A six-band stereoscopic video camera system for accurate color reproduction

M. Tsuchida[1], S. Sakai[2], M. Miura[2], K. Ito[2], T. Kawanishi[1], K. Kashino[1], J. Yamato[1], and T. Aoki[2]
1. NTT 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

In order to accurate color reproduction, a novel six-band stereoscopic video acquisition and visualization system using multi-spectral and stereo imaging has been proposed. The proposed system consists of two consumer-model digital cameras and an interference filter whose spectral transmittance is comb shaped. In the process of generating a six-band image, the Phase-Only Correlation method (POC) is used to find accurate corresponding points between the stereo image pair, and projective transformation whose parameters are estimated from the correspondence is used to correct geometric relationship between the images. The Wiener estimation method is used for color reproduction. The all image processing steps after image capture is implemented on GPUs and the frame rate of the system is 30 fps is achieved when image size is XGA.

Figure 1. Six-band stereoscopic camera system.
The interference filter is attached to the right camera.

Introduction

In digital archiving for cultural heritage preservation, in the medical field, and in some industrial fields, the 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 [1-4], all the conventional systems are multi-shot and cannot take still images of moving objects and moving pictures.

Ohsawa et al. developed a six-band HDTV camera system [5]. However, the system requires very complex and expensive customized optics, which makes this system far from practical. In order to make multi-spectrum video systems pervasive, the equipment costs must be reduced. To meet this requirement, several stereo one-shot six-band image capturing systems that also combine multi-spectrum and stereo imaging techniques have been proposed [6-8]. They are based on two consumer-model digital cameras and color filters. Shresta et al. have presented only the concept of the six-band stereoscopic camera and have shown simulations of color reproduction [7]. Their paper omitted any method for generating a six-band image from a stereo image pair. Moreover, their system needs two different color filters. On the other hand, the system proposed and constructed by Tsuchida et al. [8], which is based on a two-shot six-band system [4], needs only a sheet of color filter and has been shown to work well for not only 2-D objects like a paintings or tapestries but also 3-D objects with slight self-occlusion and gloss (e.g. portrait and landscape photo) [6, 8]. This system achieves the averaged color difference of dEab= 0.97 (24 color patches of Macbeth ColorCheker™). However, this system is only for capturing still images. Tsuchida et al. have also been developing nine-band stereoscopic camera system for improving accuracy of estimated spectral reflectance [9].This sistem consists of none video cameras and can record moving pictures. However, searching correponding point and image deformation processes for avoiding the adjustment errors among captured nine images to generate nine-band images have not archived real-time performance yet, which makes it impossible to show resultant images of color reproduction on display monitor at the same time to image capturing.

This paper presents a novel six-band stereoscopic camera system for multi-spectrum video acquisition and visualization. In the process of generating a six-band image, the Phase-Only Correlation method (POC) [10] is used to find accurate corresponding points between the stereo image pair, and projective transformation whose parameters are estimated from the correspondence is used to correct geometric relationship between the images. The computation of POC, projective transformation model are suitable for implementation on GPUs (Graphics Processing Units), since their computations can be performed in parallel. Our system can achieve the frame rate of 30 fps, i.e., real-time performance. Although this system can acquire both spectral color information and depth information at the same time, this paper focuses on using captured spectral information to achieve accurate color reproduction. Note that the process of six-band creation does not use depth information.

Six-band stereoscopic video system

Figure 1 shows the proposed six-band stereoscopic video camera. The left camera captures a normal RGB image. The right camera mounted the interference filter whose spectral transmittance is comb-shaped 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 filter cuts off short wavelengths, 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 captured three-band stereo images are combined into a six-band image for color reproduction.

Figure 3 details the image processing procedure of the proposed system. There are four main steps: (i) rectification of a stereo image pair, (ii) sub-pixel correspondence matching, (iii) geometric correction of the image to generate a six-band image, and (iv) color reproduction. All the steps are implemented on GPU to achieve real-time processing of the system. To find corresponding points between a stereo image pair, the sub-pixel correspondence matching combining the local block matching using POC and the coarse-to-fine strategy with image pyramid is used [9]. POC is a high-accuracy image matching technique using phase information in Fourier domain which can estimate translation between two images in sub-pixel accuracy. POC is also robust against illumination change, noise and color shift caused by difference of spectral sensitivity of a camera. In order to reduce the computation time, we can use 1-D POC instead of 2-D POC in the case that the stereo image pair is rectified, since the rectified stereo image pair has only horizontal translations. Details of POC are described in the following.

