

Temporal Color Consistency-Based Video Reproduction for Dichromats

Chun-Rong Huang, *Member, IEEE*, Kuo-Chuan Chiu, and Chu-Song Chen, *Member, IEEE*

Abstract—In this paper, a video re-coloring algorithm for dichromats is presented. Different from image re-coloring schemes, reproducing a video for dichromats requires maintaining temporal color consistency between frames, i.e., the same color in different frames should be re-colored to the identical new color. To achieve this goal, we extract video key colors from shots after motion estimation at first. Based on the importance of video key colors, a process order is defined to perform efficient color remapping and solve the contrast maintaining problem. Then, the remapped frame pixel values are interpolated by the re-mapped video key colors with spatial-temporal constraints. Experimental results show that our method can increase the visibility for dichromats and guarantee temporal color consistency.

Index Terms—Colorblind video, video processing.

I. INTRODUCTION

COLOR plays a vital role in the delivery of the visual information. However, people who suffer from the color vision deficiency (CVD) are unable to distinguish certain color combinations. CVD people who lack one of the three basic color mechanisms are called *dichromats* [1]. According to the missing types of colors, they are classified as *protanopia*, *deutanopia*, and *tritanopia*, respectively. Besides *dichromats*, other kinds of CVD, such as *anomalous trichromacy* [2]–[4] and *monochromacy* [1], also face a similar problem. According to [5], there are about 200 million colorblind people in the world. It would be beneficial to transfer color content to provide them with richer and distinguishable color information.

To achieve this goal, researchers have focused on re-coloring images to accommodate the accessibility of CVD people. Existing approaches can be generally classified into two categories: manual tools and automatic techniques. Manual tools aim to assist designers to manually adjust ambiguous color combinations with user interfaces [6]–[8]. When an image contains gradual

color changes or thousands of colors, these methods are troublesome in use. Thus, automatic procedures are proposed to overcome above problems. In general, the re-coloring problem is modeled as an optimization problem which minimizes the differences between colors for normal people and remapped colors for CVD people. Wakita and Shimamura [9] defined three objective constraints to describe common color effects and used the simulated annealing algorithm to find re-color results. However, the algorithm cannot always ensure to find global optimum solutions and the computational complexity is very expensive. Jefferson and Harvey [10] evaluated the color visibility by WWW consortium criteria to preserve details for dichromats. Too many parameters in their objective function tend to increase the difficulty of efficiently solving the optimization problem.

Recently, researchers tried to find a more effective and efficient way to solve the optimization problem. Re-coloring images for CVD viewers can be treated as a problem of preserving visual details in the reduced color space. By assuming the reduced space as a flat plane, Rasche *et al.* [11], [12] proposed an effective color to gray technique to reproduce images for dichromats. Kuhn *et al.* [13] proposed a mass-spring system to preserve the naturalness of the original colors and color contrast after re-coloring. To efficiently model key colors, a distribution concept is proposed in [14]. The center of each distribution group is treated as a key color. The remapping function is derived from these key colors and forms a multi-variant nonlinear problem which is hard to be efficiently solved. These re-coloring methods can be further used to measure the image accessibility when searching images for CVD people [15] or automatically point out image regions which are hardly recognized by colorblind viewers in a manually designed image [16].

In this paper, we focus on how to re-color a video for dichromats instead of a single image. A simple idea is to re-color every frame individually by using image-based methods mentioned above. However, the same color appearing in different frames cannot be guaranteed to be remapped to the identical new color, because the temporal relationship between frames is not considered. The colors of an object may become different in adjacent frames. Thus, different from image re-coloring approaches, temporal color consistency (TCC) should be considered when reproducing a video for dichromats, i.e., the same color appearing in different frames is re-colored to an identical new color.

To the best of our knowledge, there are only few works [17], [18] addressing on re-coloring videos for CVD people. In [17], Machado and Oliveira present an automatic image re-coloring technique with GPU to enhance color contrasts for dichromats. They also propose to preserve the temporal coherence for

Manuscript received August 09, 2010; revised December 24, 2010 and March 07, 2011; accepted March 09, 2011. Date of publication April 05, 2011; date of current version September 16, 2011. This work was supported by the National Science Council of Taiwan, R.O.C., under Grant NSC 100-2631-H-001-013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marcel Worring.

C.-R. Huang is with the Institute of Networking and Multimedia, and the Department of Computer Science and Engineering, National Chung Hsing University, Taichung, Taiwan (e-mail: crhuang@cs.nchu.edu.tw).

K.-C. Chiu is with the Institute of Information Science, Academia Sinica, Taipei, Taiwan (e-mail: kcchiu@iis.sinica.edu.tw).

C.-S. Chen is with the Research Center for Information Technology Innovation, and Institute of Information Science, Academia Sinica, Taipei, Taiwan (e-mail: song@iis.sinica.edu.tw).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2011.2135844

browsing an interactive 3-D graphic model and claim that their method is also suitable for video re-coloring. However, no general video re-coloring results are presented in their paper. To avoid the color inconsistency during an interactive visualization session, they use the re-color vector of the current frame for the next frame by assuming that the adjacent frames contain the same object. However, the adjacent frames of general videos may not contain the same objects due to object motions. Moreover, when shot changes occur, their assumption would be no longer satisfied. Their method might be used to re-color a video without shot changes and foreground object motions, but it could fail to make the colors consistent when a video contains multiple shots and significant object motions. Hence, their method can only achieve TCC at the shot level.

Liu *et al.* [18] divide a video into shots and enforce the same colors appearing in different frames to be identical new colors after remapping in each shot. The same colors appearing at adjacent shots are smoothly varied. Thus, they also achieve TCC at the shot level. Nevertheless, it is very common that the same colors appear in many non-adjacent shots. Hence, in their approach, the same colors in the video will have different colors in different shots, i.e., the TCC at the video level is not achieved. Moreover, they performed a rule-based local color rotation in the CIEL*a*b* color space to enhance accessibility for CVD people. Consider with the a*b* plane, they rotate colors in the left plane towards +b* and in the right plane towards -b* for the protanopia and deuteranopia. This straightforward strategy only works well when most colors in the image are red and green which is confused in the protanopia and deuteranopia vision. The remapping step is not suitable to produce discriminative colors for dichromats.

To achieve TCC at the video level, we propose a new automatic video reproduction algorithm for dichromats. The idea is to retrieve video key colors appearing in the video and remapping them at the same time to avoid color inconsistency. A video key color is a frequently visible color in the video. Retrieving video key colors is time-consuming, because a video volume usually contains millions of pixels. To increase the performance, we separate a video into shots at first. Since adjacent frames in a shot contain similar colors, the motion estimation method is used to efficiently reduce repetitive colors. Shot key colors, which are the colors frequently appearing in a shot, are obtained by clustering the residual colors after motion estimation. Then, shot key colors from different shots are clustered to obtain video key colors. For remapping video key colors, we define a processing order based on their visibility in the shots. According to the processing order, we iteratively remap these key colors to new colors so that they can be distinguished by CVD people. During each iteration, we formulate the re-coloring problem as a 1-D optimization problem, which can be solved efficiently. Based on the remapped video key colors, we remap the shot key colors. Finally, the remapped values of the pixels in frames are computed by the frame interpolation which is accelerated by taking the advantage of spatial and temporal locality properties. Different from [18], we enable the color mapping to be identical in the whole video and colors are remapped to keep the same contrast as normal people can see as possible. Besides, for computing the color mapping function of each shot, they only

use the first frame in the shot while we take into account whole shot frame sequences to get more accurate color information. The detailed method is described in Section II. Experimental results are presented in Section III to demonstrate the effectiveness of the proposed method. Section IV gives the conclusion.

