

Article

Multiple Description Coding Based on Optimized Redundancy Removal for 3D Depth Map

Sen Han ¹, Huihui Bai ^{1,*} and Mengmeng Zhang ²

¹ Institute of Information Science, Beijing Jiaotong University, No.3 Shang Yuan Cun, Hai Dian, Beijing 100044, China; 14120315@bjtu.edu.cn

² College of Information Engineering, North China University of Technology, No.5 Jinyuanzhuang Road, Shijingshan, Beijing 100144, China; zmm@ncut.edu.cn

* Correspondence: hhhbai@bjtu.edu.cn; Tel.: +86-10-5168-4108; Fax: +86-10-5168-8626

Academic Editor: Kevin H. Knuth

Received: 18 May 2016; Accepted: 24 June 2016; Published: 29 June 2016

Abstract: Multiple description (MD) coding is a promising alternative for the robust transmission of information over error-prone channels. In 3D image technology, the depth map represents the distance between the camera and objects in the scene. Using the depth map combined with the existing multiview image, it can be efficient to synthesize images of any virtual viewpoint position, which can display more realistic 3D scenes. Differently from the conventional 2D texture image, the depth map contains a lot of spatial redundancy information, which is not necessary for view synthesis, but may result in the waste of compressed bits, especially when using MD coding for robust transmission. In this paper, we focus on the redundancy removal of MD coding based on the DCT (discrete cosine transform) domain. In view of the characteristics of DCT coefficients, at the encoder, a Lagrange optimization approach is designed to determine the amounts of high frequency coefficients in the DCT domain to be removed. It is noted considering the low computing complexity that the entropy is adopted to estimate the bit rate in the optimization. Furthermore, at the decoder, adaptive zero-padding is applied to reconstruct the depth map when some information is lost. The experimental results have shown that compared to the corresponding scheme, the proposed method demonstrates better rate central and side distortion performance.

Keywords: multiple description; depth map; DCT; high-frequency

1. Introduction

3D images and video, which offer us a stereoscopic and immersive multimedia experience, have attracted increasing attention [1]. Recently, the applications of 3D images and video have become more and more popular in many fields, including cinema, TV broadcasting, streaming and smart phones [2]. Differently from the conventional 2D texture image, the depth map, as a special format of 3D image data, represents the distance information between a camera and the objects in the scene. Each pixel of the image frames of depth data streams represents the distance, which is to the camera plane from a specific object in the view of the depth sensor. Depth maps are often treated as gray-scale image sequences, which are similar to the luminance component of texture videos. However, the depth map has its own special characteristics. First, the depth map signal is much sparser than the texture image. It contains no texture, but has sharp object boundaries because the gray levels are nearly the same in most regions within an object, but change abruptly across boundaries. Furthermore, the depth map is not directly used for display, but it plays an important role in the virtual view synthesis.

During the past few years, many studies have focused on the depth map. Edge-based compression was applied to the depth map in [3], which utilizes JBIG (Joint Bi-level Image experts Group) to encode the contours and DPCM (Differential Pulse Code Modulation) to compress both the pixels around

the contours, as well as the uniform sparse sampling points of the depth map. The paper in [4] uses a symmetric smoothing filter to smooth the whole depth map before the image warping. While this method decreases the number of holes as the smoothing strength becomes stronger, some geometric distortions also become visible on regions with vertically straight lines. Instead of coding numerous views, two or three texture views with their corresponding depth maps are used in the multiview video coding (MVC) standard. Virtual views can be generated by specifying the preference viewing angles through depth-image-based rendering (DIBR) [5]. Later, many scholars applied the compressive sensing (CS) theory to the texture image and depth map coding. Compressive sensing (CS) theory has been proposed by Cands and Donoho in [6–8]. The CS applications in the image/video coding can be found in [9–12]. In [9], according to a spatially sparse signal, a new image/video coding approach is proposed, which combines the CS theory with the traditional discrete cosine transform (DCT)-based coding method to achieve better compression efficiency. In [10], using block compressed sensing, image acquisition is conducted in a block-by-block manner through the same operator. It is claimed that it can sufficiently capture the complicated geometric structures of natural images. In [11,12], compressed sensing theory is used in the depth map compression, which subsamples the depth map in the frequency domain by choosing a sub-sampling matrix in order to ensure the incoherence of the measurements. The two new papers in [13,14] propose two kinds of exponential wavelet iterative shrinkage thresholding algorithm based on the theory of compressed sensing theory. The work in [13] is only for magnetic resonance images, and [14] is with the random shift for the magnetic resonance image. Two schemes are tested by four kinds of magnetic resonance images, and they have a better reconstruction quality and faster reconstruction speed compared to state-of-the-art algorithms, such as FCSA (Fast Composite Splitting Algorithm), ISTA (Iterative Shrinkage/Threshold Algorithm) and FISTA (Fast Iterative Shrinkage/Threshold Algorithm).

Multiple description (MD) is a coding technique that has emerged as a promising approach to enhance the fault tolerance of a video delivery system [15]. In 1993, the first work on MD coding was introduced in [16]. The original video signal can be split into multiple bit streams (descriptions) using an MD encoder. Then, these MDs can be transmitted over multiple channels. There is a very low probability when all channels fail at the same time. Therefore, at the MD decoder, only one description received can reconstruct the video with acceptable quality, and the resultant distortion is called side distortion. Of course, more descriptions can produce the video with better quality. In a simple architecture of two channels, the distortion with two received descriptions is called central distortion [17].

Later, a number of methods of MDC have been presented, mainly including the MD scalar quantizer [18], the MD lattice vector quantizer [19,20], MD based on pairwise correlating transforms [21] and MD based on FEC (Forward Error Correction) [22]. All of the above methods are difficult to apply in practical applications because these specially-designed MD encoders are not compatible with the widely-used standard codec. To overcome the limitation, some standard-compliant MD video coding schemes were designed to achieve promising results [23,24]. Another significant class of MDC is based on pre- and post-processing. In preprocessing, the original source is split into multiple sub-sources before encoding, and then, these sub-sources can be encoded separately by the standard codec to generate multiple descriptions. The typical versions are MDC based on spatial sampling [25] and temporal sampling [26,27].

In this paper, considering the special characteristics of depth information, an MDC scheme is proposed based on optimized redundancy removal for the 3D depth map. Firstly, the depth map is transformed from the pixel domain into the DCT domain. According to the characteristics of DCT coefficients, at the encoder, a Lagrange optimization approach is designed to determine the amounts of high frequency coefficients in the DCT domain to be removed. Considering the low computing complexity, the entropy is adopted to estimate the bit rate in the optimization. Then, the processed depth map is transmitted, which is separated in accordance with odd and even rows or columns.

We have differently adaptive decoding methods for the two situations. Furthermore, at the decoder, adaptive zero-padding is applied to reconstruct the depth map when some information is lost.

The rest of this paper is organized as follows. The proposed scheme is presented in Section 2. In Section 3, the performance of the proposed scheme is investigated through extensive simulation in different situations. Additionally, the paper is finally concluded in Section 4.

2. Proposed Scheme

2.1. Encoder

The block diagram of the proposed encoder has been illustrated in Figure 1. Here, the depth maps of any view can be encoded as follows. At the encoder, the depth map can be firstly transformed into the DCT domain. According to the characteristics of the high frequency coefficients of DCT, a Lagrange optimization approach is designed to determine the redundancy to be removed. After inverse DCT (IDCT, Inverse Discrete Cosine Transformation), the processed depth map can be separated by the means of odd or even lines. Here, the generated sub-images can be encoded by any standard codec to produce two descriptions.

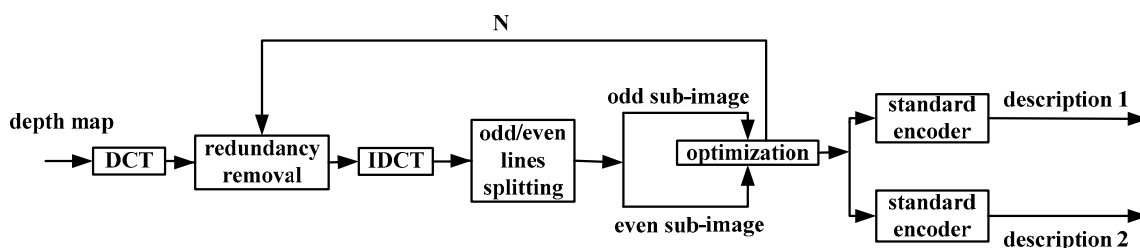


Figure 1. The diagram of the proposed encoder. DCT, discrete cosine transform; IDCT, inverse DCT.

