Color Correction for Multi-view Video Based on Background Segmentation and Dominant Color Extraction

FENG SHAO, ZONGJU PENG and YOU YANG Faculty of Information Science and Engineering Ningbo University 315211 Ningbo CHINA shaofeng@nbu.edu.cn

Abstract: - Color correction is a necessary operation in multi-view video processing, because color tends to be influenced by camera characteristic, surface reflectance or scene illumination. To achieve high-quality correction results, a new color correction method, based on the theoretical model of multi-view imaging and image restoration, is proposed in this paper. Considering the illumination problem in multi-view imaging, foreground and background regions are separated from images and dominant color extraction is used only for background regions of reference and input images, so that uniform reference surface information is used. Then, color correction is extended to video sequences with a tracking approach. Furthermore, an objective performance evaluation is proposed to evaluate the color correction. We present a variety of results for different test sequences, arguing that background-based method outperforms foreground-based method, and better subjective and objective visual effect can be achieved for images as well as videos.

Key-Words: - Multi-view video, color correction, background segmentation, dominant color extraction, principal component analysis, color difference

1 Introduction

Television (TV) has a long and successful history since it realized a human dream of seeing a distant world in real-time. It stands as the most important visual information system. However, TV shows us the same scene even if we move our viewpoints in front of the display. It is quite different from what we experience in the real world. Free viewpoint television (FTV) is expected to be a next generation visual application, which allows the user interactively controlling the viewpoint position of dynamic real-word scene in real time^[1]. The realtime FTV system covers the complete chain from capturing to display. Normally, multiple camera arrays are used to capture the scene information. If the camera density is not very dense, intermediate views can be generated by interpolation between different views using disparity estimation^[2] or 3D warping^[3].

Although FTV is capable to provide an exciting viewing experience, it is challenging to put it into practical applications. Firstly, Multi-camera capturing system for FTV might not be perfectly calibrated. Many standard geometric calibration methods had proposed for calibrating array of cameras^[4-5]. Secondly, in practical imaging, camera parameters in multi-camera capturing system may

be inconsistent, and exposure or focus may be variable for different views. The heterogeneous cameras can lead to global or local mismatches across different views when virtual views are synthesized at the client of FTV system. In addition, it is often 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 subsequent multi-view video coding (MVC)^[6] or virtual view rendering.

Many color transfer methods had been proposed. Reinhard et al. presented a pioneering work that transferred color statistic from one image to another for coping color characteristic using the mean and standard deviation^[7]. Wang et al. presented an effective algorithm for image sequence color transfer^[8]. The mean and variance used for image transfer in an image sequence were interpolated to produce in-between color transfers, and a color variation curve was used to control the interpolation across color statistics values for an in-between image sequence. Pitie et al. applied the color mapping by transforming any *N*-dimensional probability density function into another one, and reduced the grain artifact through a post-processing algorithm by preserving the gradient field of the original image^[9]. Even though color transfer can provide a new avenue for color correction of multiview images, but faces the disadvantage from lack of the correspondence in the statistic information. Moreover, it is not inconformity with human vision.

As we have known, human vision system can adapt to the change of light source and only reserve corresponding information describing intrinsic properties of object, such as surface reflectance. The human's ability is called as color constancy. The task for a color constancy algorithm is to separate spectral surface reflectance from spectral power distribution. Starting from this angle, many methods have proposed to deal with the problem, such as retinex method^[10], gamut-mapping method^[11], correlation method^[12], etc. However, it is mathematically impossible to accurately restore the spectral properties of the illumination and objects from image data. Some color correction methods were proposed with correspondence matching. Yamamoto et al. proposed to obtain color correspondence map with color pattern board^[13] or without color pattern board^[14] (detecting the correspondences by scale invariant feature transform). However, even though the pixel matching information is accurate enough for these methods, the color mapping relation may be not consistent in different matching regions, e.g. foreground and background.

