Asymmetric Coding of Multi-View Video Plus Depth Based 3-D Video for View Rendering

Feng Shao, Gangyi Jiang, Mei Yu, Ken Chen, and Yo-Sung Ho, *Senior Member, IEEE*

***Abstract*—The recent years have witnessed three-dimensional (3-D) video technology to become increasingly popular, as it can provide high-quality and immersive experience to end users, where view rendering with depth-image-based rendering (DIBR) technique is employed to generate the virtual views. Distortions in depth map may induce geometry changes in the virtual views, and distortions in texture video may be propagated to the virtual views. Thus, effective compression of both texture videos and depth maps is important for 3-D video system. From the perspective of bit allocation, asymmetric coding of the texture videos and depth maps is an effective way to get the optimal solution of 3-D video compression and view rendering problems. In this paper, a novel asymmetric coding method of multi-view video plus depth (MVD) based 3-D video is proposed on purpose of providing high-quality view rendering. In the proposed method, two models are proposed to characterize view rendering distortion and binocular suppression in 3-D video. Then, an asymmetric coding method of MVD-based 3-D video is proposed by combining two models in encoding framework. Finally, a chrominance reconstruction algorithm is presented to achieve accurate reconstruction. Experimental results show that compared with other methods, the proposed method can obtain higher performance of view rendering under the total bitrate constraint. Moreover, the perceptual visual quality of 3-D video is almost unaffected with the proposed method.**

***Index Terms*—3-D video, asymmetric coding, bit allocation, chrominance reconstruction, view rendering.**

I. Introduction

THREE-DIMENSIONAL (3-D) video has drawn increasing attention as a new multimedia technique for capturing real-world scenes, which is expected as the ensuing evolution of 2-D video [1]. 3-D video captures photorealistic texture as well as complex geometric shape of real-world scenes. Two types of 3-D video applications, three-dimensional television (3DTV) and free-viewpoint television (FTV), are envisioned. 3DTV aims to provide viewer depth perception of the scene by simultaneously rendering multiple views from different viewing angles [2]. Instead, FTV allows interactive selection of the viewpoint and direction in the scene within a certain operating range [3]. In order to promote 3DTV and FTV applications, multi-view acquisition [4], multi-view video coding (MVC) [5], and virtual view rendering [6] are being developed as the key technologies.

In order to represent a 3-D scene, several 3-D video data formats and 3-D video coding strategies currently co-exist, among which multi-view video plus depth (MVD) format has emerged as an efficient data representation for 3-D video scene and 3-D system [7]. Compared to multi-view video format which synthesizes scenes by using image interpolation, the main advantage of MVD format is that virtual views at arbitrary viewpoint positions can be conveniently generated via depth-image-based rendering (DIBR) technique for interactive application [6]. Even though generation of depth maps in MVD is somewhat complicated, MVD format is backward compatible with the existing 2-D video coding and transmission technologies. Recently, MPEG group has started the exploration work on depth estimation and view synthesis for developing the 3-D video standard [8].

However, the presence of multiple cameras as well as additional depth information significantly increases the amount of data. For texture videos, many MVC methods were developed by exploiting inter-view dependencies, which had been standardized by both joint multi-view video model (JMVM) [9] and joint multi-view video coding (JMVC) standard [10]. For depth maps, they can be compressed by considering special characteristics of depth map, such as exploiting depth smooth properties [11], or optimizing depth coding in wavelet domain [12]. In order to be backward compatible with the MVC standard, the depth maps are often treated as grey scale image sequence, which are similar to the luminance component of texture video, and can be compressed by JMVM or JMVC. Furthermore, by considering the correlations between texture and depth, several joint depth/texture coding schemes were proposed to further improve the coding efficiency [13], [14]. However, since depth maps are not directly used for display, the effects of depth map coding on view rendering should be taken into consideration in joint coding methods.

In general, quality of the rendered virtual views can be affected by compression of texture videos and depth maps [15]. Given a maximum bitrate budget to represent the 3-D scene, how to optimally distribute the bitrates between texture and depth, such that the rendering distortion is minimized, is still

Manuscript received April 05, 2011; revised July 19, 2011; accepted September 04, 2011. Date of publication September 22, 2011; date of current version January 18, 2012. This work was supported in part by the Natural Science Foundation of China (grants 60902096, 60832003, 60872094, and 61071120), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20093305120002). It was also sponsored by K.C.Wong Magna Fund in Ningbo University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ming-Ting Sun.

F. Shao, G. Jiang, M. Yu, and K. Chen are with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: shaofeng@nbu.edu.cn; jianggangyi@nbu.edu.cn; yumei@nbu.edu.cn; chenken@nbu.edu.cn).

Y.-S. Ho is with the School of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea (e-mail: hoyo@gist.ac.kr).

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

Digital Object Identifier 10.1109/TMM.2011.2169045

an open-ended problem. Results from the European ATTEST project had shown that the bitrates of depth maps were fixed at a percentage of 20% of the texture bitrates, but it cannot guarantee that the bitrate allocation is the optimal solution to practical MVD data. Daribo *et al.* proposed a rate-distortion (R-D) optimized bit allocation strategy by considering the content-specificity of sequence [14]. Morvan *et al.* used full-search method to exhaustively search the optimal bitrate trade-off between texture and depth [16]. Liu *et al.* proposed a view synthesis distortion model to seek the optimal bitrate trade-off between texture and depth [17]. Kim *et al.* proposed a new distortion metric to quantify the effect of depth coding on synthesized view quality [18]. However, these methods do not aim at establishing a quantitative bitrate ratio between texture and depth with lower complexity, and are difficult to be efficiently applied in real 3-D video systems.

From another perspective, based on the suppression theory of binocular vision, it is assumed that the binocular perception is better supported with an asymmetric quality [19]. These binocular visual characteristics have been extensively investigated in asymmetric coding [20]–[24], in which one of the two views is encoded with lower quality compared to the other one, while the perceptual quality degradation for the stereoscopic display can be negligible by human eyes. Fehn *et al.* spatially down-sampled the right view and encoded it with the lower resolution in order to reduce the total amount of data to be encoded [20]. Chen *et al.* applied an adaptive filter to generate picture-level adaptive inter-view predictors for macro-block partitions in coding [21]. Aflaki *et al.* investigated different compressed mixed-quality asymmetric stereoscopic video in order to measure the optimal asymmetric solutions [22]. Termin *et al.* suggested that human visual perception system will fuse the two images that differ in chromatic content into one perceived color stereoscopic image [23]. In our previous work prior to this research [24], in order to exploit the visual redundancy in 3-D video, all the chrominance information in selected views is discarded and the discarded chrominance information is ultimately restored at the decoder. However, these methods have not taken the optimal view rendering problem into consideration.

Therefore, asymmetric coding of the texture videos and depth maps is a significant way to get the optimal solution of 3-D video compression and view rendering problem. Up to now, there are not detailed researches on asymmetric coding of MVD-based 3-D video. The previous asymmetric coding methods of 2-D video mainly benefit the visual quality so as to reduce the bandwidth requirement. Compared with the asymmetric coding of 2-D video, asymmetric coding of 3-D video pertains to two layers of implication, that is, asymmetric coding of texture videos and depth maps, and asymmetric coding of texture videos. The first part takes the objective quality as the optimization criterion, and the second part takes the visual characteristics of binocular suppression into consideration. Besides bandwidth reduction, view rendering quality is also the most important indicator to evaluate the performance of asymmetric coding of 3-D video. Therefore, the key issue in asymmetric coding of 3-D video is how to allocate the bitrates for texture videos and depth maps appropriately, so that the objective performance of view rendering is enhanced while keeping the same perceptual visual quality.

