Hybrid Color Compensation for Virtual View Synthesis in Multiview Video Applications

Pei-Kuei Tsung, Hsin-Jung Yang, Pin-Chih Lin, Kuan-Yu Chen and Liang-Gee Chen
DSP/IC Design Lab, Graduated Institute of Electronics Engineering
National Taiwan University
Taipei, Taiwan
{iceworm, murasaki, lpg, stevenky, lgchen}@video.ee.ntu.edu.tw

***Abstract*—Multiview video (MVV) can provide viewers a complete 3D perception by its multiple-viewpoint feature. In order to support smooth viewpoint switching, virtual view synthesis is required. However, color and illumination information varies between views according to the light field conditions. This mismatch leads to ghost effects and wrong reflection phenomena. In this paper, a hybrid color compensation method is proposed. With the proposed inter-view color correspondence estimation, a smoothly changing color/illumination field is established between reference views. Furthermore, the reflection on mirror-like materials can also be detected and optimized by the proposed hybrid reflection model. As a result, virtual view frames with better perceptual quality are generated. In the objective test, the proposed algorithm also provides 0.26-0.42 dB PSNR gains.**

I. Introduction

Recently, multiview video (MVV) has drawn public attention, providing viewers a complete 3D perception by its multiple-viewpoint feature. As display technologies evolve, various multiview video applications, such as 3DTV [1] and free viewpoint TV [2], are emerging. However, some challenges on MVV applications need to be conquered. First of all, an efficient multiview video coding (MVC) method is required to deal with drastically increasing amount of data. In addition, in order to support smooth and continuous viewpoint switching, virtual view synthesis is required to generate virtual view frames between different viewpoints and fill the non-captured area, since it is impossible to capture video sequences from all viewpoints with infinite real cameras. In July 2008, MVC is standardized as the Multiview High Profile in H.264/AVC by MPEG 3D Audio/Video (3DAV) Group. [3] After finishing the MVC standardization, MPEG-FTV group is working on virtual view synthesis. The view synthesis reference software (VSRS) is released by the MPEG-FTV group as the reference software and the research platform. [4] Many design challenges are waiting to be solved, since virtual view synthesis is a newly raised research area. These challenges are mainly from the occlusion handling in virtual view frames. In a multiview sequence, the occlusion area can be filled by reference frames from viewpoints neighboring the target virtual viewpoint. However, since frames in different views are captured by different cameras at different locations, color/illumination information also changes between different views. Many previous works pay attention to color/illumination compensation in multiview video. [5][6][7] However, most of them focus on color compensation before or within the encoding step, while there is a lack of discussion about color compensation in the virtual-view-synthesis step. In this paper, a hybrid color compensation scheme that targets for virtual view synthesis in multiview video applications is proposed. With the proposed inter-view color correspondence estimation, color mismatch caused by multiple reference views can be eliminated, and smoothly changing light-field environment can be generated. Furthermore, by detecting the region of reflection and raising a proper reflection model, cases for mirror-like materials can be distinguished from general cases and modified particularly. As a result, virtual view frames for 3D and multiview video applications can be synthesized with better perceptual quality. In the objective PSNR test, the proposed method also performs 0.26-0.42 dB gains.

The remaining of this paper is organized as following: First, the view-synthesis flow and design challenges are introduced in Sec. II. Second, the proposed color compensation scheme is described in Sec. III. Then, Sec. IV shows the simulation result. Finally, Sec. V concludes this paper.

II. Design Challenges

Figure 1 shows the conventional view-interpolation flow. First, all reference views are warped to the virtual view position. Because of the depth discontinuity, each warped frame contains holes shown as the green parts in Fig. 1. Then, all warped frames are blended together and processed with iterative post-processing to generate the virtual view. In Fig. 1, there are two reference views, and the frames warped from them contain occlusions at different locations. Thus, some parts of final output frame have two references and some others have only one. At different parts of the synthesized virtual view frame, using different references may lead to color or illumination mismatch problems. The details of these design challenges are described as follows:

978-1-4244-5309-2/10/$26.00 ©2010 IEEE

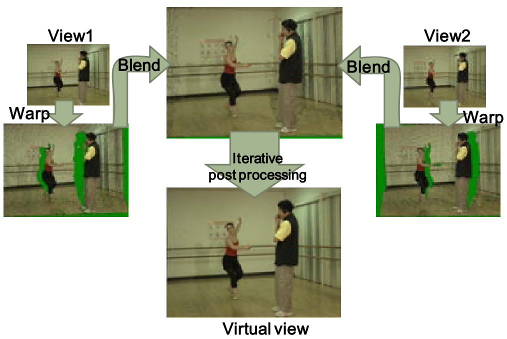

Figure 1. Illustration of view-synthesis flow.

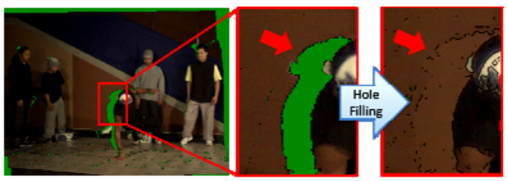

Figure 2. Color mismatch between different reference views leads to a ghost effect. The dark lines are the edge filter result to highlight the color mismatch boundary

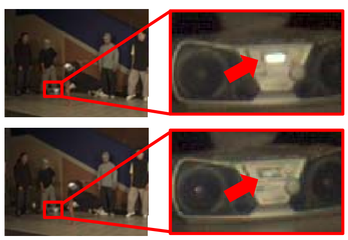

Figure 3. The reflection behavior of mirror-like materials with strong directionality.

A. Unknown Light Field Variance

Since there is no real camera at the virtual viewpoint, the actual color and illumination information in the virtual view frame is unknown. In virtual view synthesis, the information is directly copied from reference frames through warping. However, the information actually changes as the relative distance from light sources varies. In VSRS, this problem is handled by blending two reference frames to simulate the unknown light field. Unfortunately, the blending algorithm cannot be used in occlusion parts, since those regions have only one or no reference data. In the single-iteration virtual-view-synthesis flow proposed in [8], color information in most parts of the virtual view frame is copied from one of the two reference views, while the color information in the other reference is only used to fill the occlusion regions. As shown in Fig. 2, in a synthesized view without color compensation, even though the occlusion part has already been filled, it still can be observed due to the color mismatch between two references. Besides, since most information of the virtual view frames is from one of the two references, the frame-level average color/illumination cannot change smoothly when the target view point is switched from the other reference to an adjacent virtual view frame. Therefore, it is necessary to deal with the unknown light field in virtual view frames to solve above problems.

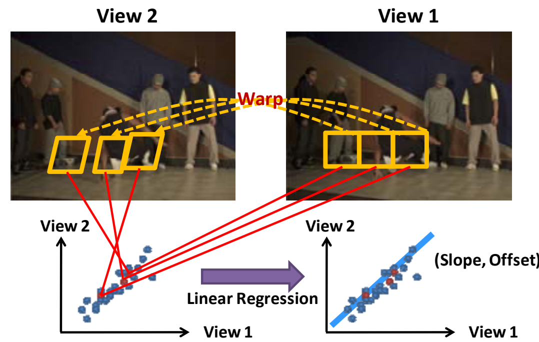

Figure 4. Proposed color-coorspondence-estimation flow.

B. Unknown Reflection Behavior

Even if the light-field varying function is well known, the virtual view frame still cannot be generated without subjective defects. A main reason is that some materials, like mirrors, have different reflection characteristics. As shown in Fig. 3, light reflected by a mirror-like material has strong directionality. Frames in Fig. 3 are captured by two cameras only 20 cm apart; however, the reflection phenomenon can only be detected in the upper frame. At the same time, all the other parts in these two frames have similar color and luminance values. Accordingly, the frame-level blending algorithm cannot deal with this case and a flexible algorithm that can support the reflection model for mirror-like materials is required.