Phase-Only Correlation (POC) function

Consider two N x 1 images, $f(n)$ and $g(n)$, where we assume that the index ranges are $n = -M, \cdots, M$ for mathematical simplicity, and hence $N = 2M + 1$. Let $F(k)$ and $G(k)$ denote the 1-D discrete Fourier transforms (DFTs) of the two images. $F(k)$ and $G(k)$ are given by

$$F(k) = \sum_n f(n) W_N^{kn} = A_F(k) e^{j\theta_F(k)}, \quad (1)$$

$$G(k) = \sum_n g(n) W_N^{kn} = A_G(k) e^{j\theta_G(k)}, \quad (2)$$

where $k = -M, \cdots, M$, $W_N = e^{-j\frac{2\pi}{N}}$, and operator $\sum_n$ denotes $\sum_{n=-M}^{M}$. $A_F(k)$ and $A_G(k)$ are amplitude components, and $e^{j\theta_F(k)}$ and $e^{j\theta_G(k)}$ are phase components.

The cross spectrum $R(k)$ between $F(k)$ and $G(k)$ is given by

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

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

$$\hat{R}(k) = \frac{F(k)\overline{G(k)}}{\left|F(k)\overline{G(k)}\right|} = e^{j\theta(k)}, \quad (4)$$

The POC function $\hat{r}(n)$ is the 1-D inverse DFT of $\hat{R}(k)$ and is given by

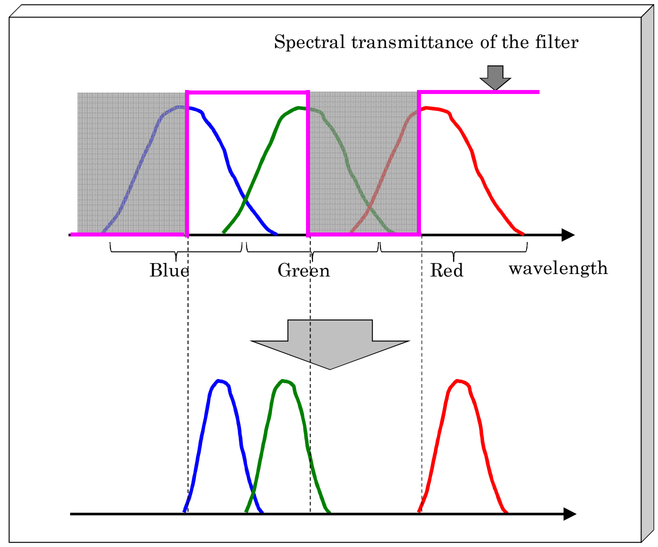

Figure 2. Principle of six-band image capture using an interference filter.

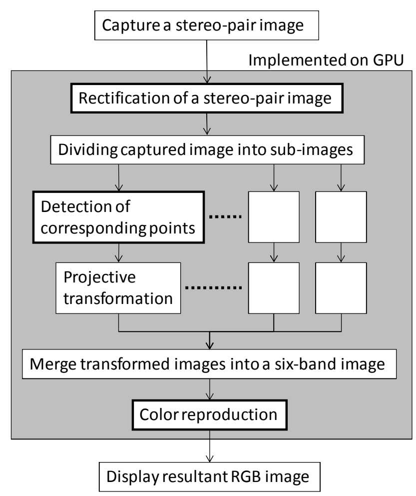

Figure 3. Diagram of image processing of the six-band stereoscopic video system

$$\hat{r}(n) = \frac{1}{N}\sum_k \hat{R}(k) W_N^{-kn}, \tag{5}$$

where $\sum_n$ denotes $\sum_{n=-M}^{M}$.

Sub-pixel image registration

Consider $f_c(x)$ as a 1-D image signal defined in continuous space with real-number index x . Let δ represent sub-pixel displacement of $f_c(x)$. That is, the displaced image signal can be represented as $f_c(x-\delta)$. Assume that $f(n)$ and $g(n)$ are spatially sampled images of $f_c(x)$ and $f_c(x-\delta)$, defined as

$$f(n) = f_c(x)\big|_{x=nT}, \tag{6}$$

$$g(n) = g_c(x-\delta)\big|_{x=nT}, \tag{7}$$

where T is the spatial sampling intervals, and index range is given by $n = -M, \cdots, M$. Let $F(k)$ and $G(k)$ be the 1-D DFTs of $f(n)$ and $g(n)$, respectively. After carefully considering the difference in properties between the Fourier transform defined in a continuous space and that defined in a discrete space, we can say that

$$G(k) \cong F(k) \cdot e^{-j\frac{2\pi}{N}k\delta}. \tag{8}$$

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

$$\hat{R}(k) \cong e^{-j\frac{2\pi}{N}k\delta}. \tag{9}$$

The POC function $\hat{r}(n)$ will be the 1-D inverse DFT of $\hat{R}(k)$, and is given by

$$\hat{r}(n) = \frac{1}{N}\sum_k \hat{R}(k) W_N^{-kn} \cong \frac{\alpha}{N}\frac{\sin\{\pi(n+\delta)\}}{\sin\{\frac{\pi}{N}(n+\delta)\}}, \tag{10}$$