We previously implemented a content-adaptive color correction method^[15]. In the method, we can automatically select color correction property for different multi-view images based on their color contents. Based on the correction property, global or preferred region matching principal component analysis (PCA) correction is performed. In this paper, we propose a technique that extends our previous method to make the algorithm more robust and to correct the colors of video frame sequences. include: The main extensions implementing background segmentation to extract uniform reference surface; performing dominant color extraction for accurate mapping information; and extending the algorithm along the temporal axis. In section II, we describe theoretical models for multiview imaging and image restoration. In section III, we describe the details of methodology. Experimental results are given and analyzed in section IV, followed by some conclusions in section V.

2 Theoretical Models for Multi-view Imaging and Image Restoration

As we have known, for a Lambertian diffuser, color image I_k is the result of a complex reflection between three components of spectral power distribution of illumination $E(\lambda)$, surface spectral reflectance of object $S(\lambda)$ and spectral sensitivity of the k-th camera sensors $R_k(\lambda)$, and it is described as

$$I_{k}(x,y) = \int_{\omega} R_{k}(\lambda) E(x,y,\lambda) S(x,y,\lambda) d\lambda$$
(1)

where λ denotes wavelengths of light in visible spectrum. x and y are the coordinates of a pixel in an image, subscript k indicates a type of color sensor of the camera, integral range ω is visible spectrum.

Estimating the influence of these three factors on the measured signal is one of the main goals of color correction. The influence of the sensors is usually known, which results from variations in camera response due to different sensor characteristics and ambient conditions like temperature, manufacturing differences, and so on^[16]. Then the remaining problem is to separate the effects of the scene properties and the influence of the illumination. In multi-view imaging, it is often impossible to capture object under perfectly constant lighting an conditions. When the scene depth is changed, the illumination intensity will be changed accordingly. In addition, Lambertian reflectance assumption is not always satisfied in the scene. Under the Lambertian reflectance assumption, the illuminated region of the surface emits the entire light equally in all direction. Therefore, the imaging is inconsistent in different scene depth even for identical objects. A simple strategy is to separate scene depth into foreground and background, and use background as a uniform reference surface between two views.

Supposing reference image $I_k^{ref}(x,y)$ in one view as benchmark image, input image $I_k^{inp}(x,y)$ in another view as degradation image, the degradation model can be expressed as

 $I_k^{inp}(x,y) = f_k^{real}(x,y) * h_k(x,y) = I_k^{ref}(x',y') * h_k(x',y')$ (2) where $f_k^{real}(x,y)$ denotes real image for input. Therefore, the purpose of color correction is to restore the real image, if we have known $h_k(x,y)$, image restoration is in fact a deconvolution process.

Many techniques, including neural network and Kalman filtering, had been used to achieve image restoration^[17,18]. Principal component analysis (PCA) is a linear transformation that transforms the data to a new orthogonal coordinate system. In the new coordinate system, the first coordinate is the first principal component, i.e., the eigenvector associated with the maximum eigenvalue. The eigenvalue indicates the importance of the eigenvector to the variance of the data. For given reference image and uncorrected input image, the PCA-based image restoration takes the following steps^[19]:

Step 1) Form an ensemble from the original image: Reshape each $N \times N$ image into a column vector \mathbf{I}_i . The ensemble is denoted as $\{\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3\}$, which is a set of three observations corresponding to three color channels. I is a column vector with N^2 scalar random variable components. Then the expectation vector $\boldsymbol{\psi}_i$ is subtracted from each \mathbf{I}_i to get $\boldsymbol{\Phi}_i$. Put $\{\boldsymbol{\Phi}_i\}$ together to form a matrix $\mathbf{A}=[\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \boldsymbol{\Phi}_3]$. A is $N^2 \times 3$.

Step 2) Construct the covariance matrix: The covariance matrix of **I** is defined as $E((\mathbf{I}-E(\mathbf{I}))(\mathbf{I}-E(\mathbf{I}))^T)$, which is estimated by the empirical covariance matrix

$$\mathbf{C} = \frac{1}{3} \mathbf{A} \mathbf{A}^{T} \tag{3}$$

Here C is $N^{2} \times N^{2}$. If the image is 256×256, then C is 65536×65536. It is computationally intractable to determine the eigenvectors and eigenvalues of a matrix of such a size. However, it has been found that the computation can be dramatically reduced by first computing the eigendecomposition of $\mathbf{A}^{T}\mathbf{A}$, which is 3×3.

