Efficient POC-based Correspondence Detection Method for Multi-channel Images

M. Tsuchida[1], S. Sakai [2], K. Ito[2], K. Kashino[1], J. Yamato[1], and T. Aoki [2]
1. Communication science laboratories, NTT Corporation / 3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa, Japan
2. Graduate School of Information Sciences, Tohoku University / 6-6-05, Aramaki Aza Aoba, Aoba-ku, Sendai-shi, Japan

Abstract

We propose a new POC (phase-only correlation)-based high-accuracy correspondence detection method for multi-channel images. There is the possibility of improving detection accuracy because conventional POC-based methods do not use color information. In the proposed method, a normalized cross spectrum (or cross-phase spectrum) and weight are calculated for each color channel in the Fourier domain. The weight is determined by the amplitude of the cross spectrum. The weighted normalized cross spectra of all color channels are combined and inverse Fourier transformation is conducted to obtain a POC function. An experimental evaluation of the matching accuracy between the conventional POC-based method and the proposed method shows that RMSE decreased approximately 25%. This paper also describes an application of the proposed method to a stereo image matching. The average detection ratio of correspondences between a stereo-pare image is improved 64% to 95%.

Introduction

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. However, accurate color reproduction of an object under arbitrary illumination conditions is difficult with current imaging systems based on three-channel (i.e., red, green, and blue) image capturing. Multispectral imaging technology, which estimates the spectrum using multi-channel data, provides accurate color reproduction. Several types of multi-channel camera systems have been developed [1-11]. Multi-channel images have richer color information than current RGB images, which enables us to improve the accuracy of image recognition. Therefore, multispectral imaging contributes to digital archiving for cultural heritage preservation and in the medical and some industrial fields. In addition, precise correspondence detection improves the accuracy of 3-D shape and image displacement estimations.

Some methods have been developed for matching between two multi-channel images with high accuracy [12-14]. Although not focused on multi-channel images, several methods using the phase-only correlation (POC) function have been proposed [15, 16], and they are highly effective due to their sub-pixel accuracy and robust performance against illumination changes, noise, and color shifts caused by differences in the spectral sensitivity of a camera (Note that the POC function is sometimes called the "phase correlation function"). POC-based methods should also work well for multi-channel images, therefore, we focus on a POC-based method in this paper.

We have used a correspondence detection technique that combines local block matching by the POC method and a coarse-to-fine strategy based on pyramid representation to find corresponding points [17]. However, conventional POC-based image matching does not use color information. Captured color images are converted into grayscale images and the obtained images are used for the image-matching process. This also contributes to the robust performance of image matching against the image-capturing condition, such as illumination variation. On the other hand, an image of each color channel should be dealt with in POC-based in the same way a grayscale image is, because color images have richer information than grayscale images. Then, color information should contribute to improving the accuracy of image matching. However, nobody has attempt to use color information to improve the accuracy of image matching.

In this paper, we propose a new POC-based correspondence detection method using color information and show its efficiency for correspondence detection between multi-channel images in experiments. In this method, the obtained normalized cross spectra (or cross-phase spectra) of each color channel image are weighted and combined in the Fourier domain. Inverse discrete Fourier transformation (DFT) is conducted on the resultant cross spectrum to obtain new POC functions and registration error between the two images is obtained. The experimental results for the proposed method are compared with those for the conventional one. The secound section describes the new POC correspondence detection method using color information. The third section shows the experimental evaluation of the matching accuracy of the proposed method. The fourth section describes an application of the method to stereo matching for stereo-pairs of multi-channel images and its performance evaluation. Finally, we conclude this paper.

Correspondence Detection Method based on POC method

Conventional POC Function

The principle of POC is briefly described below. Consider two images, $f(x,y)$ and $g(x,y)$, whose size is N_1 x N_2 and number of color channel is C (e.g., for an RGB image, $C=3$). The index ranges are $x=-M_1,\cdots,M_1$ and $y=-M_2,\cdots,M_2$, and then $N_1=2M_1+1$ and $N_2=2M_2+1$. The images of i-th color channel images is represented as $f_i(x,y)$ and $g_i(x,y)$. In the conventional method, images of all color channels are averaged as follows:

$$\bar{f}(x,y)=\sum_{i=1}^{C} f_i(x,y), \tag{1}$$

$$\bar{g}(x,y)=\sum_{i=1}^{C} g_i(x,y). \tag{2}$$

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 then given by

$$\begin{aligned} F(k_1,k_2) &= \sum_{x,y} \bar{f}(x,x) W_{N1}^{k1x} W_{N2}^{k2y} \\ &= A_F(k_1,k_2) e^{j\theta_F(k1,k2)}, \end{aligned} \tag{3}$$