As is depicted in the block diagram, by removing the N lines' high frequency coefficients in the DCT domain, approximate redundant information can be reduced. Here, N is very important for the entire coding process, which has an effect on the reconstructed image quality at the decoder. The side and central distortion will be sensitive to the value N . When N increases, both side and central distortion will increase. On the other hand, the bit rates associated with side and central distortion will decrease with the increase of N . To achieve an optimized value for N , the following optimization problem should be solved.

$$\min_N J(R(N), D_S(N), D_C(N), \lambda)$$

In the above optimization problem, we need to consider the bit rate and the distortion of reconstructed images from the side and central decoder. Furthermore, it can be rewritten as a Lagrange optimized formulation, as shown in Equation (1). Here, let $J(N)$ denote the Lagrange cost function.

$$J(N) = D_S(N) + \lambda_1 D_C(N) + \lambda_2 R(N) \tag{1}$$

$R(N)$, $D_S(N)$ and $D_C(N)$ denote the bit rate, side and central distortion of the image, respectively. Additionally, λ_1 and λ_2 are balanced parameters to play a role in $R(N)$, $D_S(N)$ and $D_C(N)$. In this paper, the side distortion $D_S(N)$ and the central distortion $D_C(N)$ will be regarded as the same significance, so $\lambda_1 = 1$. In the experimental results, we will discuss the rate distortion performance under different values of λ_2 .

Solving the optimization problem in Equation (1) is difficult in multiple description coding. Furthermore, given λ_2 , a triple $(D_S(N), D_C(N), R(N))$ needs to be computed for each mode selection. This requires actual time-consuming encoding and decoding for each mode. Taking the lower encoding

complexity into account, we use the entropy of the image instead of the bit rate $R(N)$. Here, $R_e(N)$ represents the entropy of a depth map. It is defined as follows:

$$R_e(N) = -\sum_{i=0}^k (p_i \log_2 p_i) \tag{2}$$

where p_i presents the probability of the gray level i in the depth map and k is the number of gray levels.

In this paper, we need to count the possibility of the appearance of each gray level in a depth map, which will be used in the calculation of the entropy. The equation of the possibility count is expressed as:

$$p_i = \frac{s_i}{S} \tag{3}$$

where S is the number of symbols in an image and s_i is the number of the pixels whose value is i . When we get all of the p_i of an image, we can calculate the entropy according to Equation (2).

Furthermore, the MSE (mean square error) can be used to represent to $D_S(N)$ and $D_C(N)$. Let I_1, I_2 denote two images, and then, their MSE can be obtained by the following formula:

$$MSE(I_1, I_2) = \sum_i \sum_j (I_1(i, j) - I_2(i, j))^2 \tag{4}$$

In this paper, the $D_C(N)$ is the MSE of the reconstructed image from the central decoder with the original map, and the final result of $D_S(N)$ is the average MSE of two reconstructed images from the side decoder. Here, MSE_1 is used to compute the distortion from one side decoder, and MSE_2 is used to compute the distortion from the other side decoder.

$$D_S(N) = (MSE_1 + MSE_2)/2 \tag{5}$$

In this paper, we use the iterative approach to solve N . Because the block in the encoder is set to be 16, the value of N is in the limit of 16 and $16m$, and m is different in the different sequences. While m is too small, it may not find a better value for N , and the algorithm stops; while m is too large, the value of $R_e(N)$ will be increasing, and it will lose the role of λ_2 .

The processed depth map is separated in accordance with odd and even lines in the DCT domain, then initializing N and λ_2 . We can compute the value of $R_e(N)$, $D_S(N)$ and $D_C(N)$ with the two sub-images. Then, we calculate $J(N)$ according to the front three values and circulate N . When N is not in the limit of 16 and $16m$, the optimized N is chosen in accordance with the minimum value of $J(N)$. The basic optimized algorithm is shown in Figure 2.

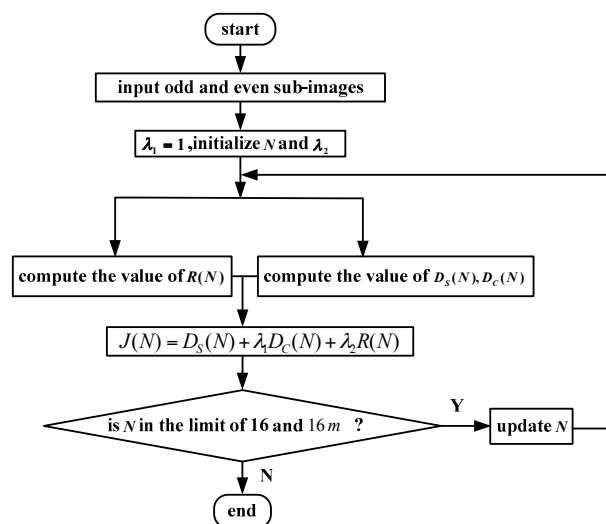


Figure 2. Optimized encoding process.

2.2. Decoder

At the decoder, there are two ways for MDC. For a depth map, the distortion with both received descriptions is called central distortion, and the distortion with only one received description is called side distortion [15]. The block diagram of the proposed decoder has been illustrated in Figure 3.

Firstly, the design of the side decoder is discussed as follows. For the side decoder, only Description 1 or 2 can be received. Since the processes for Descriptions 1 and 2 are the same, we use Description 1 as the example. After standard decoding, the decoded odd sub-image will be transformed into the DCT domain. Then, according to the resolution of the original image, adaptive zero-padding can be applied to fill the approximate zeros for the DCT coefficients of the odd sub-image. In the end, after inverse DCT, the reconstructed depth image can be achieved, which can be used to calculate the side distortion.

Then, we will discuss the design of the central decoder. For the central decoder, both descriptions can be received. Therefore, after standard decoding, Descriptions 1 and 2 are rearranged according to the odd and even lines. Then, with the same process in the side decoder, adaptive zero-padding can be used to fill zeros in the DCT domain. At last, we can obtain the reconstruction with central distortion.

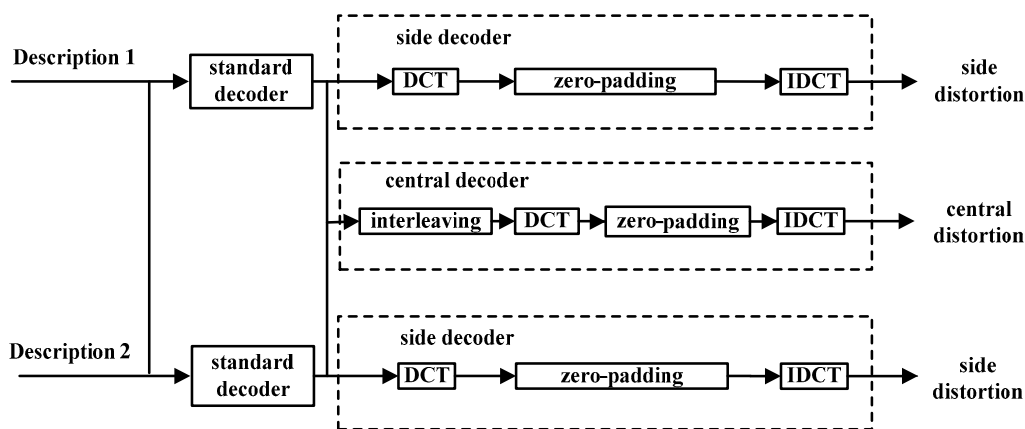


Figure 3. The diagram of the proposed decoder.

3. Experimental Results

To highlight the performance of the proposed scheme, the experiments are implemented on four standard sequences of depth maps, including *Balloons*, *Kendo*, *Newspaper*, *Dancer* and *Pantomime*, which can be download from [28]. The detailed information and the choice of some parameters of the tested sequences are shown in Table 1.

Table 1. The parameters of the tested sequences.

Sequence	Resolution	Frame Rate	λ_2 for Left View	λ_2 for Right View	m
Balloons	1024 × 768	25	0.25	0.25	25
Kendo	1024 × 768	25	0.2	0.2	25
Newspaper	1024 × 768	25	0.5	0.5	25
Pantomime	1280 × 960	25	0.15	0.04	30
Dancer	1920 × 1088	25	0.025	0.01	40

This paper focuses on the comparison of the proposed optimized scheme against the conventional scheme. To prove the universality of the experiment, some groups of data in each sequence are selected for comparison. According to the MDC quality assessment, we compare not only the rate central distortion performance when two descriptions can be received correctly, but also the rate side distortion performance when only one description can be received.

