A Perception-based Color Correction Method for Multi-view Images

Feng Shao, Gangyi Jiang, Mei Yu and Zongju Peng
Faculty of Information Science and Engineering, Ningbo University
Ningbo, China
[e-mail: {shaofeng, jianggangyi, yumei, pengzongju}@nbu.edu.cn]
*Corresponding author: Feng Shao

Received November 14, 2010; revised December 24, 2010; accepted January 18, 2011; published February 28, 2011

Abstract

Three-dimensional (3D) video technologies are becoming increasingly popular, as it can provide users with high quality and immersive experiences. However, color inconsistency between the camera views is an urgent problem to be solved in multi-view imaging. In this paper, a perception-based color correction method for multi-view images is proposed. In the proposed method, human visual sensitivity (VS) and visual attention (VA) models are incorporated into the correction process. Firstly, the VS property is used to reduce the computational complexity by removing these visual insensitive regions. Secondly, the VA property is used to improve the perceptual quality of local VA regions by performing VA-dependent color correction. Experimental results show that compared with other color correction methods, the proposed method can greatly promote the perceptual quality of local VA regions greatly and reduce the computational complexity, and obtain higher coding performance.

***Keywords*:** Free viewpoint video, multi-view images, color correction, visual attention, visual sensitivity

This research was supported by This work was supported by the Natural Science Foundation of China (grant 60872094, 60832003, 60902096), the 863 Project of China (2009AA01Z327), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20093305120002, 200816460003).

DOI: 10.3837/tiis.2011.02.009

1. Introduction

Three-dimensional (3D) video and imaging technology is an emerging trend in the development of digital video systems. New video applications such as 3D television (3DTV) and free viewpoint video (FVV) system have drawn wide attention in recent years [1]. In realizing the 3D video system with multi-view imaging, several problems need to be solved, such as multi-view acquisition [2], multi-view video coding (MVC) [3] and virtual view rendering [4], and so on. Since multi-view videos are captured by multiple cameras from different positions and orientations, illumination and color mismatch, caused by imperfect camera calibration, CCD noise, camera positions and orientations, greatly deteriorate the performances of cross view prediction and virtual view rendering.

Color correction is an effective way to compensate the color inconsistency in multi-view video. Many color correction methods were proposed, in most of which color pattern board is used to extract color mapping relations between views [5]. However, it is not easy to provide a color pattern in various imaging conditions. Yamamoto *et al*. corrected luminance and chrominance of other views by using lookup tables, the correspondences of which were detected by scale invariant feature transform [6]. Chen *et al*. used linear transformation of YUV channels, the coefficients of which were obtained by a simplified color error model [7]. Fecker *et al*. modified luminance and chrominance variations by calculating lookup tables from histograms of two views [8]. In our previous method, color correction was treated as an optimization problem, and the color mapping relationships were calculated by using dynamic programming [9].

In recent years, considerable efforts have been devoted to the research of human visual system (HVS), and some visual perceptual models have been proposed, such as visual attention (VA) model [10] and visual sensitivity (VS) model [11]. To obtain higher perceptual quality, the properties of HVS need to be better exploited. Some perceptual color correction methods were proposed for a single image. Color constancy method seeks a relationship between colors and surface illumination and only reserves corresponding information describing object intrinsic properties [12]. Rizzi *et al*. proposed an automatic color equalization model for unsupervised image color equalization, which can simulate relevant adaptation behaviors of HVS [13]. Bertalmio *et al*. proposed a perceptual-based color correction algorithm in the framework of variational techniques [14], which can produce a perceptually enhanced version for single image. Extensive studies have been carried out on this aspect [15].

From another perspective, different from single image, color correction for multiple images needs to obtain consistent color appearance with reference image. In order to simultaneously improve correction accuracy and reduce computational complexity, how to effectively utilize the above visual perceptual properties in color correction for multi-view images remains an important problem worthy of study. Stentiford *et al*. proposed an attention-based color correction that extracts color correction parameters from pairs of images [16]. The algorithm assumed that measures of attention within images were closely related to measures of similarity between images. However, this assumption is not always true for multi-view images. In general, it is a more difficult problem to achieve high correction accuracy and low computational complexity simultaneously. Therefore, the visual perceptual properties should be an inspiration for new ideas in color correction.

This paper presents a perception-based color correction for multi-view images by utilizing both VS and VA properties in HVS. The VS property aims to reduce the computational complexity and the VA property aims to improve the perceptual quality. The reminder of the paper is organized as follows. In Section 2, the color inconsistency problem in multi-view imaging is discussed. Section 3 presents the proposed perception-based color correction method. Experimental results are analyzed in Section 4, and finally conclusions are drawn.

2. Problem Description in Multi-view Imaging

Fig.1 shows a typical FVV system model [17]. On the sender side, multi-view images are captured with multiple cameras. However, the captured images may contain the geometric misalignment and color differences among the cameras, and these inconsistencies will degrade the performance of subsequent MVC or virtual view rendering. Thus, image correction, including geometric calibration and color correction, should be performed. Then, the corrected multi-view images are compressed for transmission or storage by the encoder. On the receiver side, free viewpoint images are generated by interpolating the decoded images and displayed on a 2D/3D display. Therefore, color correction is a very important process in the system.

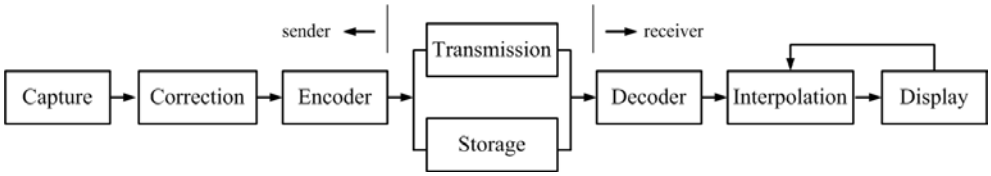


Fig. 1. FVV system model.

2.1 Color Discrepancy Model Analysis

As we have known, the image intensity I_k taken from a Lambertian surface by a digital color camera can be formulated as

$$I_k = \int_\lambda E(\lambda)S(\lambda)R_k(\lambda)d\lambda \tag{1}$$

where $E(\lambda)$ denotes spectral power distribution of the illumination, $S(\lambda)$ stands for the surface spectral reflectance of the object and $R_k(\lambda)$ is spectral sensitivity of the k-th camera sensors.

A finite dimensional linear model [18] can be used to describe the spectral function of illumination and surface reflectance as a linear combination of several basis functions

$$S(\lambda) = \sum_{j=1}^{n} \sigma_j s_j(\lambda), \quad E(\lambda) = \sum_{j=1}^{m} \varepsilon_j e_j(\lambda) \tag{2}$$

where m and n denote the numbers of basis functions corresponding to three color components. $s_j(\lambda)$ is the j-th reflectance basis function, and σ_j is its weighting coefficient. $e_j(\lambda)$ acts as the j-th basis function for the illuminant, and ε_j serves as its weighting coefficient.