$$\begin{aligned} G(k_1,k_2) &= \sum_{x,y} \bar{g}(x,y) W_{N1}^{k1x} W_{N2}^{k2y} \\ &= A_G(k_1,k_2) e^{j\theta_G(k1,k2)}, \end{aligned} \tag{4}$$

where $k_1=-M_1,\cdots,M_1$, $k_2=-M_2,\cdots,M_2$, $W_{N1}=e^{-j\frac{2\pi}{N1}}$, $W_{N2}=e^{-j\frac{2\pi}{N2}}$, and the operator $\sum_{x,y}$ denotes $\sum_{x=-M1}^{M1}\sum_{y=-M2}^{M2}$. $A_F(k_1,k_2)$ and $A_G(k_1,k_2)$ are amplitude components, and $e^{j\theta_F(k1,k2)}$ and $e^{j\theta_G(k1,k2)}$ 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

$$\begin{aligned} R(k_1,k_2) &= F(k_1,k_2)\overline{G(k_1,k_2)}, \\ &= A_F(k_1,k_2) A_G(k_1,k_2) e^{j\theta(k1,k2)}, \end{aligned} \tag{5}$$

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

$$\begin{aligned} \hat{R}(k_1,k_2) &= \frac{F(k_1,k_2)\overline{G(k_1,k_2)}}{\left|F(k_1,k_2)\overline{G(k_1,k_2)}\right|} \\ &= e^{j\theta(k1,k2)}, \end{aligned} \tag{6}$$

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

$$\hat{r}(x,y)=\frac{1}{N_1N_2}\sum_{k1,k2}\hat{R}(k_1,k_2)W_{N1}^{-k1x}W_{N2}^{-k2y}, \tag{7}$$

where $\sum_{k1,k2}$ denotes $\sum_{k1=-M1}^{M1}\sum_{k2=-M2}^{M2}$.

Proposed POC Function for Multi-channel images

Texture appearance differs among color channel images. This means that the normalized cross spectrum of each color channel image is also different. Then, we use the weighted average of the normalized cross spectrum of each color channel to obtain POC function.

Consider images of *i*-th color channel, $f_i(x_1,x_2)$ and $g_i(x_1,x_2)$, where the index ranges are $i=1,\cdots,C$. The weighted average of the normalized cross spectra $\hat{R}_W(k_1,k_2)$ is represented as

$$\hat{R}_W(k_1,k_2)=\frac{\sum_i w_i(k_1,k_2)R_i(k_1,k_2)}{\sum_i w_i(k_1,k_2)}, \tag{8}$$

where $w_i(k_1,k_2)$ denotes the weight of each frequency index. The POC function $\hat{r}(x,y)$ is the 2-D inverse DFT of $\hat{R}_W(k_1,k_2)$.

Let us consider the weight function $w_i(k_1,k_2)$. $w_i(k_1,k_2)=1$ is the simplest case, which means that $\hat{r}(x,y)$ is an average of the POC functions of each color channel. For the proposed POC function for multi-channel images, the weight function is defined using the energy of the frequency index as $w_i(k_1,k_2)=\left|F_i(k_1,k_2)\overline{G_i(k_1,k_2)}\right|$ and the weighted average of normalized cross spectra is described as

$$\begin{aligned} \hat{R}_W(k_1,k_2) &= \frac{\sum_i w_i(k_1,k_2)R_i(k_1,k_2)}{\sum_i w_i(k_1,k_2)} \\ &= \frac{\sum_i F_i(k_1,k_2)\overline{G_i(k_1,k_2)}}{\sum_i \left|F_i(k_1,k_2)\overline{G_i(k_1,k_2)}\right|}. \end{aligned} \tag{9}$$

This weight function is equivalent to the ratio of the energy of frequency index of each color channel, and it affects a color channel with large amplitude more strongly than it does one with small amplitude. The proposed POC function is given by conducting the 2-D inverse DFT of $\hat{R}_W(k_1,k_2)$.