The objective and subjective quality comparison for the sequences of *Balloons*, *Kendo*, *Newspaper*, *Dancer* and *Pantomime* is presented in Figures 4–8.

In the four figures, the horizontal axis represents the bit rate, and the vertical axis represents the PSNR values. Here, one view is chosen for each of the four sequences for comparison in (a): the first view of *Balloons*, *Kendo* and *Dancer*, the second view of *Newspaper* and the 37th view of *Pantomime*. Compared to the basic comparison scheme, our proposed depth map coding scheme can on average improve around 1.2 dB for *Balloons*, 0.4 dB for *Kendo*, 0.5 dB for *Newspaper*, 2.6 dB for *Dancer* and 3.0 dB for *Pantomime* in side distortion and around 1.2 dB for *Balloons*, 0.5 dB for *Kendo*, 0.8 dB for *Newspaper*, 3 dB for *Dancer* and 3.1 dB for *Pantomime* in central distortion at the same or a very close bit rate.

Another view is chosen for each of the above four sequences for comparison in (b): the third view of *Balloons*, *Kendo*, *Dancer*, the fourth view of *Newspaper* and the 39th view of *Pantomime*. Compared to the basic comparison scheme, our proposed depth map coding scheme can on average improve around 0.7 dB for *Balloons*, 0.6 dB for *Kendo*, 0.6 dB for *Newspaper*, 3 dB for *Dancer* and 2.8 dB for *Pantomime* in side distortion and around 0.8 dB for *Balloons*, 0.6 dB for *Kendo*, 0.8 dB for *Newspaper*, 3.3 dB for *Dancer* and 2.8 dB for *Pantomime* in central distortion at the same or a very close bit rate.

Given that the depth map is not directly used for display, the objective and subjective qualities of the rendered virtual views should be taken into account. In the aspect of objective quality, the synthesized virtual viewpoint image can be achieved by two original camera images. For example, for the tested sequences *Balloons*, *Kendo* and *Dancer*, the depth and texture from the first and third views can be used to synthesize the texture of the second view, while for the sequence *Newspaper*, the depth and texture from the second and fourth views can generate the texture of the third view, and for the sequence *Pantomime*, the depth and texture from the 37th and 39th views can be used to synthesize the texture of the 38th view. Here, the texture image is also compressed, and the comparison for synthesized images is presented in (c). Compared to the basic comparison scheme, our proposed synthesized scheme can on average improve around 1.1 dB for *Balloons*, 0.5 dB for *Kendo*, 0.4 dB for *Newspaper*, 2.2 dB for *Dancer* and 2.6 dB for *Pantomime* in side distortion and around 1.1 dB for *Balloons*, 0.6 dB for *Kendo*, 0.4 dB for *Newspaper*, 2.2 dB for *Dancer* and 2.7 dB for *Pantomime* in central distortion at the same or a very close bit rate.

Furthermore, the advantages of the proposed scheme can be more clearly seen in Figures 9–12, in which the subjective quality of the synthesized virtual viewpoint of *Kendo* and *Dancer* are presented, especially in some parts denoted by the red rectangle.

The structural similarity (SSIM) index [29] is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric. SSIM is designed to improve on traditional methods, such as peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have proven to be inconsistent with human visual perception. At the end of each figure, the SSIM values for the sequences of *Balloons*, *Kendo*, *Newspaper*, *Dancer* and *Pantomime* are reflected in the Tables 2–11; and Tables 2, 4, 6, 8 and 10 are the SSIM values of five sequences for the side decoder; Tables 3, 5, 7, 9 and 11 are the SSIM values of five sequences for central decoder. As we can see from the tables, the depth map formed by the proposed scheme and the SSIM values of the synthesized image with the uncompressed original images are close to one, and the values of the proposed scheme are greater than or equal to that of the comparison scheme. This indicates that the proposed scheme is better than the basic comparison scheme.

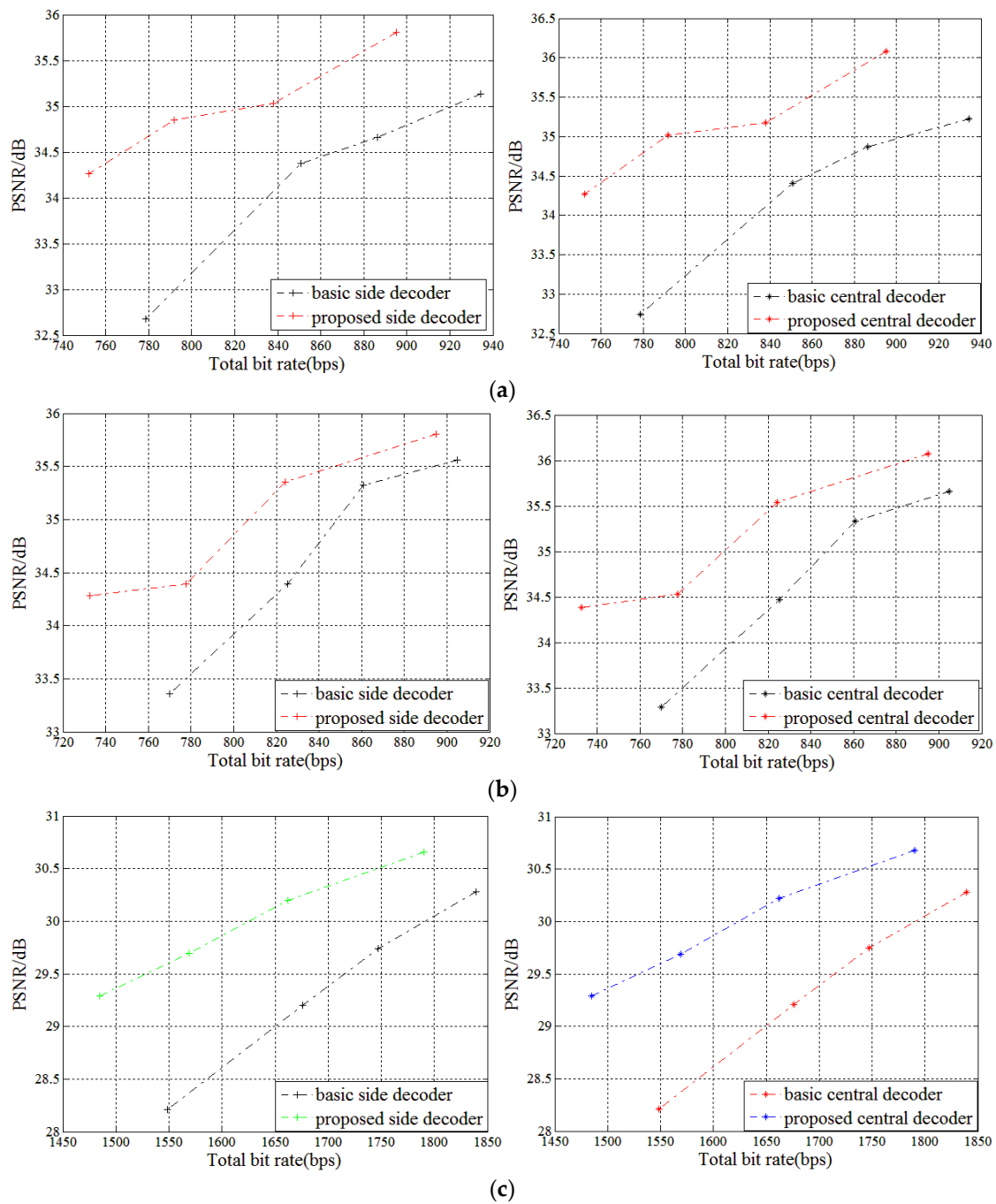


Figure 4. Objective quality comparison for the sequence *Balloons*: (a) the left view of depth map; (b) the right view of depth map; (c) the synthesized virtual viewpoint.

