Takahito Kawanishi**

**Kashino Kunio**

**Junji Yamato**

NTT Communication Science Laboratories  
Atsugi-shi, Kanagawa 243-0198, Japan

**Takafumi Aoki**

Tohoku University  
Graduate School of Information Sciences  
Sendai-shi 980-8579, Japan

---

**Abstract.** For accurate color reproduction of motion pictures or still pictures of moving objects, we developed a different one-shot six-band image-capturing and visualization system that combines multi-band and stereo imaging techniques. The proposed system consists of two consumer-model digital cameras and an interference filter whose spectral transmittance is comb-shaped. Subpixel correspondence search between the stereo image pairs is conducted and image transformation whose parameters are estimated from the correspondence is used to correct the geometric relationship between the images. The Wiener estimation method is used for color reproduction. For experiments, we have constructed two six-band camera systems to evaluate the quality of resultant images. One is for capturing high-resolution images using digital single-lens reflex cameras. The other is for capturing motion pictures using digital video cameras, for which all image processing steps after image capture are implemented on graphics processing units and the frame rate of the system is 30 fps when image size is XGA. For both systems, the average color difference between the measurement data and estimation results for 24 color patches of the Macbeth ColorChecker™ is  $dE_{a^*b^*} = 1.21$  at maximum. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JEI.22.3.033025](https://doi.org/10.1117/1.JEI.22.3.033025)]

---

## 1 Introduction

Recently, with the rapidly evolving multimedia technologies and visual telecommunications systems, color reproduction by color imaging systems is commonly required and technologies and techniques for achieving it are being explored. In visual telecommunications applications—such as electronic commerce, telemedicine, and electronic art museums—realistic color reproduction is very important as if the object is being directly observed. For this purpose, natural and high-fidelity color reproduction, high-resolution imaging, and dynamic-range enhancement are key technologies. Also important for achieving realism in archiving (e.g., cultural heritage and medical applications) are the reproduction and display of the high-fidelity colors and gloss of objects, as well as the reproduction of their texture, three-dimensional (3-D) shape, microstructure, and movement. However, it is difficult to accurately reproduce the color of an object under arbitrary illumination conditions using current imaging systems based on three-band image capturing, especially when the illumination at the image observation site is different from the illumination of the image capturing.

Multispectral imaging technology, which estimates the spectrum using multiband data, is a solution for accurate color reproduction. Although several types of multiband camera systems in the field of still imaging have been developed,<sup>1–8</sup> most of them are multi-shot-type systems, such as a monochrome camera with a rotating filter wheel, and they cannot take images of moving objects. Ohsawa et al.<sup>8</sup>

have developed a six-band HDTV camera system. However, the system requires very expensive customized equipment. In order to make multispectral technology pervasive, equipment costs must be reduced and the systems have to be able to take images of moving objects.

In this article, we present a stereo one-shot six-band image capturing system that combines multispectral and stereo imaging techniques to meet these requirements. The proposed image capturing system consists of two consumer-model digital cameras and an interference filter whose spectral transmittance is comb-shaped (the characteristics of the interference filter are described later). We have constructed two types of stereo six-band camera systems. One is for capturing high-resolution six-band still images of moving objects with two digital single-lens reflex cameras. Both cameras are synchronized by a remote controller and captured images are transferred to memory on a PC. The other is for capturing motion pictures using digital video cameras, for which all image processing steps after image capture are implemented on graphics processing units (GPUs) and the frame rate of the system is 30 fps when the image size is XGA.

The process for the proposed system mainly comprises four steps:

- Step 1 Stereo image acquisition.
- Step 2 Subpixel correspondence search between the captured stereo image pair.
- Step 3 Geometrical transformation of the captured image for generating a six-band image.
- Step 4 Spectrum-based color reproduction.

Note that the proposed system mainly deals with the diffuse reflection component. Obtaining the specular reflection component and capturing bidirectional reflectance distribution require another more complicated setup.

In related research, Shresta et al. have presented a six-band stereoscopic camera and simulations and experimental results of color reproduction.<sup>9-11</sup> Like our system, their systems consist of a stereo camera and one or two sheets of color filter. However, their systems do not address several important issues. One is the need for real-time processing of all operations, ranging from image capture to displaying the color reproduction results. A live-view function is also strongly required in various fields, such as digital archiving of moving pictures to conform the image quality including color. This is particularly true in medicine and archiving cultural heritage because the illumination condition is often constrained to avoid obstructing medical procedures surgical and to preserve target objects. Another issue is the accuracy of collecting registration errors in each stereo pair. Registration errors between two three-band images consisting of a six-band image cause pseudocolor in the resultant image of color reproduction. To avoid degrading image quality, subpixel correspondence matching and collection technique should be introduced. Third, a strategy for deciding the sensitivity of the six-band camera is important. Shresta et al. used a filter selection algorithm to select color filters among a set of filters readily available on the market. Therefore, color patches of training sets influence the filter selection and affect the results of color reproduction. In addition, the spectral sensitivity of their camera system is also



Fig. 1 Stereo six-band camera.

dependent on and limited by the set of commercially available color filters. In contrast, the filters used in our system are custom made and designed to divide the sensitivity of the camera (from 400 to 700 nm) into six parts with equal intervals. This means that the bandwidth of spectral sensitivity of the digital camera becomes almost halved. We show practical solutions for these issues in this article.

In what follows, each of the above-mentioned steps is described in detail and experimental values for each system are evaluated and discussed. The article concludes with a short summary.

## 2 Stereo Image Acquisition

Figure 1 shows the proposed six-band image capturing system. The right camera captures a normal RGB image. The left one, with the interference filter mounted in front of the lens, captures a specialized RGB image. Figure 2 shows the principle of six-band image capturing using the interference filter. The spectral transmittance of the filter is comb-shaped. The filter cuts off short wavelengths, i.e., the peaks of both the blue and red in the original spectral sensitivity of the camera. It also cuts off the long wavelength of green. The

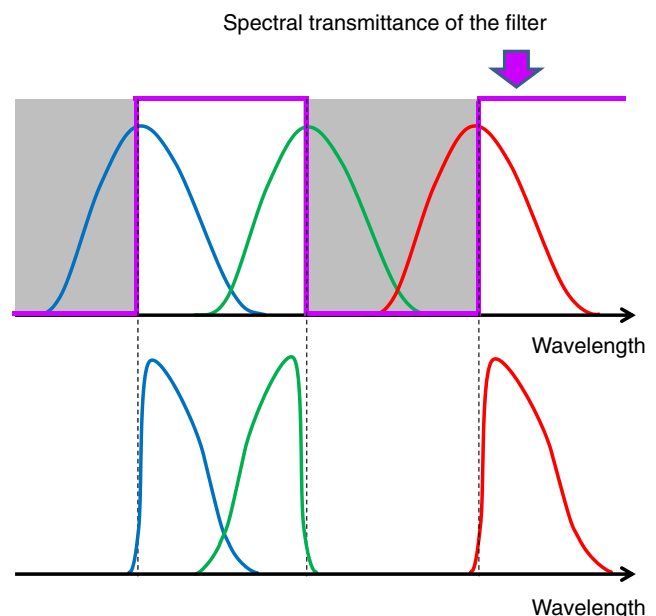


Fig. 2 Spectral sensitivity of the camera with the interference filter.

captured three-band stereo images are combined into a six-band image for color reproduction.

