IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 15, 2021, accepted May 16, 2021, date of publication May 19, 2021, date of current version May 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3081870

Color Correction for Immersive Video Applications

ADRIAN DZIEMBOWSKI, DAWID MIELOCH, SŁAWOMIR RÓŻEK, AND MAREK DOMAŃSKI, (Senior Member, IEEE)
Institute of Multimedia Telecommunications, Poznań University of Technology, 60-965 Poznań, Poland
Corresponding author: Adrian Dziembowski (adrian.dziembowski@put.poznan.pl)

This work was supported by the Ministry of Education and Science of Republic of Poland.

ABSTRACT In this paper, the color correction method developed for immersive video systems is presented. The proposed method significantly increases the consistency of color characteristics of multiview sequences, understood both as the temporal and the inter-view consistency, what highly improves the subjective quality of the synthesized virtual views presented to the final user of the immersive video system. Moreover, the proposal allows to significantly increase the quality of the depth maps calculated for natural sequences, e.g., for views with colors inconsistent due to different lighting conditions. It enables more efficient compression of natural multiview video, as the newest encoding standards highly depend on the quality of depth maps. In order to evaluate the performance of the proposal, three experiments were conducted. In the first one, the proposal was compared to state of the art in the typical immersive video application – color correction of a natural multiview sequence. In the second experiment, the performance of the proposal was tested on the Middlebury stereo dataset. In both experiments, the quality of synthesized virtual views was assessed subjectively by a group of 70 naïve viewers. The third experiment assessed the influence of color correction on the quality of estimated depth maps. All the experiments showed that the proposal significantly increases the color consistency of the multiview content. Due to the high usefulness and robustness, the proposed color correction method became the MPEG reference software for color correction. The implementation of the method is available for other researchers on the public repository.

INDEX TERMS Color correction, color refinement, immersive video, 3DoF+, inter-view consistency, virtual navigation.

I. INTRODUCTION

Immersive video [26], [4], [40] gains growing importance because of its expanding applications in virtual reality, virtual navigation, and ultra-realistic audiovisual content presentations. In general, the immersive video is an extension of 360° video, where the user is limited to only 3 degrees of freedom (change of orientation). In the immersive video, the viewer to freely control his or her position and orientation of viewing of the video content using typical monoscopic display [53] or head-mounted display (HMD) devices [4]. Immersive video may be related to both natural and computer-generated content. Here, we focus on the natural content that originates from video cameras, both perspective (producing 2D rectangular video) and omnidirectional (producing up to 360-degree video) possibly augmented by data from depth cameras [23]. Such content is sometimes described as highly realistic or ultra-realistic.

The immersive visual content is acquired either from multiple perspective cameras located around a scene, omnidirectional cameras located within a scene, or by a mixture of both. In any case, the cameras must be carefully synchronized [53]. The scene may be of very different types: a play court of a sports event, a theater stage, a wilderness scene, street environment, etc. The practical limitations (cost, portability, video processing time, system calibration complexity, etc.) imply a limited number of cameras, i.e., the cameras are often located quite distant from each other [53], [17].

For further processing, transmission, and final rendering producing the requested virtual views, a model of the dynamic scene is estimated. The following scene model types are mostly considered: multiview plus depth (MVD) [36], [59], point-based (like point clouds, e.g. [58], [24]), ray-space [50], and object-based [19], [49]. The data corresponding to such models exhibits high

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenbao Liu.

redundancy, therefore, the respective compression techniques have been developed for point clouds [41], [56], [20] and for MVD [52]. The data retrieved from the models is used to render the output virtual views that can be displayed on various devices like classic 2D displays, autostereoscopic displays, head-mounted devices, omnidirectional displays, etc.

The model of a scene is expected to be temporally consistent [51], i.e., for all voxels their parameters, i.e., color and location, change smoothly along the time axis as long no rapid change occurs in the respective real scene. Lack of temporal consistency, i.e., the temporal inconsistency, results in quality degradation of the output virtual views. By independent depth estimation in consecutive frames, the temporal inconsistency of views may result in a varying depth, even for static objects.

Typical degradations in the virtual views synthesized from inconsistent visual data include flickering, rapid changes of shape and color of objects in a scene, as well as inpainting of pieces of background or other objects into objects or background. The temporal color inconsistency of a given input view may result even from small but rapid changes of illumination (e.g., pulsing of fluorescent lamps) or rapid changes of the camera parameters (e.g., due to automatic white balancing in response to rapid scene changes) [45].

Moreover, a model should also exhibit the inter-view consistency, i.e., we require that no ambiguity in voxel colors and locations is caused by the exploitation of data from different input views with the respective depth maps [29], [30], [15], [54]. If a set of views exhibits such a property, it is called to be inter-view consistent. Such consistency may be hindered by differences in camera color reproduction models [18] and illumination differences corresponding to the individual input views. The inter-view consistency is also extremely important in depth estimation. Its lack provides problems in the search for the corresponding points in the individual views thus leading to erroneous disparity estimation [35], [8].

The goal of the paper is to propose techniques for the correction of colors in the input views in order to compensate both temporal and inter-view inconsistencies in the input views. The desirable color correction techniques need to be implementable in real time, thus there is a requirement that the color correction techniques need to have low computational cost. The efficiency of the proposed techniques is measured by the quality improvement of the output virtual video synthesized from the respective models of immersive video as well as by the improvement of compression efficiency.

II. RELATED WORKS

There are already numerous color correction techniques described in the literature. In general, they can be divided into parametric and non-parametric techniques [60]. The most commonly used parametric technique is the color transfer technique [9]. It defines a simple function allowing changing the color space by assuming there is a correlation between the color space of two images that are being compared. Some methods are focused on matching the color of views with non-linear differences that can be the result of using different settings in used cameras [70]. The most popular non-parametric color correction technique is histogram matching [21], which allows equalizing histograms of different images.

Unfortunately, these methods can be hardly used directly for the refinement of sequences captured by sparse multicamera systems because only a small field of view overlaps between more distant views [62]. A simple solution for such a kind of content is to refine colors of different views hierarchically, i.e., by comparing the color of neighboring views only and multiplying correction coefficients in the final step to equalize color characteristics of all views [38], [39].

Other methods that could be used for sparse multicamera systems, are based on the search for corresponding areas of different views and calculating differences between color characteristics only in these parts of views. Corresponding areas may be found using feature transforms, e.g., SIFT [31] or SURF [3]. The color correction techniques based on such descriptors were described e.g., in [61], [16], and [37]. While such proposals can increase not only the inter-view consistency of multi-view images but also the temporal color consistency, the computational complexity is highly increased for such approaches (even 8 seconds of processing per one frame [32], [6]). Moreover, the use of feature transforms matching does not guarantee that color inconsistencies that are visible in a small area of a view will be corrected, as not all parts of views are processed in such approaches.

Multiview color correction can be also performed by reprojecting points between different views [47] and is usually performed as a part of the virtual view synthesis process [7], [12], [67], [68], [69]. Such an approach allows to properly define corresponding areas even for systems with a very sparse arrangement of cameras. However, the temporal consistency of color characteristics in these methods is dependent on the temporal consistency of depth maps, which can be very hard to achieve for natural sequences. It can be increased using depth post-processing methods (e.g. [66]), but the use of another refinement increases the overall time required to acquire color-corrected views.

The proposed method, described in Sections IV and V, provides similar advantages as those provided by methods based on points reprojection (therefore, can be used with sparse multicamera systems), but the temporal stability enhancement is not influenced by the temporal consistency of depth maps, as is performed independently in each view before the reprojection. Moreover, as the results provided in Section VIII.D have shown, the proposal is a real-time method, unlike other methods that provide both inter-view and temporal consistency of color.

III. COLOR CORRECTION

For frame f in view I_i, image formation is described by (1) for each of the red, green, and blue channels [13]:

$$I_i(\mathbf{x},f) = G_i(\mathbf{x}) \int_{\lambda} F_i(\lambda, \mathbf{x}) S_i(\lambda) L_i(\lambda, f)\, d\lambda \qquad (1)$$

where $I_i(\mathbf{x},f)$ is intensity measured by i-th sensor at position $\mathbf{x} = (x, y)$, i.e., $\mathbf{x}$ are the coordinates on the image plane, $G_i(\mathbf{x})$ is a scaling factor due to the geometry of the rays to the pixel at position $\mathbf{x}$ in i-th view, $F_i(\lambda, \mathbf{x})$ denotes the reflectance at position $\mathbf{x}$, $S_i(\lambda)$ is the sensitivity of i-th sensor, $L_i(\lambda)$ is the radiance given off by the light source for i-th view, and λ is the light wavelength.

By the use of decomposition over the three basis functions [22], [13] one gets:

$$\mathbf{I}_i(\mathbf{x},f) = \mathbf{A}_i(\mathbf{x},f)\mathbf{F}(\mathbf{x}) \tag{2}$$

where $\mathbf{I}_i(\mathbf{x},f)$ is a $[3 \times 1]$ vector of intensities measured in f-th frame by i-th sensor at position $\mathbf{x} = (x, y)$ for the three components: red, green, and blue, $\mathbf{A}_i(\mathbf{x},f)$ is the lightning matrix [57] of the size $[3 \times 3]$ with the elements that depend on the functions $G_i(\mathbf{x})$, $L_i(\lambda)$, $S_i(\lambda)$, and $\mathbf{F}(\mathbf{x})$ is the reflectance vector with red, green, and blue components. The color of an object at position $\mathbf{x} = (x, y)$ is described by $\mathbf{F}(\mathbf{x})$ irrespective of the view.

For the sake of simplicity, we will skip the frame index f for most of the considerations.

The functions $G_i(\mathbf{x})$ may differ significantly for different cameras i due to non-Lambertian reflections, and in particular due to reflective speckles.