Step 3) *Calculate the eigenvector*: Let μ_i and v_i denote the *i*-th eigenvector and eigenvalue of the matrix $\mathbf{A}^T \mathbf{A}$, i.e.,

$$\mathbf{A}^{T}\mathbf{A}\boldsymbol{\mu}_{i}=\boldsymbol{\mu}_{i}\boldsymbol{v}_{i} \tag{4}$$

Premultiplying both sides by A, we have

$$\mathbf{A}\mathbf{A}^{T}\mathbf{A}\boldsymbol{\mu}_{i} = \mathbf{A}\boldsymbol{\mu}_{i}\boldsymbol{v}_{i} \tag{5}$$

From which we can see that $\mathbf{A}\mu_i$ is an eigenvector of $\mathbf{A}\mathbf{A}^T$ and thus of **C**. The scalar of 1/3 does not change the eigenvector.

Step 4) Compute the estimated image: The eigenvectors are often sorted by their eigenvalues in decreasing order. Supposed $\hat{\mathbf{f}} = \{\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \hat{\mathbf{f}}_3\}$, the estimation is conducted by

$$\mathbf{f}_i = \mathbf{A}\boldsymbol{u}_i + \boldsymbol{\Psi}_i \tag{6}$$

here $\mathbf{A}\mu_i$ can be interpreted as the projection of \mathbf{A} onto the *i*-th eigenvector of $\mathbf{A}^T \mathbf{A}$, and $\hat{\mathbf{f}}$ provides an estimate of the ideal image. $\mathbf{A}\mu_i$ can be interpreted as the high-frequency component, and the low frequency component is $\boldsymbol{\psi}$. μ_i can be viewed as a filter to boost the high-frequency content.

Step 5) *Restore the real image:* Suppose matrix **B** consists of eigenvector of $\mathbf{A}^T \mathbf{A}$ as column vector, since the covariance matrix of $\hat{\mathbf{f}}$ is diagonalized

whose elements are the eigenvalue of $\mathbf{A}^T \mathbf{A}$, reshaped the estimation image in the 2D matrix to get the real restored image as

$$\mathbf{f}_{real} = \mathbf{A}_{inp} \left[\mathbf{B}_{inp} \mathbf{S} (\mathbf{B}_{ref})^{-1} \right] + \boldsymbol{\Psi}_{ref}$$
(7)

Here **S** is a diagonal matrix with elements as $s_{ii} = \sqrt{v_{i ref} / v_{i inp}}$.

For more intuitive interpretation of the above result, suppose $\mathbf{I}_{inp} = [I_1^{inp}, I_2^{inp}, I_3^{inp}]^T$, $\mathbf{I}_{ref} = [I_1^{ref}, I_2^{ref}, I_3^{ref}]^T$, ψ_{inp} and ψ_{ref} denote expectation vectors, the real color vector \mathbf{I}_{real} is expressed as the deconvolution process of \mathbf{I}_{inp} .

$$\mathbf{I}_{real} = \left[\left((\mathbf{B}_{ref})^T \right)^{-1} \mathbf{S}^T (\mathbf{B}_{inp})^T \right] (\mathbf{I}_{inp} - \boldsymbol{\Psi}_{inp}) + \boldsymbol{\Psi}_{ref} \quad (8)$$

With correction matrix as $\mathbf{M} = ((\mathbf{B}_{ref})^T)^{-1} \mathbf{S}^T (\mathbf{B}_{inp})^T$.

3 Color Correction for Multi-view Image and Video

The block diagram of the proposed color correction method is given in Fig.1. First, disparity estimation with mean-removed sum of absolute differences (MRSAD) is used to separate the foreground and background from scene. Then dominant colors are extracted from background. And then PCA-based color restoration is used to achieve color correction for input view. Finally, the color correction is extended to multi-view video with keyframe and video tracking techniques.