In this paper, considering different bitrate ratios in depth/texture and the characteristics of binocular suppression in human visual system (HVS), we propose a novel asymmetric coding method of MVD-based 3-D video, which can ameliorate the performance of view rendering to the maximum extent possible. The motivation of the proposed method is inspired by two speculations: 1) No matter what coding methods are employed and the compressed errors appear in a rendered view, the target bitrates for texture videos and depth maps must be allocated in advance for the desired rendering quality. 2) Even though the visual characteristics of binocular suppression in HVS can be explained through several perceptual models, the perceptual color redundancy is particularly suitable for asymmetric coding with better depth perception and lower encoding complexity. This paper proposes a bit allocation model to characterize the view rendering distortion and a chrominance reconstruction model to characterize the binocular suppression, and finally achieves asymmetric coding scheme by combining these two parts, which aims at improving the objective quality of the rendered virtual view while keeping the same perceptual visual quality.

The rest of the paper is organized as follows. The bit allocation and chrominance reconstruction models are derived in Section II. In the model derivation, we emphasize on analyzing how the rendering distortion can be effectively represented by coding distortion, and define a new data representation format to characterize the visual masking effect. In Section III, based on the derived models, the proposed asymmetric coding method of MVD-based 3-D video is described in detail. Section IV presents the experimental results. Finally, the conclusion is drawn in Section V.

II. Bit Allocation and Chrominance Reconstruction Models

Since 3-D video is an extension of 2-D, it can be affected by the same types of visual distortions identified with 2-D video, and the additional distortions in depth maps can further adversely affect the 3-D visual perception. The framework of the MVD-based 3-D video system is illustrated in Fig. 1. On the sender side, texture videos are captured with multiple cameras, and the corresponding depth maps are generated by depth estimation method. Then, the texture videos and depth maps are compressed by MVC encoders in an independent or joint means. At the client side, virtual views are rendered from the decoded texture videos and depth maps by DIBR technique, and ultimately provide 3-D visual experiences for end users. Bandwidth reduction and view rendering quality are the two important indicators to evaluate the performance of the framework. Since MVD-based 3-D video coding must satisfy the bandwidth requirements, how to find the best compression strategy to achieve the optimal rendering quality is a vital challenge for the success of 3-D video system when total bits of texture videos and depth maps are restricted.

In order to achieve higher view rendering quality in the MVD-based 3-D video system framework, asymmetric coding strategy of 3-D video should be carefully considered. Different from the

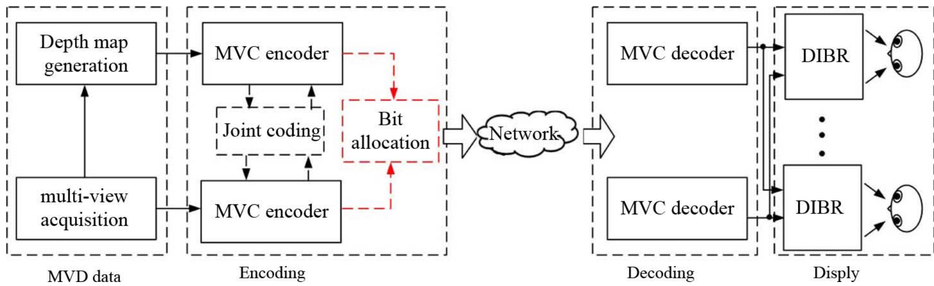


Fig. 1. System framework of the MVD-based 3-D video system.

asymmetric coding methods of 2-D video, the proposed asymmetric coding method of 3-D video mainly consists of two parts, that is, asymmetric coding of texture videos and depth maps, and asymmetric coding of texture videos. In the system framework, the first part is fulfilled by a bit allocation model which only takes the objective quality as the optimization criterion, and the second part is fulfilled by a chrominance reconstruction model which takes the visual characteristics of binocular suppression into consideration. Then, the two parts are combined into one framework by implementing appropriate bitrate allocation and rate control strategies. As a result, the objective quality of the rendered virtual view will be improved while keeping the same or nearly the same perceptual visual quality.

In the MVD-based 3-D video system, virtual views are rendered from the compressed texture videos and depth maps, and the distorted texture videos and depth maps resulting from compression can be propagated to the virtual views. Therefore, the optimal bitrate ratio problem between texture videos and depth maps is necessary to be solved in MVD-based 3-D video coding. In order to seek the optimal bitrate pair, (R_t^{opt}, R_d^{opt}), under the total bitrate constraint R_c, the problem is formulated as

$$\begin{aligned}\left(R_t^{opt}, R_d^{opt}\right) &= \underset{R_c, R_d \in Q}{\arg\min} D_v\left(R_t, R_d\right) + D_t(R_t)\\ s.t.\ R_d + R_t &\le R_c \end{aligned} \tag{1}$$

where R_t and R_d are the coding bitrates of texture videos and depth maps, respectively; they are constructed a bitrate pair, denoted as (R_t, R_d). $D_v(R_t, R_d)$ is the view rendering distortion, $D_t(R_t)$ is the coding distortion of texture videos, and Q is the candidate set of the bitrate pair. An important feature of the model is that the coding distortion $D_t(R_t)$ of texture videos is also included, because texture videos usually need to maintain higher quality for the purpose of being compatible with 2-D display.

It is assumed that the multi-view acquisition, depth generation, view rendering, and 3-D display modules in Fig. 1 are fixed; thus, the quality of the rendered virtual view may be mainly affected by the coding distortions of texture videos and depth maps. Let $D_d(R_d)$ denote coding distortion of depth maps, the view rendering distortion D_v can be represented by D_t and D_d as a function of

$$D_v = f(D_t, D_d). \tag{2}$$

In order to model the function $f(\cdot)$ in (2), let S_v denote the original texture image at the virtual view position, $\bar{S}_v$ denote the image rendered by the original texture images and the compressed depth maps, and $\hat{S}_v$ denote the image rendered by the compressed texture images and the compressed depth maps, the view rendering distortion $D_v(R_t, R_d)$ can be approximately decomposed into two components

$$\begin{aligned} D_v\left(R_t, R_d\right) &= E\left\{\left(S_v - \hat{S}_v\right)^2\right\}\\ &= E\left\{\left((S_v - \bar{S}_v) + (\bar{S}_v - \hat{S}_v)\right)^2\right\}\\ &= E\left\{(S_v - \bar{S}_v)^2\right\} + E\left\{(\bar{S}_v - \hat{S}_v)^2\right\}\\ &\quad + 2E\left\{(S_v - \bar{S}_v)(\bar{S}_v - \hat{S}_v)\right\} \end{aligned} \tag{3}$$

where $E\left\{(S_v - \bar{S}_v)^2\right\}$ represents the average view rendering distortion induced by depth compression, and $E\left\{(\bar{S}_v - \hat{S}_v)^2\right\}$ represents the average view rendering distortion induced by texture compression, and $E\left\{(S_v - \bar{S}_v)(\bar{S}_v - \hat{S}_v)\right\}$ approximates to zero [25]. In practical view rendering implementation, the virtual view is rendered from multiple adjacent views. Theoretically, the impact of different views on the same virtual view should be taken in account in the distortions $E\left\{(S_v - \bar{S}_v)^2\right\}$ and $E\left\{(\bar{S}_v - \hat{S}_v)^2\right\}$. Since the depth maps of different views have large amounts of uniform contents, the impact of different views on the same virtual view may be similar if the impact of occlusion is ignored. For simplicity, we only consider the impact from one adjacent view in the above distortions.