The functions $S_i(\lambda)$ are usually different for each color channel even within the same camera, apart from the different sensor chips [63]. Moreover, at a given time instant, the functions $S_i(\lambda)$ may be different for different cameras e.g., due to automatic white balancing. Often in contemporary cameras, especially consumer-oriented ones, sophisticated image and video preprocessing is employed [64]. Such processing depends on the individual illumination of a given camera as well as on its parameters. In practical multicamera systems, instead of professional cameras, consumer ones are often used [53], [25]. Such cameras do not allow to fully control all the acquisition parameters, e.g., white balance or exposure time. In such a case, the global color characteristics of the video may vary in time even if the lighting conditions are stable.

The functions $L_i(\lambda)$ are often different due to different illumination, in particular for the sparse distribution of cameras around a scene. Moreover, the functions $L_i(\lambda)$ may vary in time due to changes of illumination (e.g., due to pulsing of fluorescent lamps used for illumination of the scene).

In general, as mentioned above, the functions $G_i(\mathbf{x})$, $L_i(\lambda)$, $S_i(\lambda)$ may be different for each i, i.e., they may be different for individual views. Therefore, the lightning matrices $\mathbf{A}_i(\mathbf{x})$ are mostly different for various views i. Also, the functions $G_i(\mathbf{x})$, $L_i(\lambda)$, $S_i(\lambda)$ may vary in time, thus may the lightning matrices $\mathbf{A}_i(\mathbf{x})$. As the result, for the same $\mathbf{F}(\mathbf{x})$, the intensity $\mathbf{I}_i(\mathbf{x})$ may be different in different views i and at the different time instants.

Unfortunately, the camera sensors measure the intensities $\mathbf{I}_i(\mathbf{x})$, not the actual color $\mathbf{F}(\mathbf{x})$ that would be very useful for depth estimation and virtual view synthesis. In fact, for efficient depth estimation and virtual view synthesis, it is enough to know color-compensated intensities $\mathbf{I}'_i(\mathbf{x})$, i.e., the intensities:

$$\mathbf{I}'_i(\mathbf{x}) = \mathbf{A}'_i(\mathbf{x})\mathbf{F}(\mathbf{x}) \tag{3}$$

where $\mathbf{I}'_i(\mathbf{x})$ is the vector of color-compensated intensities corresponding to i-th sensor at position $\mathbf{x} = (x, y)$ for the three components: red, green, and blue. $\mathbf{A}'_i(\mathbf{x})$ is the color-compensated lightning matrix that is time-invariant (for the requirement of temporal consistency) and identical for all i (for the requirement of inter-view consistency). For practical reasons, the restrictions may be weakened for the color-compensated lightning matrix $\mathbf{A}'_i(\mathbf{x})$, i.e., its elements may vary slowly in time, and small differences between $\mathbf{A}'_i(\mathbf{x})$ and $\mathbf{A}'_j(\mathbf{x})$ for $\neq ij$ may exist due to different directions of illumination. The latter soft relaxation often results in more natural results of virtual view synthesis [65].

So, our problem is to estimate color-compensated intensities $\mathbf{I}'_i(\mathbf{x})$ from the known intensities $\mathbf{I}_i(\mathbf{x})$. This problem is a case of the fundamental problem of color constancy that is well known from colorimetry [13]. The measured intensities $\mathbf{I}_i(\mathbf{x})$ may be expressed as:

$$\mathbf{I}_i(\mathbf{x}) = \left(\mathbf{A}'_i(\mathbf{x}) + \Delta\left(\mathbf{A}_i(\mathbf{x})\right)\right)\mathbf{F}(\mathbf{x}) \tag{4}$$

where the elements of $\Delta(\mathbf{A}_i(\mathbf{x}))$ result from all the factors mentioned above that yield temporal or inter-view inconsistency. Therefore, from (3) and (4) we obtain:

$$\mathbf{I}'_i(\mathbf{x}) = \mathbf{I}_i(\mathbf{x}) - \Delta\mathbf{I}_i(\mathbf{x}) = \mathbf{I}_i(x) - \Delta\left(\mathbf{A}_i(\mathbf{x})\right)\mathbf{F}(\mathbf{x}) \tag{5}$$

where $\Delta\mathbf{I}_i(\mathbf{x})$ is the vector of corrections of the primary color components for i-th camera.

Therefore, to summarize, the problem is to compensate the influence of $\Delta(\mathbf{A}_i(\mathbf{x}))$ i.e., to estimate the vectors of color corrections $\Delta\mathbf{I}_i(\mathbf{x})$ for all i in each frame f.

These considerations may be also applied to other spaces as long they are related to the RGB color space by a linear transformation:

$$\mathbf{I}_{Cj}(\mathbf{x}) = \mathbf{C}\cdot\mathbf{I}_j(\mathbf{x}) \tag{6}$$

where $\mathbf{C}$ is the $[3 \times 3]$ color transformation matrix, and $\mathbf{I}_{Ci}(\mathbf{x})$ is the $[3 \times 1]$ vector of sample values measured by the i-th sensor at position $\mathbf{x} = (x, y)$ for three color components in the color space defined the transformation matrix $\mathbf{C}$. For video, such a color space is the common transmission color space $\mathrm{YC_BC_R}$. After left-side multiplication of both sides of (2) – (5), we get that these equations hold also for I and F expressed in each color space related to the RGB space by (6). It means that the color correction problem is the same in all such color spaces.

In the paper, we are going to propose fast and robust algorithms that estimate the color-corrected samples of the views in the immersive video representations. Sections IV and V describe such techniques, developed to reduce temporal inconsistency and inter-view inconsistency, respectively. Proposed techniques are non-iterative, exhibit low complexity, and can be easily parallelized in order to meet the real-time

processing thus can be directly used in practical implementations of immersive video systems.

In the beginning, all views are processed independently to equalize their color characteristics over time. Then, the color characteristics of all views are unified.

In further considerations, the entire refinement is performed in the RGB color space, because such independent processing of color components is very intuitive.

IV. TEMPORAL COLOR CONSISTENCY IMPROVEMENT

In the proposed technique, all input views are processed independently. The processing scheme for one view is presented in Fig. 1.

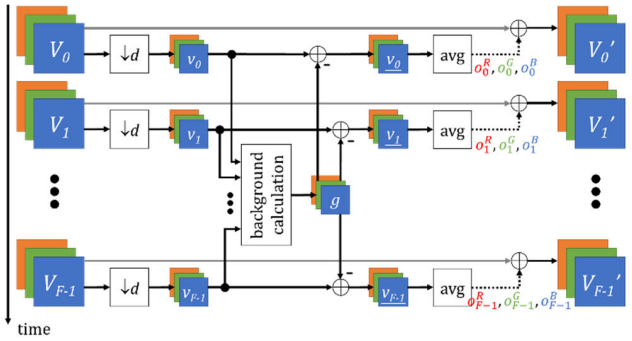

FIGURE 1. Scheme of the technique for temporal consistency improvement: white boxes denote operations, color boxes contain signal symbols, "avg" means the average over all pixels of the frame, V_f – full resolution view at frame f, d – decimation factor, v_f – reduced resolution frame (decimated by factor d), g – background frame, V_f' – full resolution view V_f after refinement. Each frame V_f consists of three color component planes: V_f^R, V_f^G and V_f^B.

The main idea of the proposed technique is to calculate the global offset between each frame and the time-invariant background frame g. This offset is calculated independently for all color components.

The correction process starts with the estimation of the background frame. In the proposed approach, each pixel of the background is calculated as a median of values of a pixel (x, y) in all the frames:

$$g^c(x,y) = \text{med}\left(v_0^c(x,y), v_1^c(x,y), \ldots, v_{F-1}^c(x,y)\right) \quad (7)$$

where $v_f^c(x, y)$ is the value of the color component c in the f-th frame in a given view and F is the total number of frames.

In order to speed up the processing, the background frame may be calculated in resolution reduced d times. In the experiments reported in this paper, the background frame has 16 times smaller width and height than each input frame. Such an approach significantly reduces computational time whereas preserving refinement efficiency.

In the second step, for each frame v_f, the average color component offset o_f^c is calculated:

$$o_f^c = \frac{1}{|E|} \sum_{(x,y)\in E} \underline{v_f^c(x,y)} \quad (8)$$

where:

$$\underline{v_f^c(x,y)} = v_f^c(x,y) - g^c(x,y) \quad (9)$$

and $|E|$ is the cardinality of set E, that is defined as:

$$E = \left\{ (x,y) : x \in \left[0, \frac{W}{d}\right]^y \in \left[0, \frac{H}{d}\right]_c^{\forall} \underline{v_f^c(x,y)} < \frac{2^b}{10} \right\} \quad (10)$$

where W and H are width and height of an input view, d – the decimation step, c denotes color component (like R, G, or B), and b is the bit depth of the input view. As denoted by a logical conjunction symbol "$\wedge$", each pixel within set E must fulfill all 3 conditions.

Set E contains only pixels that represent the same object in frame i and the background frame (i.e., pixels for which the difference between frame i and the background frame in all the color components is smaller than 10% of the bit depth, cf. Fig. 2). Such a definition of E allows the proposed algorithm to skip pixels representing foreground moving objects during the calculation of color component offset. Stationary foreground objects are treated as the background.

FIGURE 2. The 226$^{\text{th}}$ frame of the sequence SoccerArc (left), decimated background frame calculated for 250 frames (right top) and set E derived for frame 226 (right bottom); set E contains all pixels except for white areas.

Obviously, such an approach requires cameras to be stationary for at least half of the duration of the video.

In the third step, all the pixels of each full-resolution frame V_i are modified by adding corresponding color component offset using saturation arithmetic within a range from 0 to $2^b - 1$:

$$V_f^{c\prime}(x,y) = \min\left(\max\left(V_f^c(x,y) + o_f^c, 0\right), 2^b - 1\right) \quad (11)$$

After this step, the color characteristics of frames V_0', V_1', etc., are consistent in time within the entire sequence.

V. INTER-VIEW COLOR CONSISTENCY IMPROVEMENT

In the second step of the proposed color refinement, the consistency between different views is being improved. This part of the proposal concerns two major problems in sequences captured by multicamera systems, i.e., different color characteristics of camera matrices and non-Lambertian surfaces of objects in the scene.

