Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Color compensation for multi-view video coding based on diversity of cameras*

Jun-yan HUO†1, Yi-lin CHANG1, Hai-tao YANG1, Shuai WAN2
(1*State Key Lab of Integrated Service Networks, Xidian University, Xi'an 710071, China*)
(2*Shaanxi Key Lab for Information Acquisition and Processing, Northwestern Polytechnical University, Xi'an 710072, China*)
†E-mail: jyhuo@mail.xidian.edu.cn

Received Jan. 24, 2008; revision accepted Apr. 24, 2008; CrossCheck deposited Nov. 10, 2008

Abstract: A novel color compensation method for multi-view video coding (MVC) is proposed, which efficiently exploits the inter-view dependencies between views with the existence of color mismatch caused by the diversity of cameras. A color compensation model is developed in RGB channels and then extended to YCbCr channels for practical use. A modified inter-view reference picture is constructed based on the color compensation model, which is more similar to the coding picture than the original inter-view reference picture. Moreover, the color compensation factors can be derived in both encoder and decoder, therefore no additional data need to be transmitted to the decoder. The experimental results show that the proposed method improves the coding efficiency of MVC and maintains good subjective quality.

Key words: Multi-view video coding (MVC), H.264/AVC, Color compensation, Diversity of cameras
doi:10.1631/jzus.A0820075 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

Multi-view video coding (MVC) (Smolic *et al*., 2006; MPEG Video Subgroup, 2008) is a key technology that serves a wide variety of applications, including free viewpoint television, 3D television and surveillance. A new standard for MVC is under development by the Joint Video Team (JVT), which will be an extension of H.264/AVC (MPEG, 2006). Both temporal prediction and inter-view prediction are used in MVC to improve the coding efficiency. However, in some cases, there is color mismatch between views, which impairs the performance of the inter-view prediction in MVC. A few schemes have been studied to explore the inter-view dependency efficiently. In (Chen *et al*., 2006; Fecker *et al*., 2006), the histogram matching method is utilized to create the lookup tables or find the parameter factors of the linear model for color correction. However, the histogram matching method has the disadvantage of unreliability, which is especially worse when there are large parts of different occlusions between views. The preprocessed pictures can be obtained at the decoder, while these methods cannot always provide preprocessed pictures with good perceptive quality compared to the original ones. There are some other methods dedicated to improving the coding efficiency. Reference (Lee *et al*., 2006), referred to as the illumination compensation (IC) method, cancels out the difference in luminance between different views and an additive error model is employed for IC. Then, Su *et al*.(2006a) proposed both IC and color compensation (CC), in which offsets are added for Y, Cb, and Cr channels of the compensated block. A color correction method was proposed by Yamamoto *et al*.(2007), where a non-linear compensation model is used to correct RGB channels respectively. Two conversions for each pixel, YUV to RGB and RGB to YUV, are needed in (Yamamoto *et al*., 2007), which introduces large

---

* Project supported by the National Natural Science Foundation of China (No. 60772134) and the Innovation Foundation of Xidian University, China (No. Chuang 05018)

computational complexities. We proposed a CC method in (Huo *et al*., 2007) which is combined with the weighted prediction in H.264/AVC. All of these methods are implemented by modifying the reference pictures in the encoder and additional information is needed to be sent to the decoder.

Actually, the color mismatch between views is mainly caused by the diversity of cameras. In this paper, a novel CC method is proposed, which is developed in RGB channels and extended to YCbCr channels for practical use. This is due to the fact that the original mismatch between views is in RGB channels while the existing video coding standards, such as H.264/AVC, support the color format of YCbCr. Moreover, since the color mismatch between views changes very slightly over time, the CC factors can be derived from pictures which have been decoded successfully. That is to say, no additional information is needed to be transmitted in our proposed method.

