An eleven-band stereoscopic camera system for accurate color and spectral reproduction

M. Tsuchida, K. Kashino, and J. Yamato
Communication Science Laboratories, NTT Corporation / 3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa, Japan

Abstract

For accurate color and spectral reflectance reproduction, we propose a novel eleven-band acquisition system using a nine-view stereo camera. The proposed system consists of eight monochrome cameras with eight different narrow band-pass filters and an RGB camera. To generate an eleven-band image, the shapes of the nine captured stereo images are transformed to correct registration displacement caused by stereo parallax. In the process of correspondence search between stereo images, the phase-only correlation method (POC) is used. The most significant point of our method is that the captured RGB image is converted into narrow-band images for accurate correspondence search. The detected corresponding points are used for estimating parameters of image transformation and an eleven-band image is generated. By comparing with conventional method, experimental results show that the accuracy of correspondence detection and spectral reflection estimation is improved

Introduction

In digital archiving for cultural heritage preservation, in the medical field, and in some industrial fields, high-fidelity reproduction of color, gloss, texture, three-dimensional (3-D) shape, and movement is very important. Multi- or full-spectrum imaging can provide accurate color reproduction. Although several types of multi-spectral camera systems have been developed [1-6], all of them are multi-shot and cannot take still images of moving objects and moving pictures.

Ohsawa et al. developed a six-band HDTV camera system [7]. However, the system requires very complex and expensive customized optics, which makes it far from practical. The equipment costs must be reduced in order to make multi-spectrum video systems pervasive. To meet this requirement, several one-shot stereo six-band image capturing systems that combine multi-spectrum and stereo imaging technologies have been proposed [8-10]. They are based on two consumer-model digital cameras or a digital stereo-camera and color filters. These systems achieve the averaged color difference of dEa*b*= 1.21 (24 color patches of Macbeth ColorCheker[TM]). Moreover, Tsuchida et al. have also been developing a stereo nine-band camera system for improving the accuracy of estimated spectral reflectance [11]. This system consists of nine synchronized monochrome cameras with nine different narrow band-pass filters and can record moving pictures. To generate a nine-band image, the captured nine stereo images are transformed its shape to correct registration displacement caused by stereo parallax. Image transformation parameters are estimated using the detected corresponding points between the reference image and the other. However, the correspondence search among

Figure 1. Eleven-band stereoscopic camera system.

the captured images sometimes does not work well when there is a large difference in the texture appearance of each band image. The differences are caused by the narrowness of band-pass width of the filter.

In this paper, we propose a nine-view eleven-band stereo camera system consisting of eight monochrome cameras with eight different narrow band-pass filters and a RGB camera. This system can acquire an eleven-band image. The image captured by the RGB camera is regarded as a reference image and is converted into an image captured by a monochrome camera with a band-pass filter to be used for improvement of correspondence search accuracy. To show effectiveness of our method, the number of detected corresponding points obtained using the proposed method is compared with the results for a stereo nine-band camera system [11], and the estimated spectral reflectance and images of color reproduction are shown in experiments.

Eleven-band stereoscopic camera

Figure 1 shows the proposed eleven-band stereoscopic camera using nine digital cameras. The RGB camera is mounted at the center of the camera array. A different narrow-band-pass filter is attached to the lens of each monochrome camera. This system can acquire an eleven-band image in one shot; three broad-band images corresponding to a normal RGB image and eight narrow-band images. Figure 2 shows a set of images captured at 450 nm

(top) and 700 nm (bottom) using the proposed camera system. . A comparison of the two images shows that the observed texture depends on the wavelength and looks quite different.

There are four main steps for generating an eleven-band image: (1) generation of a narrow-band image from a captured RGB image, (2) sub-pixel correspondence search between captured and generated narrow-band images, (3) geometric correction of the image to generate an eleven-band image, and (4) color reproduction.

Generation of narrow band-pass image from captured RGB image

The proposed system acquires an RGB image and eight monochrome images. Let the RGB image be a reference image in the correspondence search. To improve accuracy of correspondence detection, the RGB image is converted into a monochrome image that has been captured by a monochrome camera with the band-pass filter. We applied the idea of a virtual multispectral camera to image conversion [12].