The inter-view consistency is crucial for immersive video systems as it highly influences the subjective quality of

synthesized virtual views presented to the immersive video system user – when various input views have inconsistent color characteristics, the user may notice annoying artifacts in the sequence. Moreover, color-consistent views are also necessary for good-quality depth estimation. For natural scenes, depth maps are estimated using views captured by a multiview camera system. Lack of inter-view consistency decreases the quality of depth maps, therefore, the subjective quality of sequences watched by the user is also inferior [73].

In the proposed inter-view consistency improvement technique, all input views are processed independently in order to align their color characteristics to one of the views (reference view r, e.g., the central view). The scheme of processing one frame of input view i is presented in Fig. 3.

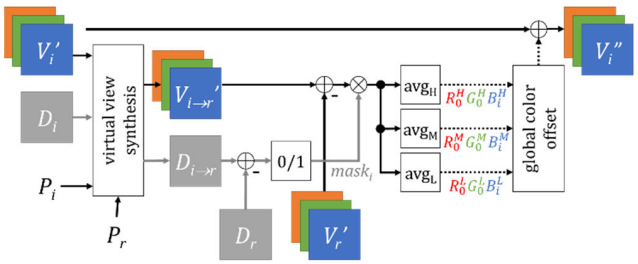

FIGURE 3. Scheme of the technique for inter-view consistency improvement: white boxes denote operations, color boxes contain signal symbols, "avg" means the average over all pixels of the frame, V_f – view at frame f, V_f'' – view V_f' after refinement. Each frame V_f consists of three color component planes: V_f^R, V_f^G and V_f^B.D_i and D_r – depth maps corresponding to views V_i and V_r; P_i and P_r – camera parameters valid for the acquisition of views V_i and V_r.

The proposed technique is based on reprojecting of each view i to the position of the reference view r. Therefore, three types of data are used for camera i: view V_i' (previously refined in the temporal consistency improvement step), depth D_i, and camera parameters P_i (intrinsic and extrinsic). These data, together with camera parameters of reference camera r (P_r) are fed into the virtual view synthesis step.

The view synthesis reprojects view and depth i to the position of the reference camera r ($V_{i\to r}'$ and $V_{i\to r}$). Then, the difference between all the pixels of views V_r' and $V_{i\to r}'$ is calculated (independently for each color component).

Proper estimation of color difference between two views requires a constraint: colors of pixels that represent different objects should not be compared. Therefore, pixels whose reprojected depth value in $D_{i\to r}$ is significantly different from the value in the reference depth map D_r are omitted. The difference between $D_{i\to r}$ and D_r is binarized and becomes a $mask_i$, which contains a value 1 for pixels representing the same object in both views and 0 otherwise. Then, differences between V_r' and $V_{i\to r}'$ are intersected with $mask_i$.

In the next step, these differences are aggregated separately for three equal ranges of each component intensity: c^L, c^M and c^H, for low, medium, and high intensity of color component c, respectively (i.e., for 10-bit components these ranges are: [0, 341), [341, 683) and [683, 1024)). For example, if component G of a pixel was equal to 250, the difference for that pixel is aggregated within G^L range. Then, the aggregated difference for each range is divided by the number of aggregated pixels to obtain an average value. At the end of this step, 3 global color offsets for each color component R, G, and B are obtained.

Therefore, in order to avoid color artifacts for pixels with a color component value close to range boundaries, calculated offsets are not simply added to the input view, but a range overlapping is performed (Fig. 4, left). In such an approach, global color offsets within an overlap are calculated as an average of color offsets of two neighboring ranges weighted by a distance to each range.

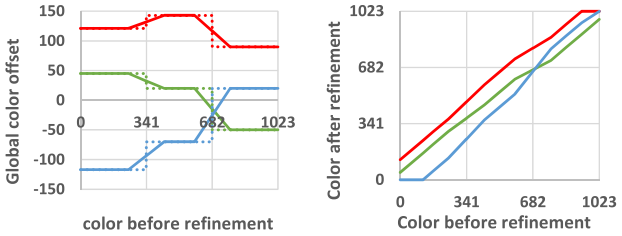

FIGURE 4. Left: exaggerated example of global color offsets for 3 color components, with (solid line) and without (dotted line) range overlapping. Right: color transform functions for 3 color components.

In the last step, all the pixels of view V_i' are modified using the color transform functions (Fig. 4, right), thus by adding proper global color offset for each color component:

$$V_i^c''(x,y) = \min\left(\max\left(V_i^{c'}(x,y) + o_i^c\left(V_i^{c'}(x,y)\right), 0\right), 2^b - 1\right) \quad (12)$$

where b is the bit depth of input view.

After this step, the color characteristics of the processed frame of view i are consistent with the reference view. To obtain color consistency for the entire sequence, this algorithm has to be performed for all the frames. However, if the common area of the scene visible in processed view i and reference view r is small, such an approach may introduce flickering artifacts (when global color offsets dramatically change between two consecutive frames). In order to reduce the possibility of flickering, global color offsets are additionally processed in the time domain. In the proposed approach, global color offsets are filtered using a simple IIR filter:

$$o_{i,f}^{c'} = \frac{o_{i,f}^c}{2} + \frac{o_{i,f-1}^{c'}}{2} \quad (13)$$

where $o_{i,f}^c$ is the global color offset for view i, color component c and frame f and $o_{i,f}^{c'}$ is the same offset after filtration.

VI. SOFTWARE IMPLEMENTATION

The above-described method is implemented as C++ software provided for use in further research and developing, testing, and promulgating technology standards developed by the ISO/IEC JTC1/SC29/WG4 MPEG VC standardization group. The software can be downloaded together with a

manual, configuration examples, and license details from the following repository:

https://gitlab.com/adziembowski/PoznanColorRefinement

VII. METHODOLOGY

Three experiments have been conducted in order to test the performance of the proposed color refinement method. In the first and the second experiment, the impact on the subjective quality of the synthesized views was assessed. In the third experiment, the influence of color refinement on depth estimation was studied.

A. QUALITY ASSESSMENT FOR IMMERSIVE VIDEO

The proposed color refinement method is meant for application in immersive video systems. Therefore, in the first experiment proposed method was tested on a set containing multiview sequences captured by multicamera systems.

The goal of any immersive video system is to allow a user to feel completely immersed in the scene. This is possible only if the subjective quality of the presented video is sufficiently good. Therefore, in order to assess the quality of synthesized views, subjective quality tests were performed.

In these tests, the PairComparison (PC) method [28], [27] was used. In each test, the participants were watching two videos containing the same virtual trajectory viewed by the user side-by-side. On the one side, the color-corrected video was presented, while on the other, the ''anchor'' – video rendered using input, non-corrected input views. Both the videos were presented in randomized order. The participants had to decide whether the left video one is better, worse, or both videos have the same quality. The total duration of the viewing session did not exceed the values described in the respective ITU recommendations [27].

In total, 70 volunteers have participated in the viewing tests. 5 of them were rejected because of the inconsistency of their results. In the tests, 49 male and 21 female volunteers participated. Their average age was 23.4 years with the standard deviation of 2.18 years. The volunteers are university students for the curriculum of electronics and telecommunications. They gained basic knowledge on general image processing, but not on immersive video and color correction algorithms, thus they can be treated as naïve viewers.

As described in [55], [33], the PC method performs better than ACR or DCR methods if there is a high diversity of quality and content between various sequences, as it allows to better differentiate the results for very miscellaneous test sets.

The proposed method was compared to two other color refinement methods based on histogram matching [21], which is the most widespread type of multi-view color-correction method [72]. The first one, described in [39], uses a hierarchical structure of reference views. The color components of input views are adjusted by the alignment of centroids of their histograms. The second method is histogram matching with a single reference view in the implementation of [46]. In this method, the Gaussian-filtered histogram is divided into three groups, whose ranges are dependent on the position of the maximum peak. In the end, these groups are used to linearly convert the color component of input views in order to match the reference view.

The proposed method was compared with [39] and [46] methods to show the advantages of the temporal consistency color refinement. The abovementioned methods provide a very high quality of refinement for still images (what was proved in the conducted experiment in Section VIII.A) but do not utilize the temporal information.

The test set contained 4 natural multiview sequences, described in Table 1.

TABLE 1. Natural multiview sequences test set.

Sequence	*Intel Frog*	*Poznan Fencing*	*Ballet*	*Soccer Arc*
Resolution	1920×1080	1920×1080	1024×768	1920×1080
Frames per second	30	25	15	25
Number of frames	300	250	100	250
Number of cameras	13	10	8	7
Camera arrangement	linear	on arc	on arc	on arc
Source	[44], [10]	[48], [10]	[5]	[43]

These sequences are widely used in research on multiview video processing and for the development of the standards on immersive video [10]. What is very important for the conducted experiment, the chosen sequences contain spatially and temporally inconsistent views.

B. QUALITY ASSESSMENT FOR MIDDLEBURY DATASET

Despite the proposed method is designed for immersive video, it can be used also for much simpler cases, e.g., stereoscopic camera systems capturing static scenes. Therefore, in the second experiment, the proposed method was tested using the Middlebury 2014 stereo dataset [34], which contains the number of color-consistent image pairs (captured in the same lighting conditions, using the same exposures, etc.), and also inconsistent views (right image captured with different exposures). Moreover, for each view, the ground-truth depth map is available, what reduces the influence of depth artifacts on the subjective quality of the synthesized virtual view.

10 stereopairs of images were used for quality assessment: *Adirondack*, *Jadeplant*, *Motorcycle*, *Piano*, *Pipes*, *Playroom*, *Playtable*, *Recycle*, *Shelves,* and *Vintage*. For each stereopair, the virtual view placed in the middle between two input views was synthesized.

The size of the test set could be increased by adding more stereopairs from older Middlebury stereo datasets, as they also contain views captured with different exposures and lighting conditions, however, the resolution of images in Middlebury 2014 datasets [34] is much higher, close to the

typical resolutions of immersive video, making this database highly relevant to the scope of the method.

For each test image, 5 virtual views were synthesized, using different input views:

1. reference (color-consistent views),
2. anchor (right view captured with different exposure),
3. refined using [39],
4. refined using [46],
5. refined using the proposed method.