Then, under finite dimensional linear model assumption, color discrepancy can be modeled as a linear transformation

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \tag{3}$$

where $\mathbf{A}$ is the 3×3 spread matrix that corresponds to illumination and surface reflectance, $\mathbf{b}$ is the 3×1 offset vector of the imaging sensors, $\mathbf{x}$ is the 3×1 vector of ideal values at a particular pixel, and $\mathbf{y}$ is the 3×1 vector containing the measured values.

Starting from Equ.(3), we can derive color discrepancy model for multiple cameras. As shown

in Equ.(4), the reference and current view images are obtained at the same time instances from various cameras for the same scene

$$\begin{aligned} \mathbf{y}_{ref} &= \mathbf{A}_1\mathbf{x}_{ref} + \mathbf{b}_1 \\ \mathbf{y}_{cur} &= \mathbf{A}_2\mathbf{x}_{cur} + \mathbf{b}_2 \end{aligned} \tag{4}$$

where $\mathbf{x}_{ref}$ and $\mathbf{x}_{cur}$ are ideal color values for current and reference view images, respectively. $\mathbf{y}_{ref}$ and $\mathbf{y}_{cur}$ are corresponding actual acquired color values from the involved cameras, respectively. Suppose that $\mathbf{x}_{ref}$ and $\mathbf{x}_{cur}$ are consistent in the matching pixels, the color discrepancy model between $\mathbf{y}_{ref}$ and $\mathbf{y}_{cur}$ can thus be expressed as

$$\begin{aligned} \hat{\mathbf{y}}_{ref} &= \mathbf{A}_1\mathbf{A}_2^{-1}(\mathbf{y}_{cur} - \mathbf{b}_2) + \mathbf{b}_1 = \mathbf{M}\mathbf{y}_{cur} + \mathbf{T} \\ \mathbf{M} &= \mathbf{A}_1\mathbf{A}_2^{-1},\ \mathbf{T} = \mathbf{b}_1 - \mathbf{A}_1\mathbf{A}_2^{-1}\mathbf{b}_2 \end{aligned} \tag{5}$$

where $\hat{\mathbf{y}}_{ref}$ denotes the matching pixels in reference view image corresponding to $\mathbf{y}_{cur}$.

2.2 Visual Perception Model Analysis

The above color discrepancy model is only suitable for a single camera, and how to extend the model for multi-view images is still a problem to be solved. In our previous works [19][20], we usually derive the color discrepancy model from global image [19] or from particular regions [20]. However, the operations omit actual visual perception in HVS. As we have known, HVS possesses an ability to interpret scenes by modeling the functionality of visual perceptual properties, and the color discrepancy model should be modified according to importance of visual cues in scene.

On the one hand, as shown in **Fig. 2**, VA plays a major functionality in HVS for the ability to filter out redundant visual information and detects the most relevant parts of visual field. The VA model endeavors to predict which parts of an image will attract our attention, which have been successfully applied to video coding [21]. If we can utilize the VA model for the correction of images, higher perceptual quality of the VA regions and higher coding efficiency can be obtained. Thus, one important issue in color correction is how to extract the VA model and derive the color discrepancy model for each VA regions.

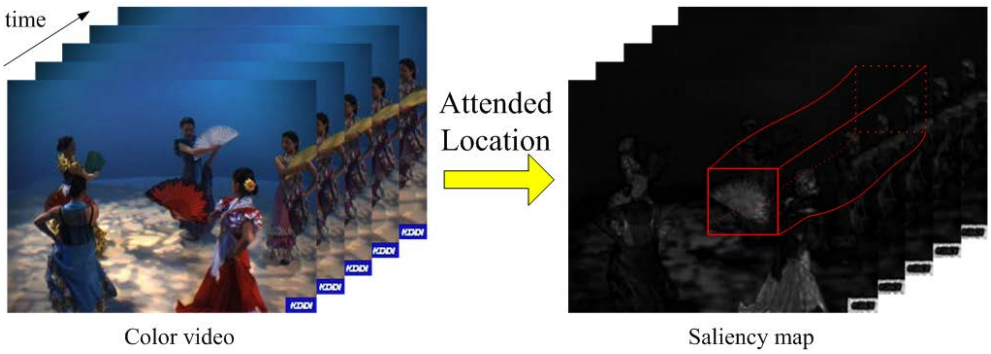


Fig. 2. Modeling VA in HVS.

On the other hand, human eyes cannot sense any changes below the just-noticeable distortion (JND) threshold. The JND provides an important cue for measuring the visibility of the HVS [11]. As shown in **Fig. 3**, the perceptual visible differences between clear and blurred images can be quantified by a JND map, on which each point represents as estimate of the magnitude of perceptual differences between the images [22]. Thus, more computational resource can be allocated to the regions with more perceptual relevance than the perceptually unimportant

regions, and this idea has been successfully applied to video coding [11]. Therefore, another important issue in color correction is how to extract these VS regions so that we can reduce the computational complexity by removing these visual insensitive regions.

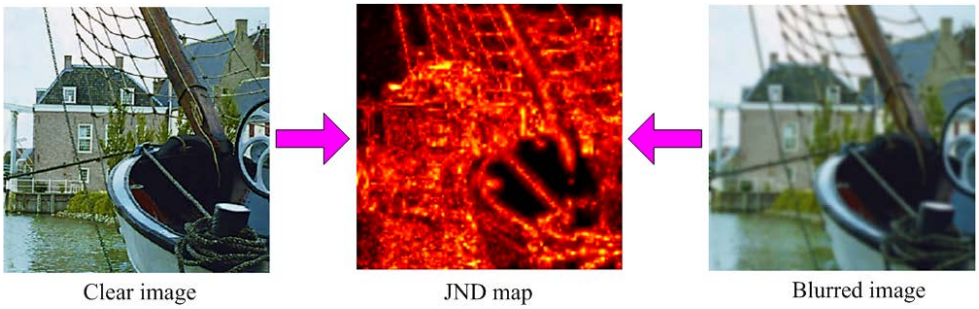


Fig. 3. Quantifying visible differences between two images.

According toour previous analyses, in order to solve perceptual quality and computational complexity problem in color correction, two critical issues have to be tackled: (i) How to improve the perceptual quality of the VA regions based on the VA model. (ii) How to reduce the computational complexity based on the VS model. Although a number of reports have appeared in the related research areas, how to tackle the above two issues effectively remains not fully investigated, let alone the combination of the two issues. The key novelty of our work is that we take full advantage of the perceptual redundancies, and derive suitable color discrepancy model to achieve overall lower computational complexity as well as higher perceptual quality.

3. The Proposed Perception-based Color Correction Method

According to the above analyses, we propose a novel perception-based color correction method for multi-view images. General framework of the proposed method is given in **Fig. 4**. In the framework, VS regions are first extracted from the input multi-view images, and then VA regions are extracted from these VS regions. Then, pixel correspondences are detected for the VA regions between multi-view images. Finally, after matching the correspondences, VA-dependent color correction is implemented for current image to obtain perceptually consistent color appearance with reference image. In the framework, four key technologies, including extraction of VS regions, extraction of VA regions, correspondence detection and VA-dependent color correction, are performed which are essentially correlative.

