Color correction and geometric calibration for multi-view images with feature correspondence*

SHAO Feng(邵枫)[1**], **JIANG Gang-yi**(蒋刚毅)[1,2], **and YU Mei**(郁梅)[1,2]
1. Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
2. the State Key Lab. of New Software Technology, Nanjing University, Nanjing 210093, China

(Received 12 January 2009)

Color and geometry inconsistency between different views is an urgent problem in multi-view imaging applications. In this paper, we present a color correction and geometric calibration method for multi-view images on the basis of feature correspondences between views. First, keypoints in views are detected by using scale invariant feature transform, and accurately matched by bi-directional feature matching between difference views. Then multiplicative and additive errors between matching keypoints are calculated to achieve color correction. In addition, an affine transformation between minimum cost matching keypoints is established to achieve geometric calibration. The experimental results verify the effectiveness of the proposed method in color correction and geometric calibration, and a higher coding efficiency is obtained.

Document code: A **Article ID:** 1673-1905(2009)03-0232-4
DOI 10.1007/s11801-009-8002-5

Free viewpoint television (FTV) is one of the most important applications of multi-view imaging and is a new type of media that can provide users stereoscopic impression and interactive experiences. Generally, to provide a smooth multi-perspective viewing experience, content producers need to capture the same scene with ideal quality from multiple viewpoints[1].

A common problem in FTV is that different camera sensors acquire different color response to an image object[2]. This problem occurs because physical factors during the imaging process introduce a variation that differs for each camera. In addition, it is practically impossible to capture an object under perfectly constant lighting conditions at different spatial positions. Simultaneously, the camera positions are variable. Those variations degrade the performance of coding efficiency, and furthermore, reduce the quality of virtual view rendering.

Many correction methods were proposed to deal with the color inconsistency, including example-based color transformation method using basic color categories[3], luminance and chrominance compensation using histogram matching[4], and content-adaptive color correction by principal component analysis (PCA)[5]. Geometric calibration had already been widely studied in computer vision, which uses camera intrinsic and extrinsic parameters to implement coordinate transformation[6].While the above color correction methods are sensitive to image scaling and rotation, the probability of disruption may be increased by occlusion, clutter, or noise. In addition, color correction and geometric calibration have not been studied together in the literature. If stable feature information can be extracted from images, the precisions of color correction and geometric calibration can be largely improved.

Recently, great progress has been made in object recognition by using local descriptors such as scale invariant feature transform (SIFT) keypoints[7]. SIFT keypoints are invariant to scale, rotation, viewpoint and illumination change, and have shown well matching ability to object motion, occlusion and noise, which consist of four stages: (1) scale-space extrema selection; (2) keypoint localization; (3) orientation assignment; (4) keypoint descriptor.

We choose one reference image, and correct other images with similar color of the reference image. Given all keypoints in the reference and input images after SIFT detection, we present a simple matching scheme based on the saliency of the keypoints. For a keypoint $P_{\mathrm{inp}}(x, y)$ in the input image, the nearest keypoint $P_{\mathrm{ref}}(x', y')$ and the second nearest keypoint $P_{\mathrm{ref}}(x'', y'')$ in the reference image can be determined by comparing the Euclidean distances in $N\times N$ window. The Euclidean distance is calculated in luminance

* This work has been supported by the National Natural Science Foundation of China (No.60672073, 60872094), the Program for New Century Excellent Talents in University (NCET-06-0537), and the Natural Science Foundation of Ningbo Province (No.2007A610037, 2008A610016).

** E-mail: jianggangyi@nbu.edu.cn

components after mean removal. Two keypoints $P_{\mathrm{inp}}(x, y)$ and $P_{\mathrm{ref}}(x', y')$ are considered to be matched if the smallest and the second smallest Euclidean distances satisfy the following relation:

$$\frac{\sum_N E(L_{\mathrm{inp}}(x,y)-\mu_{\mathrm{inp}}, L_{\mathrm{ref}}(x',y')-\mu'_{\mathrm{ref}})^2}{\sum_N E(L_{\mathrm{inp}}(x,y)-\mu_{\mathrm{inp}}, L_{\mathrm{ref}}(x'',y'')-\mu''_{\mathrm{ref}})^2} < \tau^2 \quad , \tag{1}$$

where E(A, B) denotes Euclidean distance between A and B, N denotes window size centered at keypoint, $L_{\mathrm{inp}}(x, y)$ and $L_{\mathrm{ref}}(x', y')$ are luminance values of keypoints $P_{\mathrm{inp}}(x, y)$ and $P_{\mathrm{ref}}(x', y')$, respectively, and μ_{inp} and ì$_{\mathrm{ref}}$ are corresponding luminance means in the $N\times N$ window, ô is an adjustable parameter, $0 \leqslant \tau < 1$, which controls the matching precision and the number of matched keypoints. In the experiments, we select N=7 and threshold τ=0.8 depending on experience.