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 $\alpha \le 1$ in practice. The peak position $n = -\delta$ of the 1-D POC function reflects the displacement between the two 1-D image signals. Thus, we can compute the displacement δ between signals $f(n)$ and $g(n)$ by estimating the true peak position of the 1-D POC function $r(n)$. We have also proposed the important techniques for improving the accuracy of 1-D image matching for sub-pixel correspondence matching: (i) function fitting for high-accuracy estimation of peak position, (ii) windowing to reduce boundary effects, (iii) spectral weighting for reducing aliasing and noise effects, (iv) averaging 1-D POC functions to improve peak-to-noise ratio and (v) coarse-to-fine strategy for robust correspondence search. See more details in Shibahara et al. [11].

Generation of six-band image

The image captured with the interference filter is deformed and adjusted to the other image by projective transformation. To estimate parameters for the projective transformation, the corresponding points obtained from the POC-based correspondence matching are used. Although the six-band image can be generated well in the case of 2-D objects like tapestries, several adjustment errors remain in the case of applying projective transformation to the whole image of 3-D object. The adjustment errors cause artifacts (e.g. double edges or pseudo color) to the resultant images of color reproduction. To avoid the adjustment errors, the captured images are divided into several sub-images and projective transformation is applied to each sub-image. Then, the all transformed sub-images are merged into be a six-band image.

Color reproduction using multi-spectrum data

As shown in Fig. 4, 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

$$I(\lambda) = W(\lambda) f(\lambda), \tag{11}$$

where λ is wavelength. Let us consider a situation where the reflected light is captured by an N - band sensor. Let the spectral sensitivity of the sensor be $\mathbf{S} = [S_1(\lambda), S_2(\lambda), \cdots, S_N(\lambda)]^T$. Let the matrix whose diagonal elements represent the spectral power distribution of illumination be $\mathbf{W}$. Equation (11) can then be rewritten in vector representation as

$$\mathbf{I} = \mathbf{W}\mathbf{f}. \tag{12}$$

By using the Wiener estimation method [12], the spectral reflectance is estimated from the camera signal, $\mathbf{c} = \mathbf{SWf} = \mathbf{Hf}$, as

$$\hat{\mathbf{f}} = \mathbf{Mc}, \qquad \mathbf{M} = \mathbf{RH}^t\{\mathbf{HRH}^t\}^{-1} \tag{13}$$

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.

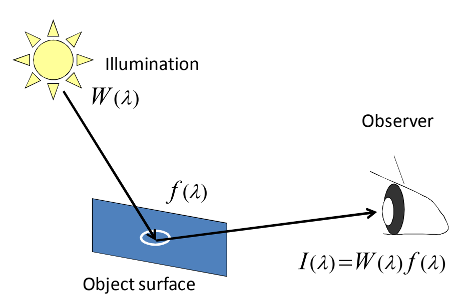

Figure 4. Geometric setup of object observation.

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

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

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

Using the estimated spectral reflectance, a spectral power distribution of illumination for observation, and tone-curves and chromaticity values of primary-colors of display monitor, output RGB signals are calculated. Even when the illumination light used in observation site is different from that for image capturing (e.g. daylight is used for image capturing and fluorescent lamp is used in 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

Two digital cameras (Grasshopper-20S4C, Point Grey Research Inc.) with IEEE 1394b (800 Mbit/s) interface are used. The camera can write out raw image data without any color correction and can take XGA-size (1024 x 768 pixels) images, each of which has 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 5 shows the spectral transmittance of the interference filters and the spectral sensitivity of the cameras used in this experiment. Note that each camera has sensitivity higher than 400 nm and lower than 730 nm since the UV- and IR-cut filters attached to the image sensor.

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 Intel Corei7-980 3.3GHz, and the size of main memory is 12 GB.

Results

The target object used in the experiment is a 3-D Japanese hina-doll. The camera array was placed 2.5 m from the object. Figure 6 shows a result of image rectification and projective transformation. The left two images were captured without the interference filter. The right top image was captured with the filter and the bottom one is the image after projective transformation. A six-band image was generated by combining the image captured without the filter and the resultant image of projective transformation. The obtained six-band image was used for color reproduction. The result of color reproduction is shown in Fig. 7. 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 two-shot type six-band camera system [4] which uses the same digital camera and the filter, reproduction of almost same image quality especially color are achieved. A Japanese-doll on a rotating table was also captured using this system as moving picture and resultant images whose color was well reproduced were displayed on a LCD monitor in real-time.