II. PROPOSED METHOD

In the following, we will introduce how to efficiently find the video key colors and their remapped colors for CVD people.

A. Shot Detection

In practice, taking all pixel values of the video volume into account requires much computation time and memory space. In order to alleviate memory redundancy, a bottom-up approach is considered by taking the advantage of the video structure. We first divide a video into shots using [19] which applied the concept of invariant local descriptors, contrast context histogram (CCH) [20], to detect the continuity of frames. If the same objects or backgrounds appear in adjacent frames, we would probably consider that there is no shot change. CCH is used to identify whether two adjacent frames belong to the same shot. In addition, by counting the number of matched CCH features, we can minimize the influence of both objects and camera motions, which can be easily mis-classified as shot transitions and successfully detect both abrupt and gradual transitions. After the shot detection, the i th shot S_i in the video can be expressed as a set of frames as follows:

$$S_i = \{f_m, f_{m+1}, \dots, f_n\} \quad (1)$$

where m and n are, respectively, the indices of the first and last frames of S_i .

B. Shot Key Color Extraction

Once a video is separated to shots, we now focus on extracting key colors of each shot. In practice, the shot volume remains too huge to gather all pixel values for clustering shot key colors. Since our goal is to find all colors appearing in a shot and then choose the primary ones, identical pixel values between adjacent frames can be eliminated in advance. As a result, we impose the motion estimation approach to prune off the repetitive colors. Since frames in the same shot are quite similar, most image blocks in the t th frame f_t can be represented by those in the previous frames f_{t-1} . Thus, only a few of memory storage is required to store the residual image blocks after deleting repetitive image blocks.

There are some standard search ways for motion estimation in the video codec standards, such as fast full search [21] and hexagon search [22]. These search methods aim to get precise estimation of motion vectors to reduce the compression bit rate as much as possible. However, they are time-consuming. To simultaneously maintain the motion estimation performance and decrease the computation time, adaptive rood pattern search (ARPS) [23] is proposed. ARPS has also been shown that outperforms many well-known search methods. The main idea of ARPS is that the motions of neighbor macro blocks are coherent in general. Thus, the motion vector of the current block can be predicted by means of the neighbor blocks. In addition, four-armed rood pattern points with adaptive sizes have

to be checked. The search size is defined as the larger absolute value of the coordinate of the predicted motion vector. This rood pattern search leads to find high probability area of a good matching. A subsequent local search with small fixed pattern is then performed. The procedure is terminated until the best matching point is at the center of the search pattern. Moreover, zero-motion prejudgment is performed at the beginning to further reduce the computational cost of zero-motion blocks.

After motion estimation, the redundant image blocks in the current frame f_t are represented by image blocks in the previous frame f_{t-1} . The remaining blocks contain the residuals in f_t . By finding the union of the residuals of each frame, we have a new subspace of the shot volume S_i . Then, minimum variance quantization (MVQ) [24] is employed to further reduce the numbers of colors in the subspace. MVQ approximates frequently appearing colors with smaller color cubes and larger ones for infrequently appearing colors in the RGB space. Here, the number of colors to be quantized in f_t is related to the ratio of residual image blocks. Let the first frame f_m of S_i be quantized into 256 colors. The number of colors to be quantized in f_t is computed as follows:

$$Q_{f_t} = \left\lfloor 256 \times \frac{\# \text{ of residual blocks in } f_t}{\# \text{ of blocks in } f_t} \right\rfloor. \quad (2)$$

The total number Q_i of quantized colors in S_i can be computed as follows:

$$Q_i = 256 + \sum_{t=m+1}^n Q_{f_t}. \quad (3)$$

In this way, we can efficiently extract the colors appearing in S_i . Nevertheless, Q_i is usually more than 1000. Among these colors, we would like to further extract shot key colors by removing duplicate or similar colors. Clustering methods such as K -means [25] can be used to extract key colors from the quantized colors. How to decide the number of clusters (key colors) remains a problem. Affinity propagation (AP) [26] gives a better way to achieve the clustering. AP is an exemplar-based clustering algorithm that finds clusters by exchanging messages between pairs of data points. Messages include responsibility and availability which reflects how suitable a candidate exemplar point is to serve as the exemplar for another point and how appropriate a point is to choose another point as its exemplar, respectively. The affinity between two data points indicates whether they should be in the same cluster. Thus, it can choose the number of clusters automatically based on the similarity. The similarity is defined as Euclidean distances between any two quantized colors in the CIEL*a*b* color space which present the perceptual difference between them. The only one parameter, inference value, is the median of the similarity by following the suggestion in [26]. After AP, we can obtain the shot key colors of S_i .

C. Video Key Color Extraction

In a video, main characters will appear in many shots but they do not always appear in adjacent shots. If the same colors of a character are not consistent in different shots, CVD viewers will feel confused because the same character has distinct colors.

To solve this problem, we propose TCC-based video reproduction approach to ensure that video key colors appearing in different shots will be remapped to the identical new colors at the video level. Since we collect shot key colors from all shots in the video, we further group them via AP to find the common colors, which are denoted as video key colors. In this way, we can reduce the computation complexity of finding video key colors directly from the video volume.

During AP clustering, we also compute the significance of each video key color. The significance is the frequency of the video key color appearing in the shots. The higher significance means that the video key color is usually appearing in different shots. If we can remap the video key colors with higher significances in the early stage, they can be distinguished more easily by CVD viewers and the remapping procedure will be more flexible. For example, if there are two shots in a video, the key colors in the first shot are green and red and the second shot contains red and yellow. Suppose that green and yellow are remapped to blue and yellow for protanopia. Then red will fail to be re-colored under the color contrast constraint, because only yellow and blue can be distinguished for protanopia. On the other hand, if we remap red which has the higher significance at first, the awkward situation can be avoided. For this reason, video key colors are sorted in the descending order according to their significances in the video.

D. Video Key Color Reproduction

After obtaining video key colors and their significances, we now focus on remapping them to new colors so that CVD people can distinguish among them. The remapping criterion is to maintain the perceptual distances between each pair of video key colors at the normal vision space and at the CVD color space. Hence, video key colors are rearranged based on color contrast maintenance within the gamut of CVD. Unfortunately, this is a multi-variant nonlinear optimization problem [14] and is complex to solve.

To simplify the optimization problem, we will introduce the CVD color space at first and remap all video key colors one by one according to their significances. The remapped video key colors are denoted as the reference colors. The unmapped video key color is computed by following the color contrast constraints comparing with the reference colors to obtain the new color. In this procedure, we can successfully reduce the optimization problem to a series of 1-D optimization ones.

1) *CVD Color Space*: It is important to model the CVD color space for dichromatic people. Since the remapped colors are presented for CVD viewers, the remapping range of the color gamut is limited to the space of a specific type of dichromats. Although Brettel *et al.* [27] have proposed a successful approach to simulate the CVD space, their nonlinear mapping functions in the CIEL*a*b* color space are implicit functions which are difficult to be applied for the analytic purpose, such as constructing an energy function. In fact, the CVD color space is a subspace of the normal vision space. In [11] and [12], they found that each type of dichromatic color gamut is a low curvature surface in the normal vision space. Based on this observation, they approximated the surface by a flat plane and extended their grey scale preserving techniques to accomplish re-coloring

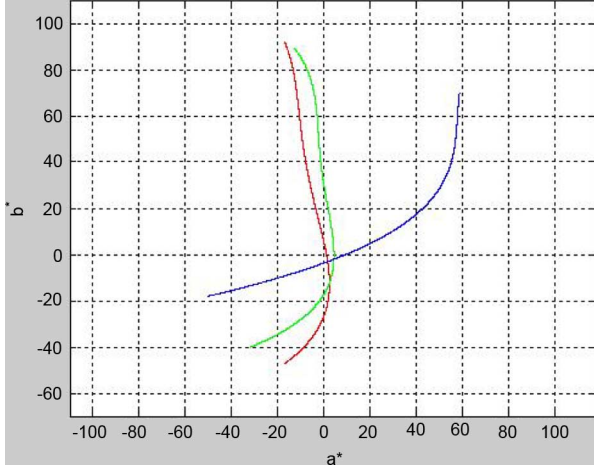


Fig. 1. Simulated gamut of 2-D collapses in CIEL * a * b* space perceived by each type of CVD viewers. Red, green, and blue lines are the simulation space for protanopia, deuteranopia, and tritanopia, respectively.

algorithms. In our approach, we propose a more precise way to fit each curvature space.