Then after establishing the unidirectional matching from keypoint $P_{\mathrm{inp}}(x, y)$ to keypoint $P_{\mathrm{ref}}(x', y')$, the disparity in the pixel (x, y) from input image to its reference image is expressed as $d_{\mathrm{inp}\to\mathrm{ref}} = (x'-x, y'-y)$. For the keypoint $P_{\mathrm{ref}}(x', y')$ in the reference image, a reverse matching scheme is performed to find a matching keypoint $P_{\mathrm{inp}}(x''', y''')$ in the input image. Thus the disparity in pixel (x', y') from the reference image to the input image is expressed as $d_{\mathrm{ref}\to\mathrm{inp}}=(x'''-x', y'''-y')$. Thereupon, $P_{\mathrm{inp}}(x, y)$ and $P_{\mathrm{ref}}(x', y')$ are regarded as a pair of matching keypoints if the disparity deviation between $d_{\mathrm{inp}\to\mathrm{ref}}$ and $d_{\mathrm{ref}\to\mathrm{inp}}$ is less than 2, that is, $|d_{\mathrm{inp}\to\mathrm{ref}}+d_{\mathrm{ref}\to\mathrm{inp}}|<2$.

It is assumed that color errors in images are mainly categorized into two classes based on the theory of image formation in machine vision - multiplicative and additive errors [8]. Then, considering color variations between different viewpoint images as the interference of multiplicative and additive errors, color correction for input image can described as

$$I_i^{\mathrm{cor}}(x,y) = a_i \cdot I_i^{\mathrm{inp}}(x,y) + e_i \quad , \tag{2}$$

where I_i^{inp} is the i-th color value of input image, I_i^{cor} is the corresponding i-th color value of corrected image, and i represents one of the three R, G and B components. The multiplicative error a_i and additive error e_i, which minimizes the residuals in the least-squares sense between I_i^{ref} and I_i^{cor}, is represented as

$$(a_i, e_i) = \arg\min_{a_i, e_i} \sum_{(x,y),(x',y')\in\Omega} (I_i^{\mathrm{ref}}, (x',y') - I_i^{\mathrm{cor}}(x,y))^2 \quad , \tag{3}$$

where Ω is a set of all the matching keypoints, and argmin(·) denotes the argument for which the function is minimized.

Supposed rigid motion between the cameras, an affine transformation is defined between two pairs of matching keypoints $P_{\mathrm{inp}}(x, y)$ and $P_{\mathrm{ref}}(x', y')$

$$\boldsymbol{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix} = \boldsymbol{Ax} + \boldsymbol{B} \quad , \tag{4}$$

where matrix $\boldsymbol{A}$ is related to the rotation and scaling, and $\boldsymbol{B}$ is related to the translation. In order to obtain six unknown parameters in the equation, at least three pairs of matching keypoints are required. In the experiments, we select three pairs of matching keypoints with the minimum Euclidean distances.

With the above analyses, the proposed color correction and geometric calibration method for multi-view images is briefly summarized as follows:

(1) Detect keypoints by using SIFT and accurately match by bi-directional feature matching between difference views.

(2) Calculate multiplicative and additive errors between matching keypoints by Eq.(2) and Eq.(3).

(3) Establish an affine transformation between minimum cost matching keypoints by Eq.(4).

In order to objectively evaluate the performance of the proposed color correction method, the color differences between matching keypoints in the reference and color corrected images are calculated. Two color difference definitions are used. The CIELAB color space is intended to be a perceptually uniform color space, so that equal distance in the color space represents equal perceived differences in appearance. The color difference in CIELAB space is defined as