For all color refinement methods, color characteristics of the right view were changed to align to the left view.

In order to assess the subjective quality, the ACR method [28] was used, however, participants rated static images instead of videos. Each of 70 participants was assessing the quality of images refined using each refinement method, reference, and anchor. 6 of 70 participants were rejected because of the inconsistency of their results.

The same group of viewers participated in both experiments. However, it presumably did not have a large impact on the results because of very different characteristics of content being watched (videos vs. static images) and the experiment itself (two different subjective quality assessment methods: PC and ACR).

C. IMPACT ON DEPTH ESTIMATION

In the last experiment, the influence of color refinement on depth map estimation was tested. Because of ground-truth depth maps availability, the Middlebury 2014 stereo dataset [34] was used.

In order to obtain good-quality depth maps, the publicly available depth estimation software [11] was used. This method is adapted to the requirements of the virtual navigation immersive systems, as it provides inter-view consistent depth maps which can be used to synthesize virtual views of satisfying quality. The method is based on superpixel segmentation, the size of segments is used to control the trade-off between the quality of depth maps and the processing time of depth estimation, making it very versatile and useful for applications described in this paper.

The quality of depth maps estimated using color-corrected views was compared to the quality of depth maps estimated using color-consistent views. The quality was estimated objectively, using 4 metrics used in Middlebury Stereo Evaluation: *avgerr* (average depth error measured in disparity levels) and *bad1*, *bad2,* and *bad4* (percentage of pixels with depth error higher than 1, 2, and 4, respectively).

The proposed color refinement method requires depth maps for reprojecting one view into another. Therefore, color refinement and depth estimation were performed iteratively 5 times. The first iteration of color refinement used depth maps D_L^0 and D_R^0 estimated using inconsistent input views. Then, these views were used for depth estimation. These depth maps (D_L^1 and D_R^1) were used for the second iteration of color refinement, etc. In order to avoid error propagation, in each iteration, the input views were being refined. The process is presented in Fig. 5.

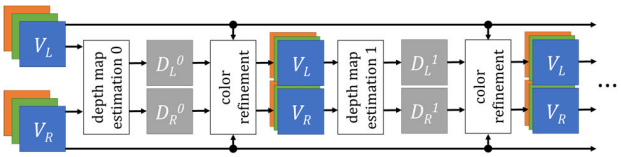

FIGURE 5. Experiment scheme: first two iterations.

VIII. EXPERIMENTAL RESULTS

As described in the previous section, three experiments were performed for testing of proposed color refinement method efficiency. Results are presented in Sections VII.A – VII.C, while in Section VII.D, the computational time required for color refinement is discussed.

A. QUALITY ASSESSMENT FOR IMMERSIVE VIDEO

In the first experiment, the performance of the proposed method in immersive video applications was tested.

Participants of subjective tests were evaluating the quality of sequences of virtual views arranged along a trajectory of virtual movement within a scene. Each virtual view was synthesized using input views refined using three color refinement methods: [39], [46], and the proposal. The quality of these sequences was compared to the quality of the sequence that contains virtual views synthesized using non-refined input views (“anchor”).

Each participant could score any sequence using a 3-step scale: $\{-1, 0, 1\}$, where 1 means that refined sequence is subjectively better than anchor, -1 indicates that anchor is better, and 0 that subjective quality of both sequences is indistinguishable.

Fig. 6 contains results averaged over all test sequences. Error bars represent the 95% confidence interval, calculated according to ITU-R recommendations [27] as:

$$CI = 1.96\frac{SD}{\sqrt{N}} \tag{14}$$

where CI is the confidence interval, SD – standard deviation, and N – number of participants.

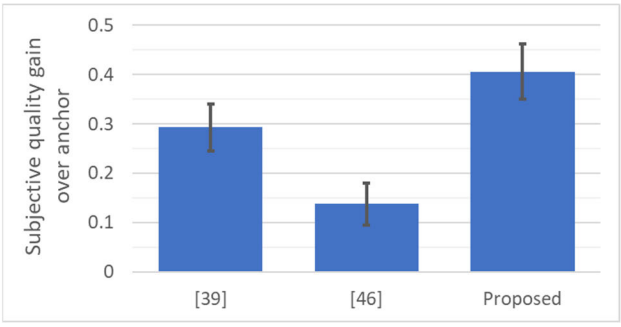

FIGURE 6. Subjective quality results evaluated over 4 multiview test sequences (PairComparison: from −1 to 1).

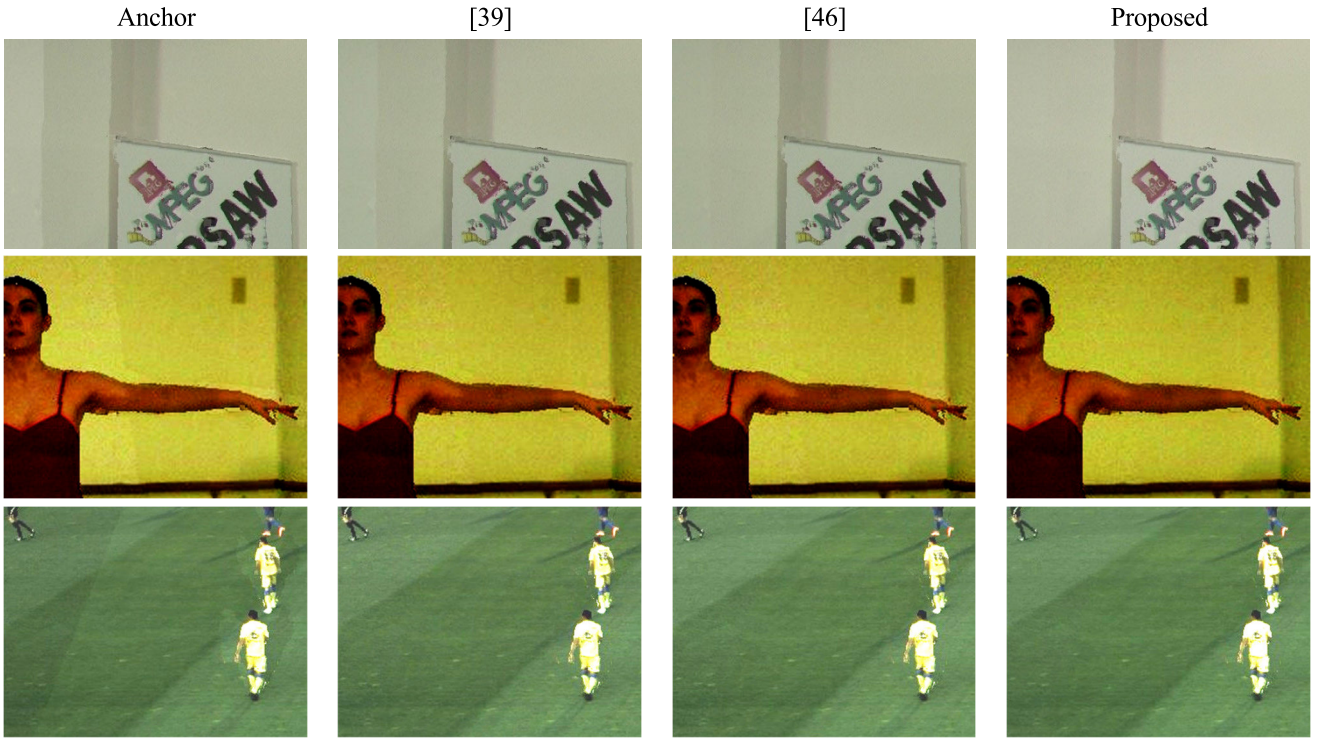

FIGURE 7. Fragment of virtual view synthesized using input views corrected by different methods; 1st row: *PoznanFencing*, 2nd row: *Ballet* (with contrast increased by 70%), 3rd row: *SoccerArc* (with contrast and brightness increased by 40%).

All tested color refinement methods allow achieving a statistically significant quality increase, as non-overlapping 95% confidence intervals indicate, that differences are statistically significant with a p-value much lower than 0.05 [75]. However, the proposed method outperforms other tested methods and the quality increase over both methods is also statistically significant.

In Fig. 7, fragments of virtual views synthesized using input views refined by different techniques are presented. On the left, in fragments of virtual views synthesized using non-refined input views, color artifacts are visible (non-existent edges between lighter and darker parts of the background, or annoying unnatural ''shadows'' of football players). In columns 2 – 4, the effects of three tested refinement methods are presented.

As shown, all methods increase color consistency, but the proposed method is the only one that allows eliminating inter-view color differences.

In Fig. 8, the temporal consistency of input views before and after color refinement is presented. Each picture in Fig. 8 consists of the left half containing information from one frame and the right half from another frame.

Pictures in the left column were created using non-refined, input views, pictures at the right – using views refined using the proposed color refinement method. In general, the vertical edge between the two halves is much less noticeable when refined views are used. Therefore, the proposed color refinement method allows to significantly increase temporal consistency of input views.

FIGURE 8. Temporal consistency improvement: left and right half of each frame contains information from different frames. Before (left column) and after proposed color correction (right). PoznanFencing (v4, frames 0/120), Ballet (v0, f. 0/96), IntelFrog (v13, f. 100/232).