Table 2. The SSIM values of the sequence *Balloons* for the side decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Balloons	0.9648	0.9658	0.9693	0.9708	0.9027	0.9087
	0.9625	0.9641	0.9674	0.9694	0.8953	0.9021
	0.9614	0.9623	0.9650	0.9675	0.8882	0.8948
	0.9563	0.9607	0.9609	0.9667	0.8729	0.8884

Table 3. The SSIM values of the sequence *Balloons* for the central decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Balloons	0.9535	0.9545	0.9584	0.9602	0.9023	0.9084
	0.9509	0.9525	0.9562	0.9572	0.8948	0.9014
	0.9466	0.9470	0.9522	0.9528	0.8875	0.8936
	0.9382	0.9449	0.9436	0.9505	0.8718	0.8876

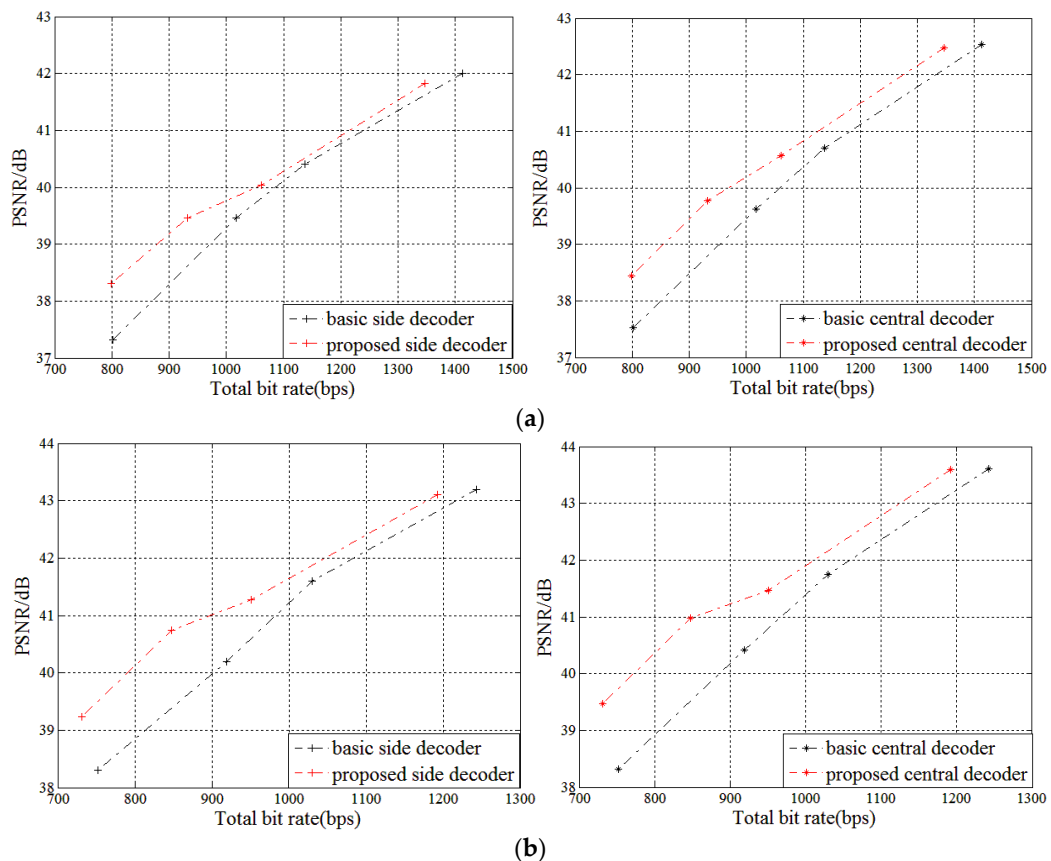


Figure 5. Cont.

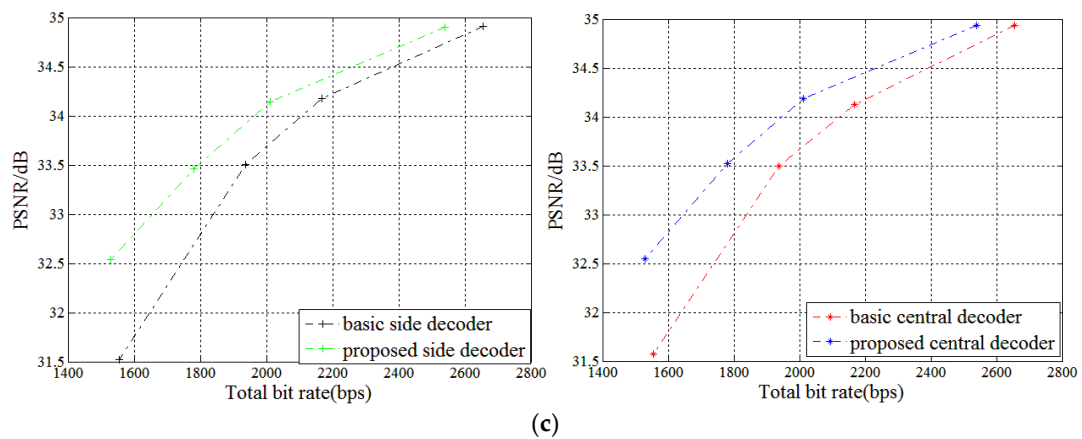


Figure 5. Objective quality comparison for the sequence *Kendo*: (a) the left view of depth map; (b) the right view of depth map; (c) the synthesized virtual viewpoint.

Table 4. The SSIM values of the sequence *Kendo* for the side decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Kendo	0.9817	0.9819	0.9852	0.9858	0.9558	0.9559
	0.9786	0.9788	0.9825	0.9829	0.9496	0.9497
	0.9759	0.9761	0.9808	0.9811	0.9444	0.9445
	0.9697	0.9729	0.9768	0.9785	0.9265	0.9352

Table 5. The SSIM values of the sequence *Kendo* for the central decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Kendo	0.9795	0.9781	0.9830	0.9852	0.9559	0.9563
	0.9738	0.9748	0.9790	0.9805	0.9494	0.9496
	0.9692	0.9699	0.9766	0.9774	0.9444	0.9444
	0.9621	0.9644	0.9707	0.9712	0.9266	0.9352

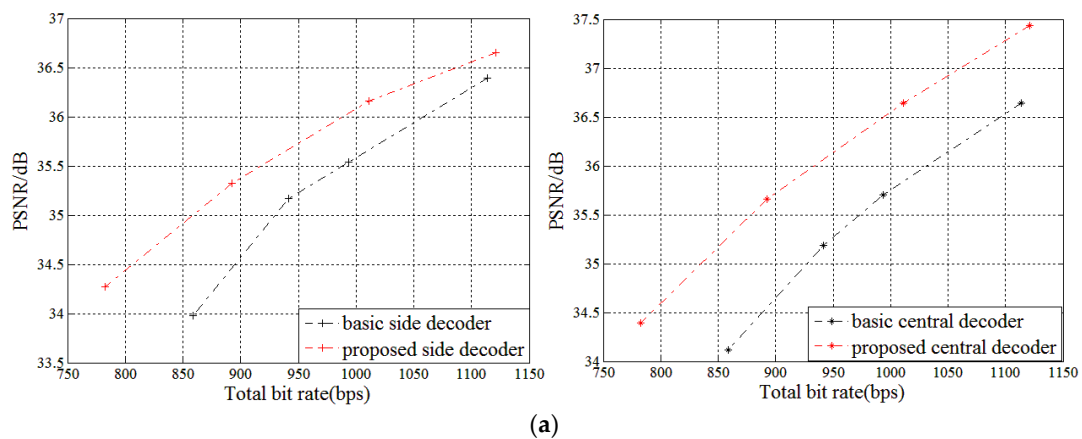


Figure 6. Cont.