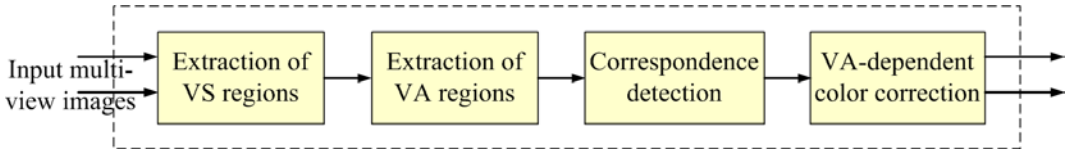


Fig. 4. Framework of the proposed color correction method.

3.1 Extraction of VS Regions

Considering human visual masking effect, human eyes are sensitive to the changes above the JND thresholds. Thus, after removing visual insensitive regions, more computational resource can be saved while the perceptual quality of image is almost unaffected. The JND model can be built based on the visual sensitivity according to luminance contrast and masking effects. In this

paper, Yang's JND model is used to extract JND map [11]. The spatial JND of each pixel can be described by

$$\mathrm{JND}(x,y) = T_l(x,y) + T_t(x,y) - C_{l,t} \cdot \min\{T_l(x,y), T_t(x,y)\} \tag{6}$$

where $T_l(x,y)$ and $T_t(x,y)$ are the visibility thresholds for background luminance masking and texture masking, respectively, and $C_{l,t}$ accounts for the overlapping effect in masking. The corresponding detailed equation can be referred in [11].

After obtaining the JND map, it is needed to quantify the perceptual differences. However, the determination of perceptual differences by subjective experiments is computationally complex [22]. Here, a threshold operation is performed in luminance component. A perceptual difference mask (PDM) is constructed as

$$PDM(x,y) = \begin{cases} 1 & if\ \ D(x,y) > \alpha\sigma_N \\ 0 & otherwise \end{cases} \tag{7}$$

where $D(x,y)$ is the JND value at pixel (x,y), σ_N is the corresponding variance, and α is a weighting factor which is set to 1 empirically. That is, if $PDM(x,y)=1$, the pixel is regarded as visual-sensitive, otherwise, it is excluded from the VS regions. The above extraction processes are performed independently for all multi-view images.

3.2 Extraction of VA Regions

In order to extract the VA regions, saliency map is used to locate the position of the most visually interesting parts. In this paper, a frequency-tuned saliency maps generation method in [23] is used. For one pixel with $PDM(x,y)=1$ and its saliency value $S(x,y)$, we use k-mean clustering algorithm to classify the saliency values and their corresponding pixels. In the algorithm, the number of Gaussian components (K) is pre-determined based on the histogram distribution in saliency map. Thus, each component is used to construct a Gaussians model as

$$\Theta = \{\omega_i, \mu_i, \sigma_i\}_{i=1}^{K} \tag{8}$$

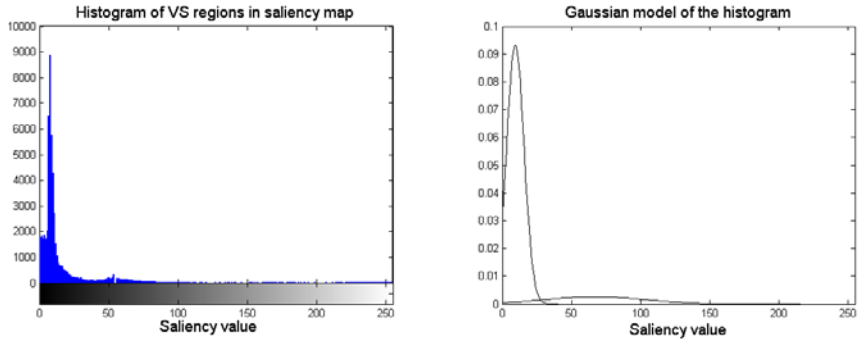
where k denotes the number of mixture components, ω_i denotes the percentage of component i so that $\sum_{i=1}^{k}\omega_i = 1$, μ_i and σ_i are the means and standard deviations of Gaussian distribution.

For each view, each pixel in VS regions are classified by maximizing the probability density of saliency values

$$\gamma = \arg\max_{\Gamma}\left[\frac{e^{-(S(x,y)-\mu_i)^2/2\sigma_i^2}}{\sum_{j=1}^{K} e^{-(S(x,y)-\mu_j)^2/2\sigma_j^2}}\right] \tag{9}$$

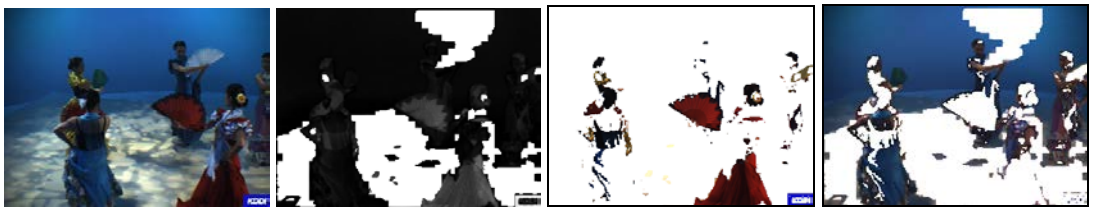
where γ is the class label for one pixel, and $\Gamma = \{\gamma^i \mid 1 \le i \le K\}$ is the set of all possible classes. The above extraction processes are only required for current corrected view.

To demonstrate the effecttiveness of the proposed classification method, we provide a classification result for '*Flamenco1*' sequence (view 2, frame #569) in **Fig. 5** and **6**. In **Fig. 5**, the histogram of VS regions in a saliency map is provided, and the corresponding Gaussian model is constructed. In **Fig. 6**, the corresponding VA regions within each class are shown. It can be observed that after the classification, class 1 corresponds to the right component with a lower probability, and class 2 corresponds to the left component with a higher probability in the model, respectively.



(a) **(b)**

Fig. 5. Saliency map at the 566-th frame in '*Flamenco1*'. **(a)** histogram; **(b)** Gaussian model.



(a) **(b)** **(c)** **(d)**

Fig. 6. Corresponding k-mean classification result of **Fig. 5**. **(a)** original image; **(b)** saliency map of VS regions; **(c)** class 1; **(d)** class 2.

3.3 Correspondence Detection

After extracting the VA regions from those VS regions, accurate correspondence must be established between reference and current views. In order to eliminate illumination change between views, we use mean-removed sum of absolute differences (MRSAD) [24] based on disparity estimation to find matching points. For a $N\times N$ block of pixels located at the position (x, y), the MRSAD is defined as

$$MRSAD(i,j)=\sum_{x=x_0}^{x_0+N}\sum_{y=y_0}^{y_0+N}|(Y^{cur}(x,y)-\mu^{cur})-(Y^{ref}(x+i,y+j)-\mu^{ref})| \tag{10}$$

where $Y^{cur}(x, y)$ and $Y^{ref}(x, y)$ are luminance values of the current and reference views, respectively, (i, j) represents a candidate disparity vector, μ^{cur} and μ^{ref} are the mean values of blocks in the current view and reference view, respectively.

