New color correction method for multi-view video using disparity vector information

Randi Fu (符冉迪), Feng Shao (邵 枫), Gangyi Jiang (蒋刚毅), and Mei Yu (郁 梅)

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211

Received April 1, 2008

Color inconsistency is an urgent problem to be solved in free viewpoint television. In this letter, a new color correction method is proposed by using disparity vector information. At first, we separate foreground and background from the scene with a method of mean-removed disparity estimation. Then the correction parameters are estimated by adopting linear fitting for foreground and background regions, respectively. Next, with expectation-maximization algorithm, we integrate correction parameters of foreground and background to get the final corrected image. Finally, video tracking technique is performed to correct multi-view video. Experimental results show that the proposed method is quite effective.

OCIS codes: 110.0110, 100.6890, 330.1690.

doi: 10.3788/COL20080609.0654.

Free viewpoint television (FTV) is an innovative visual medium that enables us to view three-dimensional (3D) scene by freely changing our viewpoints as if we were there[1]. It will bring an epochal change in the history of visual media since such a function has not yet been achieved by conventional media technology. Though FTV is capable of providing an exciting viewing experience, it is challenging to put it into practical application. Multi-view video capturing system for FTV might not be perfectly calibrated. Although standard camera calibration[2] and geometric calibration[3] methods exist for calibrating array of cameras, much less attention has been paid to color correction for multiple cameras[4,5]. Camera parameters in multi-camera capturing system may be inconsistent, so that the exposure or focus may be variable for different views. These heterogeneous cameras can cause global or local mismatches across different views when virtual views are synthesized at the client of FTV system. In addition, it is practically impossible to capture an object under perfectly constant lighting conditions at different spatial positions within an imaging environment. Those variations provide serious challenge for realization of FTV system and degrade the performance of following multi-view video coding (MVC)[6] or virtual view synthesis[7].

In this letter, a new color correction method using disparity vector information is proposed. The block diagram of the method is given in Fig. 1. Firstly, median filtering is implemented as a preprocessing for input and reference images. Next, using disparity estimation with mean-removed sum of absolution differences (MRSAD) and the foreground and background regions are separated from the scene. Then the correction parameters are estimated by adopting linear fitting for the foreground and background regions, respectively. Thus, color correction is implemented for the input image. Finally, video tracking is used to achieve color correction for multi-view video.

Image filtering is performed as a pre-processing. Median filtering is applied to reduce the noise, that is, current pixel value is substituted by the median value within a filter window, which is described as

$$I'(i,j) = \mathrm{Med}\{I(i+r, j+s), (r,s) \in A\}, \tag{1}$$

where $\mathrm{Med}(\cdot)$ denotes median operation, A denotes filter window, and a 5×5 window is used.

In order to compensate the illumination change, a MRSAD metric[8] is defined in the proposed method. MRSAD produces block correspondence with the best matched patterns after mean removal. Therefore, in the disparity estimation, the disparity vectors for each candidate block can be well preserved even with serious cross-view illumination mismatch. MRSAD is calculated by

$$\mathrm{MRSAD}(x,y) = \sum_{i=m}^{m+S-1} \sum_{j=n}^{n+T-1} \times |(I(i,j) - \mu_{\mathrm{i}}) - (R(i+x, j+y) - \mu_{\mathrm{r}})|, \tag{2}$$

where the block is with the size of $S \times T$, $I(i,j)$ and $R(i,j)$ are pixel values of input image and reference images with spatial coordinates (i,j), respectively, (x,y) represents a candidate disparity vector, μ_{i} and μ_{r} are the average values for all pixels in the input block and the reference block, respectively.