Fig.1: Block diagram of the proposed color correction method

3.1 Background segmentation

In multi-view imaging, the object depth can be estimated from the disparity between two cameras. Thus disparity vectors provide additional clues for

depth estimation of scene. By defining a threshold T_1 , a block belongs to background regions, if horizontal and vertical disparities satisfy $(d_x^2+d_y^2)^{1/2} < T_1$. Otherwise, the block belongs to foreground regions. Thus coarse results of foreground-background separation are obtained. While some blocks in the background or foreground may be isolated because of occlusion, exposure or other reasons. In order to obtain continuous background contour, we propose a smoothing mechanism with the correlation of adjacent blocks. If a current block belongs to foreground or background, for the adjacent left, top, right and bottom blocks, if at least three blocks belong to a reverse classification, we determine that the current block also belongs the reverse classification.

The conventional matching metric for disparity estimation is SAD calculation of $S \times T$ blocks. In order to compensate the illumination change, a mean-removed sum of absolute differences (MRSAD) metric is defined. MRSAD produces block correspondence with the best matched patterns after mean removal. Therefore, disparity vector for each candidate block can be well preserved even with serious cross-view illumination mismatch. And MRSAD is calculated by

$$MRSAD(x, y) = \sum_{i=m}^{m+S-1} \sum_{j=n}^{n+T-1} |(S(i, j) - \mu_s) - (R(i+x, j+y) - \mu_r)|$$
(9)

where S(i,j) and R(i,j) are pixel values for source image and reference image with spatial coordinates (i,j), respectively. μ_s and μ_r are the average values for all the pixels in the source block and the reference block, respectively.

3.2 Dominant color extraction

Although the background regions are extracted from the image, mismatching blocks are inevitable during disparity matching, especially in smooth region. Therefore, dominant color is extracted from the background so as to reduce the influence of mismatching. To extract dominant color, image segmentation which is regard as 'color naming' in this paper is performed. First cluster the CIELAB space into 11 basic color categories (BCCs)^[20] by comparing Euclidean distance using Eq.(10), and each pixel completely belongs to one of the 11 BCCs and the corresponding probability P_{xy}^{i} belonging to the *i*-th BCCs is 1. The Euclidean distance *D* is defined as

$$D = \sqrt{(L(x,y) - \overline{L}_i)^2 + (a(x,y) - \overline{a}_i)^2 + (b(x,y) - \overline{b}_i)^2}$$
(10)

where L(x, y), a(x, y) and b(x, y) are values of the three color channels in CIELAB space, and \overline{L}_i , \overline{a}_i and \overline{b}_i are color values for the *i*-th BCC, and *D* is the Euclidean distance.

Then combining with the spatial correlation, the probabilities P_{xy}^{i} belonging to the *i*-th BCC is refined as

$$P_{xy}^{i} = \frac{\sum_{(x',y')\in N} P_{x'y'}^{i} w((pix(x,y), pix(x',y')))}{\sum_{(x',y')\in N} w((pix(x,y), pix(x',y')))}$$
(11)

Here pix(x,y) denote the label of pixel in (x,y). Dist(pix(x,y), pix(x',y')) is Euclidean distance between pixel (x,y) and (x',y'), N is the set of 8neightbour pixels in (x,y), and w is a weighting function defined as $w = 1 - \frac{1}{1 + e^{-0.5(Dist(pix(x,y),pix(x',y')))-T_2)}}$, that goes to 1 as the color difference between pixel (x,y) and (x',y') becomes small, and goes to 0 if becomes large. T_1 is a threshold value.

Each pixel (x,y) is described by one of the 11 BCCs and the corresponding probabilities P_{xy}^{i} belonging to the *i*-th BCC. The percentage that the *i*th BCC occupying in the whole 11 BCCs can be described by

$$p_i = \sum_{x,y} P_{xy}^i / \left(\sum_i \sum_{x,y} P_{xy}^i \right)$$
(12)

Then $\{p_i\}$ is sorted in descending order. If the accumulative percentage $\left(\sum_{i=1}^{M} p_i\right) < T_3$, the colors with highest p_i are regarded as dominant colors. Here, *M* is the dominant color number, T_3 is a threshold controlling dominant color number and $T_3 = 0.9$ in this paper.