Other one-shot six-band camera systems<sup>8,10</sup> use two color filters and capture two specialized RGB images. On the other hand, our system captures a normal RGB image and a specialized RGB image. The sensitivity of the camera becomes almost half with the filter mounted in front of the lens. When the illumination is not bright enough for the camera with the filter (e.g., the illumination conditions are constrained for preserving target objects in archiving historical heritage), the underexposed image degrades the color accuracy of the resultant image. Even in such situations, our proposed system can guarantee the image quality reproduced from a conventional RGB camera system even though the color reproduction quality may be degraded.

### 3 Correspondence Search

The captured two images have parallax. Therefore, to generate a six-band image from the pair of images, one image should be transformed to adjust it to the other image. As a first step to do that, a search for corresponding points between two images is carried out. The detected corresponding points are used for estimating image transformation parameters to correct geometric relationships between the images. Although the two cameras take images of the same target object, the color balance between the two images is quite different because of the interference filter mounted in front of lens of one camera. General detection methods<sup>12,13</sup> cannot work well in such a case. To find corresponding points between a stereo image pair, we use a subpixel correspondence matching technique that combines local block matching by the phase-only correlation (POC) method and the coarse-to-fine strategy based on pyramid representation.<sup>14</sup> POC, a high-accuracy image-matching technique that uses phase information in the Fourier domain, can estimate translation between two images with subpixel accuracy. It is also robust against illumination changes, noise, and color shifts caused by differences in the spectral sensitivity of a camera. The computation of POC is suitable for implementation on GPUs since their computations can be performed in parallel.<sup>15</sup> Details of POC are described in the following sections.

#### 3.1 POC Function

Consider two  $N_1 \times N_2$  images,  $f(n_1, n_2)$  and  $g(n_1, n_2)$ , where we assume that the index ranges are  $n_1 = -M_1, \dots, M_1$  and  $n_2 = -M_2, \dots, M_2$  for mathematical simplicity, and hence  $N_1 = 2M_1 + 1$  and  $N_2 = 2M_2 + 1$ . Let  $F(k_1, k_2)$  and  $G(k_1, k_2)$  denote the two-dimensional (2-D) discrete Fourier transforms (DFTs) of the two images.  $F(k_1, k_2)$  and  $G(k_1, k_2)$  are given by

$$\begin{aligned} F(k_1, k_2) &= \sum_{n_1, n_2} f(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} \\ &= A_F(k_1, k_2) e^{j\theta_F(k_1, k_2)} \end{aligned} \quad (1)$$

and

$$\begin{aligned} G(k_1, k_2) &= \sum_{n_1, n_2} g(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} \\ &= A_G(k_1, k_2) e^{j\theta_G(k_1, k_2)}, \end{aligned} \quad (2)$$

where  $k_1 = -M_1, \dots, M_1$ ,  $k_2 = -M_2, \dots, M_2$ ,  $W_{N_1} = e^{-j\frac{2\pi}{N_1}}$ ,  $W_{N_2} = e^{-j\frac{2\pi}{N_2}}$ , and the operator  $\sum_{n_1, n_2}$  denotes  $\sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2}$ .  $A_F(k_1, k_2)$  and  $A_G(k_1, k_2)$  are amplitude components, and  $e^{j\theta_F(k_1, k_2)}$  and  $e^{j\theta_G(k_1, k_2)}$  are phase components.

The cross spectrum  $R(k_1, k_2)$  between  $F(k_1, k_2)$  and  $G(k_1, k_2)$  is given by

$$\begin{aligned} R(k_1, k_2) &= F(k_1, k_2) \overline{G(k_1, k_2)} \\ &= A_F(k_1, k_2) A_G(k_1, k_2) e^{j\theta(k_1, k_2)}, \end{aligned} \quad (3)$$

where  $\overline{G(k_1, k_2)}$  denotes the complex conjugate of  $G(k_1, k_2)$  and  $\theta(k_1, k_2) = \theta_F(k_1, k_2) - \theta_G(k_1, k_2)$ . On the other hand, the cross-phase spectrum (or normalized cross spectrum)  $\hat{R}(k_1, k_2)$  is defined as

$$\hat{R}(k_1, k_2) = \frac{F(k_1, k_2) \overline{G(k_1, k_2)}}{|F(k_1, k_2) \overline{G(k_1, k_2)}|} = e^{j\theta(k_1, k_2)}. \quad (4)$$

The POC function  $\hat{r}(n_1, n_2)$  is the 2-D inverse DFT of  $\hat{R}(k_1, k_2)$  and is given by

$$\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1, k_2} \hat{R}(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2}, \quad (5)$$

where  $\sum_{n_1, n_2}$  denotes  $\sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2}$ .

#### 3.2 Subpixel Image Registration

Consider  $f_c(x_1, x_2)$  as a 2-D image defined in continuous space with real-number index  $x_1$  and  $x_2$ . Let  $\delta_1$  and  $\delta_2$  represent subpixel displacement of  $f_c(x_1, x_2)$  in  $x_1$  and  $x_2$  directions, respectively. So, the displaced image can be represented as  $f_c(x_1 - \delta_1, x_2 - \delta_2)$ . Assume that  $f(n_1, n_2)$  and  $g(n_1, n_2)$  are spatially sampled images of  $f_c(x_1, x_2)$  and  $f_c(x_1 - \delta_1, x_2 - \delta_2)$ , defined as

$$f(n_1, n_2) = f_c(x_1, x_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \quad (6)$$

and

$$g(n_1, n_2) = g_c(x_1 - \delta_1, x_2 - \delta_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \quad (7)$$

where  $T_1$  and  $T_2$  are the spatial sampling intervals, and index ranges are given by  $n_1 = -M_1, \dots, M_1$  and  $n_2 = -M_2, \dots, M_2$ . Let  $F(k_1, k_2)$  and  $G(k_1, k_2)$  be the 2-D DFTs of  $f(n_1, n_2)$  and  $g(n_1, n_2)$ , respectively. Considering the difference of properties between the Fourier transform defined in continuous space and that defined in discrete space carefully, we can now say that

$$G(k_1, k_2) \cong F(k_1, k_2) \cdot e^{-j\frac{2\pi}{N_1} k_1 \delta_1} e^{-j\frac{2\pi}{N_2} k_2 \delta_2}. \quad (8)$$