To explain it, we apply the regression analysis to find a polynomial function that approximates the specific CVD space based on the simulation function provided in [27]. We transfer all of the colors in the RGB space to the CVD color space. As shown in [11], maintaining the luminance of the original color in the RGB space and modifying only the a^*b^* components are sufficient to construct the perceived colors in the CVD space. Thus, only a^*b^* space is considered. The corresponding CVD space of the protanopia, the deuteranopia, and the tritanopia are represented by red, green, and blue as shown in Fig. 1, respectively. It can be easily found that the gamut of CVD spaces cannot be well fitted by straight lines which are proposed in [11] and [12]. Because the gamut of each dichromacy is a 1-D curve in the a^*b^* plane, we observed that the point in this curve is uniquely determined when b^* is given. This implies us to represent the curve as a function of b^* ; without loss of generality, we modeled it as a d th-order polynomial $p(b^*)$ and find the coefficients of $p(b^*)$ by fitting the curve $a^* = p(b^*)$ via standard regression analysis in a least-squared-error manner. In our experiments, there is almost no difference when $d \geq 4$. Therefore, $d = 4$ is selected for our models to reduce the computational complexity. The equations of each dichromatic type are shown as follows:

Protanopia:

$$\begin{aligned} a &= F_{\text{protanopia}}(b) \\ &= 1.15 - 1.82e^{-1}b - 4.61e^{-3}b^2 \\ &\quad + 1.20e^{-4}b^3 - 7.75e^{-7}b^4 \end{aligned} \quad (4)$$

where $-47 \leq b \leq 92$ and e is the scientific notation.

Deuteranopia:

$$\begin{aligned} a &= F_{\text{deuteranopia}}(b) \\ &= 4.45 - 5.35e^{-3}b - 1.05e^{-2}b^2 \\ &\quad + 2.39e^{-4}b^3 - 1.63e^{-6}b^4 \end{aligned} \quad (5)$$

where $-40 \leq b \leq 89$.

Tritanopia:

$$\begin{aligned} a &= F_{\text{tritanopia}}(b) \\ &= 9.54 + 2.42b - 4.34e^{-2}b^2 \\ &\quad + 3.09e^{-4}b^3 - 5.54e^{-7}b^4 \end{aligned} \quad (6)$$

where $-18 \leq b \leq 70$.

2) *Optimization*: The perceptual distance between any two colors in the CIEL * a * b* color space is measured by the Euclidean distance. The remapping objective criterion is to keep the same perceptual distance between each pair of key colors at the normal vision space and at the CVD color space. So the remapped error introduced by any pair of video key colors, C_i and C_j , for maintaining the color contrast is defined as follows:

$$E_{i,j} = (\|C_i - C_j\| - \|R_i - R_j\|)^2 \quad (7)$$

where R_i and R_j are the reference colors, i.e., the remapped colors of C_i and C_j , respectively. If there are N video key colors, the color contrast error of each video key color pair can be measured by (7). By accumulating the errors, the objective function is formed as follows:

$$\begin{aligned} E &= \sum_{i=1}^N \sum_{j=i+1}^{N-1} E_{i,j} \\ &= \sum_{i=1}^N \sum_{j=i+1}^{N-1} (\|C_i - C_j\| - \|R_i - R_j\|)^2. \end{aligned} \quad (8)$$

By minimizing (8), all pairs of R_i and R_j can be found. However, it is remarkably difficult to find the global optimal solution efficiently. Many previous works [10], [14] solved the optimization problem by the gradient descent method to avoid the enormous computing load.

Instead of finding all reference colors simultaneously, we propose an iterative procedure to calculate the remapping results one by one. Let all video key colors be stored in a process queue, Ω , according to their significances in a descending order. Following the order in Ω , video key colors are remapped to the CVD space sequentially by maintaining each pair of color contrast as shown in (9):

$$\min_{R_i} \sum_{R_j \in \Omega_{\text{mapped}}} (\|C_i - C_j\| - \|R_i - R_j\|)^2 \quad (9)$$

where C_i is the i th original video key color in Ω , Ω_{mapped} denotes that the set of key colors C_j have already been processed and their corresponding reference colors R_j have already been determined. The remapped color R_1 of the first video key color C_1 in Ω is the CVD simulation color of C_1 . The following corresponding reference color R_i of the video key color C_i in Ω is determined sequentially by (9).

There are still several constraints in the optimization problem. At first, we have to maintain the luminance consistency and so the L^* component of R_i is set as the same as that of C_i . Since R_i is the remapped color for CVD viewers, it should appear in the CVD space. Therefore, we can restrict the b^* component of R_i to be positioned along $a = F_{\text{type}}(b)$ formulated in (4)–(6),

where $\text{type} \in \{\text{protanopia}, \text{deutanopia}, \text{tritanopia}\}$. Equation (9) is then further expressed as follows:

$$\begin{aligned} & \min_{b(R_i)} \sum_{R_j \in \Omega_{\text{mapped}}} (\|C_i - C_j\| - \|R_i - R_j\|)^2, \\ & \text{subject to } L(R_i) = L(C_i), \text{ and} \\ & a(R_i) = F_{\text{type}}(b(R_i)) \end{aligned} \quad (10)$$

where $L(\cdot)$, $a(\cdot)$, and $b(\cdot)$ are the mapping functions of the luminance L^* , a^* component, and b^* component of the color in the CIEL a^*b^* space, respectively. In this procedure, the dimension of the optimization problem is reduced to one which is related to the b^* component of R_i in each iteration step. Hence, it can considerably alleviate the computational complexity since the optimization solution of the 1-D problem can be solved easily. Many approaches have been proposed to solve the 1-D optimization problem [28]. In our case, the objective function of the 1-D optimization problem is a fourth-order polynomial function. The shuffled complex evolution (SCE-UA) method [29] is adopted to solve the problem because it has good performance to solve irregular objective functions.

There is still another advantage of the proposed processing order. Re-coloring for dichromats could be considered as information reduction from a 3-D space to a lower one. By following the processing order, the formerly calculated reference colors R_i are easier to maintain the color contrast between the reference colors and the original key colors because there are still few elements in Ω_{mapped} . Video key colors with lower significances have lower accessibilities for CVD viewers, because the color contrast has to be maintained with respect to more reference colors. These colors with lower significances just appear in a short period in the video; thus, the artifacts could be ignored. In this way, we can remap each video key color C_i to its new color R_i in the CVD space. The colors in the shots which are the same as the video key colors will be remapped to the identical new key colors to achieve TCC in the video level. Since the same colors from different shots are remapped to the identical colors, we can ensure TCC in the video level even across different shots.