The *i*-th dominant color mean is

$$\mu_i = \sum_{x,y} P_{xy}^i \cdot I_i(x,y) / \left(\sum_{x,y} P_{xy}^i\right)$$
(13)

Then the dominant colors is represented as

$$F = \left\{c_i, p_i, \mu_i, M\right\}$$
(14)

where c_i is the *i*-th dominant color, p_i represents its percentage value, μ_i is its color mean.

Finally, assume a set of homogeneous dominant colors $r_{ref}^{1}, \ldots, r_{ref}^{M}$ in the input image and a set of homogeneous dominant colors $r_{inp}^{1}, \ldots, r_{inp}^{M}$ in the reference image are given, then dominant colors are incorporated to blend the results

$$\mathbf{x}_{cor} = \sum_{i=1}^{M} p_i [\mathbf{M}_i (\mathbf{x}_{inp} - \mathbf{\Psi}_{inp}^i) + \mathbf{\Psi}_{ref}^i] / \sum_{i=1}^{M} p_i \quad (15)$$

3.3 Extended to multi-view video

Basically, digital video is a sequence of still images, displayed at a constant frame rate. The simplest adaptation of multi-view image color correction algorithm to multi-view video, is to perform the same algorithm frame by frame. However, the calculation complexity is usually intolerable if any temporal dependencies between adjacent frames are not taken into account. To speed up color correction for multi-view video, we present the concept of keyframe of multi-view video.

Different with the keyframe concept in computer animation^[21], the keyframe in our paper corresponds to a triggering which hints a new color correction should be performed. The span between two keyframes can be fixed, similar to the size of group of picture (GOP). If the span between two keyframes is not fixed, the keyframe triggering is related to inter-frame correlation. In this paper, we use correlation coefficient between adjacent frames as criterion.

We define correlation coefficient *R* between adjacent frames as

$$R = \frac{1}{n-1} \sum_{i=1}^{n} \left[\frac{v_i - \overline{v}}{S_v} \right] \left[\frac{k_i - \overline{k}}{S_k} \right]$$
(16)

where s_v and s_k are standard deviation of adjacent frames, \bar{v} and \bar{k} are the mean of them, respectively.

The correlation coefficient is relatively high over time as long as no sudden change occurred in the scene. Therefore, the video tracking technique in our paper is described as followed: at time t, if the correlation coefficient R is larger than a given threshold T_4 , consistent correction information at time t-1 is used, otherwise, the current frame is regarded as a new keyframe, and correction information is calculated again.

3.4 Objective performance evaluations

It is identified that defining a suitable image evaluation criterion is to be an urgent problem in the context of 3D video, especially for the color corrected image, where there is no original signal for evaluation methods. In many cases, only subjective criteria can be used for assessment purpose^[22]. In order to objectively evaluate the performance of color correction method, the color differences between reference image and color corrected image are calculated. Because CIELAB space is not uniform enough, the color differences are calculated with a quite complicated CIEDE2000 formula^[23]. The CIEDE2000 color difference ΔE_{00} formula is defined as

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right) \left(\frac{\Delta H'}{k_H S_H}\right) \quad (17)$$

where k_L , k_C , k_H , S_L , S_C , S_H , R_T are weighting factors. The detail definition of the parameters is described in literature [23].

4 Experimental Results and Analyses

The principal components are assumed to be the linear combinations of the original images. We implement the PCA-based deconvolution algorithm to validate partly restoration results using Eq.(6). Fig.2 shows two standard images including Pens, Bluegirl, and the corresponding restoration images only using the first, the second and the third principal components. Since the eigenvalue indicates the importance of the eigenvector to the variance of the data, the percentage variance can be used to a gauge to estimate the weight of each principal component. The percentage variances are 93.26%, 2.25%, 4.49% for Pens, and 95.62%, 1.15%, 3.23% for Bluegirl, respectively. It is evident that the restoration image only using the first component contains the most texture information, nevertheless the color information is lost, this explain why the second and third component should be considered simultaneously during the image restoration.