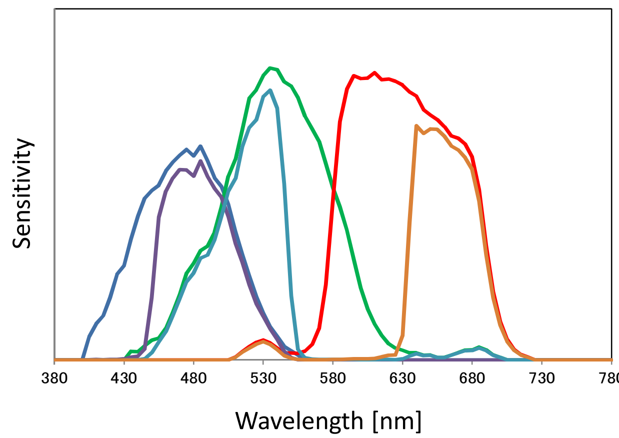

Figure 5. Spectral sensitivity of the six-band camera

Figure 6. Results of image rectification and projective transformation.
Left row: images captured without interference filter.
Right top: image captured with interference filter.

Next, total computation time is compared in the case that the size of sub-image and the sampling interval of image data for running POC and projective transformation are changed. The sub-image used for three sizes: 256 x 256, 512 x 512 and 1024 x 768

Figure 7. Output image indicating color reproduction.

pixels. The computation time is shown in Table 1. This result indicates that the sampling interval of image data is required larger than 16 pixels to achieve the frame rate of 30 fps when the image size of sub-image are 256 x 256 and 512 x 512 pixels. In addition, computation time of each processing is compared. The sampling intervals of the image data is 16 pixels. Computation time is shown in Table 2. Note that the computation time for generation of projection matrix depends on the size of sub-image and affects total computation time materially, because the number of sub-image becomes large when the size of sub-image becomes small. It is confirmed that the system can achieve the frame rate of 30 fps regardless of sub-image size.

	Image size of sub-image		
Sampling interval of image data	256 x 256	512 x 512	1024 x 768
8 pixels	58 ms	62 ms	32 ms
10 pixels	46 ms	46 ms	27 ms
16 pixels	31 ms	29 ms	20 ms
20 pixels	27 ms	25 ms	18 ms
30 pixels	24 ms	20 ms	16 ms
32 pixels	23 ms	19 ms	16 ms

Table 1. Total computation time per frame.

	Image size of sub-image		
Processing	256 x 256	512 x 512	1024 x 768
Demosaicing	3.4 ms	3.4 ms	3.4 ms
Rectification	3.0 ms	3.0 ms	3.0 ms
1-D POC	8.6 ms	11.0 ms	7.2 ms
Generation of projection matrix	9.2 ms	7.1 ms	2.7 ms
Projective transformation	4.0 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	31.0 ms	29.4 ms	20.9 ms

Table 2. Detailed computation time. Sampling interval of image data is 16 pixels

Summary

A novel six-band video acquisition and visualization system using stereo imaging has been proposed. The proposed system consists of two consumer-model digital cameras and an interference filter whose spectral transmittance is comb shaped. The all image processing steps after image capture is implemented on GPUs and the frame rate of the system is 30 fps is achieved when image size is XGA. Additional experiments achieve that this system runs at 15 fps when the image size is SXGA (1280 x 1024 pixels). This system uses projective transformation when generating a six-band image from a stereo image pair. In order to extend this system to 3-D objects which have more complex shape, a non-linear transformation method like the Thin-Plate Spline (TPS) model [13] should be implemented. Depth information obtained from detected corresponding points would also improve the quality of the generated six-band images.

One limitation of the current multi-band stereoscopic image capture system is glossy-surface objects such as car bodies. In such a case, depth information would also be effective in overcoming this problem.

References

[1] M. Yamaguchi et al., "Natural Vision: Visual Telecommunication based on Multispectral Technology.", Proc. International Display Workshop 2000,1115-1118. (2000).

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

[3] 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 (CGIV2004), 254-258. (2004)

[4] 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 (CGIV2008). (2008).

[5] 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)

[6] M. Tsuchida et al., "A stereo one-shot multi-band camera system for accurate color reproduction.", Proc. ACM Siggraph, Poster, (2010)

[7] 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)

[8] 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)

[9] M. Tsuchida, et al., "A stereo nine-band camera for accurate color and spectrum reproduction", Proc. ACM Siggraph, Poster, (2012)

[10] 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).

[11] 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)

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

[13] 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).

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.