E. Frame Color Interpolation

From previous sections, a bottom-up approach is adopted to cluster frame pixel colors to shot key colors and then cluster shot key colors to video key colors. Now, we reversely compute the remapped shot key colors from remapped video key colors and then interpolate frame pixel colors from remapped shot key colors. Let SC_k denote a shot key color that belongs to the cluster of the video key color, C_k . The corresponding shot reference color SR_k is then derived by the projection of the perceptual difference [13] as follows:

$$SR_k = R_k + (L(d), r_k a(d), r_k b(d)) \quad (11)$$

where $d = SC_k - C_k$, R_k is the reference color associated to C_k , and r_k is the ratio of the distance-weighted interpolation provide by [13] and [30]:

$$r_k = \frac{\sum_{j=1}^N \left(W_{kj} \frac{\|R_k - R_j\|}{\|C_k - C_j\| + \varepsilon} \right)}{\sum_{j=1}^N W_{kj}} \quad (12)$$

and

$$W_{kj} = \frac{1}{\|C_k - C_j\|^2 + \varepsilon} \quad (13)$$

where ε is a predefined small value to avoid zero-division and N is the number of total video key colors.

Basically, frame pixel colors can be re-colored according to shot reference colors by the same method as (11), but it is not efficient [13]. We utilize the spatial and temporal locality to accelerate the interpolation step. The spatial locality is that the re-colored pixel value at the position p is possible to be equal to its neighbor pixels at $NE(p)$ whose re-colored pixel values are already estimated. The re-colored value of p can be copied from $NE(p)$ if their original values are the same. Let FC_t and FR_t represent the t th original and re-colored frames, respectively. The original pixel value and the re-colored of p is denoted as $FC_t(p)$ and $FR_t(p)$, respectively. If $FC_t(p) = FC_t(NE(p))$, $FR_t(p) = FR_t(NE(p))$. The temporal locality considers the color similarity between adjacent frames. According to the results of motion estimation, we can reversely compute the correspondent pixel $IMV(p)$ of p in the previous frame. Again, if $FC_t(p) = FC_{t-1}(IMV(p))$, $FR_t(p) = FR_{t-1}(IMV(p))$. $IMV(p)$ means the inverse motion vector from p to its corresponding point in f_{t-1} . Since adjacent frames in the same shot are similar in general, this step is able to greatly reduce the computing load. After all, the remaining pixel values are re-colored by (11). The interpolation approach of p in the frame level is summarized as follows:

$$FR_t(p) = \begin{cases} FR_t(NE(p)), & \text{if } FC_t(p) = FC_t(NE(p)) \\ FR_{t-1}(IMV(p)), & \text{if } FC_t(p) = FC_{t-1}(IMV(p)) \\ SR_h(t) + (L(d), r_k a(d), r_k b(d)), & \text{otherwise} \end{cases} \quad (14)$$

where $SR_h(t)$ refers to the h th shot reference color. Note that $d = CF_t(p) - SC_k$ and r_k is also derived by (12) by replacing the parameters in the video level to the shot level.

III. EXPERIMENTS

A. Re-Colored Results

We applied the proposed algorithm to three video sequences collected from the Internet.¹ Fig. 2 shows the experimental results of the first sequence. As shown in Fig. 2(b), both of the red and green men become yellow men for CVD viewers with the protanopia. To avoid the visual confusion, a simple idea is

¹Demo videos are available at <http://imp.iis.sinica.edu.tw/ivclab/research/CVDVideoReproduction/index.htm>.

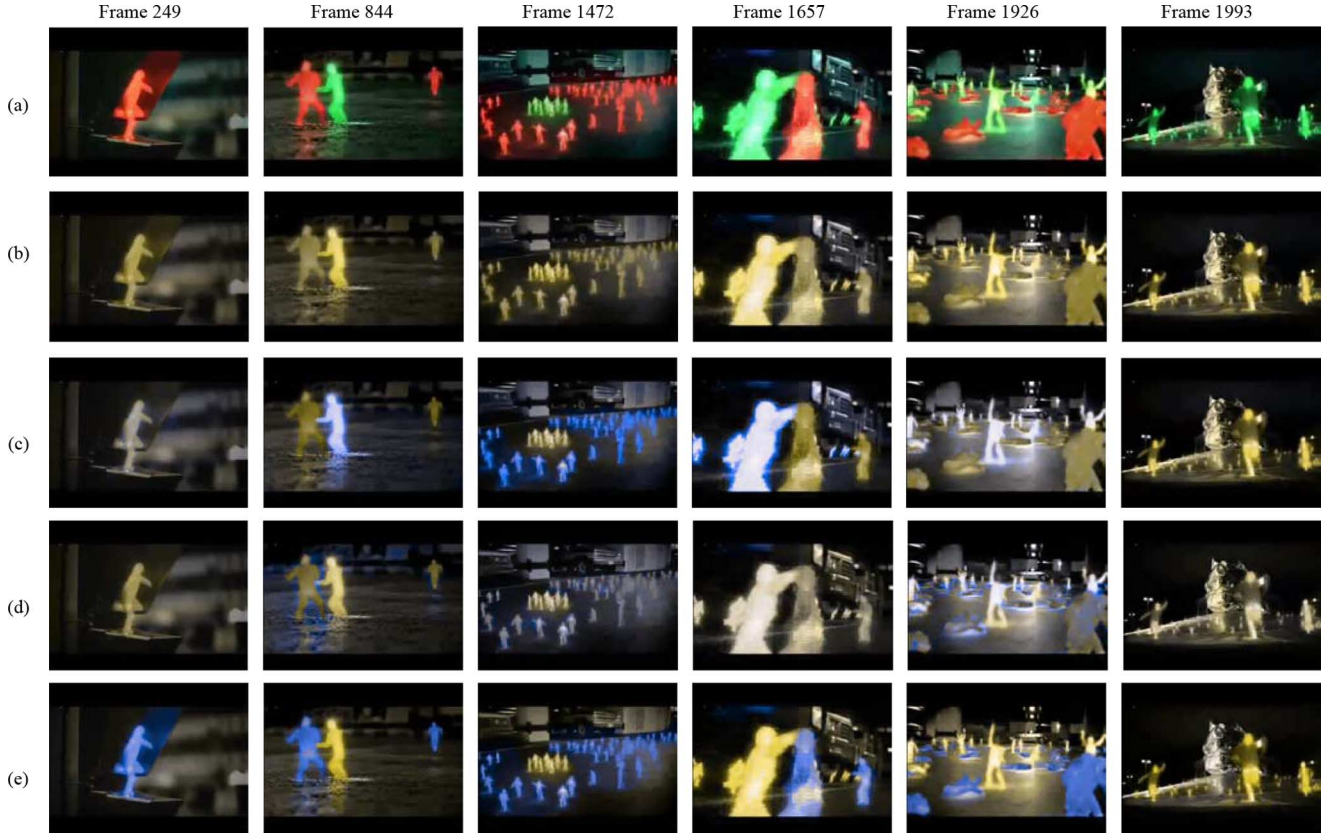


Fig. 2. Experimental results of the signal light video. The sampled images obtained from (a) the original sequence, (b) the simulation sequence of the protanopia, (c) the image-based re-coloring method, (d) Liu *et al.*'s method, and (e) the proposed method.

to re-color every frame by the image-based method [14] individually without considering TCC. As shown in Fig. 2(c), in the 249th, 844th, 1657th, and 1926th frames, the green men are remapped to the blue men and the red men are remapped to the yellow men. However, in 1472nd and 1993rd frame, the green men are remapped to the yellow men and the red men are remapped to the blue men. As a result, CVD viewers will have difficulty to differentiate between the green men and red men. Moreover, when playing the remapped video, the “color jitter” phenomenon, i.e., the colors of the same objects change frequently between adjacent frames, will make viewers uncomfortable. As shown in Fig. 2(d), Liu *et al.*'s [18] method failed to re-color two fighters in the 844th and 1657th frames. Although they consider enforcing the same colors appearing in a shot to be identical new colors, they do not consider that the same colors appearing in different shots should be also remapped to the same color. For example, the red man is re-colored to yellow in the 246th frame but is re-colored to gray in the 1472nd frame and is re-colored to yellow again in the 1657th frame. Such inconsistency of the same object between different shots will cause the confusion of the CVD viewers. The results of the image-based and Liu *et al.*'s methods [18] indicate the importance of TCC in the video level. As shown in Fig. 2(e), our method can consistently remap all the green men to the yellow men and all the red men to the blue men in the whole video. Thus, the CVD viewers can explicitly understand that the green and red men belong to two different colors.