It is supposed that the location of virtual view is known, for a particular virtual view, $E\left\{(S_v - \bar{S}_v)^2\right\}$ can be characterized by a linear model and expressed as [26]

$$\begin{aligned} E\left\{(S_v - \bar{S}_v)^2\right\} &= \omega_r^2 \cdot E\left\{\Delta P_r^2\right\} \cdot \psi_r\\ &= k_1 \cdot E\left\{\Delta P_r^2\right\} \end{aligned} \tag{4}$$

where ω_r is the weighting factor of the rendered virtual image from a particular view, ψ_r is the linear parameter which is associated with image contents, and ΔP_r acts as the warping position error, which is computed as [27]

$$\begin{aligned} \Delta P_r^2 &= \left(\frac{f_l \cdot \delta_x}{255} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right)\right)^2 \cdot D_d(R_d)\\ &= k_2 \cdot D_d(R_d) \end{aligned} \tag{5}$$

where $D_d(R_d)$ is the coding distortion of depth maps, f_l denotes the focal length of the camera in the horizontal direction,

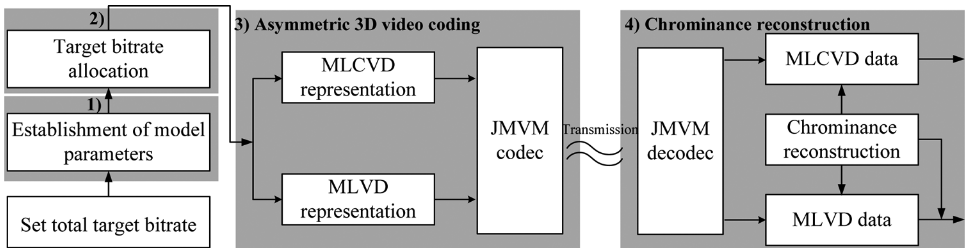


Fig. 2. Block diagram of the proposed asymmetric coding method of MVD-based 3-D video.

δ_x expresses the distance between the virtual and the particular cameras, and Z_{near} and Z_{far} are the nearest and farthest depth values. Since the location of virtual view is not fixed in the model, we select middle virtual view between two adjacent views with $\omega_r = 0.5$ as a tradeoff in the experiment. Thus, the relationship between $E\left\{(S_v - \bar{S}_v)^2\right\}$ and $D_d(R_d)$ can be approximated as

$$E\left\{(S_v - \bar{S}_v)^2\right\} \approx k_1 \cdot k_2 \cdot D_d(R_d) \tag{6}$$

where $k_1 = \omega_r^2 \cdot \psi_r, k_2 = (f \cdot \delta_x/255 \cdot (1/Z_{near} - 1/Z_{far}))^2$. Similarly, since the view rendering distortion induced by texture compression can be directly regarded as a linear combination of the coding distortions, the relationship between $E\left\{(\bar{S}_v - \hat{S}_v)^2\right\}$ and $D_t(R_t)$ can be expressed as

$$E\left\{(\bar{S}_v - \hat{S}_v)^2\right\} \approx \omega_r^2 \cdot D_t(R_t). \tag{7}$$

Thus, based on (3), (6), and (7), the optimal bit allocation problem in (1) can be simplified as

$$\begin{aligned}(R_t, R_d) &= \underset{R_c, R_d \in Q}{\arg\min} D_{total}(R_d, R_t)\\ s.t.\ R_d + R_t &\leq R_c\end{aligned} \tag{8}$$

where $D_{total}(R_d, R_t) = k_1 \cdot k_2 \cdot D_d(R_d) + (1 + \omega_r^2) \cdot D_t(R_t)$ denotes the total view rendering distortion generated by texture and depth coding distortions. It is important to be aware that the above model can be applied not only to depth/texture bit allocation, but also to joint depth/texture coding. In this study, we only consider the former case. Thus, once the parameters k_1 and k_2 are determined, it is convenient to establish the optimal bitrate pair by using the coding distortions from one adjacent view.

From another perspective, based on the characteristics of binocular suppression in HVS, the higher quality view in stereoscopic pairs will dominate the 3-D visual perception. Thus, we redefine the MVD data format as two independent data representation formats [24]. One is multi-view luminance and chrominance video plus depth (MLCVD) representation, and the other is multi-view luminance video plus depth (MLVD) representation. For MLCVD representation, the total data $\mathbf{D}_{MLCVD}$ is composed of luminance $\mathbf{\Phi}_{MLCVD}$, chrominance $\mathbf{\Psi}_{MLCVD}$, and depth $\mathbf{\Omega}_{MLCVD}$, and it can as a result be expressed as

$$\mathbf{D}_{MLCVD} = \{\mathbf{\Phi}_{MLCVD}, \mathbf{\Psi}_{MLCVD}, \mathbf{\Omega}_{MLCVD}\}. \tag{9}$$

In contrast, for MLVD representation, the total data $\mathbf{D}_{MLVD}$ is only composed of luminance $\mathbf{\Phi}_{MLVD}$ and depth $\mathbf{\Omega}_{MLVD}$, and represented as

$$\mathbf{D}_{MLVD} = \{\mathbf{\Phi}_{MLVD}, \mathbf{\Omega}_{MLVD}\}. \tag{10}$$

Then, the total data $\mathbf{D}_{total}$ for all MVD data is represented as

$$\mathbf{D}_{total} = \{\mathbf{D}_{MLCVD}, \mathbf{D}_{MLVD}\}. \tag{11}$$

It is well known that inter-view relationships can be well revealed by the supplementary depth information. Thus, the discarded chrominance information $\mathbf{\Psi}_{MLVD}$ in MLVD representation can be reconstructed from MLCVD data with the depth information. Chrominance information reconstruction can be expressed as a function in the following form:

$$\begin{aligned}&\mathbf{\Psi}_{MLVD}\\ &= g\left(\mathbf{\Phi}_{MLCVD}, \mathbf{\Phi}_{MLVD}, \mathbf{\Psi}_{MLCVD}, \mathbf{\Omega}_{MLVD}, \mathbf{\Omega}_{MLVCD}\right).\end{aligned} \tag{12}$$

In order to model the function $g(\cdot)$ in (12), a chrominance reconstruction algorithm is proposed in the next section by using the correspondences between left and right views in the proposed data representation format.

III. Proposed Asymmetric Coding Method

According to the above bit allocation and chrominance reconstruction models, we propose a novel asymmetric coding method of MVD-based 3-D video. The block diagram of the proposed method is shown in Fig. 2. It is a specific implementation of the MVD-based 3-D video coding framework in Fig. 1. In order to accurately control the encoded bitrates of 3-D video, appropriate bitrate allocation and rate control strategies are implemented in the proposed method by fusing the two models. It consists of four stages: 1) establishment of the model parameters, 2) target bitrate allocation, 3) asymmetric 3-D video coding, and 4) chrominance reconstruction. In the proposed method, the first and last stages are offline and the middle two stages are online processed. To facilitate understanding, the summary of some notations and acronyms is shown in Table I.

A. Establishment of the Model Parameters

In the 3-D video system, texture videos and depth maps are encoded by MVC method. For the sake of simplicity, we emphasize on two-view based 3-D video in this paper, and it can be easily extended to the case of multiple views. In the two-view based 3-D video, suppose that one view is regarded as left view

TABLE I
SUMMARY OF NOTATIONS AND ACRONYMS

Variables	definition
R_L	Initial bitrate of left view
R_R	Initial bitrate of right view
R_d	Initial bitrate of depth
R_t	Initial bitrate of texture
R_Φ	Initial bitrate of luminance
R_Ψ	Initial bitrate of chrominance
R_c	Total target bitrate
R_t^L	Target bitrate of texture in left view
R_t^R	Target bitrate of texture in right view
R_d^L	Target bitrate of depth in left view
R_d^R	Target bitrate of depth in right

and the other view is regarded as right view; each view is composed of a texture video and the corresponding depth map. Let α be the initial bitrate ratio between left and right views, β be the initial bitrate ratio between chrominance and luminance components in texture videos, and γ be the initial bitrate ratio between depth and texture. Under the total bitrate constraint, after encoding the left and right views, the initial bitrates R_L and R_R in left and right views, and the initial bitrates R_Φ and R_Ψ of luminance and chrominance in left view, can be obtained. Also, by applying the bit allocation model in (8) to the left view, the initial bitrates R_d and R_t of depth and texture can be obtained. Finally, the optimal model parameters can be expressed as

$$\alpha = \frac{R_L}{R_R} \quad \beta = \frac{R_\Psi}{R_\Phi} \quad \gamma = \frac{R_d}{R_t}. \tag{13}$$

It is worth noting that the approximate average bitrate ratios α, β, and γ can be estimated by calculating the statistical bitrate values after pre-encoding several frames in each view. Moreover, it is assumed that the parameters β and γ of left view are similar to those of right view. For the candidate bitrate set Q in (8), we test the bitrate ratio of $R_d/(R_d + R_t)$ ranging from 0.2 to 0.8 with a ratio interval of 0.1. More accurate bitrate ratio can be obtained with more sophisticated classification of ratio interval. Fig. 3 shows the statistical initial average bitrate distribution under various total bitrate constraints. It is clear that the average bitrate distribution between texture and depth is independent of the total bitrates, and for the given test sequence specified in this research, we can obtain the approximate average bitrate ratio after pre-encoding several frames of MVD data. In the experiment, all frames in the first two GOPs are used for pre-encoding in this stage.