Thus,  $\hat{R}(k_1, k_2)$  is given by

$$\hat{R}(k_1, k_2) \cong e^{-j\frac{2\pi}{N_1} k_1 \delta_1} e^{-j\frac{2\pi}{N_2} k_2 \delta_2}. \quad (9)$$

The POC function  $\hat{r}(n_1, n_2)$  will be the 2-D inverse DFT of  $\hat{R}(k_1, k_2)$  and is given by

$$\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \quad (10)$$

$$\cong \frac{\alpha}{N_1 N_2} \frac{\sin\{\pi(n_1 + \delta_1)\}}{\sin\left\{\frac{\pi}{N_1}(n_1 + \delta_1)\right\}} \frac{\sin\{\pi(n_2 + \delta_2)\}}{\sin\left\{\frac{\pi}{N_2}(n_2 + \delta_2)\right\}},$$

where  $\alpha = 1$ . The above equation represents the shape of the peak for the POC function for common images that are minutely displaced from each other. The peak position of the POC function corresponds to the displacement between the two images. We can prove that the peak value  $\alpha$  decreases (without changing the function shape itself) when small noise components are added to the original images. Hence, we assume  $\alpha$  in practice.

In order to reduce the computation time, we can use one-dimensional (1-D) POC instead of 2-D POC (Ref. 16) when the stereo image pair is rectified,<sup>17</sup> since the rectified stereo image pair has only horizontal translations.

#### 4 Geometrical Transformation of the Captured Image for Generating Six-Band Image

Next, the shape of the image captured with the interference filter is adjusted to that of the other image using the detected corresponding points. Projective transformation is a simple method and works well for 2-D objects. When the target object has a 3-D shape, nonlinear transformation is better. The thin-plate spline (TPS) model<sup>18</sup> was used for image transformation in this work. The resultant two three-band images are combined into a six-band image.

Although this system can acquire both spectral color information and depth information at the same time, depth information is not used in the process of generating a six-band image from the captured stereo image pair because of the computational cost to achieve real-time image processing. Depth information obtained from detected corresponding points would improve the quality of the generated six-band images.

#### 5 Spectrum-Based Color Reproduction

As shown in Fig. 3, an object's surface reflects light from an illumination source. Let the illumination spectrum and spectral reflectance be  $W(\lambda)$  and  $f(\lambda)$ , respectively. The observed spectrum,  $I(\lambda)$ , can be represented as

$$I(\lambda) = W(\lambda)f(\lambda), \quad (11)$$

where  $\lambda$  is wavelength. Let us consider a situation where the reflected light is captured by an  $L$ -band sensor. Let  $\mathbf{f}$  be the

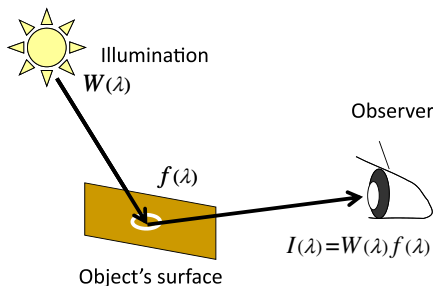


Fig. 3 Model of reflected light observation.

spectral reflectance of the object represented in a  $P$ -dimensional space,  $\mathbf{W}$  be a  $P \times P$  diagonal matrix whose diagonal elements represent the spectral power distribution of illumination, and  $\mathbf{S} = [S_1(\lambda), S_2(\lambda), \dots, S_L(\lambda)]^T$  be an  $L \times P$  matrix whose row vectors represent the spectral sensitivity of the sensor. Equation (11) can then be rewritten in vector representation as

$$\mathbf{I} = \mathbf{W}\mathbf{f}. \quad (12)$$

By using the Wiener estimation method,<sup>19</sup> the spectral reflectance is estimated from the camera signal,  $\mathbf{c} = \mathbf{S}\mathbf{W}\mathbf{f} = \mathbf{H}\mathbf{f}$ , as

$$\hat{\mathbf{f}} = \mathbf{M}\mathbf{c}, \mathbf{M} = \mathbf{R}\mathbf{H}^T\{\mathbf{H}\mathbf{R}\mathbf{H}^T\}^{-1}, \quad (13)$$

where  $\hat{\mathbf{f}}$  is the estimated spectral reflectance,  $\mathbf{M}$  is the Wiener estimation matrix obtained from  $\mathbf{H}$ , and  $\mathbf{R}$  is *a priori* knowledge about the spectral reflectance of the object, respectively.

In the Wiener estimation method, we used a correlation matrix  $\mathbf{R}$ , which is modeled on a first-order Markov process covariance matrix, in the form

$$\mathbf{R} = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{P-1} \\ \rho & 1 & \rho & \dots & \rho^{P-2} \\ \rho^2 & \rho & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{P-1} & \rho^{P-2} & \dots & \dots & 1 \end{pmatrix}, \quad (14)$$

where  $0 \leq \rho \leq 1$  is the adjacent element correlation factor; we set  $\rho = 0.999$  in our experiments, on the basis of our previous experimental experience.

Using the estimated spectral reflectance, the spectral power distribution of illumination for observation, and tone-curves and chromaticity values of primary colors of the display monitor, we calculate output RGB signals. Even when the illumination light used at an observation site is different from that for image capturing (e.g., daylight is used for image capturing and fluorescent lamp is used at observation sites), the color observed under the observation light can be reproduced as if the object is in front of observers.

#### 6 Experiments

In the experiments described below, the possibility of a stereo six-band camera system for color reproduction is confirmed. First, the relationship between the color reproduction accuracy and distance between the two cameras is evaluated. Several stereo-pair images were captured while the distance between the two cameras was changed. The estimated color and spectral reflectance of a color chart were compared with the measurement results obtained with a spectrometer. Next, 2-D and 3-D objects were captured using the proposed still camera system, and the results of color reproduction using a six-band image generated from stereo-pair images were compared with the real objects. Finally, all the steps from image acquisition to displaying the color reproduction image were implemented on GPUs. The computation time for each process and the total computation time are evaluated using the proposed camera system for motion pictures.