The rest of this paper is organized as follows. Section 2 describes the CC model after a brief discussion of the camera structure. The details of the proposed method are given in Section 3. Section 4 shows some experimental results. Finally, Section 5 concludes this paper.

COLOR COMPENSATION MODEL

To record a color image, a camera must have three sensors to capture each color component separately. In practice, the most popular primary set is RGB channels. Therefore, the color mismatch effect between views should be investigated in the RGB channels. Without loss of generality, the case of two cameras is taken as an example to explore the situation of the color mismatch. As depicted in Fig.1, the camera whose view to be encoded is called the coding camera, and its view is called the coding view correspondingly. The other camera is named as the reference camera whose view is the reference view. Let f_{cod} represent the picture of the coding view, and f_{ref} the picture of the reference view which is used as the inter-view reference picture of f_{cod}. It is also supposed that a point o in the scene projects to the pixel p in f_{ref} and the pixel q in f_{cod}. Here p and q are called the corresponding pixels.

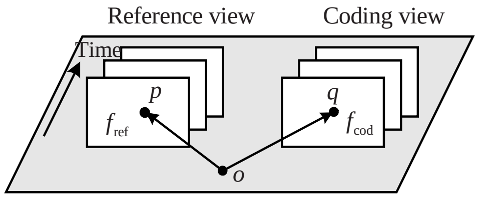


Fig.1 Illustration of the coding view and reference view

The main purpose of the CC model is to construct a modified inter-view reference picture based on f_{ref}, which should be more similar to f_{cod} than to f_{ref}. The modified reference picture is denoted as f_{ref}' and the modified pixel of p is denoted as p'. The RGB values of p are denoted as R_p, G_p, and B_p, while its YCbCr values Y_p, Cb_p, and Cr_p, respectively.

For simplicity, the input light signals of the two cameras emitted from point o are assumed equal. Then the digital signal outputs of p and q can be written approximately as follows:

$$R_p = F_{\text{R}}(k_{\text{r,ref}}r),\ \ G_p = F_{\text{G}}(k_{\text{g,ref}}g),\ B_p = F_{\text{B}}(k_{\text{b,ref}}b),\ \text{(1a)}$$

$$R_q = F_{\text{R}}(k_{\text{r,cod}}r),\ G_q = F_{\text{G}}(k_{\text{g,cod}}g),\ B_q = F_{\text{B}}(k_{\text{b,cod}}b),\ \text{(1b)}$$

where r, g, and b are the analog signals of o; $k_{\text{r,cod}}$, $k_{\text{g,cod}}$, $k_{\text{b,cod}}$ and $k_{\text{r,ref}}$, $k_{\text{g,ref}}$, $k_{\text{b,ref}}$ are the analog gains of RGB channels of the coding camera and the reference camera, respectively; $F_{\text{R}}(\cdot)$, $F_{\text{G}}(\cdot)$, and $F_{\text{B}}(\cdot)$ are the functions which convert analog signals to digital signals.

Then the RGB values of p' in f_{ref}' can be calculated by

$$R_{p'} = K_{\text{R}} R_p,\ \ G_{p'} = K_{\text{G}} G_p,\ \ B_{p'} = K_{\text{B}} B_p, \qquad (2)$$

where K_{R}, K_{G}, and K_{B} are defined as the CC factors for each color channel. They can be obtained by

$$K_{\text{R}} = R_q / R_p,\ \ K_{\text{G}} = G_q / G_p,\ \ K_{\text{B}} = B_q / B_p. \qquad (3)$$

Eq.(2) can also be represented in the form of matrix as

$$\begin{bmatrix} R_{p'} \\ G_{p'} \\ B_{p'} \end{bmatrix} = \begin{bmatrix} K_{\text{R}} & 0 & 0 \\ 0 & K_{\text{G}} & 0 \\ 0 & 0 & K_{\text{B}} \end{bmatrix} \begin{bmatrix} R_p \\ G_p \\ B_p \end{bmatrix}. \qquad (4)$$