Here, the current view is firstly divided into 8×8 blocks, and for each block, the disparity vector $\mathbf{d}_{s+1\to s}$ from the view s+1 to the view s is calculated by minimizing the *MRSAD* over a search range

$$\mathbf{d}_{s+1\to s}=\arg\min_{(i,j)\in\Omega} MRSAD(i,j) \tag{11}$$

where Ω denotes the search range of disparity estimation. However, the disparity vectors estimated with the above formula may not be the true disparity due to occlusion, mismatching or other factors. An inverse matching operation from the reference view to the current view is performed to validate the matching. If the disparity deviation between $\mathbf{d}_{s+1\to s}$ and $\mathbf{d}_{s\to s+1}$ is less than 2, that is, $|\mathbf{d}_{s+1\to s}+\mathbf{d}_{s\to s+1}|<2$, the blocks are matched.

Then, a correspondence detection mask (CDM) is computed by comparing the *PDM* of the matching points

$$CDM(x,y) = \begin{cases} 1 & if \ \ PDM^{ref}(x+dx, y+dy) = 1 \\ 0 & otherwise \end{cases} \tag{12}$$

Here, if $CDM(x,y)=1$, the pixel at (x,y) is regarded as matched, otherwise, it is excluded from the matched pixels.

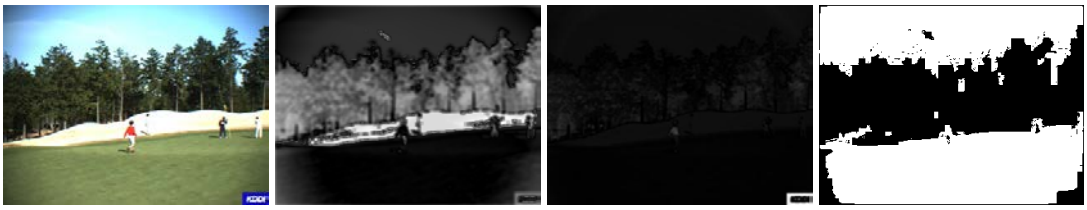
After the above operation, the accurate correspondence is established. It should be noted that the proposed method reduces the computational complexity from two aspects. Firstly, *k*-mean clustering is only needed for those pixels in VS regions, thus the number of samples is largely reduced even though it makes very few important contribution to computational complexity reduction. Secondly, correspondence detection is only needed for those pixels in VS regions, which can significantly reduces the computational complexity since performing disparity estimation is computationally expensive. **Fig.7-10** show the examples of correspondence detection for '*Flamenco1*', '*Objects2*', '*Golf2*' and '*Race1*' sequences, respectively. **Fig.7-10-(a)** show the original second viewpoint image. **Fig.7-10-(b)** and **Fig.7-10-(c)** show the corresponding JND map and saliency map, respectively. **Fig.7-10-(d)** show the *PDM* results in which black regions denote the detected VS regions in the view.

(a) **(b)** **(c)** **(d)**

Fig. 7. Example of the proposed algorithm for correspondence detection of '*Flamenco1*': **(a)** original image; **(b)** JDN map; **(c)** saliency map; **(d)** PDM result.

(a) **(b)** **(c)** **(d)**

Fig. 8. Example of the proposed algorithm for correspondence detection of '*Objects2*': **(a)** original image; **(b)** JDN map; **(c)** saliency map; **(d)** PDM result.



(a) **(b)** **(c)** **(d)**

Fig. 9. Example of the proposed algorithm for correspondence detection of '*Golf2*': **(a)** original image; **(b)** JDN map; **(c)** saliency map; **(d)** PDM result.

(a) **(b)** **(c)** **(d)**

Fig. 10. Example of the proposed algorithm for correspondence detection of '*Race1*': **(a)** original image; **(b)** JDN map; **(c)** saliency map; **(d)** PDM result.

3.4 VA-dependent Color Correction

The above color discrepancy model is suitable to RGB format data. While for YUV video format, the linear color discrepancy model is also satisfied because the transformation from RGB to YUV is linear. Then, we formulate the problem of color correction as a weighted linear sum of the current YUV values

$$\begin{bmatrix} Y^{cor} \\ U^{cor} \\ V^{cor} \end{bmatrix} = \begin{bmatrix} \alpha_{Y1} & \alpha_{Y2} & \alpha_{Y3} \\ \alpha_{U1} & \alpha_{U2} & \alpha_{U3} \\ \alpha_{V1} & \alpha_{V2} & \alpha_{V3} \end{bmatrix} \begin{bmatrix} Y^{cur} \\ U^{cur} \\ V^{cur} \end{bmatrix} + \begin{bmatrix} a_{Y4} \\ a_{U4} \\ a_{V4} \end{bmatrix} \tag{13}$$

where α_{Yi}, α_{Ui} and α_{Vi} $(1 \le i \le 4)$ are the elements in correction matrix.

In the next step, all matched pixels in the current and reference views are written in vector forms as $\mathbf{Y}^{cur}$, $\mathbf{U}^{cur}$, $\mathbf{V}^{cur}$, $\mathbf{Y}^{ref}$, $\mathbf{U}^{ref}$ and $\mathbf{V}^{ref}$. Let $\boldsymbol{\psi}=[\mathbf{Y}^{cur}\ \mathbf{U}^{cur}\ \mathbf{V}^{cur}\ 1]$ and $\boldsymbol{\Phi}=[\mathbf{Y}^{ref}\ \mathbf{U}^{ref}\ \mathbf{V}^{ref}]$, the relationship between the vectors of the matched points can be expressed as

$$\boldsymbol{\Phi} \boldsymbol{\Psi} \mathbf{C} \boldsymbol{\varepsilon} \quad + \tag{14}$$

where $\boldsymbol{\varepsilon}$ is an error vector, $\mathbf{C}$ is a 4×3 matrix. By minimizing the energy of the error vector $\boldsymbol{\varepsilon}$ in a least square sense, the correction matrix $\mathbf{C}$ can be computered by

$$\mathbf{C} \boldsymbol{\Psi} \boldsymbol{\Psi}^{T} \boldsymbol{\Phi})^{-1} \quad ^{T} \tag{15}$$

After extracting the correction matrix $\mathbf{C}_i$ $(1 \le i \le K)$ for each VA region, the most direct strategy is to perform color correction for each region with the correction matrix $\mathbf{C}_i$ independently. However, the errors may be raised because of the difference between the correction matrices, especially at the edges of each region. The reason is that two adjacent pixels may be classified into different VA regions based on the classification criterion in Equ.(9) even though their probability densities are similar. In order to solve the problem, we propose to linearly combine these different matrixes $\{\mathbf{C}_i \mid 1 \le i \le K\}$ with weights to derive the finial color correction function

$$\hat{\mathbf{C}} = \sum_{i=1}^{K} \alpha_i \mathbf{C}_i \tag{16}$$

where α_i denotes the weighting coefficients for the i-th VA region. Theoretically, it is necessary to give more weights to the correction functions which yield smaller transformation error. However, it is difficult to obtain accurate weighting coefficients for each pixel based on the transformation error criterion. Here, the weighting coefficient is defined as the probability that a pixel (x,y) belongs to i-th VA region

$$\alpha_i(x,y) = \frac{\omega_i \cdot e^{-(S(x,y)-\mu_i)^2/2\sigma_i^2}}{\sum_{j=1}^{K} \omega_j \cdot e^{-(S(x,y)-\mu_j)^2/2\sigma_j^2}} \tag{17}$$

From the above equation, it is obvious that the extraction process of weighting coefficients is

individual for each pixel, which can avoid the problem that two similar pixels have different weighting coefficients.