Sub-pixel Image Registration

Consider $f(x,y)$ as a 2-D image defined in continuous space with real-number index x and y. Let δ_1 and δ_2 represent sub-pixel displacement of $f(x,y)$ in x and y directions, respectively. The displaced image can then be represented as $f(x-\delta_1,y-\delta_2)$. Assume that $f(n_1,n_2)$ and $g(n_1,n_2)$ are

spatially sampled images of $f(x,y)$ and $f(n_1-\delta_1, n_2-\delta_2)$, defined as

$$f(n_1,n_2) = f(x,y)\big|_{x=n_1T_1, y=n_2T_2}, \quad (10)$$

$$g(n_1,n_2) = g(x-\delta_1, y-\delta_2)\big|_{x=n_1T_1, y=n_2T_2}, \quad (11)$$

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 in properties between the Fourier transform defined in continuous space and that defined in discrete space, $G(k_1,k_2)$ can be represented as

$$G(k_1,k_2) \cong F(k_1,k_2)\cdot e^{-j\frac{2\pi}{N_1}k_1\delta_1} e^{-j\frac{2\pi}{N_2}k_2\delta_2}, \quad (12)$$

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

$$\hat{R}_W(k_1,k_2) \cong e^{-j\frac{2\pi}{N_1}k_1\delta_1} e^{-j\frac{2\pi}{N_2}k_2\delta_2}, \quad (13)$$

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

$$\hat{r}_w(n_1,n_2) = \frac{1}{N_1N_2}\sum_{k1k2}\hat{R}_w(k_1,k_2)W_{N_1}^{-k_1n_1}W_{N_2}^{-k_2n_2}$$

$$\cong \frac{\alpha}{N_1N_2}\frac{\sin\{\pi(n_1+\delta_1)\}}{\sin\{\frac{\pi}{N_1}(n_1+\delta_1)\}}\frac{\sin\{\pi(n_2+\delta_2)\}}{\sin\{\frac{\pi}{N_2}(n_2+\delta_2)\}}, \quad (14)$$

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 of α decreases (without changing the function shape itself), when small noise components are added to the original images. Hence, we assume $\alpha \le 1$ in practice.

Experimental Evaluation of Matching Accuracy

Experimental Setup

Figure 1 shows the experimental setup. Images were captured with a monochrome CCD camera (GRAS-14S5M-C, Point Grey Research Inc.), which was mounted on a micro-stage that allows precise alignment of the camera position. The pixel size of the CCD sensor is 6.45 μm. The focal length of the camera lens is 12 mm. An xenon lamp (LAX-102, Asahi Spectra) with wide bandpass and narrow bandpass filters was used for illumination, and RGB (three-bands) and eight-band images were captured at each camera position. Figure 2 shows the spectral sensitivity of the image capturing system. The target object in the first experiment

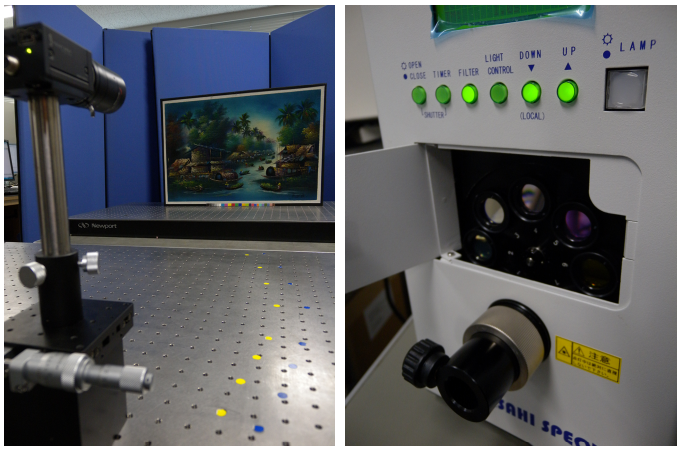


Figure 1. *Experimental setup.*
Left: Target object and camera. Right: Xenon lamp with band-pas filters.

was the oil painting shown in Fig. 3. The target object was fixed 110 cm away from the camera in such a way that it was almost parallel to the focal plane of the CCD camera. The micro-stage was moved horizontally 50 times, and the length of each movement was 0.1 mm (which is equivalent to 0.16 pixels on the