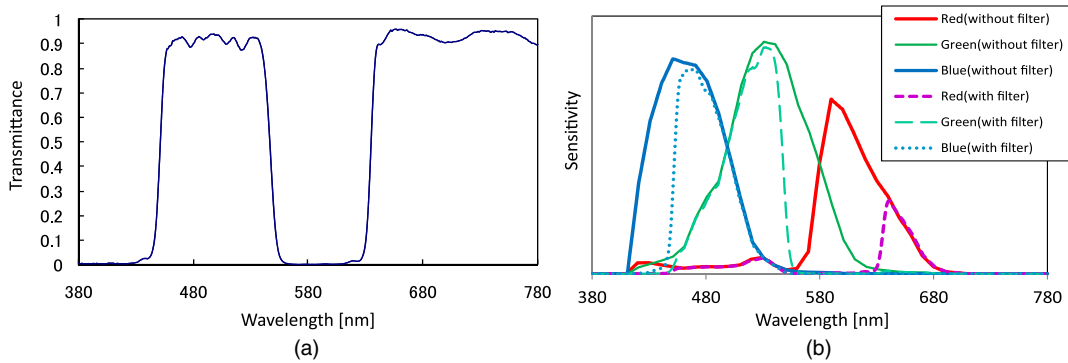


Fig. 4 Spectral transmittance of the interference filter (a) and spectral sensitivity of the six-band camera system (b).

### 6.1 Experimental Equipment for Still Image Acquisition

We used two of the same consumer-model digital cameras (D700, NIKON), which can write out raw image data without any color correction in the NEF file format. The D700 model can take 12-Mpixel images, and its bit depth is 14 bits. We analyzed the NEF file format and converted NEF files into a general raw file format. Figure 4 shows the spectral transmittance of the interference filter and spectral sensitivity of the camera. Note that it does not have sensitivity lower than 400 nm and higher than 700 nm because UV- and IR-cut filters are attached to the image sensor. For illumination, we used artificial solar lamps (SOLAX™, SERIC) whose spectral power distribution is close to natural sunlight. Before starting the experiments, characters of the display monitor (primary colors and tone curves) were also measured

to ensure the colors of the resultant images are displayed correctly.

### 6.2 Relationship Between Color Reproduction Accuracy and Distance Between the Two Cameras

We evaluate the accuracy of reproduced color and spectral reflectance when the distance between the two cameras is changed. Macbeth Color Checker™ was used as a target object. The focal length of the lens was 105 mm. The distance between camera and color chart was 2 m. As a first step, the first image was captured without the interference filter. Then the interference filter was attached in front of the camera lens and the second image was captured. Next, the camera with the filter was moved 15 cm horizontally in 1 cm intervals (see Fig. 5) and filtered images were captured at each position. The exposure setting (shutter speed, iris, etc.) of the camera was fixed. To correct registration errors between the image captured without the filter and the images captured with the filter, projective transformation was used to generate a six-band image of the color chart.

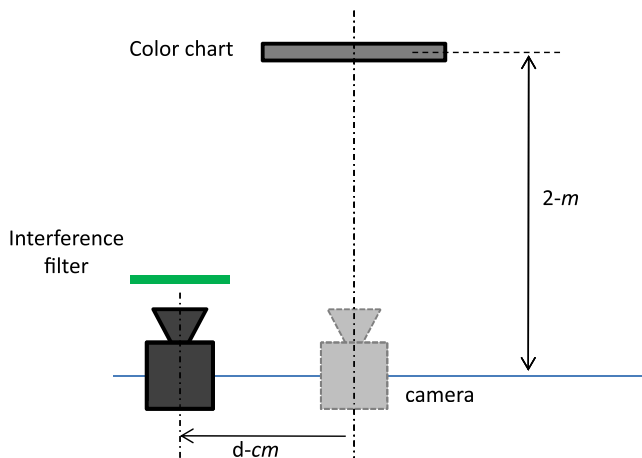


Fig. 5 Geometrical setup for evaluating the relationship between color reproduction accuracy and the position of the two cameras.

Figure 6 shows a part of the estimated spectral reflectance of Macbeth Color Checker™. The estimation results when the camera's moving distance  $d = 0, 5, 10,$  and  $15$  cm are plotted. To evaluate the estimation results, we also measured the spectral reflectance using a spectrometer, and the measurement results are plotted on the same graphs. We can see that the distance the camera is moved does not affect the estimation of spectral reflectance under this experimental geometry. Good estimation results were obtained between 400 nm and 700 nm wavelengths. There are some errors in the near UV- and near IR-wavelength domains caused by the UV- and IR-cut filter on the image sensor. (All results of estimated spectral reflectance of 24 patches of Macbeth Color Checker™ are shown in Fig. 7.)

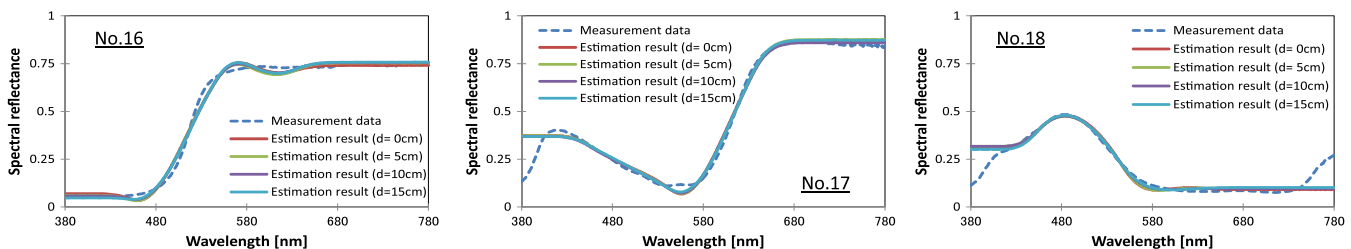


Fig. 6 Part of the estimated spectral reflectance of Macbeth Color Checker™.