III. PROPOSED COLOR COMPENSATION ALGORITHM

A. Inter-View Color Correspondence Estimation

In order to simulate the light field around a target virtual viewpoint, an inter-view color correspondence estimation flow is proposed. Under the assumption that the target virtual view has two adjacent reference views, the algorithm flow is shown in Fig. 4. First, with camera calibration matrices, the spatial relationship between view 1 and view 2 can be found. Second, the average color per-macro-block (per-MB) is calculated in view 1. Each MB in view 1 has 256 corresponding pixels in view 2, and these pixels may not form a square pattern but disperse in view 2. The third step is to calculate the average in these 256 pixels. Then, $(average_{view1}, average_{view2})$ can be seen as one sample point in linear regression process. After scanning the whole frame, the frame-level relationship between the color information in view 1 and view 2 is derived as:

$$view_2 = view_1 \times slope_{12} + offset_{12}. \quad (1)$$

In Eq. 1, $slope_{12}$ and $offset_{12}$ are results of linear regression.

This method can be easily modified to fit the multiview cases with more than two reference frames. Take a three-view case for example. First, the process performs linear regression between view 1-2 and then view 2-3. After that, the relationship between view 1 and view 3 can be derived as:

$$view_3 = (view_1 \times slope_{12} + offset_{12}) \times slope_{23} + offset_{23}. \quad (2)$$

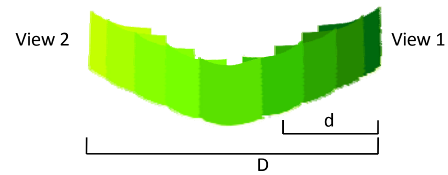

Figure 5. Linear light field between two reference views.

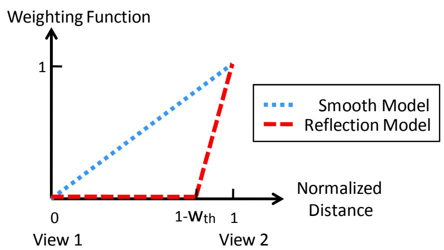

Figure 6. Proposed hybrid light field model

Based on the well defined depth maps and camera matrices as inputs for view synthesis, Eq. 2 still performs good simulation on the color correspondence between view 1 and view 3. Thus, for an N-reference case, the minimum number of execution times required for linear regression is N-1.

B. Linear Light Field Assumption

After finding the correspondence between two reference views, the frame-level light field between them can be modeled. Considering that in general the distance between two adjacent views in a multiview sequence is only about several centimeters, a linear light field model is proposed and illustrated in Fig. 5. Because of the short distance between two adjacent views, the linear approximation is enough to model the actual value of the light field. Besides, a linear model can also support a smooth light-field changing between two views. In Fig. 5, D is defined as the distance between view 1 and view 2, and d is the distance between view 1 and the target virtual view position. Both two parameters can be derived from the camera matrices, since all rotation and translation indexes are recorded in the extrinsic matrices. After calculating the distance, the pixel value in the virtual view frame, $view_{virtual}$, can be obtained as follows:

$$view_{virtual} = (1 + (slope_{12} - 1) \times w) \times view_1 + offset_{12} \times w \; ; \quad (3)$$

$$w = d/D. \quad (4)$$

In the above equations, w expresses the linear weighting at the target location. Equation 3 only takes pixels from view 1 as example. The equation for taking pixels from view 2 can be derived by a similar manner. After this step, the color intensity in all reference frames is adjusted to fit the target virtual view position. Thus, the color mismatch problem can be solved.

C. Optimization for Mirror-Like Materials

As described in Sec. II, the frame-level color compensation cannot deal with the reflection cases. In order to optimize these special cases, at first a classification algorithm is required to detect and separate the strong reflection regions from the smooth light field. In a multiview sequence, a well defined depth map is required for view synthesis. Thus, the physical correspondence between two references for each pixel is already known. With the precise physical correspondence, the outlier sample points in the color correspondence estimation can be regarded as the special cases of reflection, since there should be no outlier pixels if the depth map is well defined. In other words, a depth map with high quality rules out the possibility that two pixels from two references point to different objects, so having a different reflection model is the only reason for the outlier cases.