First, spectral reflectance is estimated from the captured RGB image. Let the spectral power distribution of illumination and spectral reflectance of the object be $W(\lambda)$ and $f(\lambda)$. The observed spectrum is represented as

$$v(\lambda) = W(\lambda)f(\lambda), \qquad (1)$$

where λ is the wavelength. Let us consider a situation where the reflected light is captured using a sensor. Let the spectral sensitivity of the sensor be $\mathbf{S}$. Let the matrix whose diagonal elements represent the spectral power distribution of illumination be $\mathbf{W}$. Equation (1) can be rewritten into vector representation as

$$\mathbf{v} = \mathbf{Wf} . \qquad (2)$$

Using the Wiener estimation method [13], spectral reflectance is estimated from the signal of the RGB camera, $\mathbf{c}_{RGB} = \mathbf{SWf} = \mathbf{Hf}$, as

$$\hat{\mathbf{f}} = \mathbf{Gc}_{RGB}, \qquad \mathbf{G} = \mathbf{RH}^t\{\mathbf{HRH}^t\}^{-1}, \qquad (3)$$

where $\mathbf{G}$ is the Wiener estimation matrix obtained from $\mathbf{H}$ and $\mathbf{R}$ is a priori knowledge about the spectral reflectance of objects. We used a correlation matrix $\mathbf{R}$, which is modeled on a first-order Markov process covariance matrix, in the form

$$\mathbf{R} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{N-1} & \rho^{N-2} & \cdots & \cdots & 1 \end{pmatrix}, \qquad (4)$$
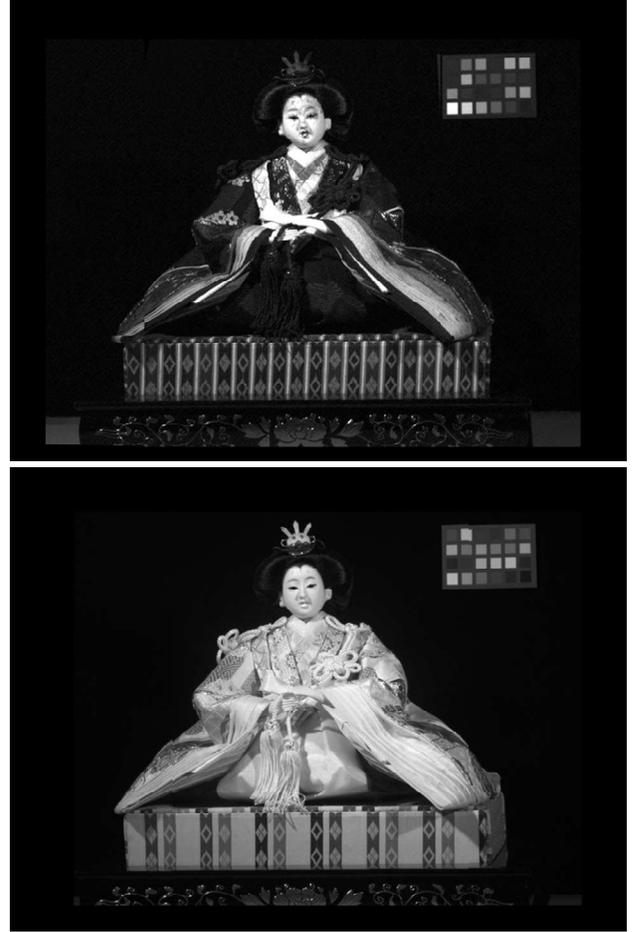

Figure 2. Sample of captured narrow-band images. (Top: 450 nm Bottom: 700nm)

where $0 \le \rho \le 1$ is the adjacent element correlation factor; we set $\rho = 0.999$ in our experiments.