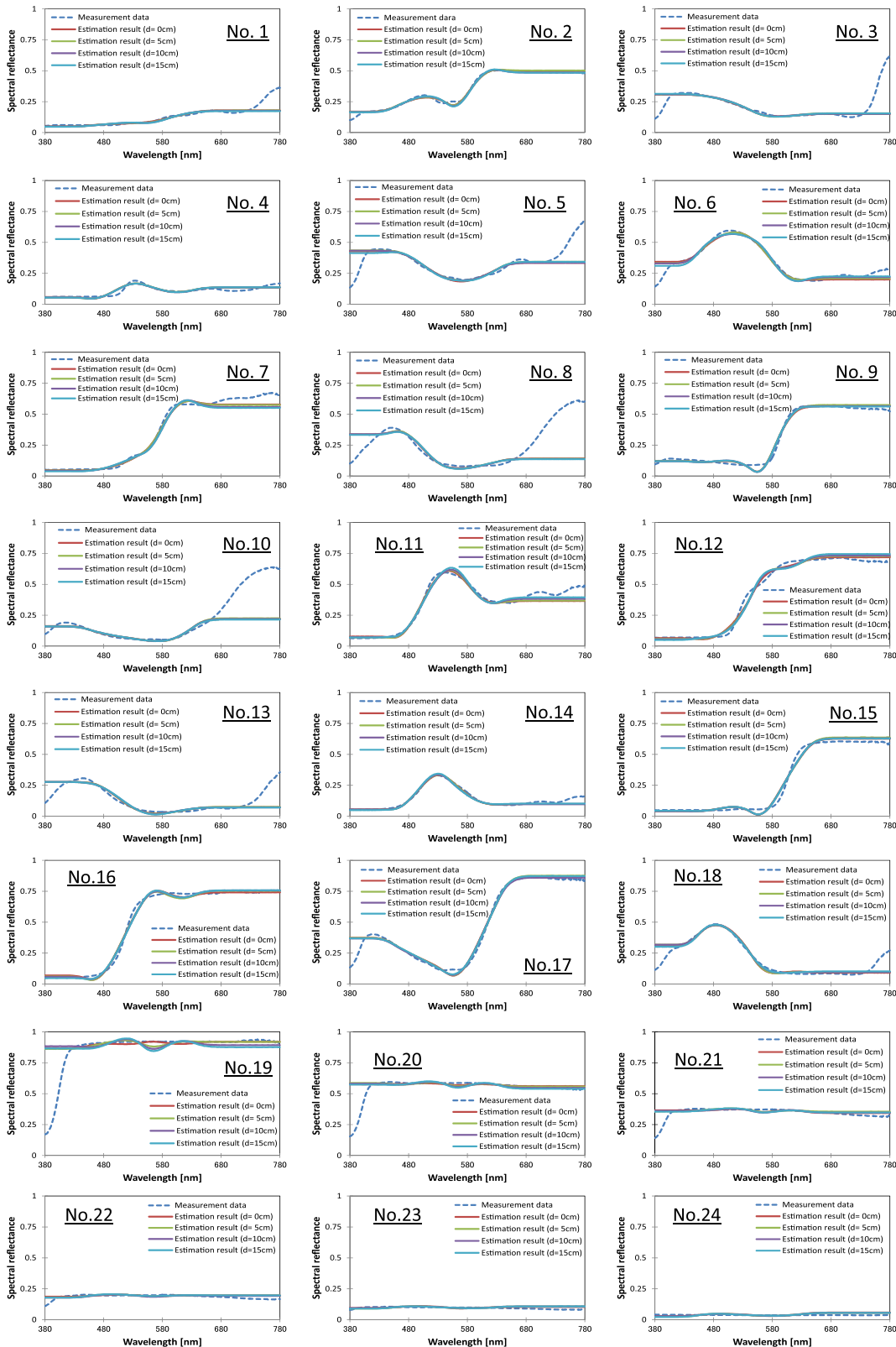


Fig. 7 Estimation results of spectral reflectance of 24 patches of Macbeth Color Checker™.

Next, color difference  $dEa^*b^*$  between measured and estimated colors was calculated. Averaged color differences of 24 color patches are  $dEa^*b^* = 0.97, 1.05, 1.15$  and  $1.21$  when  $d = 0, 5, 10,$  and  $15$  cm.

### 6.3 Experimental Results for 2-D Objects

We used old paints that had been applied on cloth as target objects in this experiment. The objects look flat, but their surface actually gently undulates and is uneven. The distance

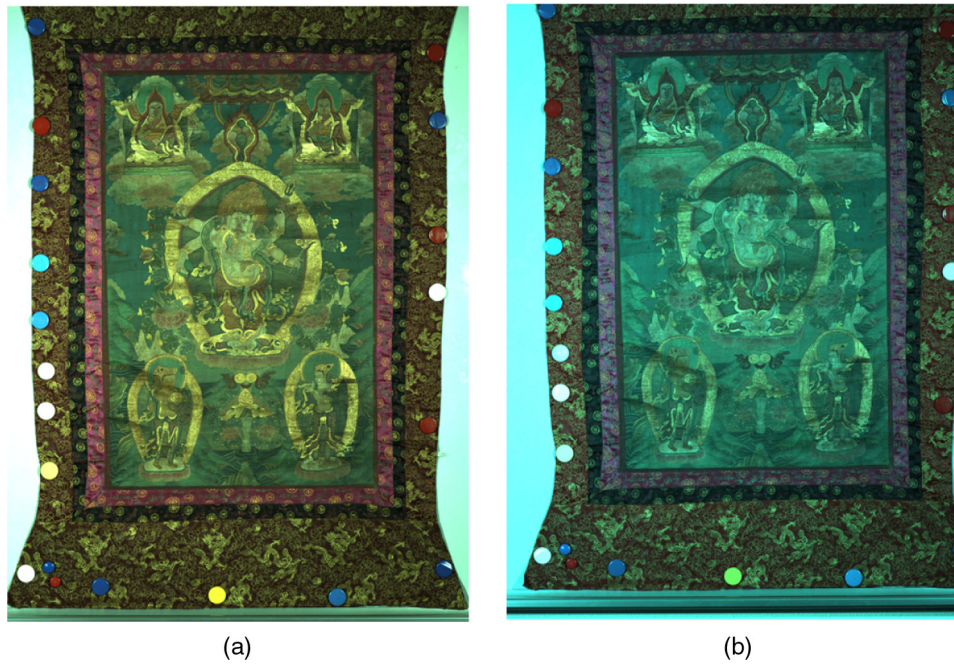


Fig. 8 Captured images without filter (a) and with filter (b).

between the center of the lenses of the two cameras was 15 cm.

First, we took two images of the objects using the proposed system at once (cameras were controlled by remote control software and remote shutter release). The exposure settings (shutter speed, iris condition, etc.) of both cameras were the same. The lenses of both cameras were also the same (in this experiment, we used a lens with a focal length of 105 mm). Figure 8 shows two captured images. Here, color balance of the image captured without the interference filter looks incorrect because we used a raw data image from the image sensor.