In Fig. 3(b), the snooker pool and the balls are undistinguishable for people with protanopia. Although the image-based

method can remap the color of the ball to blue which makes it distinguishable from the snooker pool as shown in the 600th frame of Fig. 3(c), it also remaps the color of the snooker pool to blue in the 319th frame. Such an inconsistent mapping will make the viewers confused. As shown in Fig. 3(d), Liu *et al.*'s method not only fails to remap the color of the ball in the 246th frame but also the color of the shirt of the sport man in the 1170th frame. As can be seen, our method can achieve TCC better in the video level.

In Fig. 4(b), the strawberries and their leaves are the same colors under the CVD vision. As shown in Fig. 4(c), the remapped colors of the strawberries in different frames change frequently. Although Liu *et al.*'s method can avoid the “color jitter” in a shot, their method fail to achieve TCC in the whole video. Thus, the remapped colors between different shots will gradually change. For example, the remapped colors of strawberries in the 343rd, 552nd, and 1065th frames are different as shown in Fig. 4(d). Unlike the image-based and Liu *et al.*'s approaches, our method can consistently remap the colors of the strawberries and the leaves to blue and yellow shown in Fig. 4(e), respectively. Similar results are shown in Fig. 5(e); the poinsettia flowers and their leaves are distinguishable and their colors are consistent in all frames in the video after remapping.

B. Numerical Analysis

To quantify the effectiveness of our work, we compute the contrast reduction score, the color reduction score, and the inter-frame color change rate (ICCR) score suggested in [31], respectively. Here, the contrast reduction score is the ratio of the

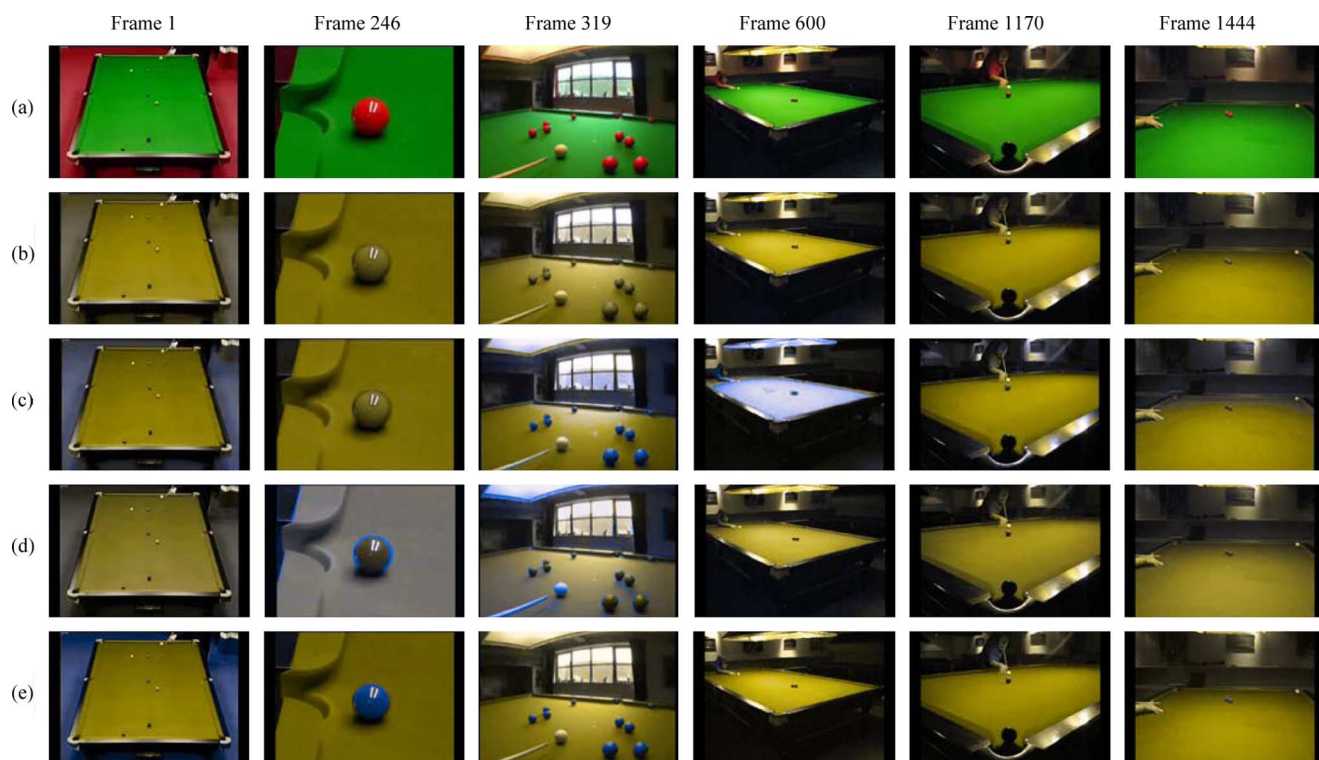


Fig. 3. Experimental results of snooker video. The sampled images obtained from (a) the original sequence, (b) the simulation sequence of the protanopia, (c) the image-based re-coloring method, (d) Liu *et al.*'s method, and (e) the proposed method.