Here, let us consider image capturing simulation using $\hat{\mathbf{f}}$, and a virtual camera whose spectral sensitivity is represented as $\mathbf{S}v$. The calculated signal of the virtual camera is represented as

$$\mathbf{c}v = \mathbf{S}v\mathbf{W}\hat{\mathbf{f}} = \mathbf{S}v\mathbf{WGc}_{RGB}. \qquad (5)$$

Equation 5 shows that RGB camera signals $\mathbf{c}_{RGB}$ can be converted into multispectral data captured using a VMSC with $\mathbf{S}v$.

Phase-Only Correlation (POC) function

To find corresponding points between a stereo image pair, we use local block matching in combination with sub-pixel correspondence matching by the phase-only correlation (POC) method [14]. POC, a high-accuracy image matching technique, uses phase information in the Fourier domain to estimate translation between two images with sub-pixel accuracy. Details of POC are described in the following.

Consider two N_1 x N_2 images, $f(n_1, n_2)$ and $g(n_1, n_2)$, where we assume that the index ranges are $n_1 = -M_1, \cdots, M_1$ and $n_2 = -M_2, \cdots, M_2$ 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 2-D discrete Fourier transforms (DFTs) of the two images. $F(k_1, k_2)$ and $G(k_1, k_2)$ are given by

$$F(k_1, k_2) = \sum_{n_1, n_2} f(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2}$$
$$= A_F(k_1, k_2) e^{j\theta_F(k_1, k_2)}, \qquad (6)$$

$$G(k_1, k_2) = \sum_{n_1, n_2} g(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2}$$
$$= A_G(k_1, k_2) e^{j\theta_G(k_1, k_2)}, \qquad (7)$$

where $k_1 = -M_1, \cdots, M_1$, $k_2 = -M_2, \cdots, M_2$, $W_{N_1} = e^{-j\frac{2\pi}{N_1}}$, $W_{N_2} = e^{-j\frac{2\pi}{N_2}}$, 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

$$R(k_1, k_2) = F(k_1, k_2) \overline{G(k_1, k_2)},$$
$$= A_F(k_1, k_2) A_G(k_1, k_2) e^{j\theta(k_1, k_2)}, \qquad (8)$$

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

$$\hat{R}(k_1, k_2) = \frac{F(k_1, k_2) \overline{G(k_1, k_2)}}{\left| F(k_1, k_2) \overline{G(k_1, k_2)} \right|}$$
$$= e^{j\theta(k_1, k_2)}, \qquad (9)$$

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

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

where $\sum_{k_1, k_2}$ denotes $\sum_{k_1 = -M_1}^{M_1} \sum_{k_2 = -M_2}^{M_2}$.

Sub-pixel correspondence search

Consider $f_c(x_1, x_2)$ as a 2-D image defined in continuous space with real-number indexes x_1 and x_2. Lets δ_1 and δ_2 represent sub-pixel 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

$$f(n_1, n_2) = f_c(x_1, x_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \qquad (11)$$
$$g(n_1, n_2) = g_c(x_1 - \delta_1, x_2 - \delta_2)|_{x_1 = n_1 T_1, x_2 = n_2 T_2}, \qquad (12)$$

where T_1 and T_2 are the spatial sampling intervals, and index ranges are given by $n_1 = -M_1, \cdots, M_1$ and $n_2 = -M_2, \cdots, M_2$. 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

$$G(k_1, k_2) \cong F(k_1, k_2) \cdot e^{-j\frac{2\pi}{N_1} k_1 \delta_1} e^{-j\frac{2\pi}{N_2} k_2 \delta_2}. \qquad (13)$$

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

$$\hat{R}(k_1, k_2) \cong e^{-j\frac{2\pi}{N_1} k_1 \delta_1} e^{-j\frac{2\pi}{N_2} k_2 \delta_2}. \qquad (14)$$

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

$$\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1 k_2} \hat{R}(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2}$$
$$\cong \frac{\alpha}{N_1 N_2} \frac{\sin\{\pi(n_1 + \delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\}} \frac{\sin\{\pi(n_2 + \delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}}, \qquad (15)$$

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 α decreases (without changing the function shape itself), when small noise components are added to the original images. Hence, we assume $\alpha \leq 1$ in practice. See more details in Shibahara et al. [15].

Generation of eleven-band image and color reproduction.