B. Target Bitrate Allocation

After establishing the model parameters, two different 3-D video representations, namely MLCVD and MLVD, are adopted for the left and right views, respectively. In the representations, the left view is composed of color texture video and its corresponding depth map, and the right view is composed of gray texture video and its corresponding depth map. The quality of

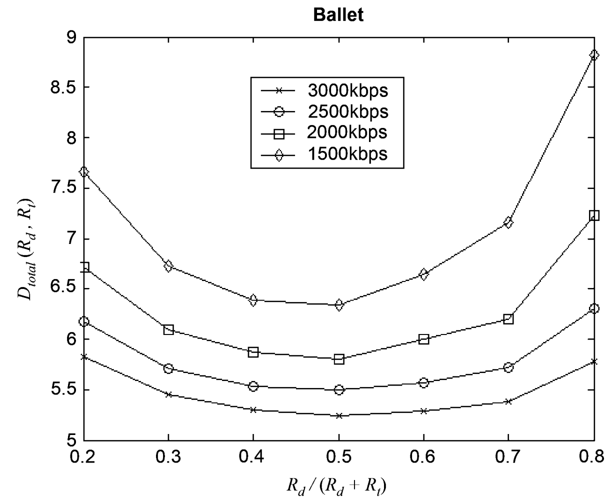


Fig. 3. Initial average bitrate distribution under different total bitrate constraints.

the rendered virtual view can be enhanced by efficiently allocating different target bitrate to each representation. Here, based on the early established bitrate ratios α, β, and γ, target bitrates for texture and depth of left and right views are allocated by considering their impacts on view rendering and the visual characteristics of binocular suppression. Under the total bandwidth R_c constraint, the target bitrates for texture and depth in the representations can be expressed as follows:

$$R_t^L = R_c \cdot \frac{\alpha}{1+\alpha} \cdot \frac{1}{1+\gamma} \tag{14}$$

$$R_d^L = R_c \cdot \frac{\alpha}{1+\alpha} \cdot \frac{\gamma}{1+\gamma} + R_c \cdot \frac{1}{1+\alpha} \cdot \frac{\beta}{1+\beta} \cdot \frac{1}{1+\gamma} \tag{15}$$

$$R_t^R = R_c \cdot \frac{1}{1+\alpha} \cdot \frac{1}{1+\beta} \cdot \frac{1}{1+\gamma} \tag{16}$$

$$R_d^R = R_c \cdot \frac{1}{1+\alpha} \cdot \frac{\gamma}{1+\gamma}. \tag{17}$$

It is noted that in the proposed target bitrate allocation strategy, the encoded bitrates for texture in the MLVD representation can be reduced by discarding partial chrominance information, and the saved bitrates from the chrominance coding are all allocated to depth coding in the MLCVD representation. Even though the quality of chrominance information in the MLVD representation is degraded, the resulting perceptual quality degradation for the stereoscopic display is not noticeable by human eyes [28]. Such a strategy will have several additional advantages. Firstly, the quality of the rendered virtual view can be improved with higher quality depth map. Secondly, more accurate pixel correspondences can be provided for the right view in the chrominance reconstruction stage.

When the target bitrates for texture and depth are established, the accurate initial quantization parameters (QPs) are necessary to be assigned in order to improve the accuracy of rate control. In the detailed configuration, only the QP values of the first encoded frames are required to be set, and the QP values for the following P-frames or B-frames can be automatically adjusted

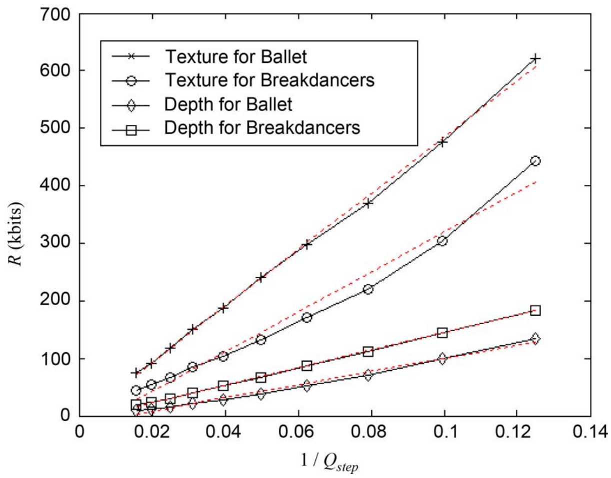


Fig. 4. Relationship between encoded bit and quantization step.

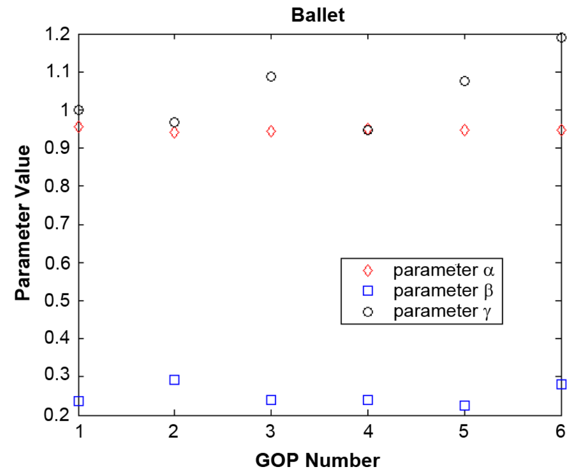


Fig. 5. Distribution of parameters α, β, and γ in different GOPs.

in the process of rate control. Thus, in the proposed method, we use only one frame to test the rate-quantization (R-Q) model. According to the statistical R-D characteristic of the first encoded frames in Fig. 4, the linear R-Q model in [29] can be used to characterize the initial QPs setting. The linear relationships between R_t^L and $1/Q_{t,step}$, and between R_d^L and $1/Q_{d,step}$, are built as

$$R_t^L = \frac{K_t}{Q_{t,step}^L} + C_t \tag{18}$$

$$R_d^L = \frac{K_d}{Q_{d,step}^L} + C_d \tag{19}$$

where K_t, K_d, C_t, and C_d are constants, and $Q_{t,step}^L$ and $Q_{d,step}^L$ are quantization steps for texture and depth in the left view, respectively. Thereupon, given the above relationship of target bitrates between R_t^L and R_d^L, and R_t^R and R_d^R in (14)–(17), the accurate initial QPs for texture and depth can be established.