In BT.601 (ITU-R, 1998), the *YCbCr* coordinate is related to the *RGB* by

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \frac{1}{256}\begin{bmatrix} 77 & 150 & 29 \\ 131 & -110 & -21 \\ -44 & -87 & 131 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix}, \quad (5)$$

and the inverse conversion matrix from *YCbCr* to *RGB* is

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{256}\begin{bmatrix} 256 & 351 & 0 \\ 256 & -179 & -86 \\ 256 & 0 & 444 \end{bmatrix}\begin{bmatrix} Y \\ Cr-128 \\ Cb-128 \end{bmatrix}, \quad (6)$$

so the *YCbCr* values of pixel p' in f_{ref}' can be derived as follows:

$$\begin{aligned}\begin{bmatrix} Y_{p'} \\ Cr_{p'} \\ Cb_{p'} \end{bmatrix} &= \frac{1}{256}\begin{bmatrix} 77 & 150 & 29 \\ 131 & -110 & -21 \\ -44 & -87 & 131 \end{bmatrix}\begin{bmatrix} R_{p'} \\ G_{p'} \\ B_{p'} \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \\ &= \frac{1}{256}\begin{bmatrix} 77 & 150 & 29 \\ 131 & -110 & -21 \\ -44 & -87 & 131 \end{bmatrix}\begin{bmatrix} K_{\mathrm{R}} & 0 & 0 \\ 0 & K_{\mathrm{G}} & 0 \\ 0 & 0 & K_{\mathrm{B}} \end{bmatrix}\begin{bmatrix} R_p \\ G_p \\ B_p \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \\ &= \frac{1}{256^2}\begin{bmatrix} 77 & 150 & 29 \\ 131 & -110 & -21 \\ -44 & -87 & 131 \end{bmatrix}\begin{bmatrix} K_{\mathrm{R}} & 0 & 0 \\ 0 & K_{\mathrm{G}} & 0 \\ 0 & 0 & K_{\mathrm{B}} \end{bmatrix} \\ &\quad \cdot\begin{bmatrix} 256 & 351 & 0 \\ 256 & -179 & -86 \\ 256 & 0 & 444 \end{bmatrix}\begin{bmatrix} Y_p \\ Cr_p-128 \\ Cb_p-128 \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix}. \quad (7)\end{aligned}$$

For clarity, the CC model in YCbCr channels Eq.(7) can be rewritten as

$$\begin{bmatrix} Y_{p'} \\ Cr_{p'} \\ Cb_{p'} \end{bmatrix} = \boldsymbol{\Phi}\begin{bmatrix} Y_p \\ Cr_p-128 \\ Cb_p-128 \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix}, \quad (8)$$

where $\boldsymbol{\Phi}$ is the CC matrix derived from K_{R}, K_{G}, and K_{B}.

NOVEL COLOR COMPENSATION METHOD

Fig.2 gives a reference prediction structure of MVC (Merkle *et al*., 2007). This scheme employs the prediction structure of hierarchical B pictures in temporal dimension. Additionally, inter-view prediction is applied to explore the inter-view correlation. For the clarity of presentation, some definitions in MVC are described in the following. As Fig.2 shows, the base view is the view in which only temporal prediction is performed. Meanwhile, the views in which inter-view prediction is used are named as non-base views. For synchronization, anchor pictures (pictures with gray background in Fig.2) are introduced, in which only inter-view prediction is allowed. Non-anchor pictures may use both temporal reference picture and inter-view reference picture.

According to Eq.(3), the CC factors can be easily obtained if the corresponding pixels are available. However, it is not feasible to search the corresponding pixels due to high computational complexity. In this study, a method with low complexity is proposed instead. As known, motion estimation is an essential module in the encoder to select the best matching blocks for each macroblock. So the pixels of the macroblocks and their best matching blocks are considered as the corresponding pixels if the blocks of the inter-view reference pictures are selected as the best matching blocks.