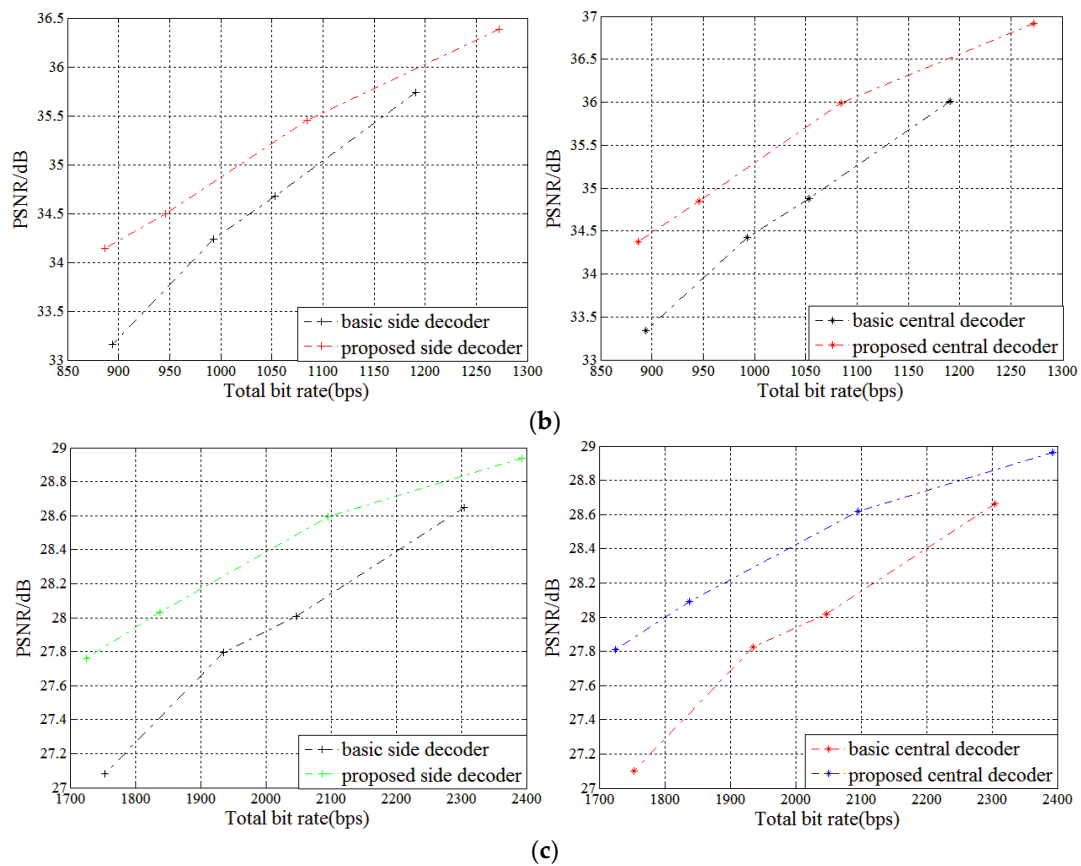


Figure 6. Objective quality comparison for the sequence *Newspaper*: (a) the left view of depth map; (b) the right view of depth map; (c) the synthesized virtual viewpoint.

Table 6. The SSIM values of the sequence *Newspaper* for the side decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Newspaper	0.9619	0.9627	0.9443	0.9475	0.8641	0.8748
	0.9591	0.9604	0.9392	0.9399	0.8482	0.8633
	0.9578	0.9579	0.9372	0.9388	0.8419	0.8482
	0.9524	0.9556	0.9270	0.9373	0.8251	0.8413

Table 7. The SSIM values of the sequence *Newspaper* for the central decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Newspaper	0.9520	0.9539	0.9349	0.9390	0.8645	0.8753
	0.9494	0.9513	0.9236	0.9277	0.8484	0.8640
	0.9445	0.9445	0.9225	0.9225	0.8421	0.8489
	0.9386	0.9386	0.9082	0.9160	0.8254	0.8420

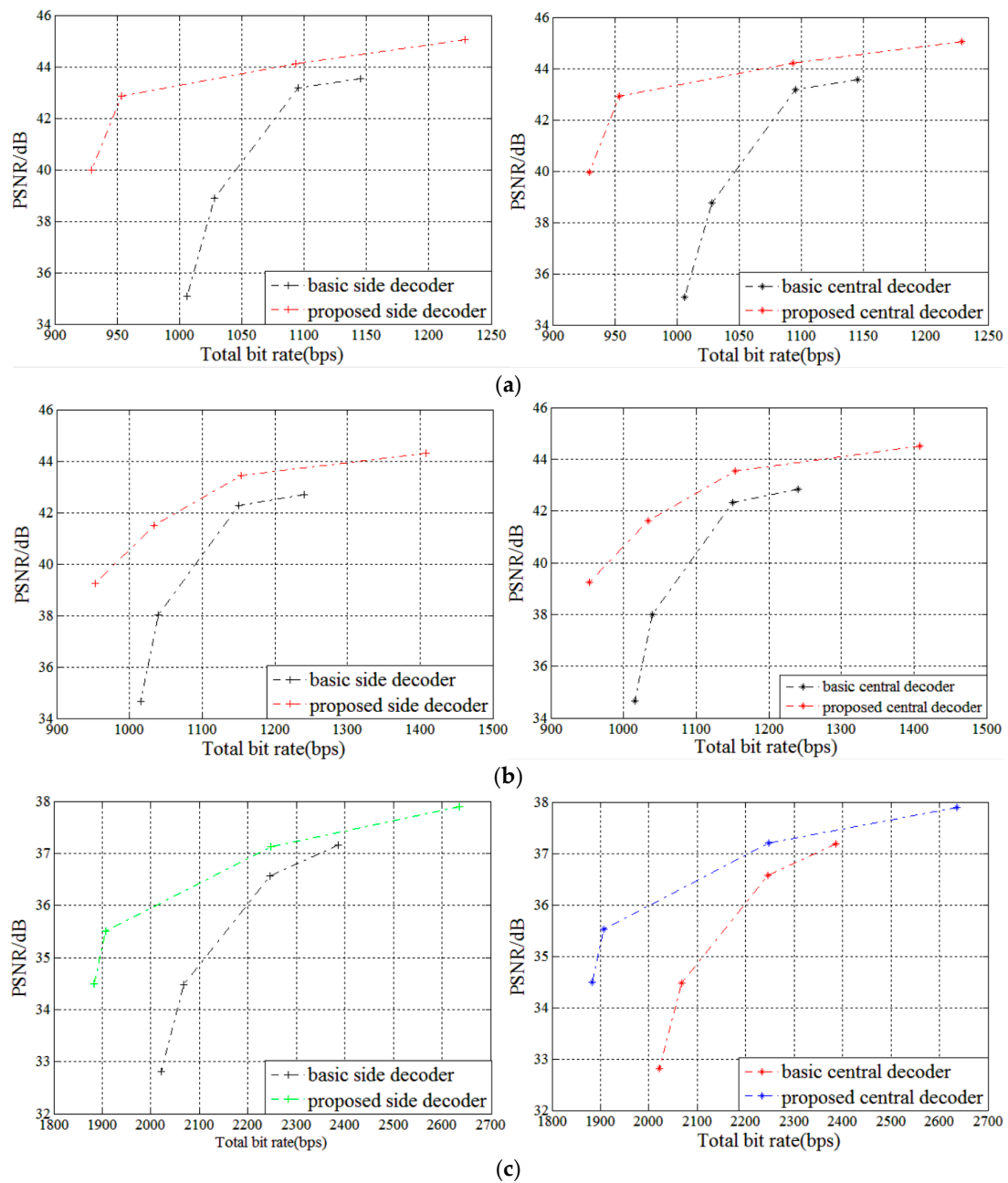


Figure 7. Objective quality comparison for the sequence *Pantomime*: (a) the left view of depth map; (b) the right view of depth map; (c) the synthesized virtual viewpoint.

Table 8. The SSIM values of the sequence *Pantomime* for the side decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Pantomime	0.9850	0.9873	0.9867	0.9914	0.9869	0.9889
	0.9836	0.9848	0.9857	0.9891	0.9833	0.9869
	0.9613	0.9807	0.9612	0.9833	0.9790	0.9812
	0.9219	0.9710	0.9203	0.9728	0.9652	0.9789

Table 9. The SSIM values of the sequence *Pantomime* for the central decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Pantomime	0.9851	0.9873	0.9870	0.9909	0.9869	0.9889
	0.9835	0.9840	0.9857	0.9892	0.9833	0.9869
	0.9603	0.9806	0.9616	0.9832	0.9790	0.9812
	0.9219	0.9704	0.9212	0.9731	0.9652	0.9789