In order to present temporal stability of color characteristics after color refinement, a simple additional comparison was performed. A 64 × 64 block of one chosen view was analyzed over time. The position of the analyzed block was arbitrarily chosen for each sequence to ensure, that it does not contain any moving objects (or their shadows). Then, for all

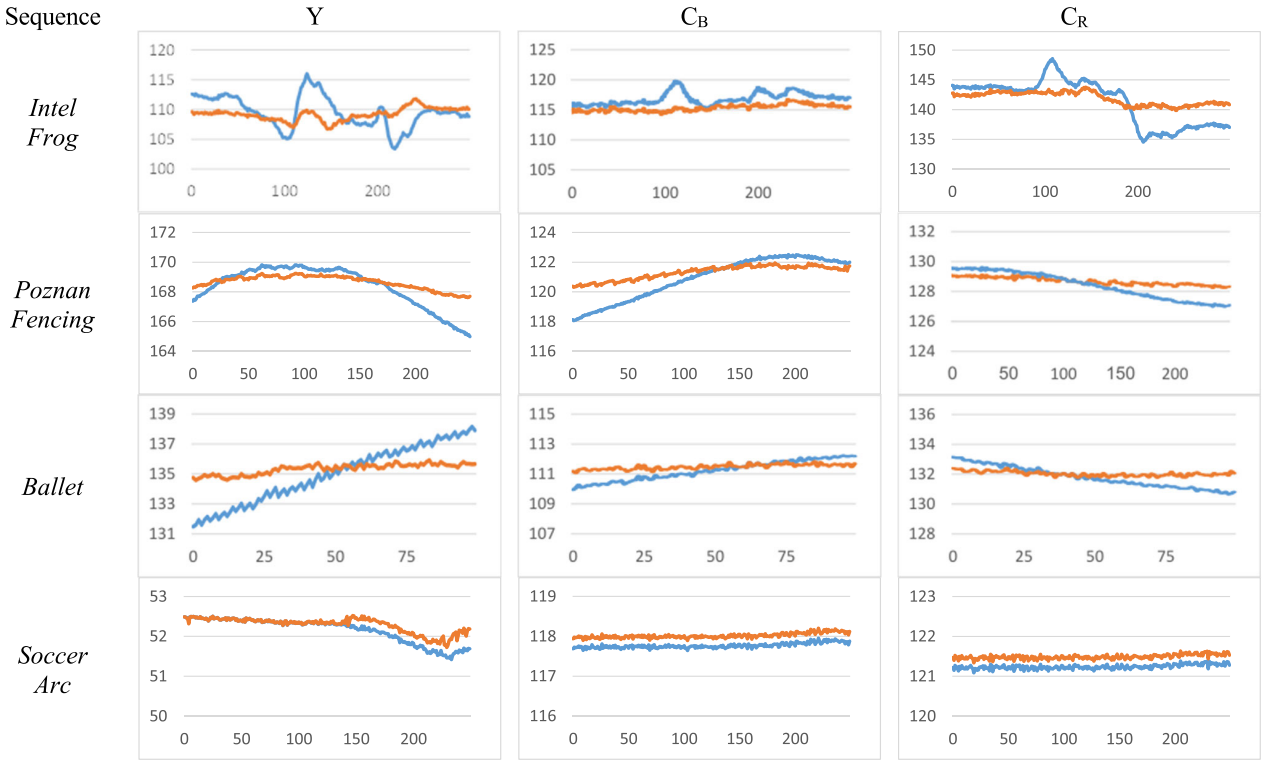

FIGURE 9. Changes of average Y, C_B, C_R color component of static background fragment over time (horizontal axes: frame number). Blue line: without refinement, orange: after proposed color refinement.

frames, each color component was averaged over the entire block.

If the color characteristics of the sequence is stable in time, the average color of an analyzed block should not change at all. Measured average color components for all test sequences are presented in Fig. 9. The input data was 8-bit, thus each pixel has a value in the range [0, 255]. Blue lines represent average colors of input views, orange ones – views refined using the proposed method. As presented, the temporal stability of refined views is significantly higher than for input views.

Table 2 contains the standard deviation for results presented in Fig. 9. In Table 3 the maximum difference of average color within the entire sequence is presented.

For 3 out of 4 test sequences, the proposed method allows improving temporal consistency of the input sequence. The only exception is the *SoccerArc* sequence, which has very temporally stable color characteristics. However, the case of *SoccerArc* shows, that the proposed method does not decrease the consistency of an already consistent sequence.

TABLE 2. The standard deviation of average color components within the entire sequence: I – input view, R – refined view.

Sequence	Y		C_B		C_R	
	I	R	I	R	I	R
IntelFrog	2.04	0.49	0.36	0.26	0.49	0.27
PoznanFencing	0.70	0.25	0.77	0.30	0.20	0.11
Ballet	1.92	0.35	0.66	0.17	0.71	0.16
SoccerArc	0.05	0.05	0.03	0.03	0.04	0.04

TABLE 3. The maximum difference of average color components within the entire sequence: I – input view, R – refined view.

Sequence	Y		C_B		C_R	
	I	R	I	R	I	R
IntelFrog	7.47	1.91	1.96	1.4	3.09	1.25
PoznanFencing	2.44	0.97	2.66	1.13	0.8	0.48
Ballet	6.65	1.32	2.29	0.77	2.48	0.64
SoccerArc	0.23	0.22	0.16	0.16	0.21	0.21

B. QUALITY ASSESSMENT FOR MIDDLEBURY DATASET

In the second experiment, the proposed color refinement method was tested on the Middlebury 2014 stereo dataset [34]. Each stereopair of the images in the dataset consists of three images: captured by the left camera (im0), captured by the right camera in the same lighting conditions and camera setup (im1), and right captured with different exposure (im1E).

Of course, because of the content type (static image pairs), only the inter-view consistency improvement technique was evaluated. In the test, the quality of the virtual view synthesized in the middle between the left and right input view was evaluated.

In Fig. 10, the subjective quality results averaged over all 10 test images are presented. Each of 70 participants was assessing image quality using a typical MOS scale (1 – 5),

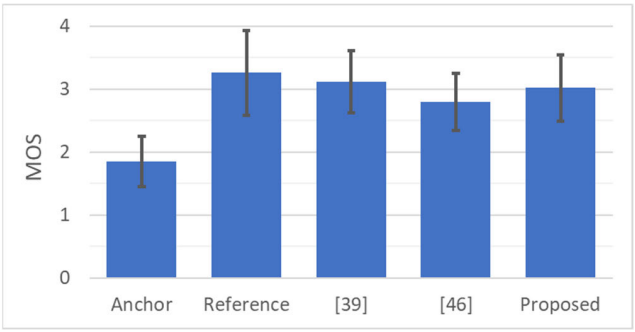

FIGURE 10. Subjective quality results evaluated over 10 test images (MOS: 1 – 5).

where 1 means bad and 5 means excellent quality. Error bars represent the 95% confidence interval.

The first bar represents the quality of ''anchor'' – images synthesized using inconsistent input views (im0 and im1E – with changed exposure). The result for consistent input views (im0 and im1 for each test image) is presented within the second bar (''reference''). Bars 3 – 5 contain results for 3 tested color refinement methods. As presented, all 3 methods perform similarly. They outperform anchor and have a similar quality to the reference (with a 5% significance level there is no statistically significant difference between these results).

In Fig. 11, the virtual views for 3 test images are presented. As shown in the first column, where many annoying color artifacts appear, color refinement methods had to deal with significantly different input views, but they were able to properly unify the color characteristics of both views.

Figs. 10 and 11 suggest that the proposed method performs similarly to two state-of-the-art methods. However, it has to be noted, that Middlebury 2014 stereo dataset is a very specific case of the immersive video system. There are only two views, both cameras were placed in parallel, close to each other. Moreover, input views are static (only one frame) and rectified (no rotations, just 1-dimensional shift between two cameras). All of these reasons make Middlebury dataset easy to process when compared to multiview sequences acquired by practical immersive video systems. And, as shown in the previous section, the proposed method performs much better for more difficult wide-baseline content that can contain also temporally unstable color characteristics.

C. IMPACT ON DEPTH ESTIMATION

The Middlebury 2014 stereo dataset was used also in the third experiment. However, instead of the subjective quality of synthesized views, in this experiment, the impact of color refinement on depth estimation was evaluated.

As mentioned in Section VII.C, in order to obtain good-quality depth maps, the publicly available superpixel-based depth estimation software [11] was used.

Depth maps estimated for 3 of 10 test images are presented in Fig. 12. Depth maps in the first column were estimated using color-consistent input views. Under each depth map, the depth error (compared to the ground-truth depth map) averaged over the entire frame is shown.

In the second column of Fig. 12, depth maps estimated using inconsistent input views are presented. These results are completely incorrect – depth for most of the scene is wrong, what would obviously affect the quality of synthesized views.

In order to estimate the depth map, the corresponding areas in different input views have to be found. This operation is much more difficult (or even impossible) if different views have significantly different color characteristics.

These depth maps were used for proposed color refinement in order to increase the inter-view consistency of both input views. Of course, because of bad quality, performing the color refinement was much harder, as only a small number of pixels were analyzed in the refinement process because of depth inconsistency. However, it still was able to significantly increase the inter-view consistency of both views.

In the third column of Fig. 12, depth maps estimated using refined views are presented. They still contain areas with wrong depth values, but their quality is much better than ones estimated using inconsistent views, therefore when used for the second iteration of color refinement, they would allow to additionally increase inter-view consistency.

In Figs. 13 and 14, changes of four quality metrics in consecutive iterations are presented. Fig. 13 contains the average depth error (averaged over the entire view). Fig. 14 presents the increase (compared to depth maps estimated using consistent views) of the number of pixels with depth error higher than 1, 2, or 4 disparity levels.

The last column of Fig. 12 contains depth maps estimated using input views refined iteratively 5 times. As presented, they are similar to ones estimated using consistent input views.

Values presented in Figs. 13 and 14 were averaged over all 10 test images. As presented, additional iterations allow to increase the quality of estimated depth maps, but even one single iteration significantly improves depth map quality. It indicates that the proposed color correction method, despite being based on the inter-view dependencies, is robust to reprojection errors which are mainly caused by the low accuracy of available depth maps.

As a result, the good-quality depth maps provide better subjective quality of synthesized views. Moreover, the newest multiview content encoding standards (e.g., MPEG Immersive Video – MIV [71]) highly depend on the quality of depth maps [73]. Therefore, the results of the third experiment indicate, that the proposed color correction technique will also allow to improve the compression ratio of the immersive video.

D. PROCESSING TIME

Besides the good quality and reliability of the refinement, the practical color refinement method should be fast in order to be usable in practical immersive video systems.