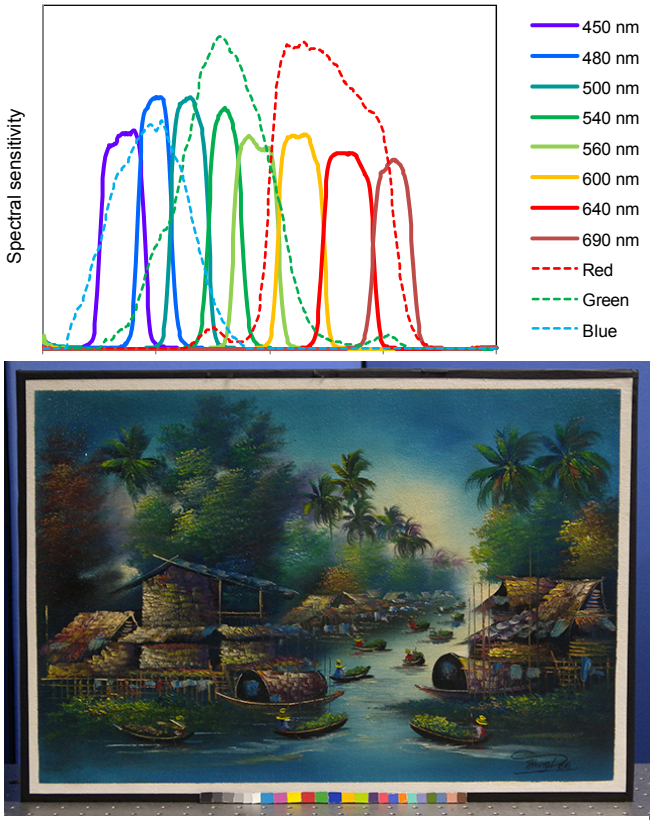


Figure 3*. Oil printing used in experiments.*

***Figure 4**. Images used for correspondence tracking.*

image).

Results

We estimated the displacement using the experimental setup described above. The reference image had been taken before the camera was moved. At every camera position, we searched for the corresponding points with sub-pixel accuracy. Let the displacement vector between the reference point in the initial image and the corresponding point in the j-th image be $\mathbf{\delta j} = (\delta_{j,1}, \delta_{j,2})$. We evaluated the performance of the image matching technique in the following manner. There are 50 sets of data containing the actual displacement Δ_j [mm] (i.e., the displacement of the micro-stage) and the estimated displacement $(\delta_{j,1}, \delta_{j,2})$ [pixel]. Ideally, the estimated displacement $\delta_{j,1}$ (or $\delta_{j,2}$) must be proportional to Δ_j . To evaluate the correspondence accuracy, we calculate a linear approximation in a least-square sense, where a_1 [pixel/mm] and a_2 [pixel/mm] are constant coefficients. The error ε_j in displacement estimation at the j-th position is calculated by

$$\varepsilon_j = \sqrt{(\delta_{j,1} - a_1 \cdot \Delta_j)^2 + (\delta_{j,2} - a_2 \cdot \Delta_j)^2} \ . \qquad (15)$$

We evaluated three algorithms: the conventional POC-based method (which does not use color information), the method using the averaged POC function of all color channels, and the proposed method. In the conventional method, grayscale images obtained by averaging the image of each color channel are used. The size of the image block is 31 x 31 pixels. Correspondence tracking was conducted for the four areas shown in Fig. 4. Figure 5 shows estimated displacements of a reference point obtained by the three algorithms. Although a linear relationship between estimated displacements of captured images and those of the micro-stage are confirmed, the differences between the three algorithms look small and are difficult to evaluate. Then, RMSEs (root means square

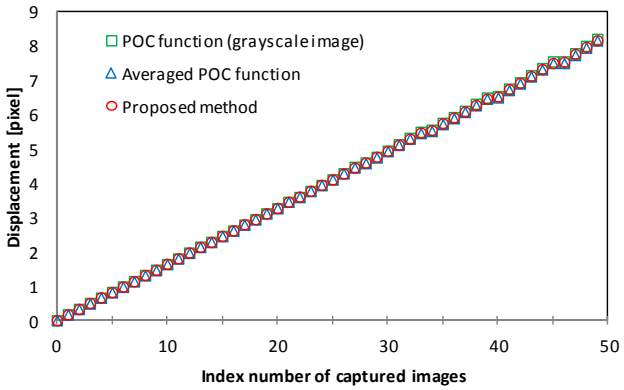
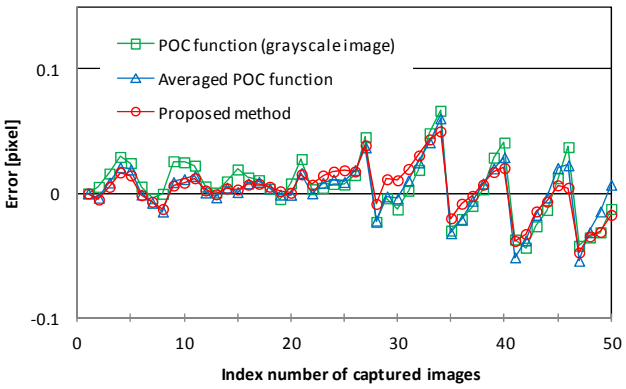