C. Asymmetric Coding of MVD-Based 3-D Video

After establishing the target bitrates and the initial QPs for left and right views, a typical hierarchical B picture (HBP) prediction structure is adopted for two-view based MVC. Generally, texture videos and depth maps are encoded using different MVC encoders. To facilitate the simultaneous controlling of the bitrates of the texture and depth, one texture image and its corresponding depth map are superposed into one combined image. Thus, the texture videos and depth maps can be encoded by using a single MVC encoder as two independent slices in the encoded frames. In the prediction structure, left view is independently encoded and right view is encoded with disparity compensated prediction technique. Since it is rather difficult to accurately control the encoding bitrates with a fixed QP, the texture videos and depth maps for left and right views are encoded with a rate control strategy. The current MVC encodes each view separately and the inter-view reference sequences are added to the encoding buffer. Thus, the rate control strategy in the proposed 3-D video coding is the same as H.264/AVC [31]. Here, the rate control strategy in the proposed method has three layers, that is, group of picture (GOP) layer, frame layer, and basic unit layer. Firstly, the target bitrate allocated to a GOP is updated based on the total target bitrates of texture in left view and the corresponding initial QPs. Then, the target bitrates for a specific frame are determined based on its buffer level, encoding complexity, and the remained bits in the GOP. Finally, the quantization step size $Q_{t,step}^L(i,j)$ of texture in left view for the jth frame in the ith GOP can be computed by the quadratic R-Q model [30]

$$T_t^L(i,j) = a_1 \times \frac{MAD(i,j)}{Q_{t,step}^L(i,j)} + a_2 \times \frac{MAD(i,j)}{Q_{t,step}^L(i,j)^2} + H(i,j) \tag{20}$$

where a_1 and a_2 are the model parameters, $T_t^L(i,j)$ is the target bitrate allocated to the jth frame in the ith GOP, $H(i,j)$ is the sum of header bits and motion bits, and $MAD(i,j)$ is a prediction of mean absolute difference (MAD) between original image data and its prediction data. Based on the relationship between QP and quantization step size Q_{step}, $Q_{step} = 2^{(QP-4)/6}$, the optimal QP can be determined. Similarly, the QPs for other representations can be computed. The corresponding equations can be referred in [31].

However, in the process of rate control, it is hard to guarantee that the estimated parameters α, β, and γ will be consistent across the whole sequence. As shown in Fig. 5, through experiments, it is found that the parameter α is relatively stable and the parameters β and γ will be changed in different GOPs. The reason is that the parameter α highly depends on the inter-view coding structure and the parameter β highly depends on the image content, and the depths of the objects have been changed in the sequence due to inaccurate depth estimation. Since chrominance information only occupies a small percentage of the total bitrates, only the parameter γ is updated, and the corresponding target bitrates in (14)–(17) are reallocated before encoding the next GOP. Since it is impossible to apply the bit allocation model with all candidate bitrate set Q in the encoding processing, the candidate bitrate set Q is restricted based on the bitrates of encoded frames in one GOP. Besides, in the encoding process, the chrominance information of texture video in right view is discarded, and the saved bitrates from the chrominance information are all allocated to depth coding in left view. Consequently, even though the total bitrates allocated to texture videos are largely reduced, the 3-D visual perception is still better supported with an asymmetric quality than a reduction of quality in both views [28].

D. Chrominance Reconstruction

At the decoder, the discarded chrominance information in right view should be reconstructed. The performance of the proposed asymmetric coding method largely depends on the accuracy of the reconstructed chrominance. With the depth information of left view, 3-D warping in DIBR is used to find pixel correspondences between left and right views, and the discarded chrominance information in right view is directly reproduced from left view. However, occlusion is inevitable in 3-D warping that only luminance information is preserved in the occluded areas. In the following step, these luminance-only pixels are colorized by referencing those already obtained chrominance values $U(x,y)$ and $V(x,y)$ [32]. By imposing the constraint that two neighboring pixels should have similar chrominance values if their luminance values are similar, we intend to minimize the reconstruction error between the chrominance value $U(x,y)$ at pixel (x,y) and the weighted average of the chrominance values at its neighboring pixels

$$J(U) = \sum_{(x,y)} \left(U(x,y) - \sum_{(x',y') \in N(x,y)} w(x,y,x',y') \hat{U}(x',y') \right)^2 \quad (21)$$

where $N(x,y)$ denotes the neighborhood of a pixel (x,y), $w(x,y,x',y')$ is a weighting function based on the normalized square difference between two intensities $w(x,y,x',y') \propto e^{-(Y(x,y)-Y(x',y'))^2/2\sigma_{xy}^2}$, and σ_{xy} is the variance of the luminance in a window around pixel (x,y).

However, the above equation is still unsolvable because the chrominance value $\hat{U}(x',y')$ is unknown. Then, it is also assumed that the chrominance value $U(x,y)$ at a pixel (x,y) is a linear function of the luminance value $Y(x,y)$, and the linear coefficients are the same for all pixels in the neighborhood of the pixel (x,y) [33]. Thus, given a set of pixels where the chrominance values are correctly obtained, we minimize $J(U)$ to obtain the optimal coefficients. Finally, the chrominance value for a pixel (x,y) can be represented as

$$\hat{U}(x,y) = a_1 Y(x,y) + a_0 \quad (22)$$

where a_1 and a_0 are the linear coefficients for U components. Similarly, the chrominance values in V component can be reconstructed.

IV. Experimental Results and Analyses

In order to objectively evaluate the performance of the proposed method, several experiments are conducted with typical 3-D video sequences of "*Ballet*", "*Breakdancers*", "*Alt Moabit*", and "*Door Flowers*". The depth maps of "*Ballet*" and "*Breakdancers*" sequences had been computed by using a stereo-matching algorithm [34], and the depth maps of "*Alt Moabit*" and "*Door Flowers*" have been estimated by depth estimation reference software (DERS) 3.0 [35]. In the experiments, the fourth and sixth views, denoted view4 and view6, in "*Ballet*" and "*Breakdancers*" are adopted as left and right views, and the virtual view, denoted view5, is rendered by view4 and view6. The eleventh and ninth views, denoted view11 and view9, in "*Alt Moabit*" and "*Door Flowers*" are adopted as left and right views, and the virtual view, denoted view10, is rendered by view11 and view9. We use view synthesis reference software (VSRS) 3.0 [36] to render the virtual view, and use the revised MVC software JMVM 7.0 [9] to implement the asymmetric coding methods. In the encoding process, the temporal GOP size is set to 15, and the total number of encoded frames in each view is 100. In order to objectively measure the performance of the proposed method, we give one traditional asymmetric coding scheme and design three schemes as follows.

Scheme-I: Traditional coding with the fixed ratio of 5:1 between texture and depth bitrates

Scheme-II: Mix-resolution asymmetric coding method

Scheme-III: The proposed asymmetric coding method without chrominance reconstruction

Scheme-IV: The proposed asymmetric coding method

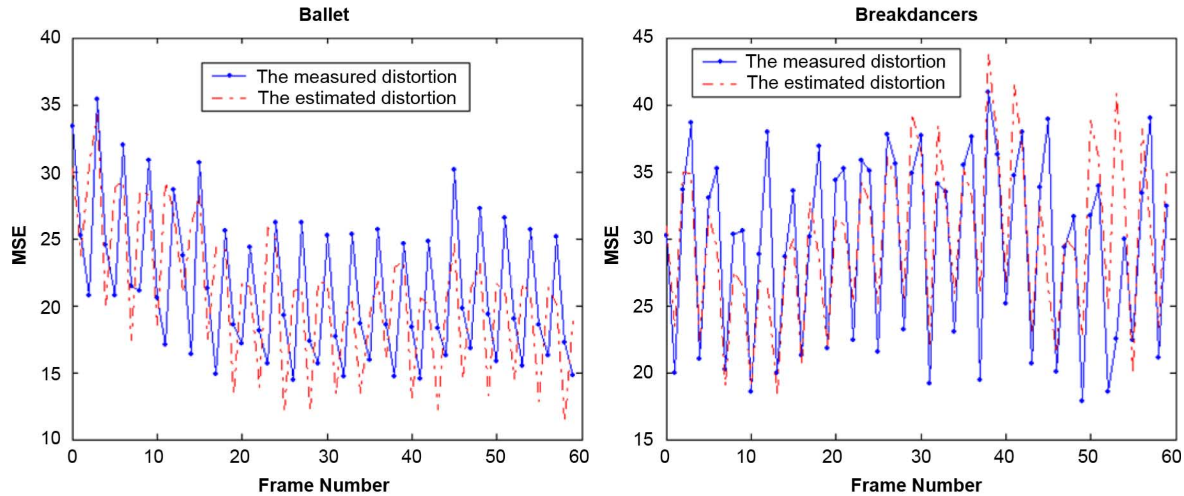
It is noted that, for Scheme-II, downsampling by a factor of two in the horizontal and vertical direction is performed, and since the mix-resolution asymmetric coding method in [20] cannot be directly applied to 3-D video, the optimal ratio between texture and depth is exhaustively searched by full-search method in [16], and other settings are similar to the proposed method.