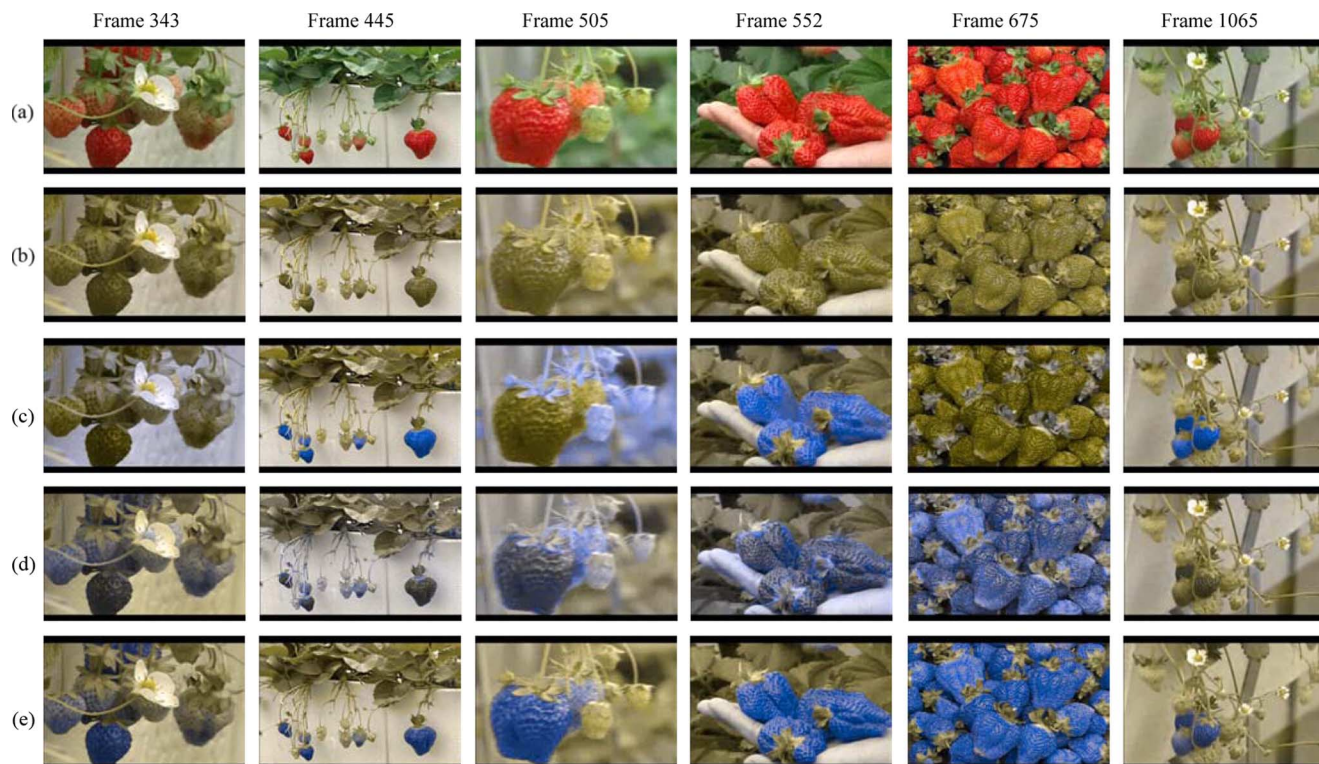


Fig. 4. Experimental results of strawberry video. The sampled images obtained from (a) the original sequence, (b) the simulation sequence of the protanopia, (c) the image-based re-coloring method, (d) Liu *et al.*'s method, and (e) the proposed method.

mean contrasts of the original and remapping versions to describe quantitatively the difference in contrast. Color contrast is computed by the perceptual distances between the pairs of a point and its neighbor point at one pixel apart. If the color contrast after remapping are close the original one, it implies

the remapping method can maintain the color contrast. The distinguishable colors mean the colors perceived by viewers. The color reduction score is the ratio of the mean number of distinguishable colors of the original and the remapping versions of videos. To evaluate the color changes between adjacent frames,

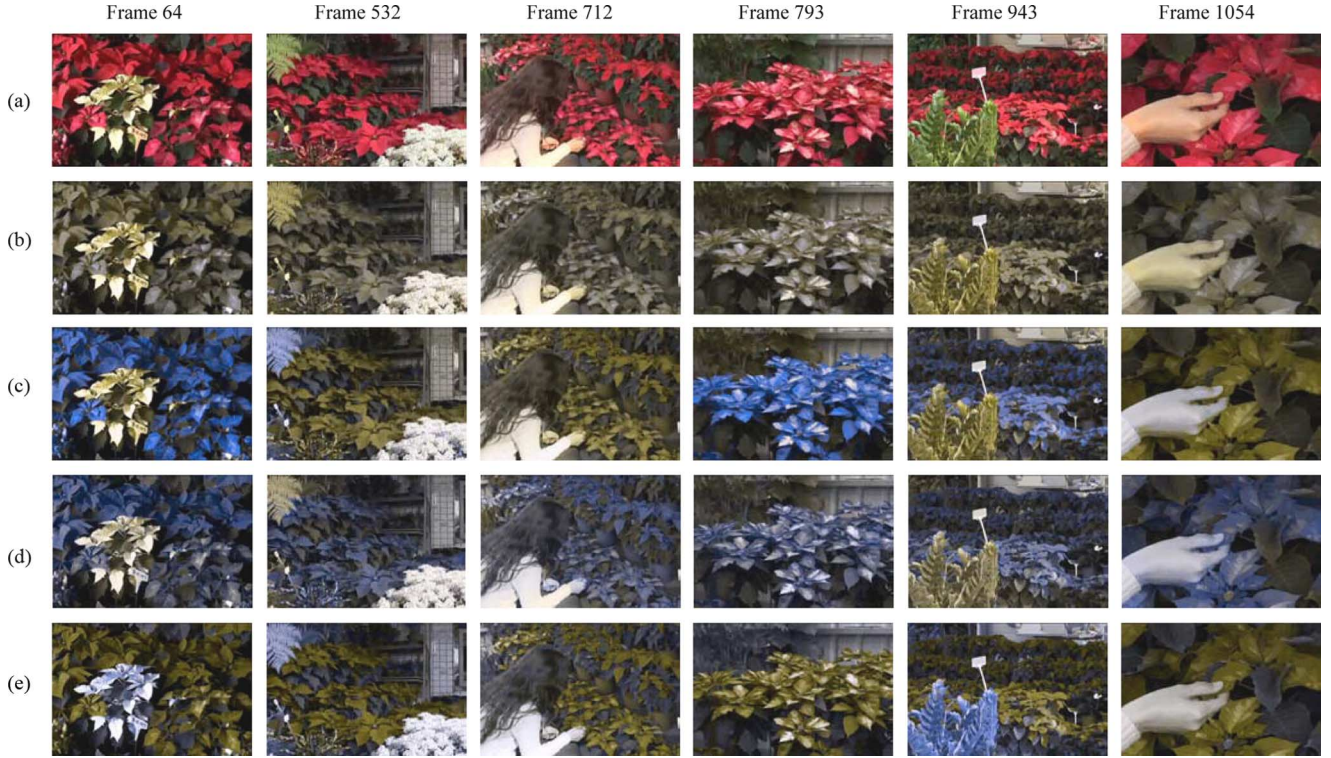


Fig. 5. Experimental results of the poinsettia video. The sampled images obtained from (a) the original sequence, (b) the simulation sequence of the protanopia, (c) the image-based re-coloring method, (d) Liu *et al.*'s method, and (e) the proposed method.

we employ the ICCR score. Let D_t be the number of distinguishable colors of the frame f_t . The ICCR score between f_t and f_{t+1} is defined as follows:

$$ICCR = \frac{D_{\text{gone}} + D_{\text{come}}}{D_t \cup D_{t+1}} \times 100\% \quad (15)$$

where $D_{\text{gone}} = D_t - D_t \cap D_{t+1}$ and $D_{\text{come}} = D_{t+1} - D_t \cap D_{t+1}$.

The statistical values of three quantified evaluations are summarized in Tables I–III, respectively. Every video is measured under the original video, the CVD simulation video of the protanopia, the video re-colored by the image-based method [14], the video re-colored by the Liu *et al.*'s method [18], and the video re-colored by the proposed method. In Table I, the mean of the color contrast is the average of the color contrast of all frames of each video. While the contrast reduction score is closer to one, the reproduced video is more similar to the original video. As shown in Table I, the proposed method has the most similar color contrast compared with the original video. Thus, the contrast reduction scores of the proposed method are close to one for all the videos. Such results reveal that our method can maintain the color contrast after reproducing the videos. The means of distinguishable colors appearing in the videos are shown in Table II. It can be found that the means of distinguishable colors in the CVD videos are the smallest because the CVD viewers cannot distinguish these colors. However, after remapping, the means of distinguishable colors obviously increase which indicates more colors can be recognized by the CVD viewers. If the color reduction score is closer to one, the reproduced video includes more similar

amount of distinguishable colors as the normal video. The proposed method performed the best results in most cases. Table III presents ICCR results. If the color changes between adjacent frames of the reproduced videos can be maintained as those of the original video, the reproduced videos have similar accessibility as the original video. Again, our method performed the best among the compared methods. Overall, our re-coloring approach gives a good way to enhance the video accessibility for dichromats.