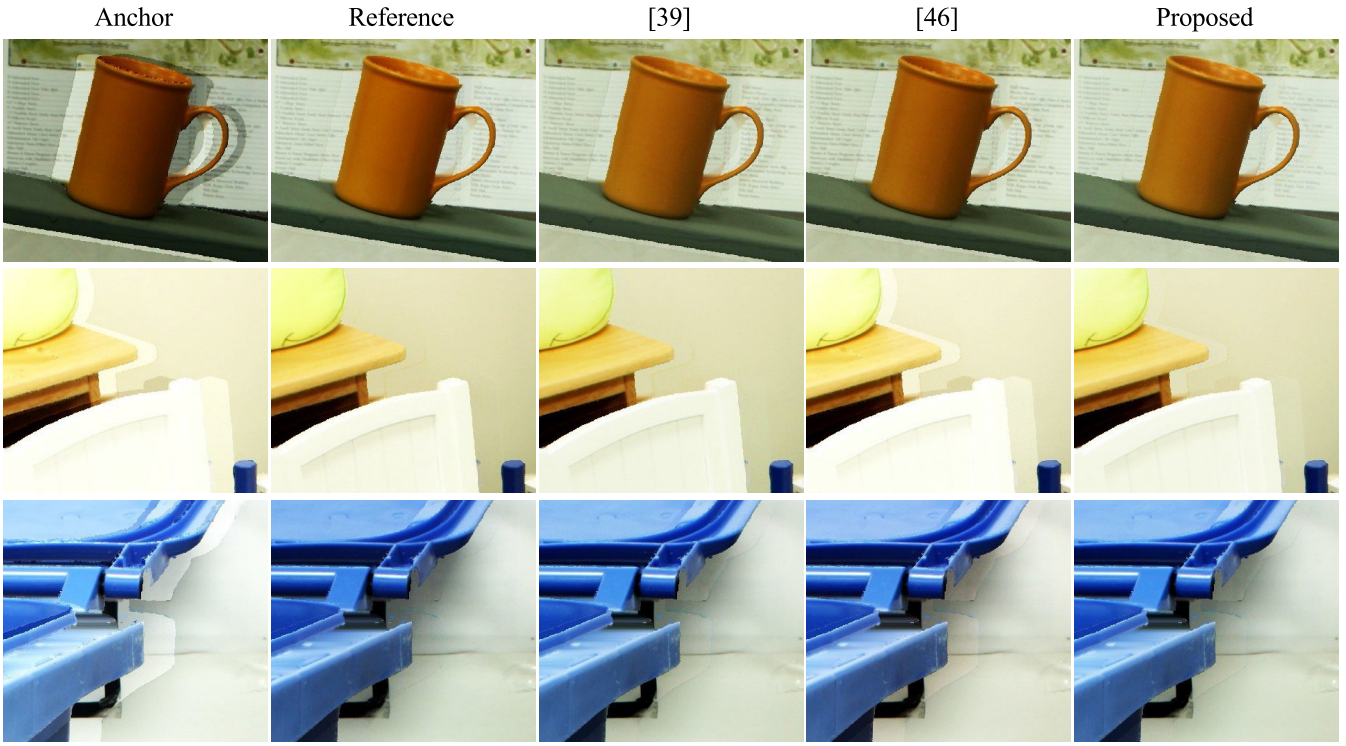

FIGURE 11. Fragments of virtual views synthesized using inconsistent input views (1st column), consistent views (2nd column), and inconsistent views corrected by different methods (columns 3 – 5); test images *Adirondack*, *Playtable*, *Recycle*.

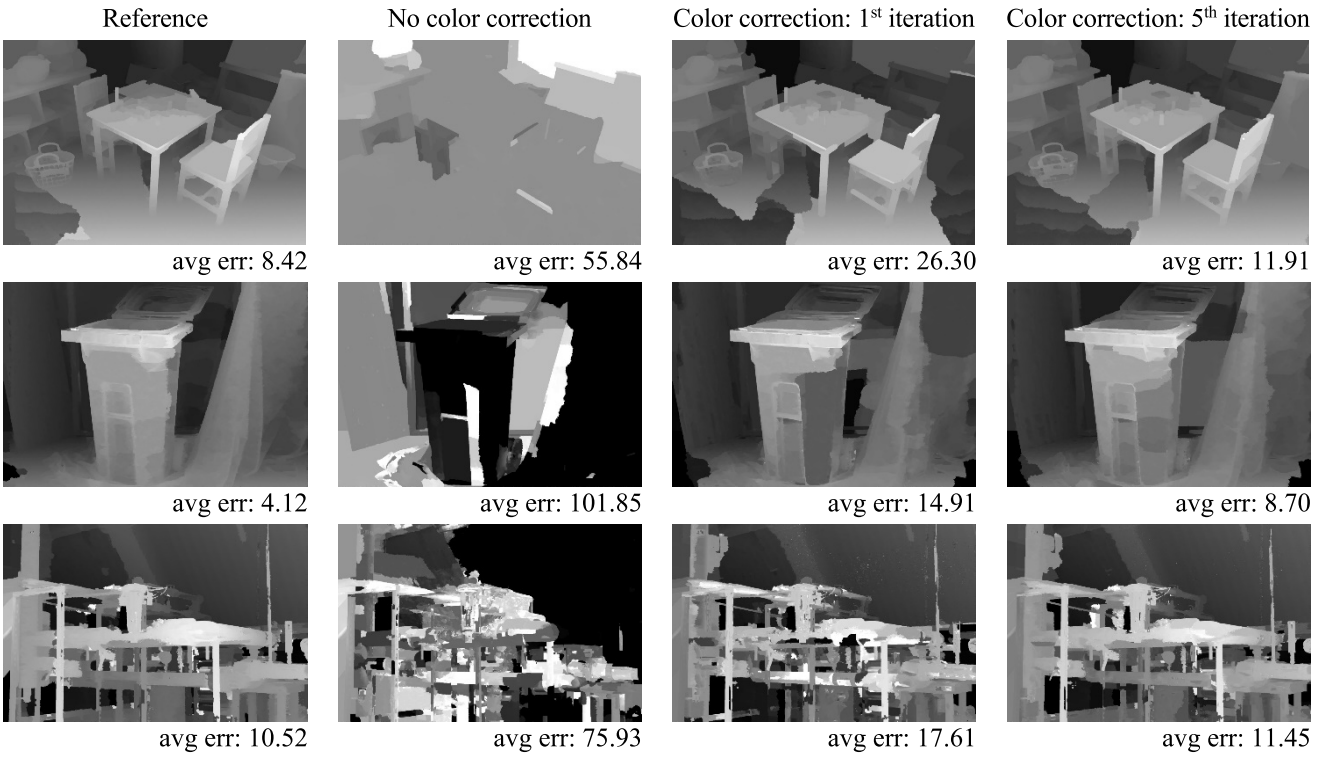

FIGURE 12. Depth maps estimated using color-consistent views (1st column) and non-consistent views before color refinement (2nd column) and after color refinement (3rd and 4th column); test images: *Playtable*, *Recycle*, *Pipes*.

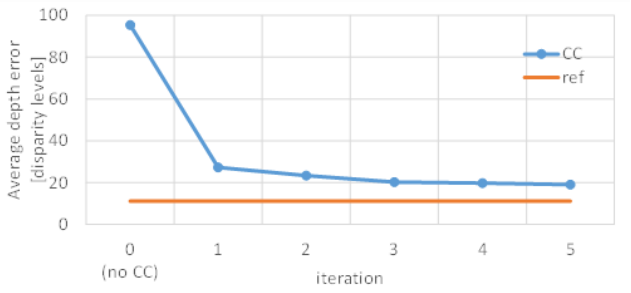

FIGURE 13. Average depth error in depth maps estimated using color-consistent views (orange line) and non-consistent views without and with color refinement (blue line).

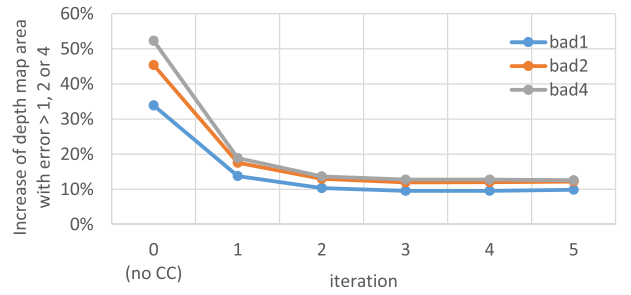

FIGURE 14. Decrease of the number of pixels with depth error higher than 1, 2, or 4 disparity levels over consecutive iterations; the number of pixels shown as an increase of area with wrong disparity compared to depth maps estimated using color-consistent views.

The proposed method meets this requirement. The average processing time of one frame of FullHD (1920 × 1080) input sequence is presented in Table 4.

TABLE 4. Mean processing time of one frame of the single view of FullHD input sequence.

Step of proposed color correction	Time [ms]
Temporal consistency improvement	9.24
View reprojection	572.26
Inter-view consistency improvement	39.08
Overall	620.59

Values in Table 4. were averaged over 100 frames of all views of 3 FullHD test sequences (*IntelFrog*, *PoznanFencing,* and *SoccerArc*). Calculations were performed on the desktop computer equipped with a 4.0 GHz CPU based on the "Skylake" microarchitecture, 32GB RAM, and SSD. The color correction software was compiled using VC16 on Windows 10.

Presented processing times were obtained for refinement of one input view. The proposed approach allows to refine all input views simultaneously, but each view is processed independently. Therefore, all the views can be refined in parallel thus presented times are valid also for the multiview case.

The highest computational time is needed for view reprojection. At this step, a non-optimized synthesis algorithm (Advanced View Syntheser [2]) was used because of its versatility (e.g., handling of perspective and omnidirectional views). However, at this moment there is virtually no omnidirectional natural content, therefore also the real-time synthesizer dedicated for perspective cameras (e.g. [14]) could be used in order to reduce view reprojection time to values smaller than 30 ms.

Assuming usage of real-time view synthesizer, processing of one frame requires ∼80 ms (2 times slower than real-time for 25 frames per second). However, if all the operations will be performed within a pipeline, each frame can be refined in less than 40 ms, thus the proposed color refinement algorithm can be considered as a **real-time algorithm**.