Second, corresponding points between the two images were detected by using the 2-D POC method. Reference points on the reference image were sampled in 50-pixel intervals, and the corresponding search was carried out at each

reference point. The local block size was  $128 \times 128$  pixels and the search range was  $\pm 32$  pixels. Using the detection results, the image captured with the interference filter was transformed according to projective transformation. It took almost 10 s for the detection and transformation processes. Then, a six-band image was generated.

And finally, the six-band image was converted into an RGB image by spectrum-based color reproduction method. With the GPU-based calculation, these color reproduction processes were run almost in real time. The resultant image is shown in Fig. 9. Artifacts or pseudocolor such as double edges caused by image registration errors cannot be observed. The resultant RGB image (Fig. 9) was compared with the real object and also with the image generated by the two-shot type six-band camera system.<sup>7</sup> It is confirmed that the resultant image generated with the proposed method is



Fig. 9 Color reproduction results for paints on cloth.



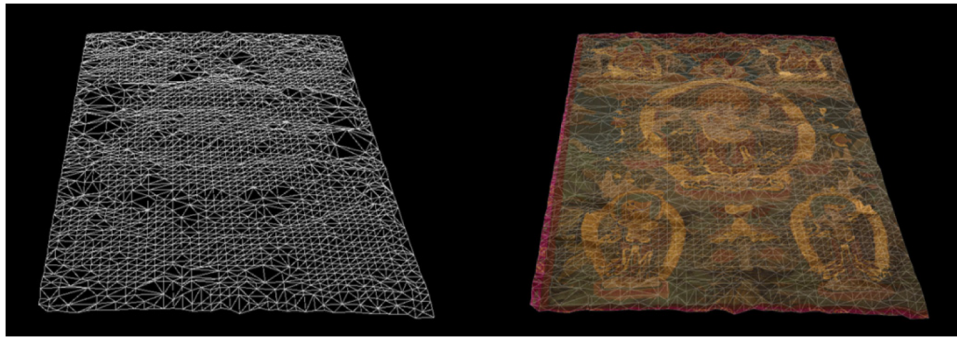


Fig. 10 Reconstructed shape of the paint.

the same color as the object, and the quality of the resultant image is the same as that for the conventional methods. No registration errors remain among the band images generated by the proposed method.

Figure 10 shows a reconstructed shape of the paint based on a Delaunay triangulation using the detected corresponding points. The obtained mesh model is well fit with the resultant image and the image looks natural. Although depth information is not used for image transformation in this article, using it would improve the quality of generated six-band images, especially for 3-D objects.

#### 6.4 Experimental Results for 3-D Objects

Next, we used a traditional Japanese kimono hung on a mannequin as a target object. The distance between the two cameras was 18 cm. The focal length of the lens was 60 mm. In this experiment, when a six-band image was generated from a stereo-pair image, the TPS model was used for image transformation instead of projective transformation. The captured images were divided into subimages whose image size was  $256 \times 256$  pixels because TPS uses a large amount of computer memory. Reference points on a subimage were

sampled in four-pixel intervals, and a corresponding search was carried out at each reference point. The local block size was  $64 \times 64$  pixels and the search range was  $\pm 16$  pixels. After correspondence matching in each subimage, all resultant images were merged into a six-band image. Figure 11 shows an example of the image transformation results. Double-edge textures (green edges) can be seen in the image before transformation. To confirm the image transformation accuracy of this camera system, we carried out experiments using distances between the object and camera of 4, 3, and 2 m. The height of the mannequin is 150 cm. Figures 12–14 show the resultant image for each capturing geometry. In each figure, the color reproduction images before image transformation is on the left, the results of color reproduction image transformation are in the center, and the grid image presenting the transformation result is on the right. The grid image shows how the captured image was transformed. Observing the resultant images in Figs. 12 and 13 (distance: 4 and 3 m), it seems that the image transformation works very well and good color reproduction quality can be obtained. On the other hand, concerning Fig. 14 (distance: 2 m), there are some areas with



Fig. 11 Image transformation and color reproduction results for 3-D object.



Fig. 12 Experimental results (distance from camera: 4 m).



Fig. 13 Experimental results (distance from camera: 3 m).



Fig. 14 Experimental results (distance from camera: 2 m).

mis-transformation results, especially around relatively hard edges with self-occlusion (e.g., around the sleeve of the left arm). This indicates that the distance limit for image capturing is 2 m when the camera setup is the same as that used in this experiment.

### 6.5 Experimental Equipment for Moving Picture Acquisition

Two digital cameras (Grasshopper-20S4C, Point Grey Research Inc.) with the IEEE 1394b (800 Mbit/s) interface

were used. This model can write out raw image data without any color correction and can take XGA-size ( $1024 \times 768$  pixels) images, each of which has a bit depth of 16 bits at 30 fps. The baseline length of the two cameras is 44 mm, which makes it possible to reduce the influence of image parallax between the two cameras in six-band image generation. Figure 15 shows a photo of the camera system and its spectral sensitivity used in this experiment. Note that each camera has sensitivity higher than 400 nm and lower than 730 nm since UV- and IR-cut filters are attached to the image sensor.

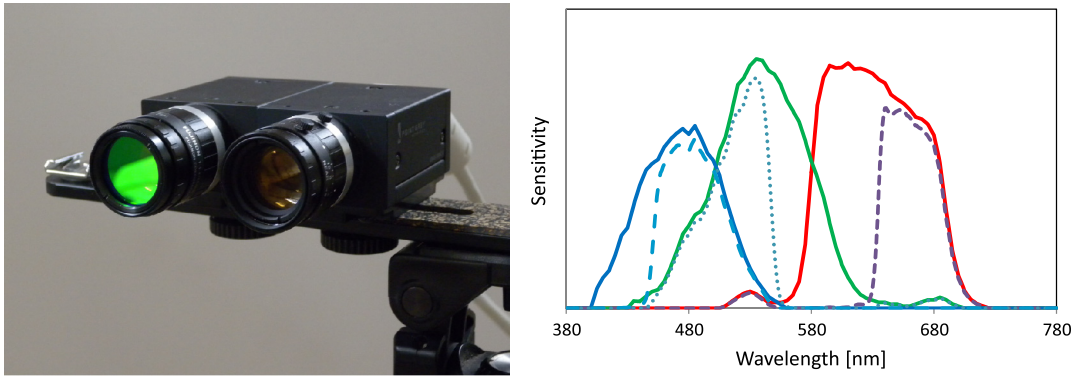


Fig. 15 Stereo six-band video system.