The proposed algorithm was implemented by using Matlab on a PC with Intel four cores 2.66-G CPU. The computation time of each step is summarized in Table IV. As we can see, a larger frame size requires more computation time in general. Key color remapping time is independent of the frame size which is proportional to the number of video key colors.

C. Subjective Study

During the experiment, 11 subjects including one protanopia subject reviewed the original videos at first. Then, they gave their assessments to every re-colored video as shown from Figs. 2–5. The subjects, who do not know the proposed method, are able to compare the results of different methods appearing in randomized order for assessments, and thus, they do not know which method is applied to a particular video in advance. Three kinds of assessments, contrast, naturalness, and performance were made. Contrast indicates the distinguish ability of colors of the re-colored video. Naturalness shows how the re-colored video can preserve the colors which are discriminative in the original video for the CVD people. Performance is the overall visual rating of the re-colored video. In order to

TABLE I
COLOR CONTRAST SCORE

Video	Version	Mean	Contrast reduction score
Signal light	Original	11.17	1.00
	Protanopia	9.44	1.18
	Image-based	10.12	1.10
	Liu	9.21	1.21
	Proposed	10.41	1.07
Snooker	Original	7.39	1.00
	Protanopia	6.64	1.11
	Image-based	11.99	0.62
	Liu	5.99	1.23
	Proposed	6.77	1.09
Strawberries	Original	9.11	1.00
	Protanopia	8.33	1.09
	Image-based	7.99	1.14
	Liu	7.88	1.16
	Proposed	8.57	1.06
Poinsettia	Original	11.95	1.00
	Protanopia	8.66	1.38
	Image-based	9.92	1.20
	Liu	9.42	1.27
	Proposed	11.82	1.01

TABLE II
COLOR REDUCTION SCORE

Video	Version	Mean	Color reduction score
Signal light	Original	6764.3	1.00
	Protanopia	2429.7	2.78
	Image-based	2446.9	2.76
	Liu	2850.4	2.37
	Proposed	3627.3	1.86
Snooker	Original	12490.8	1.00
	Protanopia	4657.4	2.68
	Image-based	3396.1	3.68
	Liu	4625.4	2.70
	Proposed	4315.5	2.89
Strawberries	Original	15050.1	1.00
	Protanopia	3924.2	3.84
	Image-based	4165.8	3.61
	Liu	4248.5	3.54
	Proposed	5135.8	2.93
Poinsettia	Original	13104.8	1.00
	Protanopia	2843.7	4.61
	Image-based	4434.3	2.96
	Liu	3994.7	3.28
	Proposed	4043.6	3.24

TABLE III
ICCR SCORE

Video	Version	ICCR
Signal light	Original	53.94
	Protanopia	36.66
	Image-based	56.63
	Liu	41.95
	Proposed	48.96
Snooker	Original	24.70
	Protanopia	16.03
	Image-based	44.83
	Liu	17.29
	Proposed	19.34
Strawberries	Original	19.36
	Protanopia	14.00
	Image-based	28.62
	Liu	14.58
	Proposed	18.97
Poinsettia	Original	14.62
	Protanopia	12.88
	Image-based	37.26
	Liu	15.85
	Proposed	14.01

TABLE IV
COMPUTATION TIME IN SECONDS

Video	Signal light	Snooker	Strawberries	Poinsettia
Frame size	320×240	720×480	720×480	720×480
Number of frames	2430	1444	1066	1217
Key color extraction time	1481.98	2036.10	2874.65	2511.11
Key color reproduction time	490.55	92.55	103.62	125.24
Frame interpolation	864.70	956.35	1284.72	1239.15
Average analysis time per frame	1.17	2.14	4.00	3.18

TABLE V
SUBJECTIVE RESULTS OF CONTRAST ASSESSMENT

Video	Image-based	Liu	Proposed
Signal light	1.73	1.82	3.36
Snooker	1.36	2.00	2.91
Strawberries	1.45	2.27	3.64
Poinsettia	2.36	2.54	3.18

obtain a quantitative assessment, each assessment is rated from 0 to 4, where 0 means the worst and 4 means the most satisfied. Then, an average score of each assessment is calculated from all of the test videos. The results of each assessment with respect to videos are shown in Tables V–VII. Generally, Liu *et al.*'s method and our method show better results compared with the image-based method because it cannot achieve TCC at shot and video levels. The naturalness assessment usually contains the relatively worse results compared with the contrast and performance assessments in Liu *et al.*'s and our methods. This is because a video contains richer colors than a single image, and thus, the re-colored video is relatively hard to preserve the colors which are discriminative originally. The proposed method owns the highest scores in all of the three

assessments. Moreover, to the protanopia subject, the average assessment scores of the four videos in contrast, naturalness, and performance with respect to our method are 3, 2, and 2.5, respectively, higher than 1.75, 1.5, and 1.75 of the Liu's method and 2.5, 0.5, and 1 of the image-based method. The promising results show that the proposed method can re-color videos with better visual quality.

D. Discussion

As shown in the experiments, our method can successfully achieve the TCC at the video level. However, there are some limitations. At first, our method only focuses on re-coloring videos for dichromats. Re-coloring videos for people with anomalous trichromacy [2]–[4] and monochromacy [1] are not considered.

TABLE VI
SUBJECTIVE RESULTS OF NATURALNESS ASSESSMENT

Video	Image-based	Liu	Proposed
Signal light	1.45	2.00	2.45
Snooker	1.55	1.91	2.45
Strawberries	1.91	1.82	2.36
Poinsettia	1.91	1.73	2.55

TABLE VII
SUBJECTIVE RESULTS OF PERFORMANCE ASSESSMENT

Video	Image-based	Liu	Proposed
Signal light	1.32	1.86	3.36
Snooker	1.18	1.91	2.91
Strawberries	1.32	2.14	3.09
Poinsettia	1.45	2.36	3.18

One of the possible solutions is to modify our polynomial equations for trichromacy people and then apply the proposed technique to re-color the videos. As for monochromacy people, once the reduced monochromacy space with respect to the original color space is derived, the videos can also be re-colored by the iterative procedure. Second, when a long video contains a huge amount of colors, the number of the clustered video key colors might increase from dozens to hundreds. Although these key colors can be mapped to their new colors according to the processing order, it is not guaranteed that the remapped new colors are distinguishable for dichromats. In other words, our method might fail to arrange appropriately the key colors for a long and very colorful video. One of the possible solutions is to reduce the quantization levels after motion estimation. As a result, the number of video key colors is reduced which implies a better remapping results for these key colors; however, since the collected key colors are reduced, the re-colored video will lose more of the colorful information. Thus, handling long and colorful videos for dichromats still remains an open problem. Finally, since our method needs to collect the video key colors from the whole video, it cannot be applied to live streaming.

IV. CONCLUSION

We have proposed a novel video reproduction approach for dichromats. Compared with image-based re-coloring and Liu *et al.*'s approaches, our method can ensure TCC in the whole reproduced video. Our method can make the objects of the same colors consistent for CVD viewers to avoid jittering of the remapped colors. Moreover, to enhance the accessibility for dichromats, we iteratively select video key colors according to their visibility in the video and formulate 1-D optimization problems to remap each video key color according to the model of the CVD gamut. By utilizing the hierarchical structure of videos and the motion estimation technique, we reduce the memory usage and speed up reproduction tasks. Experimental results and evaluation demonstrate the effectiveness of our method.