A. Estimation Accuracy of the Distortion Model

In order to verify the estimation accuracy of the proposed bit allocation model in (8), we compare the estimated and the measured view rendering distortions for view4 under total bitrate 600 kbps in Fig. 6. The view rendering distortion in the model is measured between the rendered virtual image and the original image. The quality of the rendered virtual view will be also affected by the detailed view blending and hole filling algorithms in DIBR. In order to eliminate these influences and objectively measure the effects of coding distortion on view rendering, the virtual view, view5, is rendered only from view4, and the occluded areas are excluded in the comparison. From the results, it can be observed that the estimated curve shows the similar trend with the measured curve. Since the proposed model uses some approximations, the distortion may be overestimated or underestimated so that the estimated curve is lower or higher than the measured curve in some frames. However, since the proposed bit allocation model only involves identification of the optimal bitrate ratio, the minimum view rendering distortion can be correctly differentiated from different texture/depth bitrate ratio combinations even though there is a certain degree of inaccuracy in the estimated distortion.

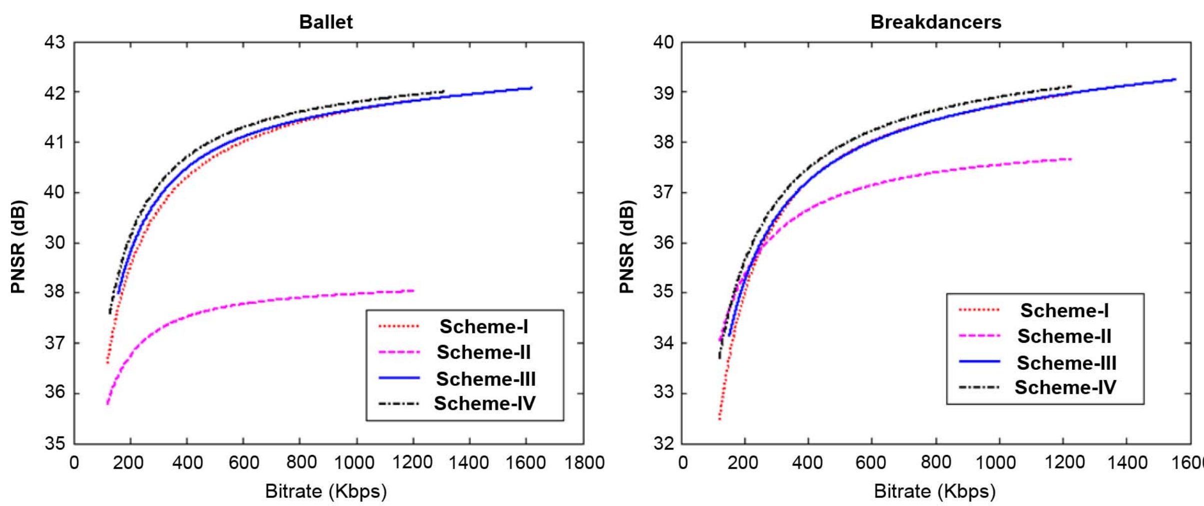
B. Coding Performance Comparison Results

Fig. 7 shows the R-D curves of the four schemes for texture video only in right view. The ordinate in each sub-figure is the average luminance peak signal-to-noise ratio (PSNR), while the abscissa corresponds to the bitrates designated for the right view. As depicted by the results, the R-D performance



(a) (b)

Fig. 6. Comparison between the estimated and the measured distortions for view4. (a) Ballet. (b) Breakdancers.



(a) (b)

Fig. 7. RD performance comparison of "*Ballet*" and "*Breakdancers*". (a) Ballet. (b) Breakdancers.

of Scheme-IV is better than other schemes due to the reduced amount of data allocated to texture video by discarding the chrominance information. Scheme-III reveals the consistent change trends of R-D curves with Scheme-I owing to the same encoding prediction structure used. However, for Scheme-II, even though the total amount of data is largely reduced by downsampling, the reconstructed quality is also greatly deteriorated after upsampling, especially at high bitrate side. Therefore, the mix-resolution asymmetric coding scheme is especially suitable for low bitrate visual communications such as mobile 3DTV. In other words, the proposed asymmetric coding of 3-D video is not an R-D optimization problem, and view rendering quality or visual perceptual quality may be more suitable to evaluate the performance of the proposed method.

C. View Rendering Comparison Results

In order to objectively evaluate the performance of view rendering, suppose that the virtual view rendered with original texture videos and depth maps is regarded as reference, and the average weighted PSNR is measured with coefficients 0.8, 0.1, and 0.1 in YUV components, respectively. The detailed bit allocation results for texture and depth, and the corresponding Bjonteggrad delta PSNR (BDPSNR) [37] for Scheme-II, Scheme-III, and Scheme-IV are tabulated in Table II, in which Scheme-I is supposed as the comparison basis. It is obvious that compared to traditional coding scheme with the fixed ratio of 5:1 between texture and depth bitrates, the other three schemes can achieve better view rendering performances. Moreover, for Scheme-IV, by allocating the saved bitrate in MLVD representation to depth coding, the view rendering performance can be further enhanced, which again confirms that the quality of the compressed depth maps has a big impact on the quality of rendered views. However, for Scheme-II, the view rendering performance is slightly lower than the proposed method. The reason is that the mix-resolution asymmetric coding method aims to reduce the amount of data by providing consistent perceptual quality, while the proposed asymmetric coding method