Tables 5 and 6 contain processing times of steps of temporal and inter-view consistency improvement techniques.

TABLE 5. Mean time of temporal consistency improvement of one frame of a single view of FullHD input sequence.

Step of the temporal consistency improvement	Time [ms]
Background calculation*	0.38
Refinement	8.86
Overall	9.24

TABLE 6. Mean time of inter-view consistency improvement of one frame of the single view of FullHD input sequence.

Step of the inter-view consistency improvement	Time [ms]
Offset calculation	30.34
Color transform function calculation	< 0.01
IIR filtration	< 0.01
Refinement	9.05
Overall	39.08

It has to be noted, that background calculation is performed as a preprocessing step, once for the entire sequence. Time presented in Table 5 was calculated as a time of background calculation, divided by the number of frames. To be clear, background calculation for a 100-frame sequence requires 38 ms on average, what means a one-frame delay in the real-time processing.

Calculation of color transform function and IIR filtration are negligible (both required several microseconds). Much higher computational time is needed for offset calculation and refinement. However, both operations can be easily parallelized. Values presented in Table 6 were obtained for the case when the entire frame is being processed on one thread. If the frame will be divided into slices (operations performed for each pixel are independent), the processing time will be significantly reduced (e.g., for 4 slices it will drop to ∼10 ms).

IX. CONCLUSION

The paper describes the effective method of color correction for immersive video applications. The proposal improves the consistency of color characteristics of multiview video in real time, focusing on correction of both temporal and inter-view consistencies, what makes is it particularly useful for immersive video applications. Moreover, the proposed temporal stability enhancement is not influenced by the temporal consistency of depth, as is performed independently in each view, making the proposal robust to typical low-quality depth maps of natural test sequences.

In order to evaluate the performance of the proposed method, understood as its influence on color stability and immersive video encoding efficiency, three experiments were conducted.

In the first experiment, the proposal was tested in its main application – the correction of natural multiview sequences – to assess the temporal and inter-view consistency improvement. In the second experiment, the inter-view consistency was evaluated using the Middlebury stereo database. In both experiments, a group of 70 participants assessed the subjective quality of synthesized virtual views. The proposal was compared to two state-of-the-art methods for the color correction of a multiview content. The results of subjective tests show that the proposal allows to highly improve the color consistency of synthesized virtual views. Moreover, for the multiview, immersive content, the proposed algorithm outperformed both tested state-of-the-art approaches due to the much higher temporal stability of the virtual views synthesized from the input views corrected using the proposal.

The influence of color correction of input images on depth estimation was also tested. The experiment showed that the use of the proposed method can significantly increase the quality of depth maps estimated for color-inconsistent input views. Therefore, the proposed color correction method improves not only the subjective quality of synthesized views but also the immersive video coding efficiency, as it highly depends on the quality and consistency of depth maps [71].

Due to its high performance and flexibility, the proposed algorithm was approved by the experts of the ISO/IEC MPEG group [42] as the MPEG reference software for color refinement for immersive video applications [1]. The implementation of the proposed method is also available on the public repository [74], allowing it to be a useful reference for future works on color correction and refinement in multiview systems.

REFERENCES

[1] *Manual of Color-Correction Tool*, document ISO/IEC JTC1/SC29/ WG11 MPEG/N19141, Brussels, Belgium, Jan. 2020.

[2] A. Dziembowski, D. Mieloch, O. Stankiewicz, M. Domanski, G. Lee, and J. Seo, "Virtual view synthesis for 3DoF+ video," in *Proc. Picture Coding Symp. (PCS)*, Ningbo, China, Nov. 2019, pp. 1–5.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[4] M. Domanski, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media—MPEG-I: 360 video, virtual navigation and beyond," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Poznań, Poland, May 2017, pp. 1–9.

[5] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600–608, Aug. 2004.

[6] B. Ceulemans, S.-P. Lu, P. Schelkens, and A. Munteanu, "Globally optimized multiview video color correction using dense spatio-temporal matching," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (DTV-CON)*, Lisbon, Portugal, Jul. 2015, pp. 1–4.

[7] B. Ceulemans, S.-P. Lu, G. Lafruit, and A. Munteanu, "Robust multiview synthesis for wide-baseline camera arrays," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2235–2248, Sep. 2018.

[8] O. Stankiewicz, K. Wegner, and M. Domanski, "Nonlinear depth representation for 3D video coding," in *Proc. IEEE Int. Conf. Image Process.*, Melbourne, VIC, Australia, Sep. 2013, pp. 1752–1756.

[9] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 31–41, Sep. 2001.

[10] *Common Test Conditions for Immersive Video*, document ISO/IEC JTC1/SC29/WG11 MPEG/N18789, Geneva, Switzerland, Oct. 2019.

[11] D. Mieloch, O. Stankiewicz, and M. Domanski, "Depth map estimation for free-viewpoint television and virtual navigation," *IEEE Access*, vol. 8, pp. 5760–5776, 2020.

[12] A. Dziembowski and M. Domański, "Adaptive color correction in virtual view synthesis," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Helsinki, Finland, Jun. 2018, pp. 1–4.

[13] M. Ebner, *Color Constancy*. Hoboken, NJ, USA: Wiley 2007.

[14] A. Dziembowski and J. Stankowski, "Real-time CPU-based virtual view synthesis," in *Proc. Int. Conf. Signals Electron. Syst. (ICSES)*, Kraków, Poland, Sep. 2018, pp. 78–82.

[15] E. Ekmekcioglu, V. Velisavljevic, and S. T. Worrall, "Content adaptive enhancement of multi-view depth maps for free viewpoint video," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 352–361, Apr. 2011.

[16] S. A. Fezza, M.-C. Larabi, and K. M. Faraoun, "Feature-based color correction of multiview video for coding and rendering enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1486–1498, Sep. 2014.

[17] G. Lafruit, M. Domański, K. Wegner, T. Grajek, T. Senoh, J. Jung, P. Kovács, P. Goorts, L. Jorissen, A. Munteanu, B. Ceulemans, P. Carballeira, S. García, and M. Tanimoto, "New visual coding exploration in MPEG: Super-MultiView and free navigation in free viewpoint TV," in *Proc. 27th IST Electron. Imag., Stereoscopic Displays Appl.*, San Francisco, CA, USA, 2016, pp. 1–9.

[18] P. Green and L. MacDonald, *Colour Engineering*. Hoboken, NJ, USA: Wiley, 2002.

[19] G. Miller, J. Starck, and A. Hilton, "Projective surface refinement for free-viewpoint video," in *Proc. 3rd Eur. Conf. Vis. Media Prod. (CVMP)*, London, U.K., 2006, pp. 153–162.

[20] *G-PCC Test Model V9 User Manual*, document ISO/IEC JTC 1/SC 29/WG 11 MPEG/N19083, Brussels, Belgium, Jan. 2020.

[21] U. Fecker, M. Barkowsky, and A. Kaup, "Histogram-based prefiltering for luminance and chrominance compensation of multiview video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, pp. 1258–1267, Sep. 2008.

[22] J. Ho, B. V. Funt, and M. S. Drew, "Separating a color signal into illumination and surface reflectance components: Theory and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 966–977, Oct. 1990.

[23] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," *Mach. Vis. Appl.*, vol. 27, no. 7, pp. 1005–1020, Oct. 2016.

[24] F. Hu, L. Ye, W. Zhong, L. Fang, and Q. Zhang, "Deep auxiliary learning for point cloud generation," *IEEE Access*, vol. 8, pp. 18538–18545, 2020.

[25] M. Domanski, A. Dziembowski, T. Grajek, A. Grzelka, K. Klimaszewski, D. Mieloch, R. Ratajczak, O. Stankiewicz, J. Siast, J. Stankowski, and K. Wegner, "Demonstration of a simple free viewpoint television system," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Sep. 2017, pp. 4589–4591.

[26] F. Isgro, E. Trucco, P. Kauff, and O. Schreer, "Three-dimensional image processing in the future of immersive media," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 288–303, Mar. 2004.

[27] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-9, ITU-R, Nov. 1998.