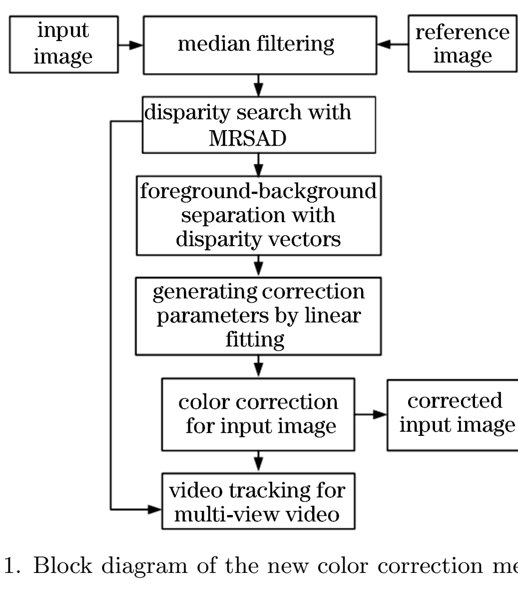

Fig. 1. Block diagram of the new color correction method.

As we know, a color image is the result of a complex reflection among three components of the optical properties of the scene, the illumination source, and the sensor[9]. It is often impossible to capture an object under perfectly constant lighting condition, and then the spectral power distribution of illumination is variable in different scene depths. In addition, under the Lambertian reflectance assumption, the illuminated region of the surface emits the entire light equally in all directions, but the Lambertian reflectance assumption is not always satisfied in object surfaces. Therefore, the multi-view imaging is inconsistent in different scene depths even for identical objects. It is necessary to compensate the color change in difference depths.

Here we propose a disparity-based foreground-background separation method. In multi-view imaging, horizontal disparity d_x and depth z satisfy the relation of $d_x = F \times L/z$, where F is the focus length of camera, L is the baseline length between two parallel cameras. Thus disparity vectors provide additional clues for scene depth estimation. In the proposed method, only two classes of scene, foreground and background, are used. By defining a threshold T_b, a block with disparities satisfying with $(d_x^2 + d_y^2)^{1/2} < T_\mathrm{b}$ belongs to the background region, otherwise the block belongs to foreground regions. After obtaining initial foreground-background separation results, some blocks in the background or foreground may be isolated because of occlusion or exposure. In order to obtain continuous background contour, we propose a smoothing mechanism adopting the correlation of adjacent blocks. When a current block belongs to foreground or background, for its adjacent left, top, right, and bottom blocks, if at least three blocks belong to a reverse classification, we consider that the current block also belongs to the reverse classification.

For one pixel (x, y) and its corresponding disparity vector $d(x, y)$, we adopt expectation-maximization algorithm[10] based on the Gaussian mixture model to refine the probability belonging to foreground or background regions. Let $\theta_k = (\mu_k, \Sigma_k)$ denote mean and variance of Gaussian distribution, and $p_k(d(x, y)|\theta_k)$ denote probability density function of disparity vector $d(x, y)$ with θ_k. Then, the probability that $d(x, y)$ belongs to the foreground ($k = 0$) or background ($k = 1$) is calculated by

$$p(k|d(x, y), \theta_k) = \frac{p_k(d(x, y)|\theta_k)}{\Sigma_{j=1}^{2} p_j(d(x, y)|\theta_k)}. \quad (3)$$

Supposed $I_i(x, y)$ is the pixel value of input view in the ith channel, $R_i(x + d_x, y + d_y)$ is the corresponding pixel value of the reference view, Ω is a set about disparity vectors, d_x and d_y are horizontal and vertical disparities for spatial coordinates (x, y), respectively. Based on finite dimensional linear model of surface spectral reflectance[11], a linear relation between different views can be drawn, and then least square metric is used to linear fitting the correction parameters by

$$(a_{ij}, e_i) = \arg\min_{a_i, e_i} \sum_{d_x, d_y \in \Omega} \mathrm{abs}(e_i + \sum_{j=1}^{3} a_{ij} I_j(x, y) - R_i(x + d_x, y + d_y)), \quad (4)$$

where a_{ij} and e_i denote multiplicative and additive correction parameters for the ith component.

Finally, we integrate the correction parameters of the foreground and background to get the final corrected image $C_i(x, y)$ as

$$C_i(x, y) = \sum_{k=0}^{1} p(k|d(x, y), \theta_k) \cdot (\sum_{j=1}^{3} a_{ij}^k I_j(x, y) + e_i^k). \quad (5)$$

Because of the high temporal correlation existing between two consecutive frames, video tracking mechanism is used to correct multi-view video. The frame difference image between two consecutive frames is firstly obtained and the edge contour can be extracted. Then the disparity vectors in the previous frame are used as initial disparity vectors of current frame. For each block in the current frame, if the edge contour is within the block, MRSAD-based disparity estimation is performed to substitute the disparity vectors, otherwise, disparity vectors in the previous frame are appointed. Finally, foreground-background separation and color correction are implemented using the same operation as that for the previous frame.