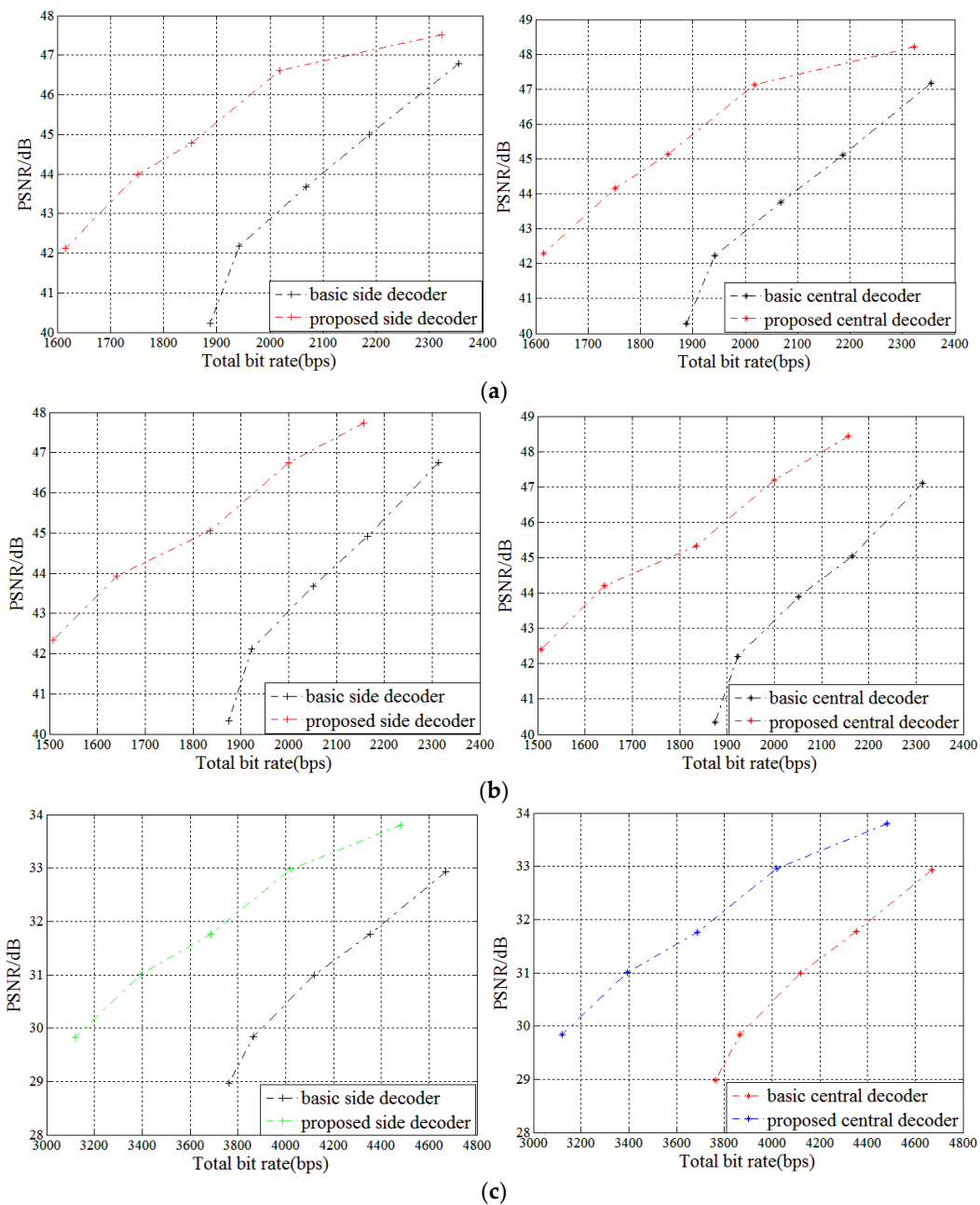


Figure 8. Objective quality comparison for the sequence *Dancer*: (a) the left view of depth map; (b) the right view of depth map; (c) the synthesized virtual viewpoint.

Table 10. The SSIM values of the sequence *Dancer* for the side decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Dancer	0.9996	0.9996	0.9996	0.9996	0.9962	0.9967
	0.9993	0.9995	0.9993	0.9996	0.9957	0.9963
	0.9991	0.9993	0.9991	0.9993	0.9952	0.9957
	0.9986	0.9991	0.9986	0.9992	0.9986	0.9986
	0.9972	0.9987	0.9976	0.9987	0.9932	0.9943

Table 11. The SSIM values of the sequence *Dancer* for the central decoder.

Sequence	SSIM for Left View of Depth Map		SSIM for Right View of Depth Map		SSIM for the Synthesized Virtual Viewpoint	
	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme	Basic Scheme	Proposed Scheme
Dancer	0.9996	0.9997	0.9996	0.9997	0.9962	0.9967
	0.9994	0.9996	0.9993	0.9996	0.9957	0.9963
	0.9991	0.9994	0.9991	0.9994	0.9952	0.9957
	0.9986	0.9992	0.9986	0.9992	0.9943	0.9952
	0.9973	0.9988	0.9976	0.9987	0.9933	0.9943

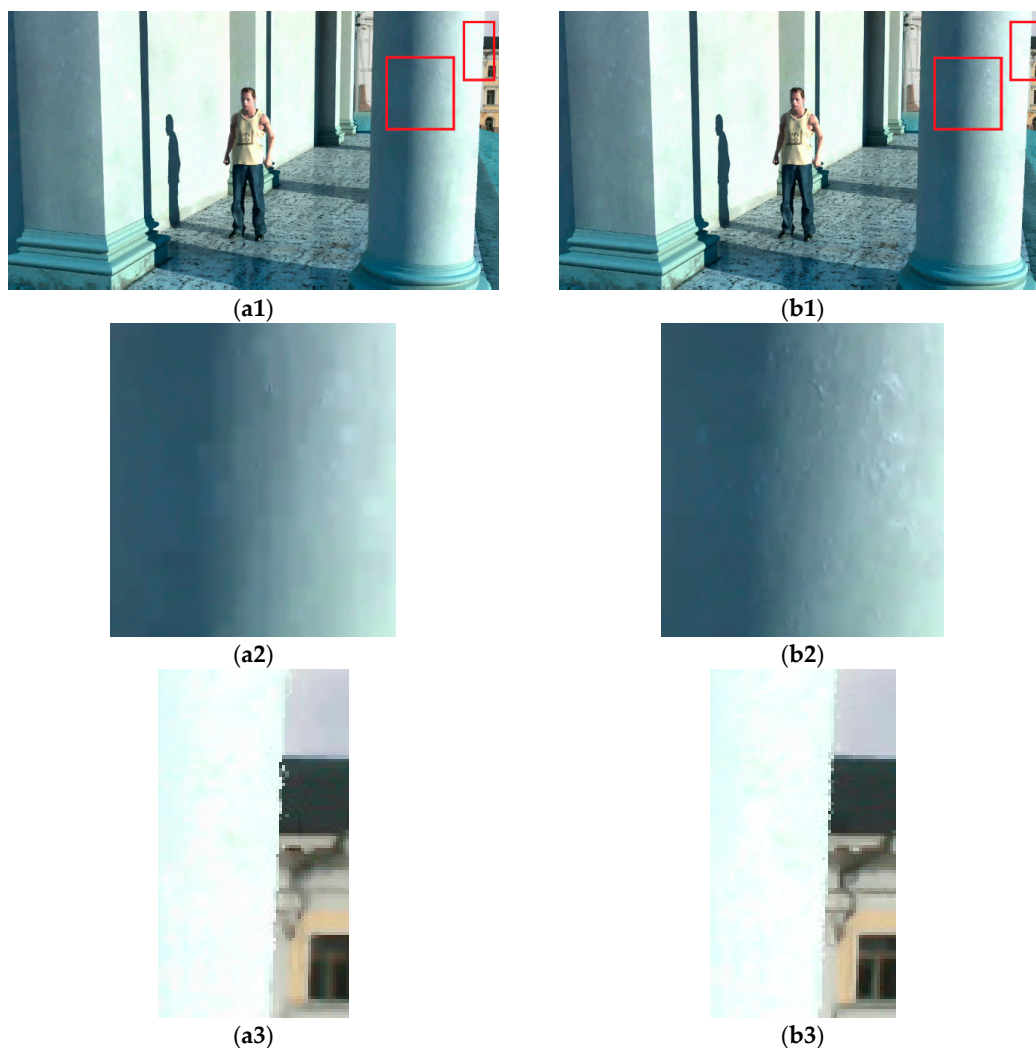


Figure 9. Subjective quality comparison of the side synthesized virtual viewpoint for *Dancer*. (a1,a2,a3) basic comparison scheme; (b1,b2,b3) the proposed scheme.