Fig.2: PCA-based partly restoration results. (a) original image; (b) the restoration image using the first principal component; (c) the restoration image using the second principal component; (d) the restoration image using the third principal component.

The proposed method will have well application in color style adjustment for nature images with different scenes. In this case, background segmentation can be omitted. Fig.3 shows the color correction results for two different images. Here, city of Francfort in thunderstorm shown in Fig.3(a) is as reference image, and Castle Sant'Angelo in Fig.3(b) as input image. Since it is difficult to establish accurate mapping by preferred region matching, the visual effects of the corrected image in Fig.3(c) is not satisfied. Fig.3(d) and (e) show the color naming result for reference image and input image, respectively. Fig.3(f) shows the result achieved by the proposed method. Since corresponding matching in the bridge and cloud regions is not accurate enough, the correction effect in Fig.3(c) is not evident for such regions, while our dominant color extraction method can eliminate this influence.



Fig.3: Color correction results for nature images. (a) reference image; (b) input image; (c) the corrected image with content-adaptive method in [15]; (d) color naming result for reference image; (e) color naming result for input image; (f) the corrected image with the proposed method.

We have also used multi-view sequences, 'golf2' and 'flamenco1', provided by KDDI Corp, and 'Uli', provided by HHI, are used as the test sets to evaluate the effect of the proposed method. The size of 'golf2' and 'flamenco1' is 320×240, and the images are taken by a horizontal parallel camera configuration with eight viewpoints and 200mm camera interval^[24]. The size of 'Uli' is 1024×768, with eight viewpoints and 20cm camera distance^[25]. Fig.4 shows eight original viewpoint images of the

'golf2', 'flamenco1' and 'Uli' multi-view video sequence. Clearly, the color consistency among these eight original viewpoint images is poor. Thus, the color correction is necessary if these multi-view images will be used to render new virtual arbitrary viewpoint image.



Fig.4: Eight original viewpoint images of 'golf2', 'flamenco1' and 'Uli'.

Figs.5-6(a) and (b) show the reference images and input images of 'golf2' and 'flamenco1', respectively. Figs.5-6(c) show the horizontal disparity images, which the block size is set as 8, the maximum horizontal and vertical disparities are set as 40 and 4, respectively. Figs.5-6(d) show the final background segmentation results with $T_1=8$ and T_1 =25 for 'golf2' and 'flamenco1', respectively. For 'golf2', it is an outdoor scene, uniform reference surface is satisfied for foreground and background, and thus the corrected images obtained with the global PCA correction method in [15], and only with foreground or background PCA-based image restoration, as shown in Figs.5(e)-(g), are almost consistent. For 'flamencol', the influence of nonuniform reference surface for foreground can not be neglected, the corrected images with respect to global PCA correction, preferred region matching in [15] and foreground PCA-based image restoration, as given in Figs.6(e)-(g) have evident inconsistent color appearance with reference image, mainly in the floor in Fig.6(e) and Fig.6(g), and in the fan in Fig.6(f). By contrast, the corrected image obtained with the proposed method, as shown in Fig.6(h), does not have such phenomena.





(i) (j)
Fig.5: Color correction results for 'golf2' test set. (a) reference image; (b) input image; (c) horizontal disparity image; (d) background segmentation results; (e) the corrected image with global PCA correction in [15]. (f) the corrected image with foreground; (g) the corrected image after 70 frames; (i) input image after 70 frames; (j) the corrected







Fig.6: Color correction results for 'flamenco1' test set. (a) reference image; (b) input image; (c) horizontal disparity image; (d) background segmentation results; (e) the corrected image with global PCA correction; (f) the corrected image with preferred region matching in [15]; (g) the corrected image with foreground; (h) the corrected image with background; (i) reference image after 70 frames; (j) input image after 70 frames; (k) the corrected image of Fig.6(j) using the proposed method.