In experiments, a representative multi-view video sequence ‘flamenco1’, provided by KDDI Corp., Japan, is used as the test sequences. Figures 2(a) and (b) show the reference and input images at the 550th frame, corresponding to the first and second views, respectively.

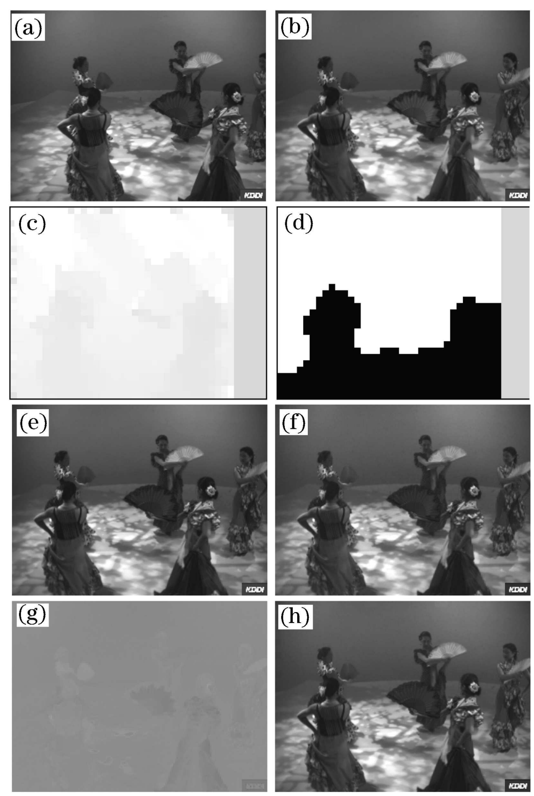

Fig. 2. Color correction for “flamenco1” multi-view images. (a) Reference image; (b) input image; (c) horizontal disparity image; (d) foreground-background separation result; (e) corrected image only with foreground correction parameters; (f) corrected image only with background correction parameters; (g) residual image of Figs. 2(e) and (f); (h) corrected image with the proposed method.

Figure 2(c) shows the horizontal disparity image from input image to reference image, where the block size is 8×8, and the maximum horizontal and vertical disparities are set as 40 and 8, respectively. Figure 2(d) shows the final foreground-background separation result for the input image, with the threshold $T_{\mathrm{b}} = 25$. Figures 2(e) and (f) show the corrected images only with foreground or background correction parameters, respectively. Figure 2(g) shows the corresponding residual image between Figs. 2(e) and (f), which is formed by adding 128 to the residual value for each color channel. The color deviation appears between foreground and background, which explains the fact that the imaging is inconsistent in foreground and background. In fact, for 'flamenco1' red channel, $a_{11} = 0.42$, $a_{12} = 0.77$, $a_{13} = -0.37$, and $e_1 = 3.39$ for the foreground, and $a_{11} = 0.63$, $a_{12} = 0.32$, $a_{13} = -0.14$, and $e_1 = -0.37$ for the background. Figure 2(h) shows the corrected image with the proposed method, which can overcome this shortcoming with full consistence with their reference images in color appearance.

For multi-view video, the video tracking mechanism is used. Figures 3(a) and (b) show the reference image and input image at the 551th frame. Figure 3(c) shows the frame difference image for input images between the 550th and 551th frames. Figure 3(d) shows the horizontal disparity image from the input image to reference image at the 551th instant, and Fig. 3(e) shows the corresponding foreground-background separation result for the input image. Figure 3(f) shows the corrected image obtained with the proposed video tracking mechanism. The subjective correction quality is quite well, meanwhile, the efficiency of the proposed video correction mechanism is high.