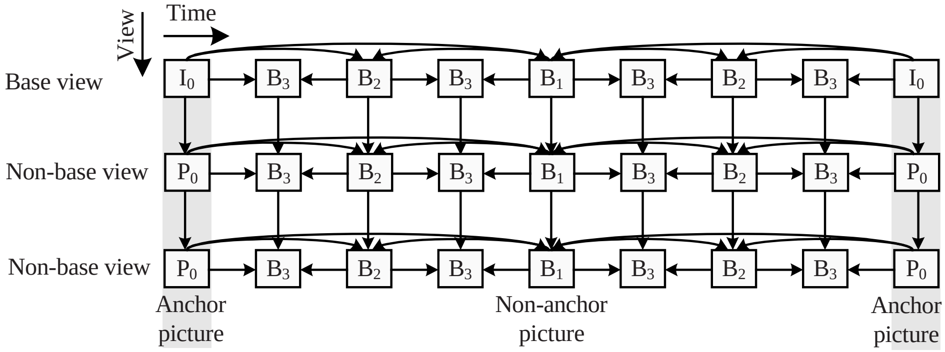


Fig.2 Example of prediction structure for multi-view video coding

Fig.3 gives a flowchart of MVC where the proposed CC method is depicted by the shaded modules. The proposed method is a loop with GOP (group of pictures). For a GOP, the CC factors are calculated in the anchor picture and implemented in the non-anchor picture. The proposed method can be accomplished as follows:

Step 1: At the beginning of encoding a view, K_R, K_G, and K_B are initialized to 1.

Step 2: For the anchor picture of each GOP, the inter-view reference picture is modified according to Eq.(8) with the CC factors of the previous GOP. For the first anchor picture of the view, the initial values of CC factors are used.

Step 3: The modified inter-view reference is used to encode the anchor picture.

Step 4: Those macroblocks in the anchor picture, which adopt the inter 16×16, inter 16×8 and inter 8×16 as the final prediction modes and their best matching blocks, are selected for CC factor calculation. The CC factors are obtained by

$$K_R = \frac{1}{N}\sum_{i=1}^{N}\frac{R_{\text{cod}}}{R_{\text{ref}}},\ K_G = \frac{1}{N}\sum_{i=1}^{N}\frac{G_{\text{cod}}}{G_{\text{ref}}},\ K_B = \frac{1}{N}\sum_{i=1}^{N}\frac{B_{\text{cod}}}{B_{\text{ref}}}, \tag{9}$$

where N is the total number of pixels in the macroblocks which satisfy the above-mentioned condition; R_{cod}, G_{cod}, B_{cod} and R_{ref}, G_{ref}, B_{ref} are the RGB values of the corresponding pixels in the reconstructed macroblocks and the reference blocks, respectively.

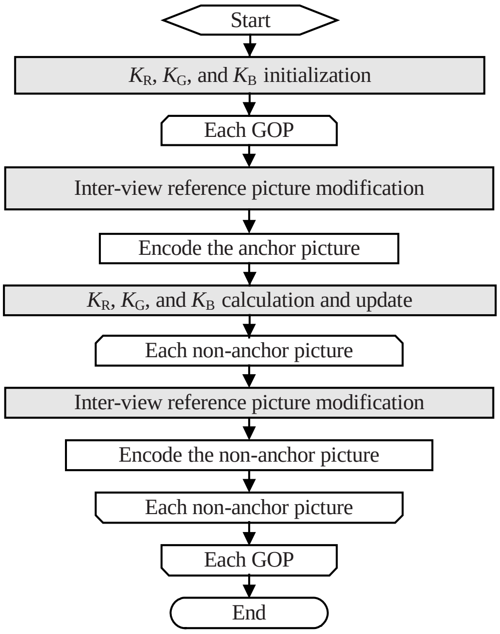


Fig.3 Flowchart of MVC with the proposed method

Step 5: The inter-view reference pictures of the following non-anchor pictures of the current GOP are modified using K_R, K_G, and K_B obtained in Step 4. Then both the modified inter-view reference and the temporal reference are used to encode the picture.