***Figure 5**. Result of displacement estimation.*



	RMSE[pixel]
POC function (grayscale image)	0.02417
Averaged POC function	0.02140
Proposed method	0.01946

***Figure 6**. Estimated displacement errors of an image (top) and the averaged RMSE of the four images (bottom).*

***Figure 7**. Stereo camera used in experiments.*

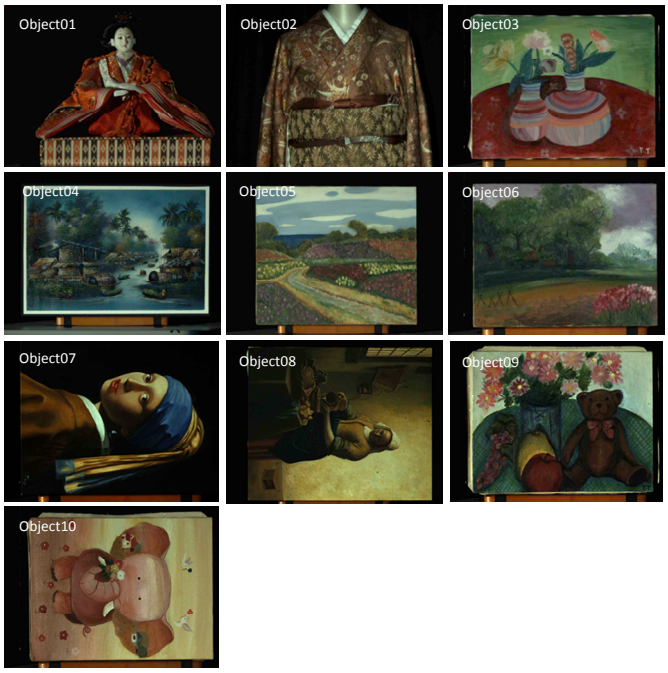


***Figure 8**. Target objects.*

errors) of estimated displacement using for the three algorithms were calculated. Figure 6 shows estimation errors for an image at each camera position and the averaged RMSE of estimated displacement. The graph shows the error in the estimated displacement of the proposed method is totally the smallest. The graph shows the estimated displacement error increases as the length of camera movement increases. Theoretically, the POC-based method can detect displacement whose length is 25% of the image-block size. The image-block size in this experiment was 31 pixels, which means that the theoretical limitation of detectable displacement is approximately 7.5 pixels. Then, the error tends to become larger when the displacement approaches the limitation. Even so, the estimation accuracy is kept under ±0.1 pixels. The results of the averaged RMSE of four images show that the proposed method improved accuracy of estimated displacement approximately 25% compared to the conventional POC-based method using grayscale images.

Stereo Matching for Multi-channel Images

Several stereoscopic multi-channel image capturing systems that combine multi-spectrum and stereo imaging technologies have been developed [5-7]. In this experiment, we implemented our proposed method for multi-channel stereo image matching. Figure 7 shows the stereo monochrome camera used in this experiment. The focal length of the camera lens is 25 mm. The distance between the cameras was 4.4 cm and the distance between the cameras and the object was 4 m. The effect of the stereo disparity was therefore small. Eight-band images were captured by combining this stereo monochrome camera setup and the xenon lamp system mentioned above.