4. Experimental Classification Results and Analysis

In experiments, multi-view video sequences '*Flamenco1*', '*Objects2*', '*Golf2*' and '*Race1*' are used as the test sequences [25]. The size of the sequences is 320×240, and the multi-view images are taken by a horizontal parallel camera configuration with eight viewpoints. We implemented our coding experiments in JMVM7.0 [26]. The test condition is set as follows: four basis QPs 22, 27, 32, 37 are used, the temporal GOP size is 15, and the total number of encoded frames in each view is 60.

In order to objectively measure performances of the proposed method in correction accuracy and computational complexity, three color correction schemes are discussed

Scheme 1: The global color correction method [19]

Scheme 2: The proposed color correction method without VS function

Scheme 3: The proposed color correction method

Note that, for 'scheme 1', spatial color correction method for keyframe frames in [19] is used. For 'scheme 2', extraction of VS regions is removed from the proposed framework in **Fig. 4**, and other functions are preserved.

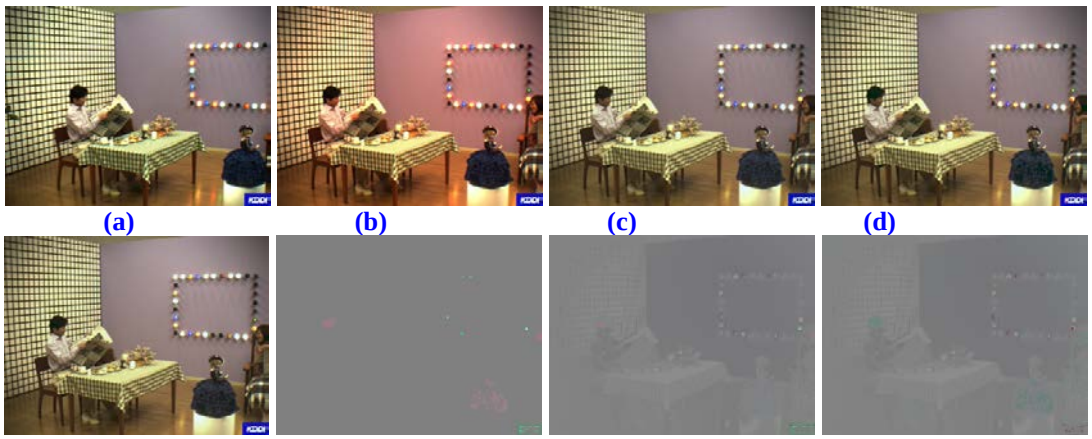
4.1 Objective and Subjective Correction Performance Comparison

Fig. 11-14-(a)-(b) show the reference image and current image in the view 1 and view 2 at time 569 of '*Flamenco1*', '*Objects2*', '*Golf2*' and '*Race1*'. Clearly, the color consistency among the views is poor. Thus, color correction is necessary if the multi-view images will be used to render new virtual arbitrary views. **Fig. 11-14-(c)-(e)** show the corrected images with scheme 1, scheme 2 and scheme 3, respectively. Although there are no obvious differences between the overall appearances of the corrected images, the partial details are enhanced with the proposed scheme, as shown in the corresponding error images in **Fig. 11-14-(f)-(h)**. We enlarge the parts in **Fig. 11-12**, and show the detail examples in **Fig. 15-16**. It can be seen more clearly that compared to scheme 1, the proposed scheme can not only correct the color of current image to that of reference one, but also promote the perceptual quality of local VA regions. Since the differences between the corrected image with scheme 2 and scheme 3 are not obvious, the effect of VS function to perceptual quality is almost negligible. Moreover, the proposed scheme is particularly effective for indoor scenes, such as '*Flamenco1*' and '*Objects2*'. This is because the change of VA region is relatively obvious for the indoor scenes, and the effect of VA-dependent color correction will be more significant.

(a) **(b)** **(c)** **(d)**

(e) **(f)** **(g)** **(h)**

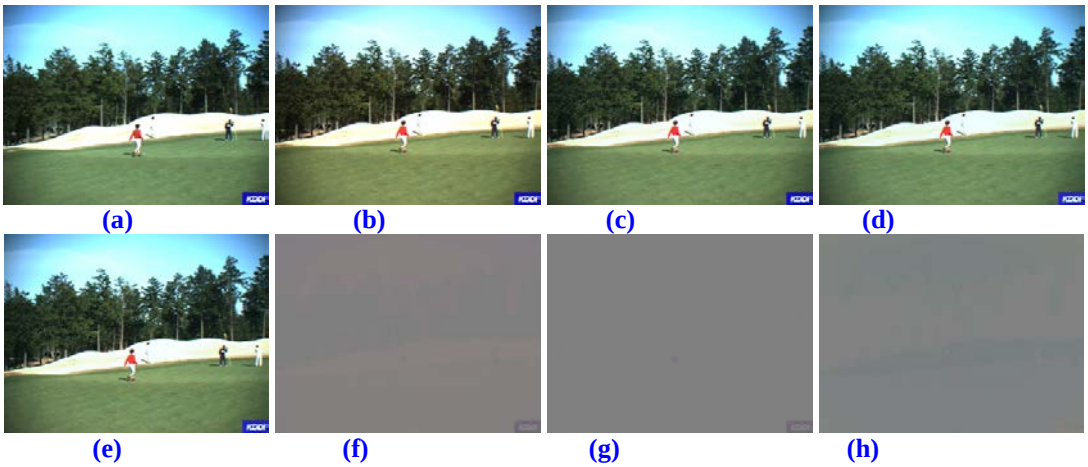
Fig. 11. Color correction results of '*Flamenco1*'. **(a)** reference image (view 1, frame #569); **(b)** current image (view 2, frame #569); **(c)** the corrected image with scheme 1; **(d)** the corrected image with scheme 2. **(e)** the corrected image with scheme 3; **(f)** the error image between **(c)** and **(d)**; **(g)** the error image between **(c)** and **(e)**; **(h)** the error image between **(d)** and **(e)**.



(a) **(b)** **(c)** **(d)**

(e) **(f)** **(g)** **(h)**

Fig. 12. Color correction results of '*Objects2*'. **(a)** reference image (view 1, frame #569); **(b)** current image (view 2, frame #569); **(c)** the corrected image with scheme 1; **(d)** the corrected image with scheme 2. **(e)** the corrected image with scheme 3; **(f)** the error image between **(c)** and **(d)**; **(g)** the error image between **(c)** and **(e)**; **(h)** the error image between **(d)** and **(e)**.



(a) **(b)** **(c)** **(d)**

(e) **(f)** **(g)** **(h)**

Fig. 13. Color correction results of '*Golf2*'. **(a)** reference image (view 1, frame #569); **(b)** current image (view 2, frame #569); **(c)** the corrected image with scheme 1; **(d)** the corrected image with scheme 2. **(e)** the corrected image with scheme 3; **(f)** the error image between **(c)** and **(d)**; **(g)** the error image between **(c)** and **(e)**; **(h)** the error image between **(d)** and **(e)**.