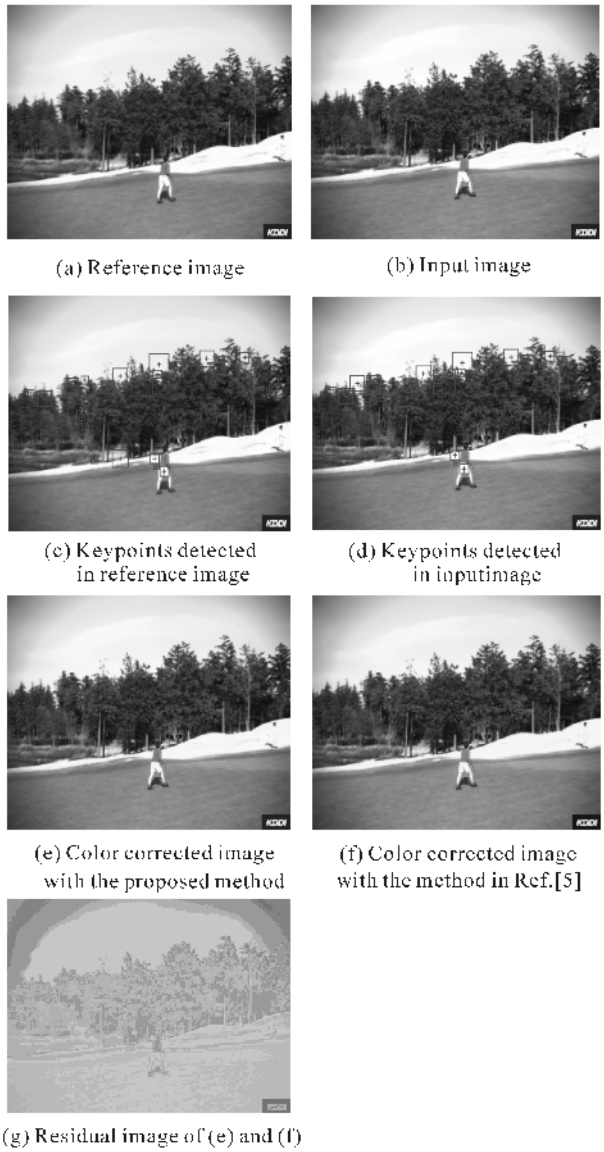
$$\Delta E_{ab}^{*} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad , \tag{5}$$

where ΔL^*, Δa^* and Δb^* are the differences in lightness L, chroma a and chroma b, respectively. The color differences are also calculated with a quite complicated CIEDE2000 formula[9], because CIELAB space is not uniform enough. The CIEDE2000 color difference ΔE_{00} formula is defined as

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)} \quad , \tag{6}$$

where $\Delta L'$, $\Delta C'$ and $\Delta H'$ are the differences in lightness, chroma and hue for a pair of samples. k_L, k_C, k_H, S_L, S_C and S_H are weighting factors, and R_T is a function that accounts for the interaction between chroma and hue differences in the blue region.

We select a representative multi-view video sequence, 'golf 2' (320×240, 4:2:0 YUV format, and 8 viewpoints), for evaluating the performance of the proposed method. Fig. 1(a) and (b) show the reference and input images, respectively. Fig.1(c) and (d) show the final matching keypoints for reference and input images, in which the size of square block denotes SIFT scale. Fig.1(e) shows the color corrected image with the proposed method. Fig.1(f) shows the color corrected image with the content-adaptive method by PCA in Ref.[5]. In PCA, a 3×3 matrix as functions of translation, rotation and scaling is used to achieve color correction. Fig.1(g) shows the corresponding residual image between Fig.1(e) and (f), in which only a slight color deviation appears in forest and glass regions. It is indicated that the correction performances are similar with the proposed method and PCA method, however, PCA is sensitive to image scaling, rotation and other factors, and the matching keypoints information is more beneficial to geometric calibration.



(a) Reference image

(b) Input image

(c) Keypoints detected in reference image

(d) Keypoints detected in inputimage

(e) Color corrected image with the proposed method

(f) Color corrected image with the method in Ref.[5]

(g) Residual image of (e) and (f)

Fig.1 Color correction results for 'golf2'

Fig.2 shows the color difference comparison results for matching keypoints between reference image and input image, and the same matching keypoints between reference image and corrected image. From the figure, it is noted that color correction can achieve smaller color difference regardless of different metrics used, which is consistent with the subjective color appearance evaluation in Fig.1.

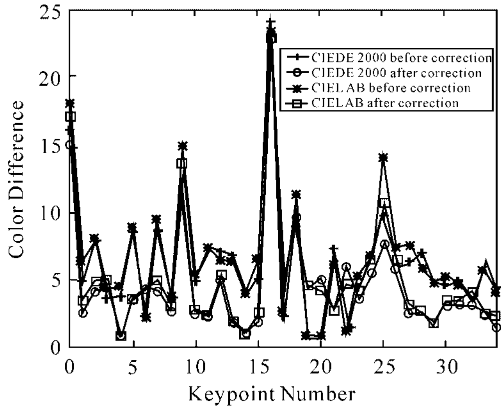


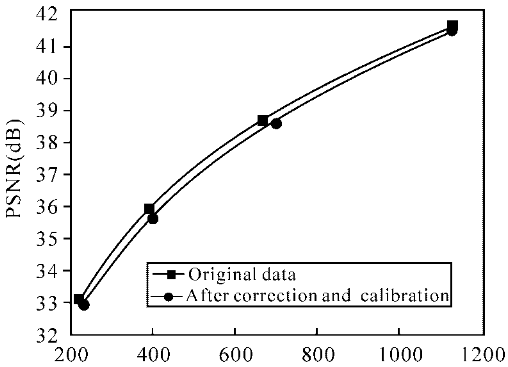
Fig.2 Color difference comparisons for 'golf2'