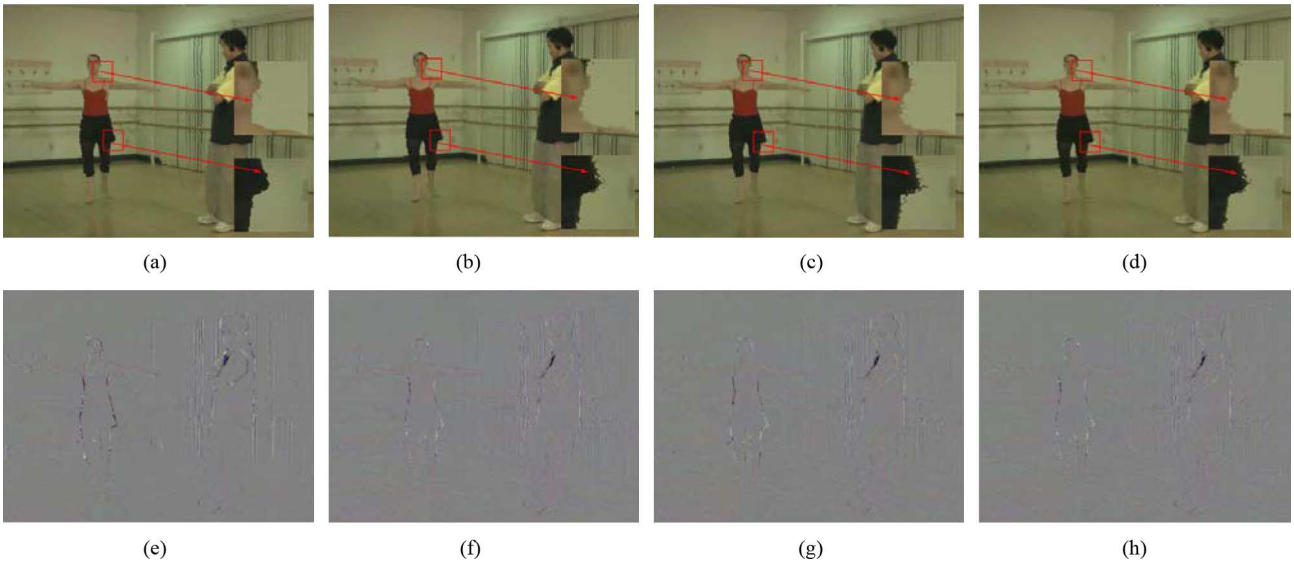


Fig. 8. Subjective view rendering results for "*Ballet*" under total bitrate 600 kbps constraint. (a) Rendered texture image with Scheme-I (32.89 dB). (b) Rendered texture image with Scheme-II (34.53 dB). (c) Rendered texture image with Scheme-III (34.61 dB). (d) Rendered texture image with Scheme-IV (34.80 dB). (e) Error image between (a) and the original rendered image. (f) Error image between (b) and the original rendered image. (g) Error image between (c) and the original rendered image. (h) Error image between (d) and the original rendered image.

TABLE II
COMPARISON RESULTS OF BITRATE ALLOCATION FOR TEXTURE AND DEPTH AND THE CORREPONDING BDPSNR

Sequence	Target bitrate (kbps)	Scheme-I (kbps)		Scheme-II (kbps)		Scheme-III (kbps)		Scheme-IV (kbps)		BDPSNR (dB)		
		texture	depth	texture	depth	texture	depth	texture	depth	II	III	IV
Ballet	600	505.57	106.48	195.73	433.06	313.64	313.17	283.54	345.03	1.91	2.03	2.24
	1200	1013.47	209.83	390.17	939.65	628.10	620.52	566.69	682.81			
	2600	2360.29	449.64	858.3	1920.82	1375.02	1340.77	1245.64	1467.72			
	6000	5334.56	1034.25	2075.37	4300.12	3292.36	3062.46	2983.95	3357.28			
Breakdancers	600	508.29	102.10	194.4	414.56	304.75	305.38	273.85	336.54	1.10	1.20	1.30
	1200	1028.82	203.37	389.56	831.32	611.78	608.18	548.58	669.15			
	2600	2247.61	441.65	847.98	1811.82	1338.58	1317.39	1202.08	1453.52			
	6000	5187.45	1013.06	1952.36	4286.53	3130.64	3040.90	2800.59	3333.41			
Alt Moabit	600	508.31	105.29	200.63	431.12	323.68	293.07	293.10	330.07	0.28	0.37	0.40
	1200	1023.42	216.13	404.28	857.12	648.21	600.18	581.96	667.68			
	2600	2168.70	445.10	845.97	1964.89	1405.60	1217.44	1262.98	1447.62			
	7200	6078.52	1269.75	2377.15	4951.36	3885.09	3439.71	3605.01	3651.45			
Door Flower	600	502.09	107.85	195.06	432.16	302.75	315.68	279.87	343.55	0.66	0.73	0.75
	1200	1002.94	213.17	397.16	845.02	603.03	631.68	557.53	692.81			
	2600	2172.52	450.24	825.33	1851.2	1305.55	1360.08	1201.59	1492.37			
	7200	6006.41	1246.19	2258.24	5002.58	3609.66	3649.27	3342.58	3909.33			

aims to improve view rendering quality by reasonable bit allocation, even though the optimal ratio between texture and depth is exhaustively searched by full-search method. The subjective view rendering results for "Ballet" under total bitrate 600 kbps constraint in Fig. 8 can clearly reflect this phenomenon. Even though the object boundary in Fig. 8(a) is more smooth, the contour deformation has occurred, and over high texture/depth bitrate ratio in Scheme-I and over low texture/depth bitrate ratio in Scheme-II cannot produce good view rendering results.

D. Chrominance Reconstruction Results

In order to appraise the performances of the proposed chrominance reconstruction algorithm, tests are conducted with the decoded texture images. Fig. 9(a) shows the original decoded texture image in right view with Scheme-III under total bitrate 600 kbps, respectively. Fig. 9(b) shows the initial warped texture image in right view with Scheme-IV, where chrominance information in some pixels is still discarded. Fig. 9(c) shows the reconstructed texture image in right view obtained by the proposed algorithm, and Fig. 9(d) shows the corresponding error image of chrominance component U between (c) and (a). It can be observed that the PSNR values of U and V components between the reconstructed and the original images readily exceed 32 dB, indicating high image quality. Since HVS can tolerate a certain degree of quality degradation in one of the views and is more sensitive to luminance than to chrominance, the chrominance degradation can be negligible by human eyes when stereoscopic video is displayed on stereoscopic display.

In order to evaluate how the chrominance degradation affects the perceived stereoscopic effect by assigning different quality chrominance information in right view, a subjective experiment is conducted. Three kinds of stereoscopic image pairs are used

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

Fig. 9. Subjective chrominance reconstruction results for "*Ballet*". (a) Decoded texture image in right view with Scheme-III. (b) Initial warped texture image in right view with Scheme-IV. (c) Proposed reconstructed texture image in right view. (d) Reconstruction error between (a) and (c) of U component.

TABLE III
DEPTH PERCEPTION AND COLOR PERCEPTION RESULTS

Test images	Subjective depth perception		Subjective color perception	
	Pair-B	Pair-C	Pair-B	Pair-C
Ballet	0.15	0.00	1.63	0.16
Breakdancers	0.21	0.00	1.47	0.35
Alt Moabit	0.28	0.00	1.12	0.20
Door Flower	0.25	0.00	1.52	0.05

in the experiment. The first kind of stereoscopic image pairs are created with the decoded texture image in left view and the rendered virtual texture image with Scheme-III (denoted Pair-A). The second kind of stereoscopic pairs are created with the decoded texture image in left view and the rendered virtual texture image with Scheme-IV before chrominance reconstruction (denoted Pair-B). The third kind of stereoscopic pairs are created with the decoded texture image in left view and the rendered virtual texture image with Scheme-IV after chrominance reconstruction (denoted Pair-C). The Pair-A is regarded as reference, and the experimental settings are similar to [28]. Table III shows the detailed difference mean opinion scores (DMOS) evaluation results of the depth and color perception for each stereoscopic image pair using ITU-R Recommendation BT.500 on a scale of 0–10. It is obvious that depth perception is almost unaffected by different stereoscopic image pairs, and slight quality degradation in the perception of color is inevitable in Pair-B, but color perception remains almost intact in Pair-C.

V. CONCLUSIONS

This paper has presented a novel asymmetric coding method of MVD-based 3-D video. In the proposed method, a bit allocation model is first established to characterize the view rendering distortion, and a chrominance reconstruction model is proposed to characterize the binocular suppression in human visual system. Then, the proposed asymmetric 3-D video coding method is obtained by combining the two models. Experimental results show that the proposed method can greatly improve the performance of view rendering under the total bitrate constraint. In this work, the proposed method encodes the texture video of right view only considering perceptual color redundancy. In further work, we will focus on the following two aspects. Firstly, other perceptual models, such as binocular just noticeable difference or just noticeable difference in depth, can be used to drive further performance improvement. Secondly, subjective perceptual experiments should be implemented to validate the visual redundancy for further target bit allocation scheme design.

REFERENCES

[1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, "Three-dimensional video postproduction and processing," *Proc. IEEE*, vol. 99, no. 4, pp. 607–625, Apr. 2011.

[2] C. Fehn, R. de la Barré, and S. Pastoor, "Interactive 3-DTV-Concepts and key technologies," *Proc. IEEE*, vol. 94, no. 3, pp. 524–538, Mar. 2006.