[28] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document Rec. ITU-T P.910, ITU-T, Apr. 2008.
[29] M. Kurc, O. Stankiewicz, and M. Domanski, "Depth map inter-view consistency refinement for multiview video," in *Proc. Picture Coding Symp.*, Kraków, Poland, May 2012, pp. 137–140.
[30] M. Kurc, "Hybrid techniques of depth map estimation and their application in three-dimensional video systems," Ph.D. dissertation, Fac. Electron. Telecommun., Poznań Univ. Technol., Poznań, Poland, 2019. [Online]. Available: http://www.multimedia.edu.pl/?page=publications_phd
[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[32] S.-P. Lu, B. Ceulemans, A. Munteanu, and P. Schelkens, "Spatio-temporally consistent color and structure optimization for multiview video color correction," *IEEE Trans. Multimedia*, vol. 17, no. 5, pp. 577–590, May 2015.
[33] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," *Comput. Graph. Forum*, vol. 31, no. 8, pp. 2478–2491, Dec. 2012.
[34] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. German Conf. Pattern Recognit. (GCPR)*, Münster, Germany, Sep. 2014, pp. 31–42.
[35] D. Mieloch, O. Stankiewicz, and M. Domański, "Depth map estimation for free-viewpoint television and virtual navigation," *IEEE Access*, vol. 8, pp. 5760–5776, 2020.
[36] K. Müller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proc. IEEE*, vol. 99, no. 4, pp. 643–656, Apr. 2011.
[37] Y. Niu, X. Zheng, T. Zhao, and J. Chen, "Visually consistent color correction for stereoscopic images and videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 697–710, Mar. 2020.
[38] M. Domanski, A. Dziembowski, D. Mieloch, A. Luczak, O. Stankiewicz, and K. Wegner, "A practical approach to acquisition and processing of free viewpoint video," in *Proc. 31st Picture Coding Symp. (PCS)*, Cairns, QLD, Australia, May 2015, pp. 10–14.
[39] J. Stankowski, K. Klimaszewski, O. Stankiewicz, K. Wegner, and M. Domański, *Preprocessing Methods Used for Poznan 3D/FTV Test Sequences*, document ISO/IEC JTC1/SC29/WG11 MPEG/M17174, Kyoto, Japan, Jan. 2010.
[40] O. Oyman, R. Koenen, P. Higgs, C. Johns, R. Mills, and M. O'Doherty, "Virtual reality industry Forum's view on state of the immersive media industry," *SMPTE Motion Imag. J.*, vol. 128, no. 8, pp. 91–96, Sep. 2019.
[41] L. Cui, R. Mekuria, M. Preda, and E. S. Jang, "Point-cloud compression: Moving picture experts Group's new standard in 2020," *IEEE Consum. Electron. Mag.*, vol. 8, no. 4, pp. 17–21, Jul. 2019.
[42] *Summary on MPEG-I Visual Activities*, document ISO/IEC JTC1/SC29/WG11 MPEG/N18994, Brussels, Belgium, Jan. 2020.
[43] P. Goorts, "The barcelona dataset in real-time adaptive plane sweeping free viewpoint navigation in soccer scenes," Ph.D. dissertation, Hasselt Univ., Hasselt, Belgium, Jul. 2014, pp. 181–186.
[44] B. Salahieh, B. Marvar, M. M. Nentedem, A. Kumar, V. Popvic, K. Seshadrinathan, O. Nestares, and J. Boyce, *Kermit Test Sequence for Windowed 6DoF Activities*, document ISO/IEC JTC1/SC29/WG11 MPEG/M43748, Ljubljana, Slovenia, Jul. 2019.
[45] C. Poynton, *Digital Video and HD*. San Mateo, CA, USA: Morgan Kaufmann, 2012.
[46] T. Senoh, N. Tetsutani, and H. Yasuda, *[MPEG-I Visual] Proposal of Trimming and Color Matching of Multi-View Sequences*, document ISO/IEC JTC1/SC29/WG11 MPEG/M47170, Geneva, Switzerland, Mar. 2019.
[47] D. Shin and Y. S. Ho, "Color correction using 3D multi-view geometry," *Proc. SPIE*, vol. 9395, Feb. 2015, Art. no. 93950O.
[48] M. Domański, A. Dziembowski, A. Grzelka, A. Łuczak, D. Mieloch, O. Stankiewicz, and K. Wegner, *Multiview Test Video Sequences for Free Navigation Exploration Obtained Using Pairs of Cameras*, document ISO/IEC JTC1/SC29/WG11 MPEG/M38247, Geneva, Switzerland, Jun. 2016.
[49] A. Smolic, K. Müller, P. Merkle, M. Kautzner, and T. Wiegand, "3D video objects for interactive applications," in *Proc. 13th Eur. Signal Process. Conf.*, Antalya, Turkey, 2005, pp. 1–4.
[50] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "FTV for 3-D spatial communication," *Proc. IEEE*, vol. 100, no. 4, pp. 905–917, Apr. 2012.
[51] N. Vretos and P. Daras, "Temporal and color consistent disparity estimation in stereo videos," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 3798–3802.
[52] *Test Model 4 for Immersive Video*, document ISO/IEC JTC 1/SC 29/WG 11 MPEG/N19002, Brussels, Belgium, Jan. 2020.
[53] O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, and J. Samelak, "A free-viewpoint television system for horizontal virtual navigation," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2182–2195, Aug. 2018.
[54] H.-C. Shih and H.-F. Hsiao, "A depth refinement algorithm for multi-view video synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2010, pp. 742–745.
[55] A. M. van Dijk and J.-B. Martens, "Subjective quality assessment of compressed images," *Signal Process.*, vol. 58, no. 3, pp. 235–252, May 1997.
[56] *V-PCC Test Model V9*, document ISO/IEC JTC 1/SC 29/WG 11 MPEG/N19085, Brussels, Belgium, Jan. 2020.
[57] B. A. Wandell, "The synthesis and analysis of color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 1, pp. 2–13, Jan. 1987.
[58] K. C. Wei, Y. L. Huang, and S. Y. Chien, "Point-based model construction for free-viewpoint TV," in *Proc. IEEE 3rd Int. Conf. Consum. Electron. Berlin (ICCE-Berlin)*, Berlin, Germany, Sep. 2013, pp. 220–221.
[59] P. Wu, Y. Liu, M. Ye, J. Li, and S. Du, "Fast and adaptive 3D reconstruction with extensively high completeness," *IEEE Trans. Multimedia*, vol. 19, no. 2, pp. 266–278, Feb. 2017.
[60] W. Xu and J. Mulligan, "Performance evaluation of color correction approaches for automatic multi-view image and video stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, Jun. 2010, pp. 263–270.
[61] K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, "Multiview video coding using view interpolation and color correction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1436–1449, Nov. 2007.
[62] S. Ye, S.-P. Lu, and A. Munteanu, "Color correction for large-baseline multiview video," *Signal Process., Image Commun.*, vol. 53, pp. 40–50, Apr. 2017.
[63] G. Finlayson, M. M. Darrodi, and M. Mackiewicz, "Rank-based camera spectral sensitivity estimation," *J. Opt. Soc. Amer.*, vol. 33, no. 4, pp. 589–599, 2016.
[64] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, "Color image processing pipeline," *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 34–43, Jan. 2005.
[65] L. Fang, W. Zhong, L. Ye, R. Li, and Q. Zhang, "Light field reconstruction with a hybrid sparse regularization-pseudo 4DCNN framework," *IEEE Access*, vol. 8, pp. 171009–171020, 2020.
[66] D. Mieloch and A. Grzelka, "Segmentation-based method of increasing the depth maps temporal consistency," *Int. J. Electron. Telecommun.*, vol. 64, no. 3, pp. 293–298, 2018.
[67] L. Yao, Y. Han, and X. Li, "Fast and high-quality virtual view synthesis from multi-view plus depth videos," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 19325–19340, 2019.
[68] I. Ganelin and P. Nasiopoulos, "Virtual view color estimation for free viewpoint TV applications using Gaussian mixture model," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 3898–3902.
[69] Y. Qiao, L. Jiao, S. Yang, B. Hou, and J. Feng, "Color correction and depth-based hierarchical hole filling in free viewpoint generation," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 294–307, Jun. 2019.
[70] R. G. Rodríguez, J. Vazquez-Corral, and M. Bertalmio, "Color matching images with unknown non-linear encodings," *IEEE Trans. Image Process.*, vol. 29, pp. 4435–4444, 2020.
[71] *Text of ISO/IEC CD 23090-12 MPEG Immersive Video*, document ISO/IEC JTC1/SC29/WG11 MPEG/N19482, Jul. 2020.
[72] C. Ding and Z. Ma, "Multi-camera color correction via hybrid histogram matching," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Nov. 17, 2020, doi: 10.1109/TCSVT.2020.3038484.
[73] D. Mieloch, A. Dziembowski, and M. Domanski, "Depth map refinement for immersive video," *IEEE Access*, vol. 9, pp. 10778–10788, 2021.
[74] *Git Repository*. Accessed: May 19, 2021. [Online]. Available: https://gitlab.com/adziembowski/PoznanColorRefinement
[75] I. MacGregor-Fors and M. E. Payton, "Contrasting diversity values: Statistical inferences based on overlapping confidence intervals," *PLoS ONE*, vol. 8, no. 2, Feb. 2013, Art. no. e56794.

ADRIAN DZIEMBOWSKI was born in Poznań, Poland, in 1990. He received the M.Sc. and Ph.D. degrees from the Poznan University of Technology, in 2014 and 2018, respectively. Since 2019, he has been an Assistant Professor with the Institute of Multimedia Telecommunications. He authored and coauthored about 30 articles on various aspects of immersive video, free navigation, and free viewpoint television systems. He is actively involved in ISO/IEC MPEG activities towards future MPEG immersive video coding standard.

DAWID MIELOCH received the M.Sc. and Ph.D. degrees from the Poznań University of Technology, in 2014 and 2018, respectively. He is currently an Assistant Professor with the Institute of Multimedia Telecommunications. He is actively involved in ISO/IEC MPEG activities, where he contributes to the development of the immersive media technologies. He has been involved in several projects focused on multiview and 3-D video processing. His professional interests include free-viewpoint television, depth estimation, and camera calibration.

SŁAWOMIR RÓŻEK was born in 1994. He received the first M.Sc. degree in electronics and telecommunication and the second M.Sc. degree in automatics and robotics from the Poznań University of Technology, in 2018 and 2020, respectively. He is involved in ISO/IEC MPEG activities towards future MPEG video coding for machines standard. His professional interests include multiview and 3-D video, microcontrollers, and electronic circuits design.

MAREK DOMAŃSKI (Senior Member, IEEE) received the M.Sc., Ph.D., and Habilitation degrees from the Poznań University of Technology, Poland, in 1978, 1983, and 1990, respectively. Since 1993, he has been a Professor with the Poznań University of Technology, where he leads the Institute of Multimedia Telecommunications. He coauthored one of the very first AVC decoders for tv set-top boxes, in 2004, highly ranked technology proposals to MPEG for scalable video compression, in 2004, 3-D video coding, in 2011, and immersive video coding, in 2019. He authored three books and more than 300 articles in journals and conference proceedings. His contributions were mostly on image, video and audio compression, virtual navigation, free-viewpoint television, image processing, multimedia systems, 3-D video and color image technology, digital filters, and multidimensional signal processing. He served as a member of various steering, program, and editorial committees of international journals and international conferences. He was the General Chairman/the Co-Chairman and the Host of several international conferences, including Picture Coding Symposium, PCS 2012; IEEE International Conference on Advanced Vedio and Signal Based Surveillance, AVSS 2013; European Signal Processing Conference, EUSIPCO 2007; 73rd and 112nd Meetings of MPEG; International Workshop on Signals, Systems and Image Processing, IWSSIP 1997 and 2004; and International Conference Signals and Electronic Systems, ICSES 2004.