Step 6: Repeat Steps 2~5 until the end of the multi-view sequence.

The calculation of the CC factors is based on the reconstructed macroblocks and their best matching blocks, which can also be obtained at the decoder. Therefore there is no additional information to be sent to the decoder in the proposed method.

EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method was implemented based on the MVC reference software JMVM 4.0 (Pandit *et al.*, 2007). Four test sequences were used whose properties are listed in Table 1, and the prediction structures are given in Fig.4. The common test conditions for MVC (Su *et al.*, 2006b) developed by JVT was employed in our experiments to evaluate the performance of the proposed method. Specifically, four fixed QPs (22, 27, 32, 37) as defined in (Su *et al.*, 2006b) were used to get four rate-distortion data points, and the Bjontegaard measure (Bjontegaard, 2001) was used to calculate the average PSNR/bitrate differences between R-D curves of the proposed method and JMVM. For 'Ballroom' and 'Rena', the cameras were rectified before capturing, so there was no color mismatch between views. However, for 'Race1' and 'Flamenco2', color mismatch existed and was visible to the eyes.

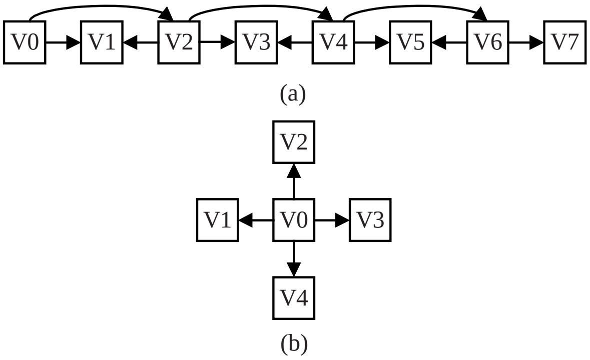


Fig.4 Prediction structures of test sequences. (a) Ballroom, Race1, Rena; (b) Flamenco2

Table 1 Properties of the test sequences

Test sequence	Number of views	Camera arrangement	Number of pictures*	Rectified
Ballroom	8	1D parallel	121	Yes
Rena	8	1D parallel	151	Yes
Race1	8	1D parallel	151	No
Flamenco2	5	2D parallel	151	No