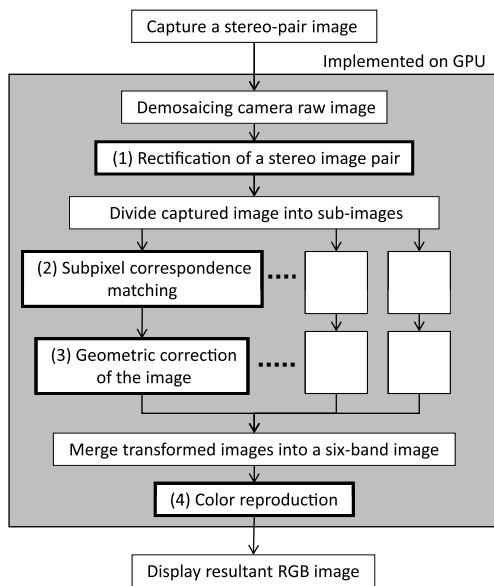


Fig. 16 Diagram of image processing of the six-band stereoscopic video system.

The spectral transmittance of the interference filter is same to that of the still camera system (Fig. 4).

Two graphics cards (nVidia GeForce GTX580) were installed into a PC and used for real-time image processing. The CPU on the mother board is an Intel Core i7-980 3.3GHz, and the size of the main memory is 12 GB.

### 6.6 Experimental Results for Six-Band Video System

The target object used in the experiment was a 3-D Japanese doll on a rotating table. The camera array was placed 2.5 m from the object. 1-D POC for correspondence search and projective transformation were used for generating six-band images. Figure 16 shows the image processing procedure of the system. There are four main steps: (1) rectification of a stereo image pair, (2) subpixel correspondence matching, (3) geometric correction of the image to generate a six-band image, and (4) color reproduction. All the steps are implemented on GPUs. Although the six-band image can be generated well in the case of 2-D objects like tapestries, several adjustment errors remain in applying projective transformation to the whole image of a 3-D object. The adjustment errors cause artifacts (e.g., double edges or pseudo



Fig. 17 Color reproduction results.

color) in the resultant images of color reproduction. To avoid the adjustment errors, the captured images were divided into several subimages and projective transformation was applied to each subimage. Then, all transformed subimages were merged into be a six-band image.

The results of color reproduction are shown in Fig. 17. Few artifacts (e.g., double edges or pseudo color) caused by image transformation error are observed. Comparing the resultant image to the real object confirms that the color of the object is well reproduced. Comparing the resultant images obtained from this system and a two-shot type six-band camera system,<sup>7</sup> which uses the same digital camera and filter, it can be seen that reproduction of almost the same image quality, especially color, is achieved. The resultant moving pictures were displayed on an LCD monitor in real time.

Next, we compared the total computation times when the size of the subimage and the sampling interval of image data for running POC and projective transformation were changed. The subimage sizes were  $256 \times 256$ ,  $512 \times 512$ , and  $1024 \times 768$  pixels. The sampling intervals of the reference points were 8, 10, 16, and 20 pixels and the computation time was evaluated for each interval. The local block

**Table 1** Total computation time/frame.

	Image size of subimage		
	$256 \times 256$	$512 \times 512$	$1024 \times 768$
Sampling interval of correspondence			
8 pixels	43 ms	45 ms	32 ms
10 pixels	34 ms	33 ms	27 ms
16 pixels	23 ms	21 ms	20 ms
20 pixels	20 ms	19 ms	18 ms

**Table 2** Detailed computation time. Sampling interval of correspondence is 16 pixels.

Processing	Image size of subimage		
	$256 \times 256$	$512 \times 512$	$1024 \times 768$
Demosaicing	3.4 ms	3.4 ms	3.4 ms
Rectification	3.0 ms	3.0 ms	3.0 ms
1-D POC	7.8 ms	7.3 ms	6.6 ms
Generation of projection matrix	3.3 ms	3.0 ms	2.7 ms
Projective transformation	3.0 ms	2.1 ms	1.8 ms
Color reproduction	0.8 ms	0.8 ms	0.8 ms
Display on monitor	2.0 ms	2.0 ms	2.0 ms
Total time/frame	23.3 ms	21.6 ms	20.3 ms

size was 32 pixels and the search range was  $\pm 8$  pixels. Table 1 shows the results, which indicate that the sampling interval of correspondence search larger than 16 pixels is required in order to achieve the frame rate of 30 fps when the subimage sizes are  $256 \times 256$  and  $512 \times 512$  pixels. Table 2 shows the computation times of each processing. The sampling interval of the correspondence search was 16 pixels. Note that the computation time for projection matrix generation depends on the subimage size and affects the total computation time materially because the number of subimages becomes large when the size of the subimage becomes small. It is confirmed that the system can achieve the frame rate of 30 fps regardless of subimage size. Additional experiments confirmed that this system runs at 15 fps when the image size is SXGA ( $1280 \times 1024$  pixels).

## 7 Conclusion

A novel six-band image acquisition and real-time color reproduction system using stereo imaging have been proposed. The system consists of two consumer-model digital cameras and an interference filter whose spectral transmittance is comb-shaped. It works well for 2-D objects that have a wavy structure like a tapestry. In order to extend this system to 3-D objects, the TPS model, a kind of nonlinear transformation method, was implemented, and it worked well for generating a six-band image from a stereo-pair image. Moreover, all image processing steps after image acquisition to display color reproduction results are implemented on GPUs and the frame rate of the system is 30 fps when image size is XGA. Although this six-band video system uses 1-D POC and projective transformation for reducing computational time for generating a six-band image from a stereo image pair, dividing captured images into some subimages enables the system to work well even when the target object has a 3-D shape.

Depth information is not used in the proposed system. It can be estimated from the result of the correspondence search and would improve the quality of the generated six-band images. There is a problem that accurate six-band information cannot be obtained from a captured stereo-pair image when the target object has glossy surface, such as a car body, because the appearance of gloss captured by each camera is different. In such a case, depth information would also be effective in overcoming this problem.