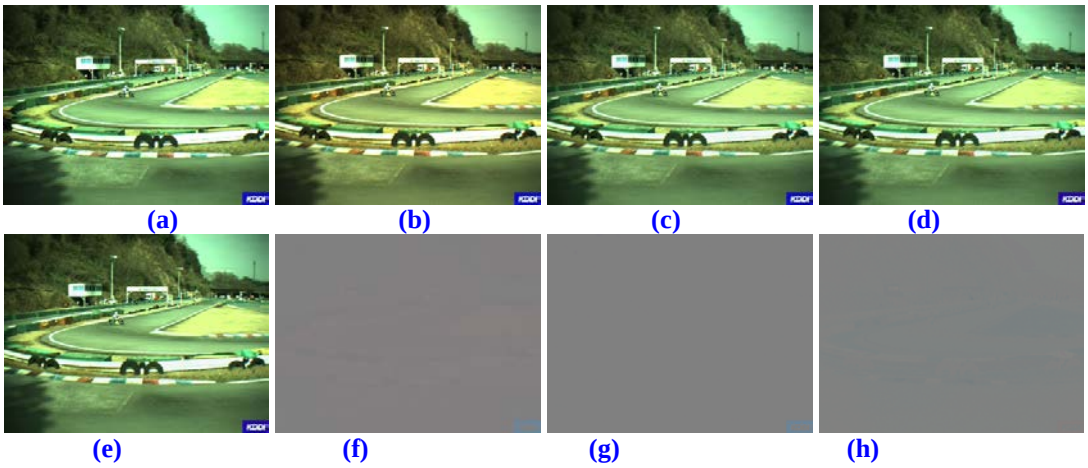


Fig. 14. Color correction results of '*Race1*'. **(a)** reference image (view 1, frame #569); **(b)** current image (view 2, frame #569); **(c)** the corrected image with scheme 1; **(d)** the corrected image with scheme 2. **(e)** the corrected image with scheme 3; **(f)** the error image between **(c)** and **(d)**; **(g)** the error image between **(c)** and **(e)**; **(h)** the error image between **(d)** and **(e)**.

Fig. 15. Enlarged part in **Fig. 11**. **(a)** reference image; **(b)** current image; **(c)** the corrected image with scheme 1; **(d)** the corrected image scheme 3.

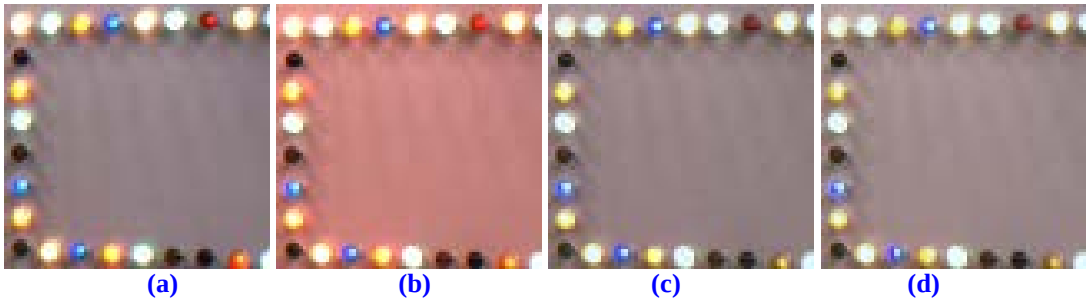


Fig. 16. Enlarged part in **Fig. 12**. **(a)** reference image; **(b)** current image; **(c)** the corrected image with scheme 1; **(d)** the corrected image scheme 3.

Then, in order to subjectively compare the performances of different schemes, we calculate the weighted PSNR (WPSNR) [19] between the corrected images and the actual images in one GOP. Since there is no actual reference for the corrected images, the corrected images with scheme 1 are supposed as the benchmark. Then, the WPSNR value between the corrected and the benchmark images is calculated. As shown in **Fig. 17**, the overall differences between the corrected images with scheme 1 and scheme 3 are not obvious, especially for '*Golf2*' and "*Race1*'. Moreover, compared with scheme 3, the corrected image with scheme 2 is more similar with the benchmark image, because more global information is retained in computing

the correction coefficients. It should be noted that the PSNR evaluation is not an effective means in quality evaluation, and perceived quality measurement should be developed through subjective perceptual experiment.

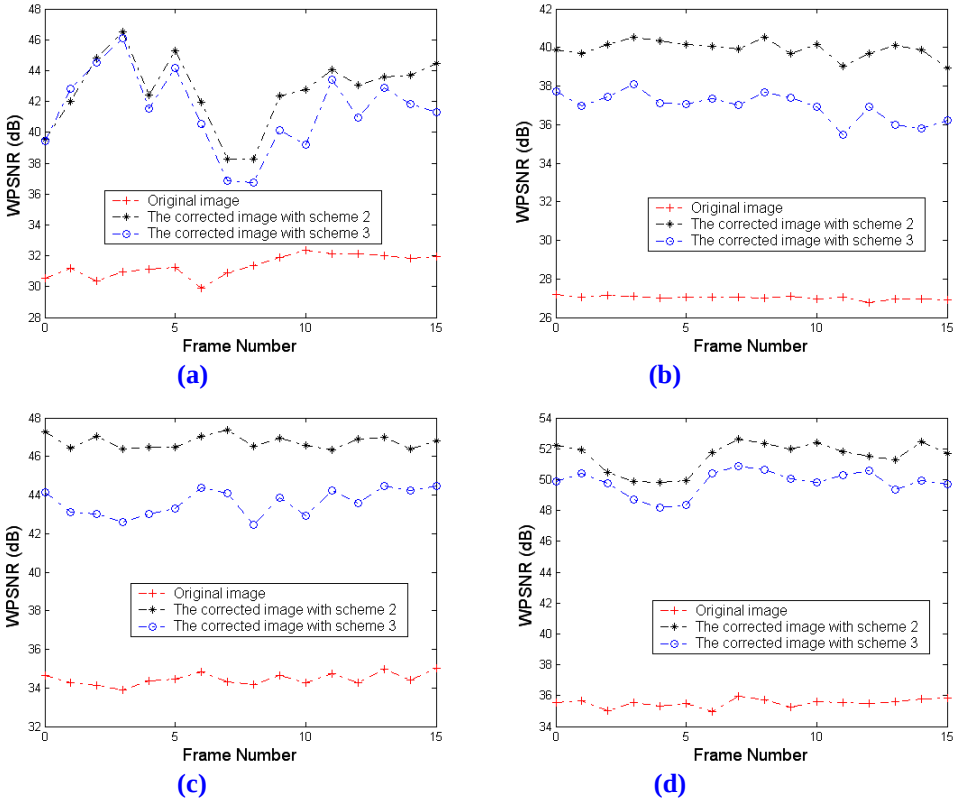


Fig. 17. WPSNR comparison results. **(a)** *Flamenco1*; **(b)** *Objects2*; **(c)** *Golf2*; **(d)** *Race1*.