The monochrome images captured with the band-pass filter are transformed and adjusted to the RGB image. To estimate parameters for image transformation, the corresponding points obtained from the POC-based correspondence matching are used. Projective transformation is a simple method and works well for two-dimensional (2-D) objects. However, when the target object has a 3-D shape, several adjustment errors remain when projective transformation is applied to the whole image. The adjustment errors cause artifacts (e.g. double edges or pseudo color) to the resultant images of color reproduction. To avoid the adjustment errors, there are two approaches. One is to divide the captured images into several sub-images and apply projective transformation to each sub-image. Then, all transformed sub-images are merged into the whole image of the band. Projective transformation models are suitable for implementation on Graphics

Processing Units (GPUs), since their computations can be performed in parallel. The second approach is to adapt nonlinear transformation like the thin-plate spline (TPS) model [16] to the image transformation. For 3-D objects, which have a more complex shape, TPS-based image transformation can work better than projective transform [17]. However, computational cost becomes much larger.

Spectral reflectance of the object is estimated from the obtained eleven-band image using the Wiener estimation method. A color image for display ion a monitor is obtained using the estimated spectral reflectance, spectral power distribution of illumination for observation, and tone-curves and chromaticity values of primary-colors of the display monitor. Even when the illumination light used at the observation site is different from that for image capturing (e.g. daylight is used for image capturing and a fluorescent lamp is used at the observation site), the color observed under the observation light can be reproduced as if the object is in front of observers.

Experiments

Characteristics of the experimental equipment

A color digital camera and eight monochrome digital cameras (GRAS-20S4C and GRAS-20S4M, Point Grey Research Inc.) with the IEEE 1394 b (800 Mbit/s) interface are used. These camera can write out raw image data without any color correction and can take 2-M pixels images (1600 x 1200 pixels), each of which has bit depth of 10 bits. The frame rat e of image transfer is 15 fps at maximum for a full-size image, but 30 fps can be achieved when image size is reduced to 800 x 600 pixels (XGA). All cameras are synchronized by using a function generator and images of moving objects can be taken as still- or moving images. The baseline length of the two neighboring cameras is 55 mm.

Figure 3 shows the spectral sensitivities of the nine cameras used in this experiment. Visible wavelength is divided in to eight with little overlap. Note that each camera has sensitivity higher than 400 nm and lower than 730 nm since the UV- and IR-cut filters are attached to the image sensor. Before starting the experiments, the spectral sensitivity of the cameras, the illumination spectrum, and monitor characteristics (primary color and tone curves) were measured.

Results

First, we confirmed the accuracy of the estimated spectral reflectance by comparing it with that measured with a spectrometer. As a target object, we used Macbeth ColorChecker[TM]. Figure 4 shows the measured and estimated spectral reflectance. As reference data, the estimation results obtained using a two-shot six-band system are also shown. Although some errors are found in the short-wavelength domain between 380 and 420 nm, spectral reflectance seems to be well estimated using the proposed camera system. These errors were caused by a lack of camera sensitivity in this wavelength domain. For the 24 color patches of Macbeth ColorChecker[TM], this eleven-band camera system has achieved an average color difference of $\Delta Ea*b*=1.8$, almost the same as that for the six-band camera system. However, the RMSE of spectral reflectance

estimated from the eleven-band image was smaller than that of six-band. The average RMSE values for Macbeth's 24 color patches were 0.0038 (eleven-band) and 0.0043 (six-band). This means that an eleven-band image is better for analyzing spectral reflectance.