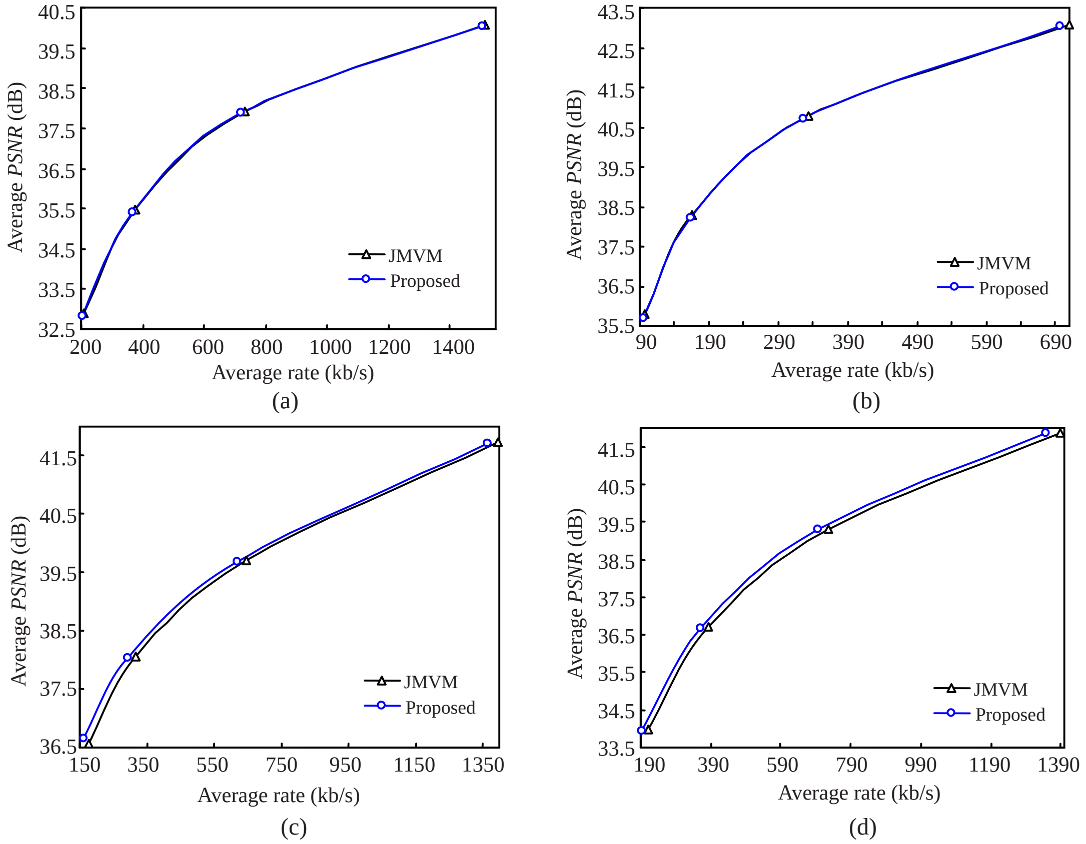
* All from 10 GOPs

In our experiments, both subjective and objective measures were exploited to verify the performance of the proposed algorithm. The objective measurement, $PSNR_{\text{YCbCr}}$, is introduced to evaluate the performance of Y, Cb, and Cr components together, given as

$$PSNR_{\text{YCbCr}} = 10\lg\left(\frac{6}{\dfrac{4}{10^{PSNR_{\text{Y}}/10}}+\dfrac{1}{10^{PSNR_{\text{Cb}}/10}}+\dfrac{1}{10^{PSNR_{\text{Cr}}/10}}}\right), \quad (10)$$

where $PSNR_{\text{Y}}$, $PSNR_{\text{Cb}}$, and $PSNR_{\text{Cr}}$ are the $PSNR$ values of Y, Cb, and Cr components, respectively. Fig.5 gives the rate-distortion curves of the proposed method and the JMVM. It can be seen that the proposed method has a similar performance to JMVM for the rectified test sequences and that a better coding efficiency can be achieved for the 'Race1' and 'Flamenco2'. Our complete results on $\Delta PSNR$ and $\Delta bitrate$ are shown in Table 2. The bitrate reductions for 'Race1' and 'Flamenco2' are 3.69% and 4.97%, respectively, while 'Ballroom' and 'Rena' have smaller benefits (0.67% and 0.33%, respectively) because the CC factors are always around 1. It is observed that the proposed CC method works well for the sequences with the existence of color mismatch.

For the comparison of subjective quality, the modified reference pictures and the original reference pictures of 'Race1' and 'Flamenco2' are shown in Fig.6. Figs.6a and 6d are the coding pictures of the two sequences and Figs.6b and 6e are the original inter-view reference pictures. It is obvious that color



(a) (b)

(c) (d)

Fig.5 Rate-distortion curves for test sequences. (a) Ballroom; (b) Rena; (c) Race1; (d) Flamenco2