After the regions of mirror-like materials are found, a hybrid reflection model is proposed to optimize the color information in these regions. As shown in Fig. 6, the reflection model is merged from a smooth term and a reflection term. The smooth term is a linear model described in the previous sections; the reflection term is simulated by a two-step linear function to present the strong directionality. Under this hybrid model, the effect of reflection can only be observed in the virtual views that are sufficiently near (with normalized distance smaller than W_{th}) the reference frame with reflection. The color value of pixels in the reflection region is therefore described as:

$$view_{virtual} = Smooth(view_1) + Reflection\,(view_2, view_1) \quad (5)$$

Function *Smooth()* in Eq. 5 means the linear smooth function similar to Eq. 3; function *Reflection()* means the two-step function. In *Reflection()*, pixel values from two references are both required, because the pixel pair is the outlier of the linear regression and cannot apply the linear relation formed previously. After getting two pixel values, the color/illumination difference is calculated and weighted by the reflection model weighting function shown in Fig. 6.

IV. Simulation Result

In this paper, two multiview sequences, “ballet” and “breakdancers” published by Microsoft, are taken as test sequences. The view-synthesis tool is implemented from the previous work in [8]. In order to test the change in light field between different views, 9 virtual views distributed evenly between two adjacent reference views are generated. The location of each virtual view is calculated by interpolating the translation terms in the extrinsic matrix. Figure 7 shows the view-by-view light field variance between two reference views. The vertical axis in Fig. 7 is the frame-level average luminance, and the horizontal axis illustrates the normalized distance between two reference views. As shown in Fig. 7, without color compensation, the color/illumination information in virtual views almost maintains constant, since pixels in virtual view frames are copied directly from view 1. However, there exist a color/illumination mismatch between view 1 and view 2. Thus, a sudden illumination jump occurs when the target position is switching to view 2 in [8]. On the contrary, the light field between view 1 and view 2 changes smoothly under the frame-level blending in VSRS and the proposed linear regression model respectively.

Figure 8 shows the result of color mismatch reduction. In Fig. 8 (a), the synthesized virtual view and its occlusion region is shown. This occlusion part can be filled by other reference views in a multiview sequence. However, as mentioned in Sec. II, a direct filling scheme without color compensation leads to the ghost effect shown in Fig. 8 (b) and (c). Figure 8 (b) and (c) show the virtual view frame with the highlight color mismatched edges from reference [8] and VSRS respectively. Figure 8 (d) shows the

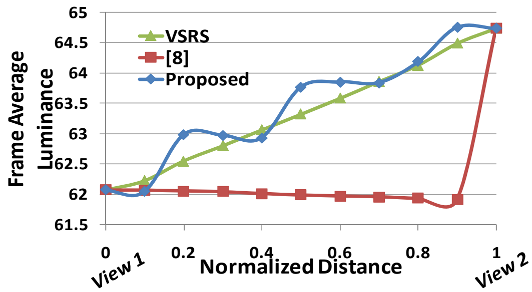

Figure 7. Variation in frame-level average luminance.

Figure 8. Color mismatch is reduced by the proposed color compensation. (a) virtual view with un-filled occlusion, (b) occlusion filled by [8], and (c) VSRS, (d) occlusion filled with proposed color compensation. The dark lines are the edge filter result to highlight the color mismatch boundary.

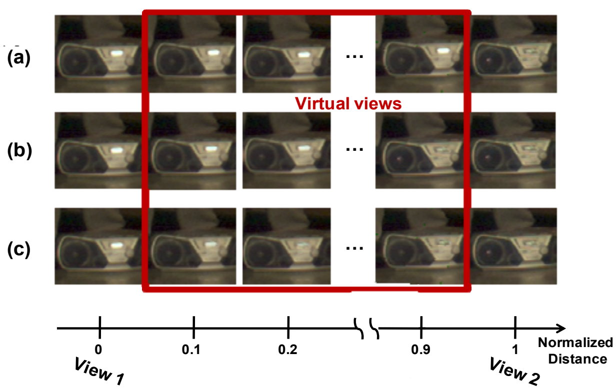

Figure 9. Hybrid color compensation result. (a) and (b) show the virtual view frames by [8] and VSRS respectively. (c) shows virtual view frames under the proposed color compensation with both smooth and reflection terms. Only the reflection in (c) is well modeled. W_{th} is set as 5 cm.