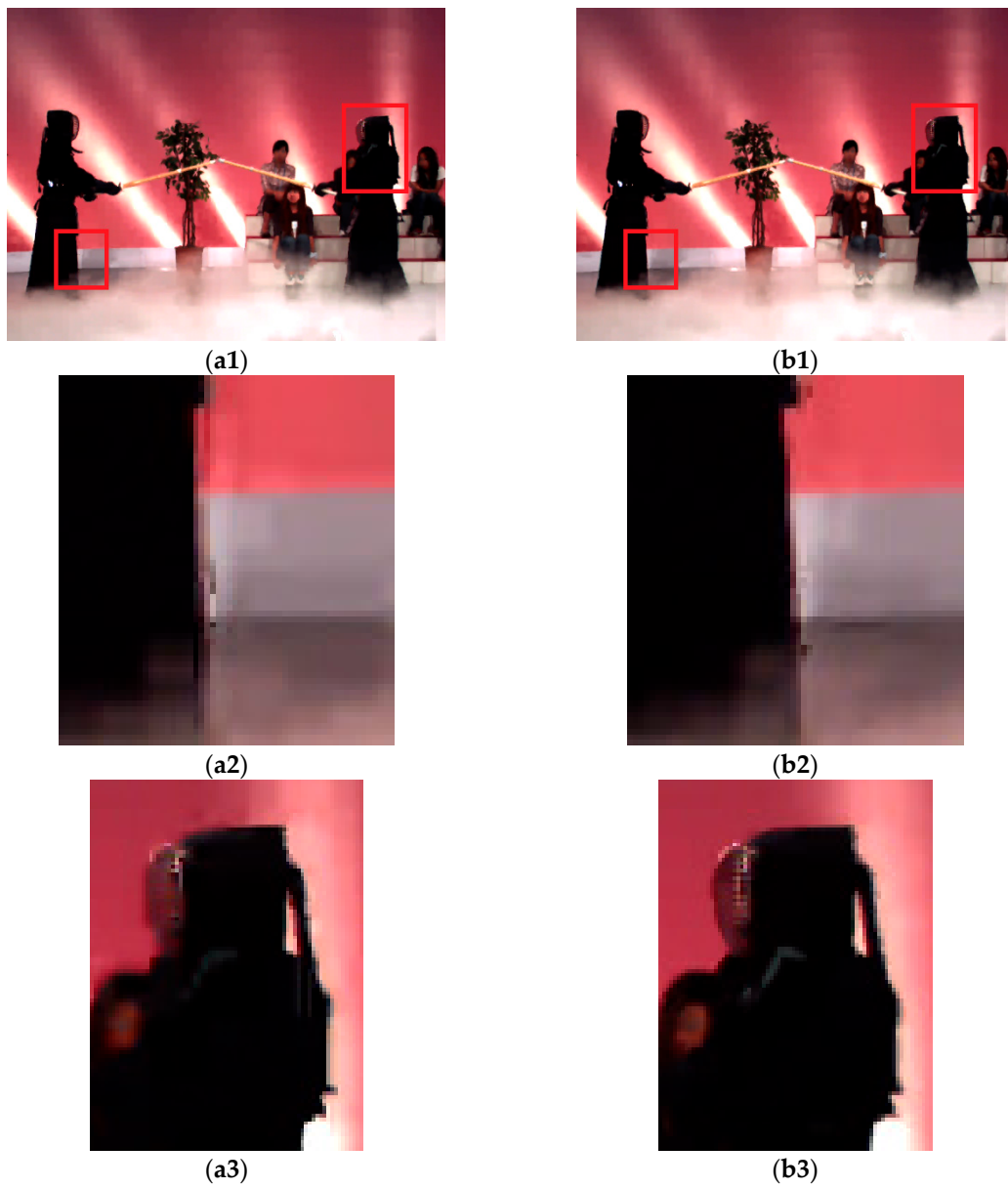


Figure 10. Subjective quality comparison of the side synthesized virtual viewpoint for *Kendo*. (a1,a2,a3) basic comparison scheme; (b1,b2,b3) the proposed scheme.

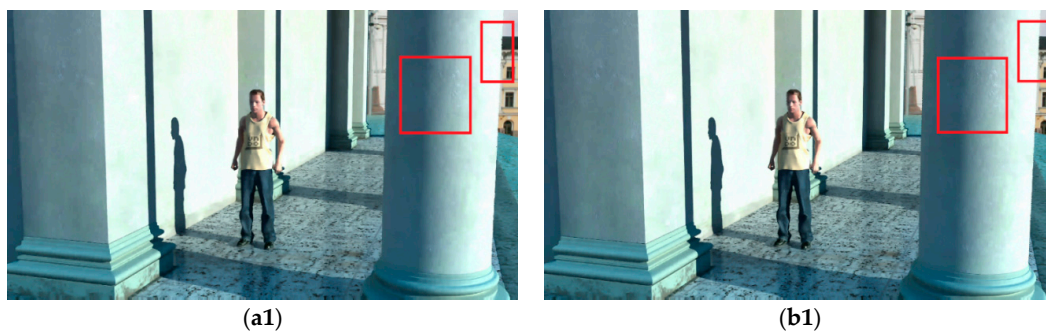


Figure 11. Cont.

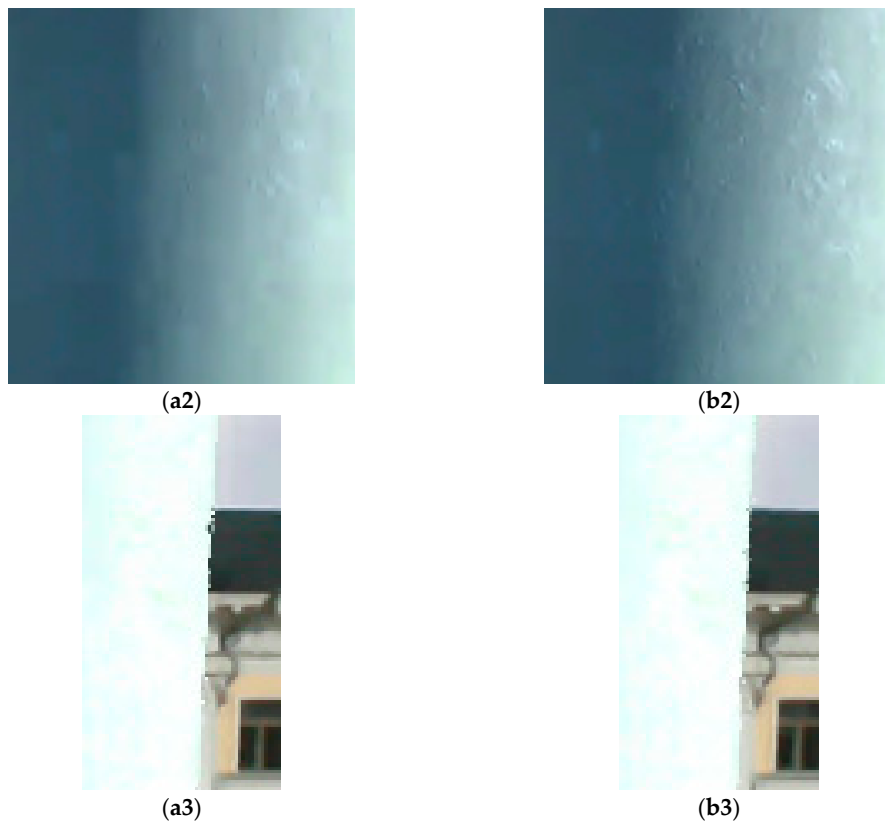


Figure 11. Subjective quality comparison of the central synthesized virtual viewpoint for *Dancer*. (a1,a2,a3) basic comparison scheme; (b1,b2,b3) the proposed scheme.