4.2 Computational Complexity Comparison

To evaluate the computation complexity of the proposed method, we implemented these three schemes on PC and compared the software run-time costs. The execution time for one GOP is summarized in **Table. 1**, in which *TS* indicates the average time saving and it is defined by

$$TS = \frac{T_{Scheme1} - T_{Proposed}}{T_{Scheme1}} \times 100[\%] \tag{18}$$

where $T_{Scheme1}$ and $T_{Proposed}$ are the total computational time for scheme 1 and the proposed scheme, respectively. Here, the maximum disparity search ranges in the experiments are 40 and 5 at the horizontal and vertical directions, respectively. It is obvious that the proposed scheme can save the computational time, ranging from 6.75% to 29.98% when compared with scheme 1, and ranging from 9.90% to 32.55% when compared with scheme 2 for different test sequences. This time saving may be important for low complexity applications. Moreover, this time savings is not at the cost of reducing perceptual performance.

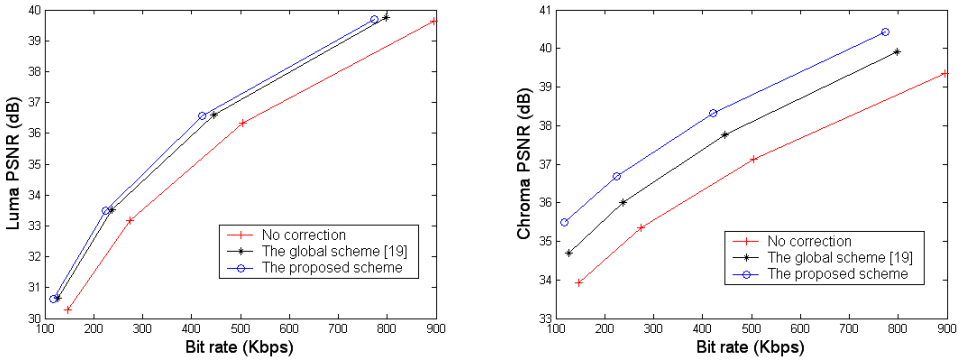
Table. 1. Speedup performance comparison of different schemes.

	$T_{Scheme1}$[s]	$T_{Scheme2}$[s]	$T_{Scheme3}$[s]	$TS_{3->1}$[%]	$TS_{3->2}$[%]
Flamenco1	37.500	38.812	34.968	6.75	9.90

Objects2	37.312	38.734	26.125	29.98	32.55
Golf2	37.156	38.796	29.625	20.27	23.64
Race1	37.375	39.031	32.218	13.80	17.46

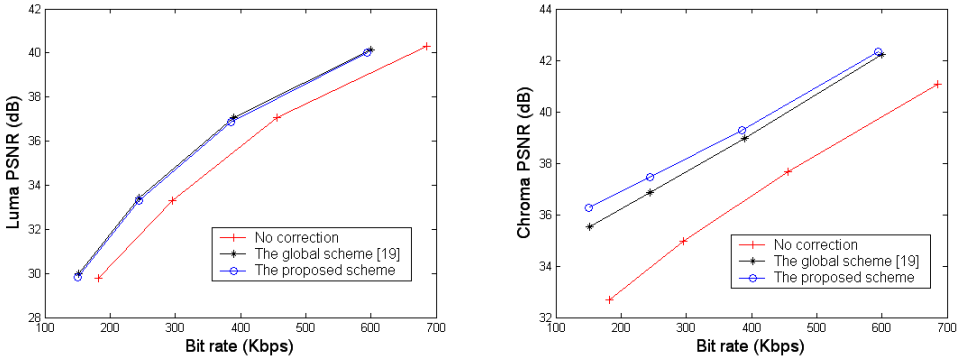
4.2 Coding Performance Comparison

We compare the coding performance of the proposed scheme with 1) compressing the original video; 2) the global scheme [19]. **Fig. 18-21** show the rate-distortion performance comparisons of '*Flamenco1*', '*Objects2*', '*Golf2*' and '*Race1*' for the second view, respectively. The vertical axis in each sub-figure is the PSNR of luma or chroma channel, while the horizontal axis corresponds to the sum of the bitrates used for encoding. The Y component carries the luma channel information, and the chroma PSNR is the average PSNR of the U and V components. As can be seen from the results, color correction can result in larger PSNR gains compared to compressing the original data. Besides, for '*Flamenco1*' and '*Objects2*', the coding performances of the proposed scheme are much higher than that of the global scheme, especially in chroma channel. The reason lies in that the proposed scheme aims to promote the perceptual quality of local regions, which can increase the local correlations in MVC. While for outdoor scenes, such as '*Golf2*' and "*Race1*', the coding performances are almost consistent with the global and the proposed schemes.



(a) **(b)**

Fig. 18. Coding performance comparisons of '*Flamenco1*'. **(a)** luma channel; **(b)** chroma channel.



(a) **(b)**

Fig. 19. Coding performance comparisons of '*Objects2*'. **(a)** luma channel; **(b)** chroma channel.

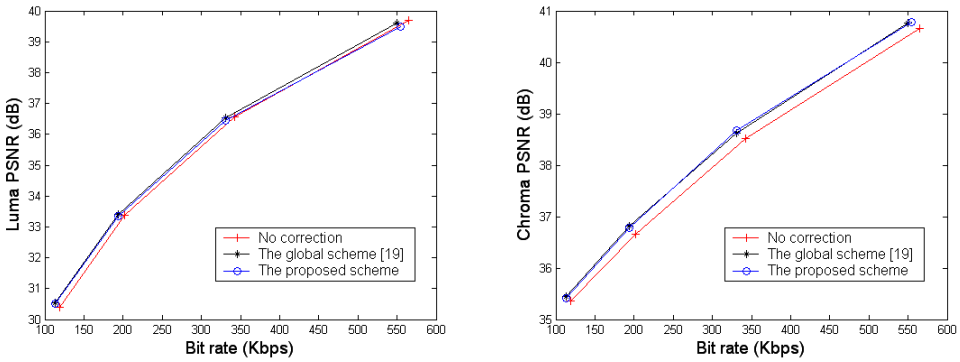


Fig. 20. Coding performance comparisons of '*Golf2*'. **(a)** luma channel; **(b)** chroma channel.

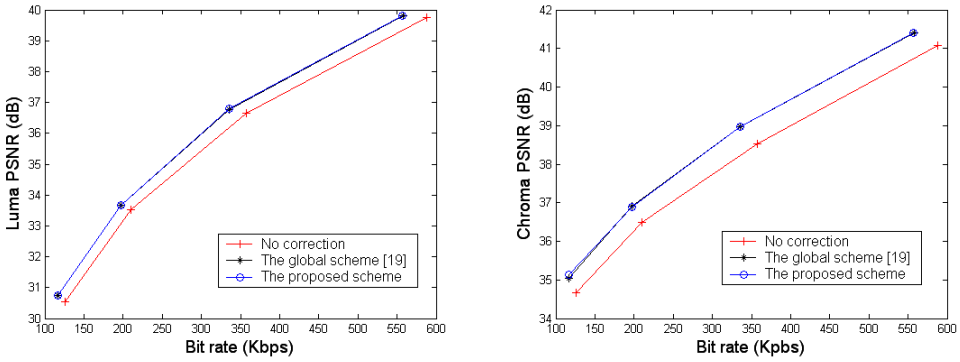


Fig. 21. Coding performance comparisons of '*Race1*'. **(a)** luma channel; **(b)** chroma channel.