virtual view after the proposed color compensation. The ghost effect is eliminated since both two reference frames are optimized with parameters obtained from linear regression. Here note that even the VSRS provides better linear light field modeling, it cannot deal with the color mismatches in the occlusion parts and thus the proposed scheme has better visual quality. In addition, no blending or other pixel-based interpolation algorithm is used here. All virtual view pixels under linear reflection model are modified from only one pixel. Thus, the single iteration characteristic proposed in [8] is still maintained.

Figure 9 indicates the performance of the proposed hybrid reflection model. A mirror-like material is highlighted in Fig. 9. If only the frame-level color compensation is used, the reflection behavior cannot be observed. Figure 9 (a) and (b) denote this situation. If view 1 is chosen as the primary reference in the view-synthesis flow in [8], a continuous strong reflection and a sudden dark is observed and vice versa. In VSRS, the frame-level blending also causes a wrong reflection behavior. The result after the proposed optimization is illustrated in Fig. 9 (c). By detecting the mirror-like material and adopting the hybrid color compensation, a more suitable reflection behavior is established. Here, the W_{th} is set as 5 cm and the normalized distance is around 0.3.

TABLE I. PSNR COMPARISON

	PSNR (without color compensation)	PSNR (with color compensation)
"Breakdancers"	32.53 dB	32.79 dB
"Ballet"	28.52 dB	28.94 dB

In order to supply the objective PSNR data rather than the subjective perceptual result, another virtual view setting is considered. The distance of reference views is set to 2. For example, view 1 and view 3 is taken as the reference views and the virtual view is set at the same position as view 2. Thus, the original captured data in view 2 can be used as the golden answer for the view synthesis. Table 1 shows the PSNR result. The proposed color compensation improves the PSNR quality about 0.26-0.42 dB than virtual views without color compensation.

V. CONCLUSION

In this paper, a hybrid color compensation scheme is proposed for virtual view synthesis in multiview video applications. Based on the proposed inter-view color correspondence estimation, a model with linear and smooth light field is established. As a result, color mismatch and ghost effects in the synthesized virtual view frames are eliminated. Besides, the region with strong reflection is detected from the outliers in linear regression and optimized by hybrid reflection model. Thus, a proper reflection behavior is shown. Compared with virtual views without color compensation, the proposed method improves the PSNR result by about 0.26-0.42 dB.

REFERENCES

[1] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technologies," *Proceedings of the IEEE, vol. 93, no. 1, pp 33-36, Jan. 2005*

[2] M. Tanimoto, "Free viewpoint television - FTV," *in Proceedings of Picture Coding Symposium, 2004*

[3] Joint Video Team, "Joint Draft 8.0 on Multiview Video Coding" *Number ISO/IEC JTC1/SC29/WG11 JVT-AB204, July, 2008, Hannover, Deutsche.*

[4] MPEG-FTV Group, "LDV Virtual View Rendering Software" *ISO/IEC JTC1/SC29/WG11 MPEG2008/M16040, Feb. 2009*

[5] K. Yamamoto et al, "Multiview Video Coding Using View Interpolation and Color Correction," *Circuits and Systems for Video Technology, IEEE Transactions on, Vol. 17, Issue 11, pp. 1436-1449, Nov. 2007*

[6] J. H. Kim et al, "New Coding Tools for Illumination and Focus Mismatch Compensation in Multiview Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on, Vol. 17, Issue 11, pp. 1519-1535, Nov. 2007*

[7] L. Hou et al, "A novel ray-space based color correction algorithm for multi-view video," *in proceedings of IEEE International Symposium on Circuits and Systems 2009 (ISCAS 2009), May 2009, pp. 2381-2384*

[8] P. K. Tsung et al, "Single iteration view interpolation for multiview video applications," *in Proceedings of 3DTV conference 2009, May 2009, pp. 1-4*