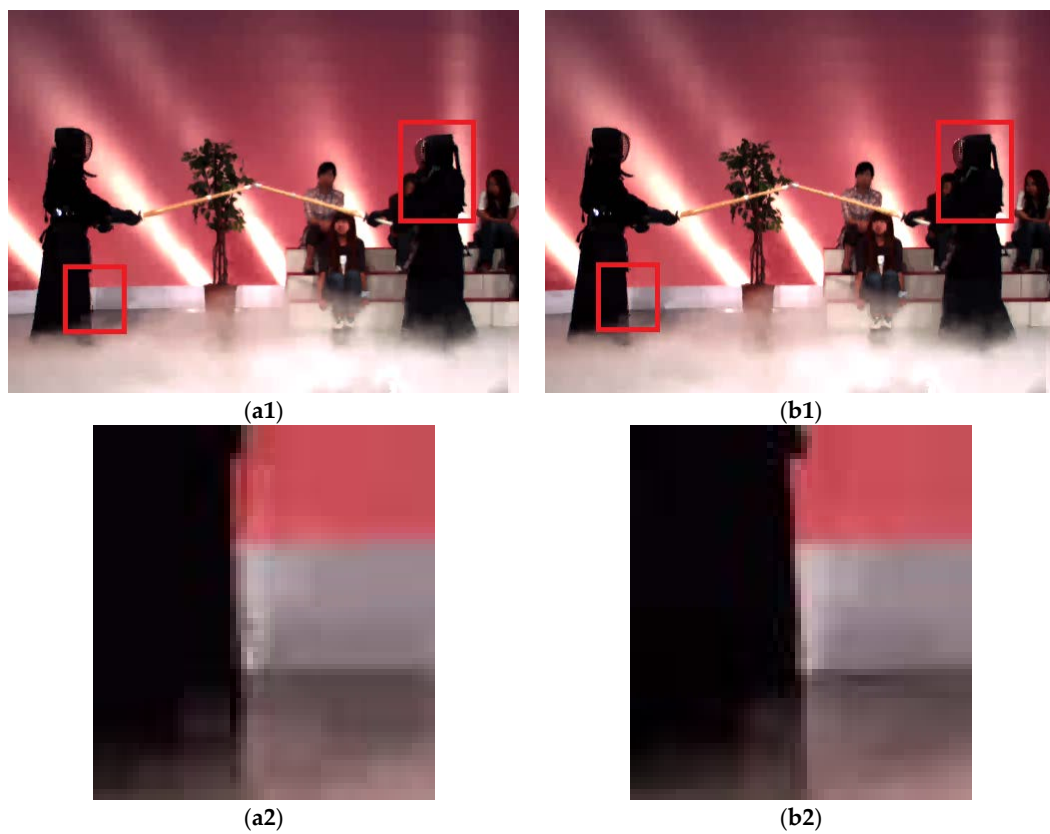


Figure 12. *Cont.*

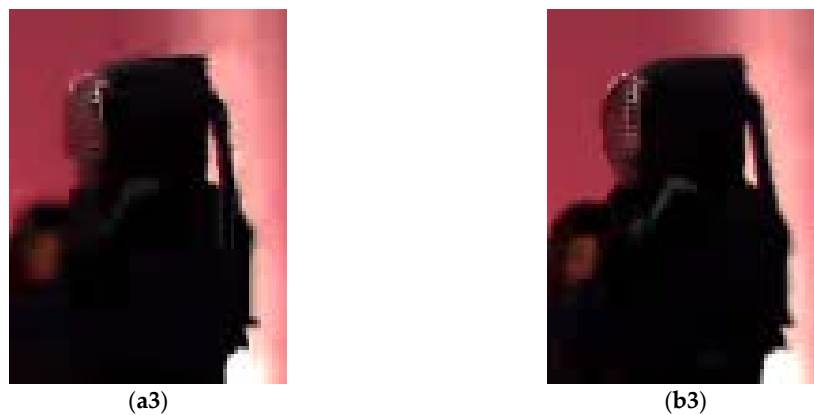


Figure 12. Subjective quality comparison of the central synthesized virtual viewpoint for *Kendo*. (a1,a2,a3) basic comparison scheme; (b1,b2,b3) the proposed scheme.

4. Conclusions

In this paper, the depth map is removed as a part of the high-frequency component in the DCT domain. An effective and adaptive optimization scheme encoding was accommodated in the proposed scheme to achieve a better rate and central/side distortion performance. It can be seen from the experiment that the proposed scheme also has a better subjective quality. Therefore, our proposed scheme is clearly a worthy choice for depth map coding.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China (No. 61272051, No. 61370111), the Beijing Higher Education Young Elite Teacher Project (YETP0543), the Fundamental Research Funds for the Central Universities of China (2014JBM028) and CCF (China Computer Federation)-Tencent Open Fund.

Author Contributions: Sen Han and Huihui Bai conceived of and designed the experiments. Sen Han performed the experiments and analyzed the data. Mengmeng Zhang contributed materials/analysis tools. Sen Han and Huihui Bai wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhu, C.; Zhao, Y.; Yu, L.; Tanimoto, M. *3D-TV System with Depth-Image-Based Rendering*; Springer: New York, NY, USA, 2013.
2. Chen, Y.; Hannuksela, M.; Suzuki, T.; Hattori, S. Overview of the MVC+D 3D video coding standard. *J. Vis. Commun. Image Represent.* **2014**, *25*, 679–688. [[CrossRef](#)]
3. Gautier, J.; le Meur, O.; Guillemot, C. Efficient depth map compression based on Lossless edge coding and diffusion. In Proceedings of the Picture Coding Symposium, Krakow, Poland, 7–9 May 2012; pp. 81–84.
4. Tam, W.J.; Alain, G.; Zhang, L.; Martin, T.; Renaud, R. Smoothing depth maps for improved stereoscopic image quality. In Proceedings of the International Society for Optical Engineering Conference Three-Dimensional TV, Video, and Display III, Philadelphia, PA, USA, 25 October 2004; Volume 5599, pp. 162–172.
5. Kauff, P.; Atzpadin, N.; Fehn, C.; Müller, M.; Schreer, O.; Smolic, A.; Tanger, R. Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability. *Signal Process. Image Commun.* **2007**, *22*, 217–234. [[CrossRef](#)]
6. Candes, E.J.; Wakin, M. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 21–30. [[CrossRef](#)]
7. Candes, E.J. Compressive sampling. In Proceedings of the International Congress of Mathematicians, Madrid, Spain, 22–30 August 2006; pp. 1433–1452.
8. Donoho, D.L. Compressive sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]

9. Zhang, Y.; Mei, S.; Chen, Q.; Chen, Z. A novel image/video coding method based on compressed sensing theory. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 1361–1364.
10. Gan, L. Block compressed sensing of natural images. In Proceedings of the IEEE International Conference on Digital Signal Processing, Wales, UK, 1–4 July 2007; pp. 403–406.
11. Sarkis, M.; Diepold, K. Depth map compression via compressed sensing. In Proceedings of the IEEE International Conference on Image Processing, Cairo, Egypt, 7–10 November 2009; pp. 737–740.
12. Duan, J.; Zhang, L.; Liu, Y.; Pan, R.; Sun, Y. An improved video coding scheme for depth map sequences based on compressed sensing. In Proceedings of the IEEE International Conference on Multimedia Technology, Hangzhou, China, 26–28 July 2011; pp. 3401–3404.
13. Zhang, Y.; Dong, Z.; Phillips, P.; Wang, S.; Ji, G.; Yang, J. Exponential wavelet iterative shrinkage thresholding algorithm for compressed sensing magnetic resonance imaging. *Inf. Sci.* **2015**, *322*, 115–132. [[CrossRef](#)]
14. Zhang, Y.; Wang, S.; Ji, G.; Dong, Z. Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging. *IEEE Trans. Electr. Electron. Eng.* **2015**, *10*, 116–117. [[CrossRef](#)]
15. Wang, Y.; Reibman, A.R.; Lin, S. Multiple description coding for video delivery. *Proc. IEEE* **2005**, *93*, 57–69. [[CrossRef](#)]
16. Vaishampayan, V. Design of multiple description scalar quantizers. *IEEE Trans. Inf. Theory* **1993**, *39*, 821–834. [[CrossRef](#)]
17. Goyal, V.K. Multiple description coding: Compression meets thenetwork. *IEEE Signal Process. Mag.* **2001**, *18*, 74–93. [[CrossRef](#)]
18. Lin, C.; Zhao, Y.; Zhu, C. Two-stage diversity-based multiple description image coding. *IEEE Signal Process. Lett.* **2008**, *15*, 837–840. [[CrossRef](#)]
19. Bai, H.; Zhu, C.; Zhao, Y. Optimized multiple description lattice vector quantization for wavelet image coding. *IEEE Trans. Circuits Syst. Video Technol.* **2007**, *17*, 912–917. [[CrossRef](#)]
20. Liu, M.; Zhu, C. M-description lattice vector quantization: Index assignment and analysis. *IEEE Trans. Signal Process.* **2009**, *57*, 2258–2274.
21. Reibman, A.R.; Jafarkhani, H.; Wang, Y. Multiple-description video coding using motion-compensated temporal prediction. *IEEE Trans. Circuits Syst. Video Technol.* **2002**, *12*, 193–204. [[CrossRef](#)]
22. Chang, S.; Cosman, P.C.; Milstein, L.B. Performance analysis of channel symmetric FEC-based multiple description coding for OFDM networks. *IEEE Trans. Image Process.* **2011**, *20*, 1061–1076. [[CrossRef](#)] [[PubMed](#)]
23. Lin, C.; Tillo, T.; Zhao, Y.; Jeon, B. Multiple description coding for H.264/AVC with redundancy allocation at macro block level. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 589–600.
24. Xu, Y.; Zhu, C. End-to-end rate-distortion optimized description generation for H.264 multiple description video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1523–1536.
25. Wang, A.; Zhao, Y.; Bai, H. Robust multiple description distributed video coding using optimized zero-padding. *Sci. China Ser. F Inf. Sci.* **2009**, *52*, 206–214. [[CrossRef](#)]
26. Zhu, C.; Liu, M. Multiple description video coding based on hierarchical B pictures. *IEEE Trans. Circuits Syst. Video Technol.* **2009**, *19*, 511–521. [[CrossRef](#)]
27. Zhang, M.; Bai, H. Adaptive multiple description video coding and transmission for scene change. *EURASIP J. Wireless Commun. Netw.* **2012**, *265*. [[CrossRef](#)]
28. Mobile 3DTV. Available online: <http://sp.cs.tut.fi/mobile3dtv/stereo-video/> (accessed on 27 June 2016).
29. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]