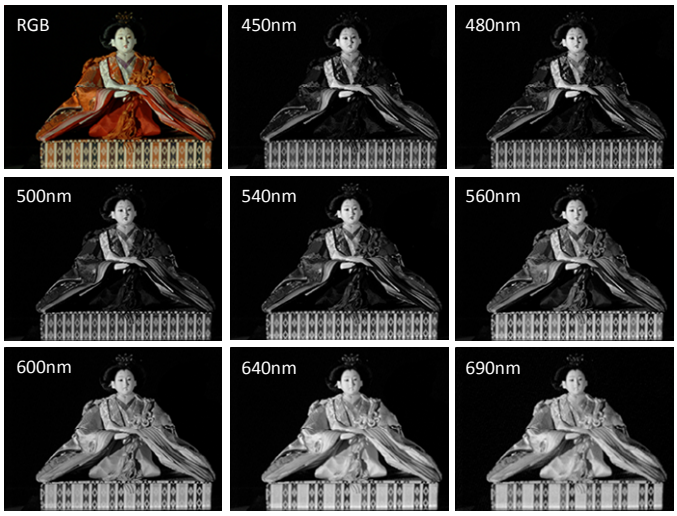


***Figure 9**. Examples of captured multi-channel images.*

	Conventional method (RGB image)	Conventional method (8-band image)	Proposed method
Object 01	4980 (72.3%)	4998 (72.6%)	6703 (97.3%)
Object 02	1981 (28.8%)	2400 (34.8%)	6736 (97.8%)
Object 03	3932 (57.1%)	4087 (59.3%)	6381 (92.7%)
Object 04	5449 (79.1%)	5629 (81.7%)	6654 (96.6%)
Object 05	4491 (65.2%)	4930 (71.6%)	6673 (96.9%)
Object 06	5279 (76.7%)	5370 (78.0%)	6814 (98.9%)
Object 07	5814 (84.4%)	6066 (88.1%)	6786 (98.5%)
Object 08	2721 (39.5%)	2777 (40.3%)	6425 (93.3%)
Object 09	3161 (45.9%)	3216 (46.7%)	6085 (88.4%)
Object 10	3178 (46.1%)	4698 (68.2%)	6402 (93.0%)
Average detective ratio	59.5%	64.1%	95.3%

***Figure 10**. Number of detected correspondences.*

Figure 8 shows photographs of the target objects: a Japanese doll, a traditional Japanese kimono hung on a mannequin, and oil paintings. The total number of the target objects was ten. Figure 9 shows examples of multi-channel images of the Japanese doll. We conducted correspondence search using the conventional and the proposed POC-based methods. The size of captured images was XGA (1024 x 768 pixels). We identified reference points on each of the reference images in 10 pixel intervals, and the range of search was ±16 pixels. The threshold value on the POC function was 0.3.

Before conducting the experiment, we checked the number of detectable correspondences using an image captured from the same viewpoint and confirmed that the average for the target objects was 6887. Results of the correspondence detection using eight-band stereo images are shown in Fig. 10. The number of

detected correspondences was divided by that of detectable correspondences at the same viewpoint, and the detection ratio was obtained. Comparing results for the three POC-based methods, we can show that the average detection ratio is improved from 64% to 95% when the proposed method is used.

Conclusion

We proposed a new efficient POC-based high-accuracy correspondence detection method for multi-channel image. Eight-band images were used in the experiment to evaluate the proposed method. We evaluated the RMSE of displacement estimations between the conventional and proposed methods, and we proved that the proposed method using color information improves the accuracy of estimated displacement approximately 25% compared to the conventional POC-based method. We also implemented the proposed method for multi-channel stereo image matching. The proposed method improved the average detection ratio from 65% to 95%.

References

[1] P.D. Burns and R.S. Berns, "Analysis of multispectral image capture," Proc. 4th Color Imag. Conf. (CIC4), pp. 19-22. (1996).

[2] S. Tominaga, "Multichannel vision system for estimating surface and illuminant functions," J. Opt. Soc. Am. A, 13(11), pp. 2163-2173. (1996).

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

[4] K. Ohsawa et al., "Six-band HDTV camera system for spectrum-based color reproduction.", J. of Imaging Science and Technology, 48, 2, pp. 85-92. (2004).

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

[8] 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).

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

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

[11] Y. Murakami et al., "Hybrid-resolution multispectral imaging based on color filter array: Basic principles and computer simulation", Proc. CGIV2012, pp. 259-265. (2012).

[12] S. Mattoccia, et al., "Efficient templete matching for multi-channel images", Pattern Recognition Letters, vol. 32, pp. 694-700. (2011).

[13] M. Doi, et al., "Accurate stereo matching based on multiband imaging", Proc. SPIE, vol. 7864, pp. 786408-1 -786408-9. (2011).

[14] S. Saleem, et al., "A Modified SIFT Descriptor for Image Matching under Spectral Variations.", Springer LNCS, vol. 8156, pp. 652-661. (2013).

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

[16] C.D. Kuglin et al., "The phase correlation image alighnment method.", Proc. Int. Conf. on Cybernetics and Society, pp. 163-165, (1975).

[17] K. Takita et al., "A Sub-Pixel Correspondence Search Technique for Computer Vision Applications.", IEICE Trans. Fundamentals, Vol. E-87-A, No. 8, pp. 1913-1923. (2004).

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 researcher at Ritsumeikan University, Kyoto.