Fig.7 shows the statistic results of correlation coefficients of the first and second viewpoints video of 'golf2' and 'flamenco1', and third and fourth viewpoints video of 'Uli'. It is evident that there are strong correlations between adjacent frames since most of the coefficients are larger than 0.9. Thus, we first select threshold $T_4 = 0.9$ in the experiments. Fig.5(h), Fig.6(i) and Fig.5(i), Fig.6(j) show the reference image and input image of 'golf2' and 'flamenco1' after 70 frames, respectively. Using the proposed color correction mechanism for multiview video, the visual effect of the corrected images in Fig.5(j) is satisfied, while the corrected image in Fig.6(k) has a little inconsistent color appearance

with reference image in Fig.6(i) due to the influence of illumination change. Therefore, larger threshold T_4 should be selected.



Fig.7. Statistic results of correlation coefficient.

For 'golf2' test sequence, there is almost no change in the natural light source during imaging, therefore, the video correction results are satisfied even after a large keyframe interval. While for 'flamenco1' test sequence, the light source is frequently changed during imaging, thus the dominant color information in background is also changed frequently. In such circumstance, the keyframe should be updated frequently, i.e., larger T_4 should be adopted. The illumination condition is comparatively fixed in another multi-view sequence 'Uli'. Figs.7(a)-(b) and Figs.7(c)-(d) show the reference image and input image at t=0 and t=200, respectively. Fig.7(e) shows the corrected image of Fig.7(b) at *t*=0 using the proposed color correction method. Fig.7(f) shows the corrected image of Fig.7(d) at the time t=200 using the consistent correction information with respect to the time t=0, the visual effect is satisfied even after 200 frames. It is concluded that the efficiency of the proposed video correction mechanism will be high for the fixed imaging conditions during video capturing, while if the imaging condition is changed frequently, the efficiency is comparatively low because of more keyframes adoption.





Fig.7. Multi-view video correction results for 'Uli' test set. (a) reference image at t=0; (b) input image at t=0; (c) reference image at t=200; (d) input image at t=200; (e) the corrected image of Fig.7(b) using the proposed image correction method at t=0; (f) the corrected image of Fig.7(d) using the proposed video correction mechanism at t=200.

Then we use the color difference defined in Eq.(15) to compare the color consistency of the original images without correction, the corrected images with respect to the content-adaptive method in [15] and the proposed method for successive 70 frames. Fig.8 shows the color difference comparison results. It is seen that for 'golf2', even though the subjective color difference is not significant, the objective evaluation indicates that the proposed method can achieve smallest color difference. For 'flamenco1', the proposed method also achieves smallest color difference, which means that it performs well than the content-adaptive method.



Fig.8 (a): Color difference comparison of 'golf2'



Fig.8 (b): Color difference comparison of 'flamencol'

5 Conclusion

We have presented an algorithm that achieves fast and effective color correction for multi-view video. Our method is especially applicable to multi-view images, and is more accurate than our previous method by improving the mapping between reference and input images. We obtain visually plausible results, even though we do not consider more accurate background segmentation method. We believe our algorithm makes significant process in comparison with previous method, but some future work remains. We would like to improve precision of the background segmentation. In this paper, we use disparity vectors as our segmentation criterion. However, if scene content in different view changes quickly, then the disparity estimation capability is reduced. To avoid the problem, we are trying to develop a fast background segmentation method to improve segmentation precision. We also would like to develop a more objective color consistency evaluation criterion.

Acknowledgment This work was supported by the Natural Science Foundation of China (grant 60672073), the Program for New Century Excellent Talents in University (NCET-06-0537), the Scientific Research Fund of Zhejiang Provincial Education Department (grant 20070962) and the Natural Science Foundation of Ningbo (grant 2008A610016). It was also sponsored by K.C.Wong Magna Fund in Ningbo University.

References:

[1] M. Tanimoto, Overview of free viewpoint television, *Signal Processing: Image*

Communication, vol. 21, no. 6, 2006, pp. 454-461.