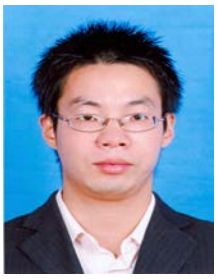
5. Conclusion

Color correction is an important issue for multi-view video coding and virtual view synthesis in three dimensional television (3DTV) and free viewpoint video (FVV) applications. In this paper, a perception-based color correction method for multi-view images is proposed. In the proposed method, human visual attention (VA) and visual sensitivity (VS) properties are incorporated in correction process. The proposed method has the following characteristics. (i) The VS model aims to reduce the computational complexity by removing the perceptual redundancy. (ii) The VA model aims to improve the perceptual quality of the VA regions. Experimental results show the effectiveness of the proposed method. In this paper, we only consider the case of multi-view images. If extending the algorithm to multi-view video, the following aspects should be considered in the future research. Firstly, temporal properties should be considered in the VA and VS model. Furthermore, stereoscopic just noticeable distortion (JND) model and region-of-interest (ROI) model should be incorporated into the final algorithm. Finally, subjective perceptual experiment and perceptual video coding might be required to evaluate the performance of the algorithm.

References

[1] Y. Morvan, D. Farin and P. H. N. de With, "System architecture for free-viewpoint video and 3D-TV," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 925-932, 2008. Article (CrossRef Link)
[2] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 10-21, 2007. Article (CrossRef Link)
[3] P. Merkle, A. Smolic, K. Muller and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1461-1473, 2007. Article (CrossRef Link)
[4] K. Muller, A. Smolic, K. Dix, P. Merkle, P. Kauff and T. Wiegand, "View synthesis for advanced 3D video system," *EURASIP Journal on Image and Video Processing*, vol. 2008, Article ID 438148, 11 pages, 2008. Article (CrossRef Link)
[5] S. H. Lee and J. H. Choi, "Design and implementation of color correction system for images captured by digital camera," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 268-276, 2008. Article (CrossRef Link)
[6] K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura and Y. Yashima, "Multiview video coding using view interpolation and color correction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1436-1449, 2007. Article (CrossRef Link)
[7] Y. Chen, J. Chen and C. Cai, "Luminance and chrominance correction for multi-view video using simplified color error model," in *Proc. of Picture Coding Symposium*, Beijing, China, April 2006.
[8] U. Fecker, M. Markowsky and A. Kaup, "Histogram-based pre-filtering for luminance and chrominance compensation of multi-view video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1258-1267, 2008. Article (CrossRef Link)
[9] F. Shao, G. Jiang and M. Yu, "Multi-view video color correction using dynamic programming," *Journal of System Engineering and Electronics*, vol. 19, no. 6, pp. 1115-1120, 2008. Article (CrossRef Link)
[10] Z. Lu, W. lin, X. Yang, E. Ong and S. Yao, "Modelling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1928-1942, 2005. Article (CrossRef Link)
[11] X. Yang, W. Lin, Z. Lu, E. Ong and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 6, pp. 742-751, 2005. Article (CrossRef Link)
[12] K. Barnard, V. Cardei and B. Funt, "A comparison of computational color constancy algorithms – Part 1: methodology and experiments with synthesized data," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 972-983, 2002. Article (CrossRef Link)
[13] A. Rizzi and C. Gatta, "From retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalization," *Journal of Electronics Imaging*, vol. 13, no. 1, pp. 75-84, 2004. Article (CrossRef Link)
[14] M. Bertalmio, V. Caselles, E. Provenzi and A. Rizzi, "Perceptual Color correction through variational techniques," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1058-1072, 2007. Article (CrossRef Link)
[15] R. Palma-Amestoy, E. Provenzi, M. Bertalmio and V. Caselles, "A perceptually inspired variational framework for color enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 458-474, March 2009. Article (CrossRef Link)
[16] F. W. M. Stentiford and A. Bamidele, "Attention-based colour correction," in *Proc. of the SPIE*, vol. 6057, pp. 158-167, 2006. Article (CrossRef Link)
[17] ISO/IEC JTC1/SC29/WG11, "Available Technologies for FTV," Doc. M15088, Antalya, Turkey, January 2008.
[18] T. Lanrence and A. Brian, "Color constancy: A method for recovering surface spectral reflectance," *Journal of Optical Society of America(A)*, vol. 3, no. 1, pp. 29-33, 1986. Article (CrossRef Link)
[19] F. Shao, G. Jiang, M. Yu and Y. Ho, "Fast color correction for multi-view video by modeling spatio-temporal variation," *Journal of Visual Communication and Image Representation*, vol. 21, no. 5-6, pp. 392-403, 2010. Article (CrossRef Link)

[20] F. Shao, G. Jiang, M. Yu and Y. Ho, "Highlight-detection-based color correction method for multi-view images," *ETRI Journal*, vol. 31, no. 4, pp. 448-450, 2009. Article (CrossRef Link)
[21] Y. Zhang, G. Jiang, M. Yu, Y. Yang, Z. Peng and K. Chen, "Depth perceptual region-of-interest based multiview video coding," *Journal of Visual Communication and Image Representation*, vol. 21, no. 5-6, pp. 498-512, 2010. Article (CrossRef Link)
[22] J. Lubin, M. Brill and R. Crane, "Vision model-based assessment of distortion magnitudes in digital video," Available: http://www.mpeg.org/MPEG/JND.
[23] R. Achanta, S. Hemami, F. Estrada and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. of International Conference on Computer Vision and Pattern Recognition*, Miami, U.S.A., 2009. Article (CrossRef Link)
[24] J. Hur, S. Cho and Y. Lee, "Adaptive local illumination change compensation method for H.264/AVC-based multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1496-1505, 2007. Article (CrossRef Link)
[25] ISO/IEC JTC1/SC29/WG11, "KDDI multi-view video sequences for MPEG 3DAV use," Doc. M10533, Munich, Germany, 2004.
[26] ISO/IEC JTC1/SC29/WG11, "Joint multiview video model (JMVM) 7.0," JVT-Z207, Antalya, Turkey, Jan. 2008.



Feng Shao received the B.S. and Ph.D degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in Electronic Science and Technology. He is currently an associate professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests include image/video coding, image processing, and image perception.

Gangyi Jiang received his M.S. degree from Hangzhou University in 1992, and received his Ph.D. degree from Ajou University, Korea, in 2000. He is now a professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include digital video compression and communications, multi-view video coding and image processing.

Mei Yu received her M.S. degree from Hangzhou Institute of Electronics Engineering, China, in 1993, and Ph.D. degree form Ajou University, Korea, in 2000. She is now a professor in Faculty of Information Science and Engineering, Ningbo University, China. Her research interests include image/video coding and video perception.

Zongju Peng received the B.S. degrees from Sichuan University, Sichuan, China, in 1998. He is currently a Ph.D. student in Institute of Computing Technology, Chinese Academy of Science, China. His research interests include image/video coding, image processing, and image perception.