## References

1. P. D. Burns and R. S. Berns, "Analysis of multispectral image capture," in *Proc. 4th Color Imaging Conf. (CIC4)*, Scottsdale, Arizona, pp. 19–22 (1996).
2. S. Tominaga, "Multichannel vision system for estimating surface and illuminant functions," *J. Opt. Soc. Am. A* **13**(11), 2163–2173 (1996).
3. M. Yamaguchi et al., "Natural color reproduction in the television system for telemedicine," *Proc. SPIE* **3031**, 482–489 (1997).
4. S. Tominaga and R. Okajima, "Object recognition by multi-spectral imaging with a liquid crystal filter," in *Proc. Conf. Pattern Recognition*, Barcelona, Vol. 1, pp. 708–711, IEEE (2000).
5. S. Helling, E. Seidel, and W. Biehlig, "Algorithms for spectral color stimulus reconstruction with a seven-channel multispectral camera," in *Proc. Second European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, Aachen, Germany, pp. 254–258 (2004).
6. J.-I. Park et al., "Multispectral imaging using multiplexed illumination," in *Proc. IEEE Int'l. Conf. on Computer Vision (ICCV)*, Rio de Janeiro, pp. 1–8, IEEE (2007).
7. M. Hashimoto, "Two-shot type 6-band still image capturing system using commercial digital camera and custom color filter," in *Proc. Fourth European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, Terrassa, Barcelona, España, pp. 538–541 (2008).

8. K. Ohsawa et al., "Six-band HDTV camera system for spectrum-based color reproduction," *J. Imaging Sci. Technol.* **48**(2), 85–92 (2004).
9. R. Shrestha, J. Y. Hardeberg, and A. Mansouri, "One-shot multispectral color imaging with a stereo camera," *Proc. SPIE* **7876**, 797609 (2011).
10. R. Shrestha, A. Mansouri, and J. Y. Hardeberg, "Multispectral imaging using a stereo camera: concept, design and assessment," *EURASIP J. Adv. Signal Process.* **57**, 1–15 (2011).
11. R. Shrestha and J. Y. Hardeberg, "Simultaneous multispectral imaging and illuminant estimation using a stereo camera," *Lec. Notes Comput. Sci.* **7340**, 45–55 (2012).
12. D. Lowe, "Distinctive image features from scale invariant keypoints," *Int. J. Comput. Vision* **60**(2), 91–110 (2004).
13. M. Shimizu and M. Okutomi, "Sub-pixel estimation error cancellation on area-based matching," *Int. J. Comput. Vision* **63**(3), 207–224 (2005).
14. K. Takita et al., "High-accuracy image registration based on phase-only correlation," *IEICE Trans. Fundam.* **E86-A**(8), 1925–1934 (2003).
15. M. Miura et al., "GPU implementation of phase-based stereo correspondence and its application," in *Proc. 19th IEEE International Conference on Image Processing (ICIP)*, Orlando, Florida, pp. 1697–1700 (2012).
16. T. Shibahara et al., "A sub-pixel stereo correspondence technique based on 1D phase-only correlation," in *Proc. Int. Conf. Image Processing*, San Antonio, Texas, pp. 221–224, IEEE (2007).
17. A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Mach. Vision Appl.* **12**, 16–22 (2000).
18. F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(6), 567–585 (1989).
19. W. K. Pratt and C. E. Mancill, "Spectral estimation techniques for the spectral calibration of a color image scanner," *Appl. Opt.* **15**(1), 73–75 (1976).



**Masaru Tsuchida** received the BE, ME, and PhD degrees from the Tokyo Institute of Technology, Tokyo, in 1997, 1999, 2002, respectively. In 2002, he joined NTT Communication Science Laboratories, where his research areas included color science, three-dimensional image processing, and computer vision. His specialty is color measurement and multiband image processing. From 2003 to 2006, he worked at the National Institute of Information and Communication Technology (NICT) as a researcher for the "Natural Vision" project.



**Shuji Sakai** received the BE degree in information engineering, and the MS degree in information sciences from Tohoku University, Sendai, Japan, in 2010 and 2012, respectively. He is currently working toward the PhD degree of the Graduate School of Information Sciences at Tohoku University. His research interest includes signal and image processing and computer vision.



**Mamoru Miura** received the BE degree in information engineering, and the MS degree in information sciences from Tohoku University, Sendai, Japan, in 2010 and 2012, respectively. He is currently working toward the PhD degree of the Graduate School of Information Sciences at Tohoku University. His research interest includes signal and image processing and computer vision.



**Koichi Ito** received the BE degree in electronic engineering and the MS and PhD degree in information sciences from Tohoku University, Sendai, Japan, in 2000, 2002, and 2005, respectively. He is currently an assistant professor of the Graduate School of Information Sciences at Tohoku University. From 2004 to 2005, he was a research fellow of the Japan Society for the Promotion of Science. His research interests include signal and image processing and biometric authentication.



**Takahito Kawanishi** is senior research scientist, Research Planning Section at NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation. He received the BE degree in information science from Kyoto University, Kyoto, the ME and the PhD degree in information science from Nara Institute of Science and Technology, Nara, in 1996, 1998, and 2006, respectively. He joined NTT Laboratories in 1998. From 2004 to 2008, he worked at Plala Networks Inc. (now NTT Plala) as a technical manager and developer of commercial IPTV and VoD systems. He is currently engaged in R&D of online media content identification, monitoring and search systems. He is a senior member of IEICE and a member of IPSJ and JSIAM.



**Kashino Kunio** received the BE, ME, and PhD degrees from University of Tokyo in 1990, 1992, and 1995, respectively. In 1995, he joined NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, where he is currently a senior research scientist and supervisor. He has been working on multimedia information retrieval and music recognition. His research interests include acoustic signal processing, Bayesian information integration, and sound source separation. He was awarded the IEEE Transactions on Multimedia Paper Award in 2004.



**Junji Yamato** received the BE, ME, and PhD degrees from the University of Tokyo in 1988, 1990, and 2000, respectively, and the SM degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 1998. His areas of expertise are computer vision, pattern recognition, human-robot interaction, and multi-party conversation analysis. He is currently executive manager of the Media Information Laboratory, NTT Communication Science Laboratories. He is a visiting professor of Hokkaido University and Tokyo DENKI University. He is a member of IEEE, IEICE, and the Association for Computing Machinery.



**Takafumi Aoki** received the BE, ME, and DE degrees in electronic engineering from Tohoku University, Sendai, Japan, in 1988, 1990, and 1992, respectively. He is currently a professor of the Graduate School of Information Sciences (GSIS) at Tohoku University. In April 2012, Aoki was appointed as the vice president of Tohoku University. His research interests include theoretical aspects of computation, computer design and organization, LSI systems for embedded applications, digital signal processing, computer vision, image processing, biometric authentication, and security issues in computer systems. He has received more than 20 academic awards, including the IEE Ambrose Fleming Premium Award (1994), the IEE Mountbatten Premium Award (1999), the IEICE Outstanding Transaction Paper Awards (1989 and 1997), the IEICE Inose Award (1997), the Ichimura Award (2008), as well as many outstanding paper awards from international conferences and symposiums such as ISMVL, ISPACS, SASIMI, and COOL Chips.