- [2] H. Sadeghi, P. Moallem, S. A. Monadjemi, Feature based dense stereo matching using dynamic programming and color, *International Journal of Computational Intelligence*, vol. 4, no. 3, pp. 179-186.
- [3] M. Tanimoto, T. Fujii, K. Suzuki, Experiment of view synthesis using multi-view depth, *ISO/IEC JTC1/SC29/WG11*, Doc. M14889, Shenzhen, China. October 2007.
- [4] Z. Y. Zhang, A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, 2002, pp. 1330-1334.
- [5] N. Joshi, B. Wilburn, V. Vaish, et al, Automatic color calibration for large camera arrays, UCSD CSE Tech. Rep. CS2005-0821, May 2005.
- [6] M. Flierl, A. Mavlankar, B. Girod, Motion and disparity compensated coding for multiview video, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, Nov. 2007, pp. 1474-1484.
- [7] E. Reinhard, M. Ashikhmin, B. Gooch, et al, Color transfer between images, *IEEE Computer Graphics, and Application*, vol. 21, no. 4, 2001, pp. 34-41.
- [8] C. M. Wang, Y. H. Huang, M. H. Huang, An effective algorithm for image sequence color transfer, *Mathematical and Computer Modelling*, vol. 44, no. 7-8, 2006, pp. 608-627.
- [9] F. Pitie, A. Kokaram, R. Dahyot, Automated colour grading using colour distribution transfer, *Computer Vision and Image Understanding*, vol. 107, no. 1-2, 2007, pp. 123-137.
- [10] R. Kimmel, D. Shaked, M. Elad, A variational framework for retinex, *International Journal of Computer Vision*, vol. 52, no. 1, 2003, pp. 7-23.
- [11] D. Forsyth, A novel algorithm for color constancy, *International Journal of Computer Vision*, vol. 5, no. 1, 1990, pp. 5-36.
- [12] G. D. Finlayson, S. D. Hordley, P. M. Hubel, Color by correlation: A simple, unifying framework for color constancy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, 2001, pp. 1209-1221.
- [13] K. Yamamoto, T. Yendo, T. Fujii, et al, Colour correction for multiple-camera system by using correspondences, *The Journal of the Institute of*

Image Information and Television Engineers, vol. 61, no. 2, 2007, pp. 213-222.

- [14] K. Yamamoto, M. Kitahara, H. Kimata, et al, Multiview video coding using view interpolation and color correction, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, 2007, pp. 1436-1449.
- [15] F. Shao, G. Y. Jiang, M. Yu, et al, A contentadaptive multi-view video color correction algorithm, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 15-20, 2007, Honolulu, Hawaii, pp. 969-972.
- [16] J. C. Russ, The image processing handbook, CRC Press, 2006, pp.235-245.
- [17] E. Salari, S. Zhang, Neural network-based image restoration using scaled residual with space-variant regularization, *International Journal of Imaging System and Technology*, vol. 16, no. 6, 2003, pp. 247-253.
- [18] Y. K. Chee, Y. C. Soh, A robust kalman filter design for image restoration, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, Salt Lake City, Utah, pp. 1825-1828.
- [19] D. L. Li, R. M. Mersereau, S. Simske, Atmospheric turbulence-degraded image restoration using principal component analysis, *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, 2007, pp. 340-344.
- [20] A. Mojsilovic, A computational model for color naming and describing color composition of images, *IEEE Transactions on Image Processing*, vol. 14, no. 5, 2005, pp. 690-699.
- [21] A. Agarwala, A. Hertzmann, D. H. Salesin, et al, Keyframe-based tracking for rotoscoping and animation, *ACM Transactions on Graphics*, vol. 23, no. 3, 2004, pp. 584-591.
- [22] L. M. J. Meessters, W. A. IJsselsteijn, P. J. H. Seuntiens, A survey of perceptual evaluation and requirements of three-dimensional TV, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, 2004, pp. 381-391.
- [23] M. R. Luo, G. Cui, B. Rigg, The development of the CIE2000 colour-difference formula: CIEDE2000, *Color Research and Application*, vol. 26, no. 5, 2001, pp. 340-350.
- [24] R. Kawada, KDDI multi-view video sequences for MPEG 3DAV use, *ISO/IEC*

JTC1/SC29/WG11/MPEG2004, Doc. M10533, Munich, Germany, 2004.

[25] R. Tanger, T. Wiegand, Fraunhofer HHI test data sets for MVC, *ISO/IEC JTC1/SC29/WG11/MPEG2005*, Doc. M11894, Busan, Korea, 2005.