REFERENCES

- [1] R. W. Burnham, R. M. Hanes, and C. J. Bartleson, *Color: A Guide to Basic Facts and Concepts*. New York: Wiley, 1963.
- [2] S. Yang, Y. M. Ro, E. K. Wong, and J.-H. Lee, "Quantification and standardized description of color vision deficiency caused by anomalous trichromats-Part I: Simulation and measurement," *EURASIP J. Image Video Process.*, 2008.
- [3] S. Yang, Y. M. Ro, E. K. Wong, and J.-H. Lee, "Quantification and standardized description of color vision deficiency caused by anomalous trichromats-Part II: Modeling and color compensation," *EURASIP J. Image Video Process.*, 2008.
- [4] J. Nam, Y. M. Ro, Y. Huh, and M. Kim, "Visual content adaptation according to user perception characteristics," *IEEE Trans. Multimedia*, vol. 7, no. 3, pp. 435–445, Jun. 2005.
- [5] L. T. Sharpe, A. Stockman, H. Jagle, and J. Nathans, *Color Vision: From Genes to Perception*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [6] C. Rigden, "The eye of the beholder-designing for colour-blind users," *British Telecommun. Eng.*, vol. 17, pp. 2–6, Jan. 1999.
- [7] F. Vienot, H. Brettel, and J. D. Mollon, "Digital video colourmaps for checking the legibility of displays by dichromats," *Color Res. Appl.*, vol. 24, no. 4, pp. 243–252, 1999.
- [8] L. Jefferson and R. Harvey, "An interface to support color blind computer users," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '07)*, 2007, pp. 1535–1538.
- [9] K. Wakita and K. Shimamura, "Smartcolor: Disambiguation framework for the colorblind," in *Proc. 7th Int. ACM SIGACCESS Conf. Computers and Accessibility (Assets '05)*, 2005, pp. 158–165.
- [10] L. Jefferson and R. Harvey, "Accommodating color blind computer users," in *Proc. 8th Int. ACM SIGACCESS Conf. Computers and Accessibility (Assets '06)*, 2006, pp. 40–47.
- [11] K. Rasche, R. Geist, and J. Westall, "Detail preserving reproduction of color images for monochromats and dichromats," *IEEE Comput. Graph. Appl.*, vol. 25, no. 3, pp. 22–30, May/June 2005.
- [12] K. Rasche, R. Geist, and J. Westall, "Re-coloring images for gamuts of lower dimension," in *Proc. EuroGraphics*, 2005, vol. 24, no. 3, pp. 423–432.
- [13] G. R. Kuhn, M. M. Oliveira, and L. A. Fernandes, "An efficient naturalness-preserving image-recoloring method for dichromats," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1747–1754, Nov./Dec. 2008.
- [14] J.-B. Huang, C.-S. Chen, T.-C. Jen, and S.-J. Wang, "Image recolorization for the colorblind," in *Proc. 2009 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'09)*, 2009, pp. 1161–1164.
- [15] M. Wang, B. Liu, and X.-S. Hua, "Accessible image search," in *Proc. 17th ACM Int. Conf. Multimedia (ACMMM'09)*, 2009, pp. 291–300.
- [16] M. Wang, Y. Sheng, B. Liu, and X.-S. Hua, "In-image accessibility indication," *IEEE Trans. Multimedia*, vol. 12, no. 4, pp. 330–336, Jun. 2010.
- [17] G. M. Machado and M. M. Oliveira, "Real-time temporal-coherent color contrast enhancement for dichromats," *Comput. Graph. Forum*, vol. 29, no. 3, pp. 933–942, 2010.
- [18] B. Liu, M. Wang, L. Yang, X. Wu, and X. Hua, "Efficient image and video re-coloring for colorblindness," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME'09)*, 2009, pp. 906–909.
- [19] C.-R. Huang, H.-P. Lee, and C.-S. Chen, "Shot change detection via local keypoint matching," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1097–1108, Oct. 2008.
- [20] C.-R. Huang, C.-S. Chen, and P.-C. Chung, "Contrast context histogram-An efficient discriminating local descriptor for object recognition and image matching," *Pattern Recognit.*, vol. 41, no. 10, pp. 3071–3077, 2008.
- [21] Y. Lin and S. Tai, "Fast full-search block-matching algorithm for motion-compensated video compression," *IEEE Trans. Commun.*, vol. 45, no. 5, pp. 527–531, May 1997.
- [22] C. Zhu, X. Lin, and L. P. Chau, "Hexagon-based search pattern for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 5, pp. 349–355, May 2002.
- [23] Y. Nie and K. K. Ma, "Adaptive rood pattern search for fast block-matching motion estimation," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1442–1449, Dec. 2002.
- [24] P. Heckbert, "Color image quantization for frame buffer display," in *Proc. 9th Annu. Conf. Computer Graphics and Interactive Techniques (SIGGRAPH '82)*, 1982, pp. 297–307.
- [25] S. P. Lloyd, "Least square quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

- [26] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.
- [27] H. Brettel, F. Vionot, and J. D. Mollon, "Computerized simulation of color appearance for dichromats," *J. Optic. Soc. Amer. A*, vol. 14, no. 10, pp. 2647–2655, 1997.
- [28] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Electrical Applications*. New York: Springer-Verlag, 2007.
- [29] Q. Y. Duan, V. K. Gupta, and S. Sorooshian, "Shuffled complex evolution approach for effective and efficient global minimization," *J. Optim. Theory Appl.*, vol. 76, no. 3, pp. 501–521, 1993.
- [30] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proc. 1968 23rd ACM National Conf. (ACM '68)*, 1968, pp. 517–524.
- [31] V. A. Kovalev, "Mining dichromatic colours from videos," in *Proc. Industrial Conf. Data Mining 4065*, 2006, pp. 431–443.



Chun-Rong Huang (M'05) received the B.S. and Ph.D. degrees in the electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 1999 and 2005, respectively.

In 2005, he joined the Institute of Information Science, Academia Sinica, Taipei, Taiwan, where, since 2005, he has been a Postdoctoral Fellow. In 2010, he became an Assistant Professor with both the Institute of Networking and Multimedia, and the Department of Computer Science and Engineering, National Chung Hsing University, Taichung, Taiwan. His research interests include computer vision, computer graphic, multimedia signal processing, image processing, and medical image processing.

Dr. Huang is a member of the IEEE Circuits and Systems Society and the Phi Tau Phi honor society.

Kuo-Chuan Chiu received the B.S. degree in electrical engineering from National University of Kaohsiung, Kaohsiung, Taiwan, in 2004 and the M.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2006.

He joined the Institute of Information Science, Academia Sinica, Taipei, as a research assistant in 2009. His research interests include computer vision and image processing.



Chu-Song Chen (S'94–M'96) received the B.S. degree in control engineering from National Chiao-Tung University, Taiwan, in 1989 and the M.S. and Ph.D. degrees from the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, in 1991 and 1996, respectively.

He is now a deputy director of the Research Center for Information Technology Innovation (CITI) and a research fellow of the Institute of Information Science (IIS), Academia Sinica, Taipei, Taiwan. He is

also an adjunct professor of the Graduate Institute of Networking and Multimedia, National Taiwan University. His research interests include computer vision, pattern recognition, signal/image processing, and multimedia analysis. In 2007–2008, he served as the Secretary-General of the IPPR Society, Taiwan, which is one of the regional societies of the International Association of Pattern Recognition (IAPR).

Dr. Chen has served as the program co-chair of the conferences *ICDAT 2005* and *ICDAT 2006*, theme chair of *PSIVT 2009*, and area chair of *ACCV 2009*, *ACCV 2010* and *NBiS 2010*. He is on the editorial board of the *Journal of Multimedia* (Academy Publisher), *Machine Vision and Applications* (Springer), and *IPSI Transactions on Computer Vision and Applications*.