Next, images of a 3-D object (a Japanese doll) were taken with the proposed system. The image captured for each channel is shown in Fig. 5. Figure 6 shows the grey-scale image generated from the reference RGB image (top) and the captured and the generated images (bottom) for 690 nm. Comparing the captured image and the image generated from the RGB image, it seems that the generated image is well reproduced. The correspondence search based on POC was conducted under the condition that the interval of reference point was 10 pixels. Before conducting the
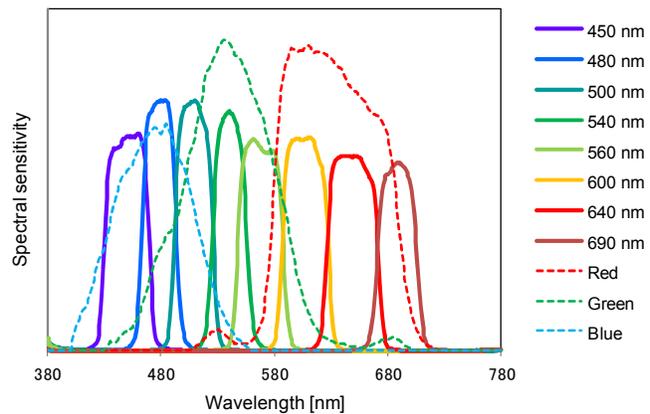

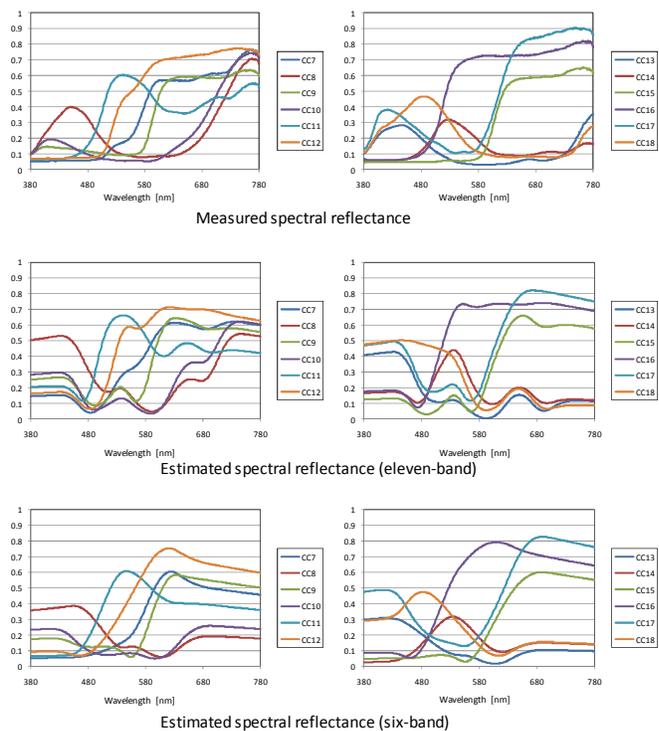*Figure 3. Spectral sensitivity of the eleven-band camera.*

Figure 4. A part of measured and estimated spectral reflectance of Macbeth Color Checker[TM]

Figure 5. Images captured with the proposed system.

Figure 6. Image generated from the reference RGB image.

experiment, we checked the number of detectable corresponding points using the RGB image and confirmed that the number was 7400. The results of the correspondence search of all bands are shown in Fig. 7. Comparing the results of the proposed method with the results using the RGB image itself as a reference image, detection ratio of the correspondences became almost twice. The averaged detective ratio improved 33% to 56%. These results show that the accuracy of correspondence search for a multi-band image was improves using the proposed method.

Final images after image transformations and color reproduction of the images of the 3-D objects are carried out. We compared the color of the real object and that of image displayed on a LCD monitor by eye, and confirmed that the object color was well reproduced.

Summary and future work

A novel eleven-band acquisition system using a nine-view stereo camera has been proposed. The proposed system consists of eight monochrome cameras with eight different narrow band-pass filters and an RGB camera. The captured RGB image is converted into narrow-band images to use for accurate correspondence search. The detected corresponding points are used for estimating parameters of image transformation and an eleven-band image is generated. Experimental results show that correspondence detection and spectral reflection estimation can work more accurately than the nine-band stereo camera system comprising nine monochrome cameras [11].