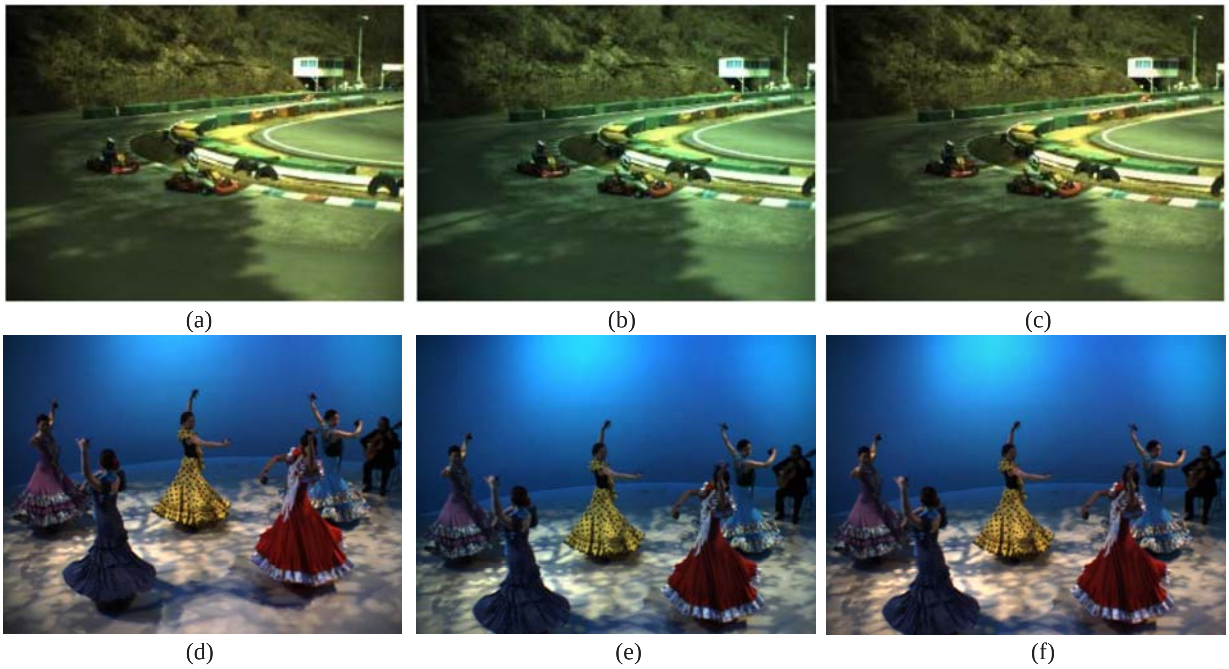


Fig.6 Illustration of the subjective quality of the proposed method. (a) Race1, coding picture; (b) Race1, original inter-view reference picture; (c) Race1, modified inter-view reference picture; (d) Flamenco2, coding picture; (e) Flamenco2, original inter-view reference picture; (f) Flamenco2, modified inter-view reference picture

Table 2 Performance evaluation of the proposed method

Sequence	$\Delta PSNR$ (dB)	$\Delta bitrate$ (%)
Ballroom	0.024	−0.67
Rena	0.012	−0.33
Race1	0.137	−3.69
Flamenco2	0.223	−4.97

mismatch exists between the coding pictures and the reference pictures. Figs.6c and 6f are the modified inter-view reference pictures using the proposed method. The modified reference pictures are more similar to the coding pictures than the original ones. In this way, the inter-view correlation can be utilized more efficiently and the coding efficiency of MVC can be improved.

Finally, the complexity analysis of the proposed method is discussed. Compared with the method in which CC factors are transmitted explicitly, the decoder using the proposed method needs to derive CC factors in the anchor picture of each GOP. The derivation of CC factors can be accomplished in two steps—the extraction of the correspondences and the calculation of CC factors. The complexity of the former step can be negligible because the macroblock and its best matching blocks, which are considered as the correspondences, can be directly extracted from the reconstructed picture and its reference picture. In the latter step, the calculation of CC factors, as shown in Eq.(9), requires $3(N+1)$ division operations and $3(N-1)$ addition operations for each GOP. Therefore, the decoder complexity increased by using the proposed method is acceptable. Moreover, the encoding complexity is identical to the decoding complexity because no additional data need to be transmitted and the same CC process is performed in both the encoder and the decoder.