For 'golf2', the Epipolar line between two cameras is not absolutely orthogonal to optical axis, and geometric calibration is necessary. Fig.3(a) shows the original first and second viewpoint images, where the view direction among the three best matching keypoints is not horizontally calibrated. Fig.3(b) shows the original first viewpoint image and geometrically calibrated second viewpoint image. It is obvious that the Epipolar line between the two images is completely horizontal after geometric calibration.

Fig.3(a) Original viewpoint images for 'golf2'

Fig.3(b) Geometric calibration results for 'golf2'

Although the proposed method primarily aims at improving the quality of virtual view rendering, they can be easily used to improve the coding efficiency in multi-view video coding. JSVM version 6.5 is modified that only inter-view prediction is used, in order to focus on the efficiency for inter-view prediction. We encode 100 frames multi-view video in four different basicQP parameters (24, 28, 32, 36). Fig.4 demonstrates the rate-distortion performance with IPPP structure in inter-view prediction, in which the horizontal and vertical axes represent bitrate and PSNR, respectively. For luminance component *Y*, about 0-0.1 dB PSNR gain is decreased under the same bit rate with the proposed method, while 0.6-0.7 dB and 0.4-0.5 dB PSNR gains for *U* and *V* components can be achieved with the proposed method, respectively.



(a) Y component

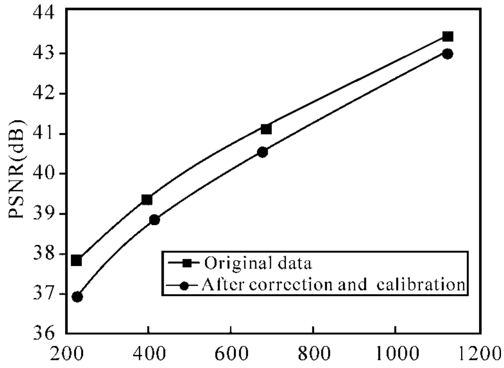
In conclusion, a color correction and geometric calibration method with feature correspondence is proposed. Experimental results show the effectiveness of the proposed method in color correction and multi-view video coding. In future work, we will research on how to achieve effective view rendering.

References

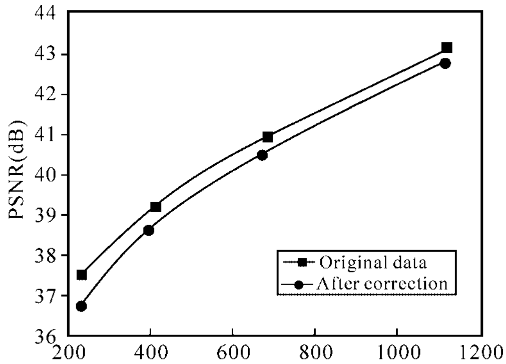
[1] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, IEEE Trans. on Circuits and Systems for Video Technology, **17** (2007), 1461.

[2] Ai-hong Zhang, Ting-ting Wang, and Zhi-gang Fan, Opt. Lett., **3** (2007), 0315.

[3] Y. Chang, S. Saito, and M. Nakajima, IEEE Trans. on Image



(b) U component



(c) V component

Fig.4 Rate-distortion performance comparison.

Processing, **16** (2007), 329.

[4] U. Fecker, M. Barkowsky, and A. Kaup, IEEE Trans. on Circuits and Systems for Video Technology, **18** (2008), 1258.

[5] Feng Shao, Gang-yi Jiang, and Mei Yu, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, (2007), 969.

[6] Ying Wang, Guang-jun Zhang, and Jian-lin Wang, Journal of Optoelectronics. Laser, **19** (2008), 754. (in Chinese)

[7] D. G. Lowe, International Journal of Computer Vision, **60** (2004), 91.

[8] Y. C. Chang, J. K. Reid, IEEE Trans. on Image Processing, **5** (1996), 1414.

[9] M. R. Luo, G. Cui, B. Rigg, Color Research and Application, **26** (2001), 340.