[3] M. Tanimoto, "Overview of free viewpoint television," *Signal Process.: Image Commun.*, vol. 21, no. 6, pp. 454–461, Jul. 2006.

[4] F. Shao, G. Y. Jiang, M. Yu, and H. S. Ho, "Fast color correction for multi-view video by modeling spatio-temporal variation," *J. Vis. Commun. Image Represent.*, vol. 21, no. 5–6, pp. 392–403, Jun. 2010.

[5] A. Smolic, K. Müller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllides, and A. Koz, "Coding algorithms for 3DTV—a survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1606–1621, Nov. 2007.

[6] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," *Signal Process.: Image Commun.*, vol. 24, no. 1–2, pp. 65–72, Jan. 2009.

[7] ISO/IEC JTC1/SC29/WG11, Joint Multiview Video Model (JMVM) 7.0, Doc. JVT-Z207. Antalya, Turkey, 2008.

[8] ISO/IEC JTC1/SC29/WG11, Description of Exploration Experiments in 3D Video Coding, Doc. N11274. Dresden, Germany, 2010.

[9] ISO/IEC JTC1/SC29/WG11, WD 3 Reference Software for MVC, Doc. JVT-AC207. Busan, Korea, 2008.

[10] ISO/IEC MPEG & ITU-T VCEG, Draft Reference Software for MVC, Doc. JVT-AE207. London, U.K., 2009.

[11] Y. Morvan, D. Farin, and P. H. N. de With, "Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview image," in *Proc. IEEE Int. Conf. Image Processing*, San Antonio, TX, Sep. 2007, pp. 105–108.

[12] M. Maitre and M. N. Do, "Depth and depth-color coding using shape-adaptive wavelet," *J. Vis. Commun. Image Represent.*, vol. 21, no. 5–6, pp. 513–522, Jul. 2010.

[13] G. Cernigliaro, F. Jaureguizar, A. Ortega, J. Cabrera, and N. Garcia, "Fast mode decision for multiview video coding based on depth maps," in *Proc. SPIE,* vol. 7257, article no. 72570N, Jan. 2009.

[14] I. Daribo and C. Tillier, "Motion vector sharing and bitrate allocation for 3D video-plus-depth coding," *EURASIP J. Adv. Signal Process.*, pp. 1–13, Jan. 2009.

[15] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P. H. N. de With, and T. Wiegand, "The effect of multiview depth video compression on multiview rendering," *Signal Process.: Image Commun.*, vol. 24, no. 1–2, pp. 73–88, Jan. 2009.

[16] Y. Morvan, D. Farin, and P. H. N. de With, "Joint depth/texture bit-allocation for multi-view video compression," in *Proc. Picture Coding Symp.*, Lisbon, Portugal, Nov. 2007.

[17] Y. W. Liu, Q. M. Huang, S. W. Ma, D. B. Zhao, and W. Gao, "Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model," *Signal Process.: Image Commun.*, vol. 24, no. 8, pp. 666–681, Sep. 2009.

[18] W. S. Kim, A. Ortega, P. L. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," in *Proc. IEEE Int. Conf. Image Processing*, Cairo, Egypt, Nov. 2009, pp. 721–724.

[19] P. Seuntiëns, L. Meesters, and W. IJsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Trans. Appl. Percept.*, vol. 3, no. 2, pp. 95–109, Apr. 2006.

[20] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, and J. Kim, "Asymmetric coding of stereoscopic video for transmission over T-DMB," in *Proc. 3DTV Conf.*, Kos Island, Greece, May 2007, pp. 1–4.

[21] Y. Chen, Y. K. Wang, M. M. Hannuksela, and M. Gabbouj, "Picture-level adaptive filter for asymmetric stereoscopic video," in *Proc. IEEE Int. Conf. Image Processing*, San Diego, CA, Oct. 2008, pp. 1944–1947.

[22] P. Aflaki, M. M. Hannuksela, J. Häkkinen, P. Lindroos, and M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video," in *Proc. IEEE Int. Conf. Image Processing*, Hong Kong, Sep. 2010, pp. 4021–4024.

[23] Y. Termin and G. A. Kaminka, "Color stereoscopic images requiring only one color image," *Optic. Eng.*, vol. 46, no. 8, p. 087003, Aug. 2007.

[24] F. Shao, G. Y. Jiang, M. Yu, and J. Y. You, "Asymmetric multi-view video coding based on chrominance reconstruction," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Singapore, Jul. 2010, pp. 205–208.

[25] ISO/IEC JTC1/SC29/WG11, Distortion Model of Virtual Views Based on Texture and Depth Compression, Doc.M171880. Kyoto, Japan, 2010.

[26] P. Ramanathan and B. Girod, "Rate-distortion analysis for light field coding and streaming," *Signal Process.: Image Commun.*, vol. 21, no. 6, pp. 462–475, Jul. 2006.

[27] P. Lai, A. Ortega, C. C. Dorea, P. Yin, and C. Gomila, "Improving view rendering quality and coding efficiency by suppressing compression artifacts in depth-image coding," in *Proc. SPIE—Visual Communication and Image Processing,* article no. 72570O, San Jose, CA, Jan. 2009.

[28] F. Shao, G. Y. Jiang, X. Wang, and M. Yu, "Stereoscopic video coding with asymmetric luminance and chrominance qualities," *IEEE Trans. Consum. Electron.*, vol. 56, no. 4, pp. 2460–2468, Nov. 2010.

[29] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[30] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[31] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, Adaptive Basic Unit Layer Rate Control for JVT, Doc. JVT-G012. Pattaya, Thailand, 2003.

[32] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, Aug. 2004.

[33] A. Torralba and W. T. Freeman, "Properties and applications of shape recipes," in *Proc. IEEE Computer Vision and Pattern Recognition*, Madison, WI, Jun. 2003, vol. 2, pp. 383–390.

[34] 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, 2004.

[35] ISO/IEC JTC1/SC29/WG11, Depth Estimation Reference Software (DERS) 3.0, Doc. M16390. Maui, HI, 2009.

[36] ISO/IEC JTC1/SC29/WG11, 3DV/FTV EE2: Report on VSRS Extrapolation, Doc. M18356 Guangzhou, China, 2010.

[37] ITU-T SG16/Q6, Calculation of Average PSNR Differences Between RD-curves, Doc. VCEG-M33. Austin, TX, 2001.

Feng Shao received the B.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in electronic science and technology.

He is currently an Associate Professor in the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. His research interests include video coding, image processing and perception, etc.

Gangyi Jiang received the M.S. degree from Hangzhou University, Hangzhou, China, in 1992, and the Ph.D. degree from Ajou University, Suwon, Korea, in 2000.

He is now a Professor in the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. His research interests mainly include video compression, multi-view video coding, etc.

Mei Yu received the M.S. degree from Hangzhou Institute of Electronics Engineering, Hangzhou, China, in 1993, and the Ph.D. degree form Ajou University, Suwon, Korea, in 2000.

She is now a Professor in the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. Her research interests include image/video coding and video perception.

Ken Chen received the M.S. degree from University of Akron, Akron, OH, in 1996, and the Ph.D. degree from West Virginia University, Morgantown, WV, in 2000.

Currently, he is serving as an Associate Professor in the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. His research interests include image and video processing, intelligent control, etc.

Yo-Sung Ho (SM'06) received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, Korea, in 1981 and 1983, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990.

He joined Electronics and Telecommunications Research Institute (ETRI), Daejon, Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, NY, where he was involved in development of the Advanced Digital High-Definition Television (AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korean DBS digital television and high-definition television systems. Since 1995, he has been with Gwangju Institute of Science and Technology (GIST), Gwangju, Korea, where he is currently a Professor of the Information and Communications Department. His research interests include digital image and video coding, image analysis and image restoration, advanced video coding techniques, digital video and audio broadcasting, three-dimensional video processing, and content-based signal representation and processing.