In order to objectively evaluate the color correction performance, we calculate the following color difference ΔE_{ab} between reference image and input image, and compare it with the color difference between the reference image and the corrected image. In this way, the color data in red, green, and blue (RGB) space is transformed into the data in CIELAB color space, and color difference ΔE_{ab} is calculated by[12]

$$\Delta E_{\mathrm{ab}} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}, \qquad (6)$$

where ΔL^*, Δa^*, and Δb^* represent the average differences of L, a, and b channels in CIELAB color space. The values of ΔE_{ab} are reduced from 6.78 to 1.07 at the 550th frame and 6.82 to 1.03 at the 551th frame, respectively, which indicates that the degree of color consistency is improved after color correction.

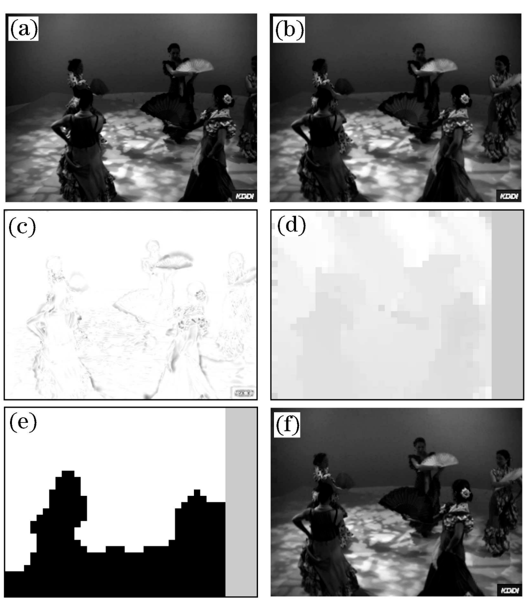

Fig. 3. Color correction for "flamenco1" multi-view video. (a) Reference image; (b) input image; (c) frame difference image; (d) horizontal disparity image; (e) foreground-background separation result; (f) corrected image with the proposed video tracking mechanism.

In conclusion, color correction is an important issue for virtual view synthesis and MVC in FTV. In this letter, a new color correction method for multi-view images using disparity vector information is proposed. Experimental results show the effectiveness of the proposed method. Here, we have assumed that the color values after correction are consistent with the reference, so that the linear fitting technique can be used. But this assumption is not always satisfied during the multi-view imaging. In future work, we will do further research on how to derive the degradation model between cameras and the high-level differential relations among different depths in complicated imaging conditions.

This work was supported by the National Natural Science Foundation of China (No. 60672073), the Program for New Century Excellent Talents in University (NCET-06-0537), the Scientific Research Fund of Zhejiang Provincial Education Department (No. 20070962 and 20071067), and the Natural Science Foundation of Ningbo (No. 2007A610037 and 2008A610016). M. Yu is the author to whom the correspondence should be addressed, her e-mail address is yumei@nbu.edu.cn.

References

1. M. Tanimoto, Signal Processing: Image Commun. **21,** 454 (2006).
2. C. Sun, X. Zhang, and Y. Qu, Chin. Opt. Lett. **3,** 585 (2005).
3. Z. Chen and P. Shi, Chin. Opt. Lett. **3,** 69 (2005).
4. F. Shao, M. Yu, and G. Jiang, Chin. Opt. Lett. **5,** 449 (2007).
5. K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, IEEE Trans. Circ. Syst. Video Technol. **17,** 1436 (2007).
6. A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, IEEE Trans. Circ. Syst. Video Technol. **17,** 1606 (2007).
7. S. Yea and A. Vetro, in *Proceedings of IEEE International Conference on Image Processing 2007* **1,** 209 (2007).
8. J. H. Kim, P. Lai, J. Lopez, A. Ortega, Y. Su, P. Yin, and C. Gomila, IEEE Trans. Circ. Syst. Video Technol. **17,** 1519 (2007).
9. G. Unal, A. Yezzi, S. Soatto, and G. Slabaugh, IEEE Trans. Pattern Anal. Machine Intell. **29,** 1322 (2007).
10. P. Guo and H. Lu, Acta Opt. Sin. (in Chinese) **22,** 1479 (2002).
11. L. T. Maloney, J. Opt. Soc. Am. A **3,** 1673 (1986).
12. Z. Wang and H. Xu, Acta Opt. Sin. (in Chinese) **27,** 1344 (2007).