CONCLUSION

In this paper, a novel CC model based on the diversity of cameras is proposed. Based on a thorough analysis, the CC model is built in RGB channels and further extended to the YCbCr channels, which makes the proposed method compatible with current video coding standards. The parameters of the CC model can be obtained according to the macroblocks and their best matching blocks. The experimental results show that the modified inter-view reference pictures are more similar to the coding pictures than the original inter-view reference pictures; therefore the coding efficiency for MVC is improved by the proposed method.

References

Bjontegaard, G., 2001. Calculation of Average PSNR Differences Between RD-curves. VCEG-M33, http://ftp3.itu.ch/av-arch/video-site/0104_Aus/

Chen, Y., Chen, J., Cai, C., 2006. Luminance and Chrominance Correction for Multi-view Video Using Simplified Color Error Model. Proc. Picture Coding Symp., Beijing.

Fecker, U., Barkowsky, M., Kaup, A., 2006. Improving the Prediction Efficiency for Multi-view Video Coding Using Histogram Matching. Proc. Picture Coding Symp., Beijing.

Huo, J., Yang, H., Chang, Y., Lin, S., Gao, S., Xiong, L., 2007. Weighted Prediction for MVC Using Color Compensation. JVT-X055, http://ftp3.itu.ch/av-arch/jvt-site/2007_06_Geneva/

ITU-R, 1998. Studio Encoding Parameter of Digital Television for Standard 4:3 and Wide-screen 16:9 Aspect Ratios. Recommendation ITU-R BT.601-5.

Lee, Y., Hur, J., Kim, D., Lee, Y., Cho, S., Hur, N., Kim, J., Su, Y., 2006. CE11: Illumination Compensation. JVT-U052, http://ftp3.itu.ch/av-arch/jvt-site/2006_10_Hangzhou/

Merkle, P., Smolic, A., Mueller, K., Wiegand, T., 2007. Efficient prediction structures for multi-view video coding. *IEEE Trans. on Circuits Syst. Video Technol.*, **17**(11): 1461-1473. [doi:10.1109/TCSVT.2007.903665]

MPEG, 2006. Request for Amendment 14496-10:2006 Amd.4 Multiview Video Coding. ISO/IEC JTC1/SC29/WG11, N8017.

MPEG Video Subgroup, 2008. Introduction to Multiview Video Coding. ISO/IEC JTC1/SC29/WG11, N9580.

Pandit, P., Vetro, A., Chen, Y., 2007. JMVM 4 Software. JVT-W208, http://ftp3.itu.ch/av-arch/jvt-site/2007_04_SanJose/

Smolic, A., Mueller, K., Merkle, P., Fehn, C., Kauff, P., Eisert, P., Wiegand, T., 2006. 3D Video and Free Viewpoint Video—Technologies, Applications and MPEG Standards. IEEE Int. Conf. on Multimedia and Expo, p.2161-2164. [doi:10.1109/ICME.2006.262683]

Su, Y.P., Yin, P., Gomila, C., Kim, J., Lai, P., Ortega, A., 2006a. Thomson's Response to MVC CfP. ISO/IEC JTC1/SC29/WG11.

Su, Y.P., Vetro, A., Smolic, A., 2006b. Common Conditions for MVC. JVT-U211, http://ftp3.itu.ch/av-arch/jvt-site/2007_06_Geneva/

Yamamoto, K., Kitahara, M., Kimata, H., Yendo, T., Fujii, T., Tanimoto, M., Shimizu, S., Kamikura, K., Yashima, Y., 2007. Multi-view video coding using view interpolation and color correction. *IEEE Trans. on Circuits Syst. Video Technol.*, **17**(11):1436-1449. [doi:10.1109/TCSVT.2007.903802]