The proposed system can acquire motion pictures. We are now developing a stereo six-band video system in which image processes are implemented on GPUs [10]. This system will be extended to the nine-view camera system for application to moving objects.
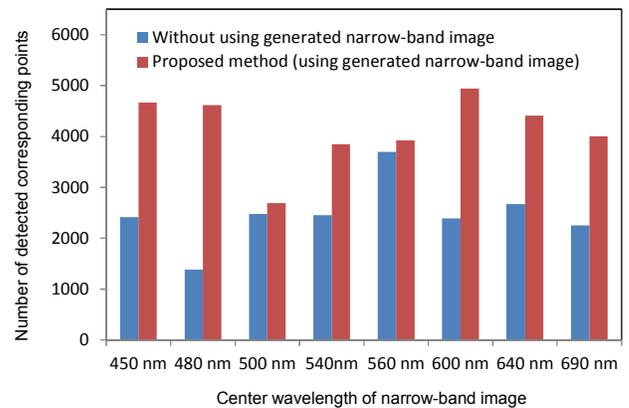

Figure 7. Results of correspondence search.

To apply the proposed system for objects with more complicated shapes or natural scenes, depth information estimated from a stereo image captured by the proposed camera system should be used. In addition, foreground and background objects should be separated when image transformation is performed.

References

[1] P.D. Burns and R.S. Berns, "Analysis of multispectral image capture," Proc. 4th Color Imaging Conf. (CIC4), 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, R. Iwama, Y. Ohya, T. Obi, N. Ohyama, Y. Komiya, and T. Wada, "Natural color reproduction in the television system for telemedicime", Proc. SPIE 3031, 482-489 (1997).

[4] S. Tominaga et al., "Object Recognition by Multi-Spectral Imaging with a Liquid Crystal Filter.", Proc. Conf. on Pattern Recognition, vol.1, 708-711 (2000).

[5] S. Helling et al., "Algorithms for spectral color stimulus reconstruction with a seven-channel multispectral camera.", Proc. Second European Conference on Colour in Graphics, Imaging, and Vision (CGIV), 254-258 (2004).

[6] M. Hashimoto, "Two-Shot type 6-band still image capturing system using Commercial Digital Camera and Custom Color Filter.", Proc. Fourth European Conference on Colour in Graphics, Imaging, and Vision (CGIV), 538-541 (2008).

[7] K. Ohsawa et al., "Six-band HDTV camera system for spectrum-based color reproduction.", J. Imaging Sci. and Technol., 48, 2, 85-92 (2004)..

[8] M. Tsuchida et al., "A stereo one-shot multi-band camera system for accurate color reproduction.", Proc. ACM Siggraph, Poster No. 66 (2010).

[9] R. Shrestha et al., "One-shot multispectral color imaging with a stereo camera. ", Proc. SPIE-IS&T Electronic Imaging, 7876, 797609-1 - 797609-11 (2011).

[10] M. Tsuchida et al., "A six-band stereoscopic video camera system for accurate color reproduction. ", Proc. 20th Color and Imaging conf., 117-122 (2012).

[11] M. Tsuchida et al., "A stereo nine-band camera for accurate color and spectrum reproduction", Proc. ACM Siggraph, Poster No. 18, (2012).

[12] T. Uchiyama et al., "A Method for the Unified Representation of Multispectral Images with Different Number of Bands", J. Imaging Sci. and Technol., Vol. 48, No. 2, 120- 124 (2004).

[13] W. K. Pratt et al., "Spectral estimation techniques for the spectral calibration of a color image scanner.", Applied Optics, 15, 73–75. (1976).

[14] H. Takita et al., "High-accuracy image registration based on phase-only correlation.", IEICE Trans. of Fundamentals, Vol. E86-A, no.8, pp.1925-1934. (2003).

[15] T. Shibahara et al., "A sub-pixel stereo correspondence technique based on 1D phase-only correlation". Proc. Int'l Conf. Image Processing, 221–224. (2007).

[16] F. L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 11, No. 16, pp.567-585. (1989).

[17] M. Tsuchida, et al., "Evaluating Color Reproduction Accuracy of Stereo One-shot Six-band Camera System. ", Proc. 19th Color and Imaging conf., 326-331.(2011).

Author Biography

Masaru Tsuchida received the B.E., M.E and Ph.D. degrees from the Tokyo Institute of Technology, Tokyo, in 1997, 1999, 2002, respectively. In 2002, he joined NTT 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. Since 2011, he has been a visiting professor at Ritsumeikan University, Kyoto.