Entropy **2014**, *16*, 6590-6601; doi:10.3390/e16126590

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Depth Image Coding Using Entropy-Based Adaptive Measurement Allocation

Huihui Bai [1,*], Mengmeng Zhang [2], Meiqin Liu [1], Anhong Wang [3] and Yao Zhao [1]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China; E-Mails: mqliu@bjtu.edu.cn (M.L); yzhao@bjtu.edu.cn (Y.Z.)

[2] College of Information Engineering, North China University of Technology, Beijing 100144, China; E-Mail: zmm@ncut.edu.cn

[3] Electronic Information and Engineering College, Taiyuan University of Science and Technology, Taiyuan 030024, China; E-Mail: wah_ty@163.com

* Author to whom correspondence should be addressed; E-Mail: luckybhh@gmail.com; Tel.: +86-10-5168-4108; Fax: +86-10-5168-8626.

External Editor: Kevin H. Knuth

Received: 20 October 2014; in revised form: 7 December 2014 / Accepted: 12 December 2014 / Published: 17 December 2014

Abstract: Differently from traditional two-dimensional texture images, the depth images of three-dimensional (3D) video systems have significant sparse characteristics under the certain transform basis, which make it possible for compressive sensing to represent depth information efficiently. Therefore, in this paper, a novel depth image coding scheme is proposed based on a block compressive sensing method. At the encoder, in view of the characteristics of depth images, the entropy of pixels in each block is employed to represent the sparsity of depth signals. Then according to the different sparsity in the pixel domain, the measurements can be adaptively allocated to each block for higher compression efficiency. At the decoder, the sparse transform can be combined to achieve the compressive sensing reconstruction. Experimental results have shown that at the same sampling rate, the proposed scheme can obtain higher PSNR values and better subjective quality of the rendered virtual views, compared with the method using a uniform sampling rate.

Keywords: depth image coding; entropy; 3D video system; compressive sensing

---

1. Introduction

Three-dimensional (3D) video can provide the viewers a high-quality and immersive multimedia experience, which has drawn increasing attention among industry and academic researchers [1]. Two typical 3D applications have appeared in the form of three-dimensional television (3DTV) [2] and free-viewpoint television (FTV) [3]. In 3DTV applications, multiple views from different viewing angles can be rendered for depth perception of the scene while in FTV applications, arbitrary viewpoints within a certain range can be selected interactively by viewers.

The basic format of 3D video is a multiview representation which is usually captured simultaneously by multiple cameras with slightly displaced positions [4]. However, with an increasing number of the views, the huge amount of data from multiview video poses great challenge for 3D applications, such as data compression and transmission. In order to solve this problem, the multiview video plus depth (MVD) format has emerged as an efficient data representation for 3D systems. Compared to the pure multiview video format without depth information, the main advantage of the MVD format is that desired virtual views at arbitrary viewpoint positions can be conveniently synthesized via the depth-image-based rendering (DIBR) technique [5].

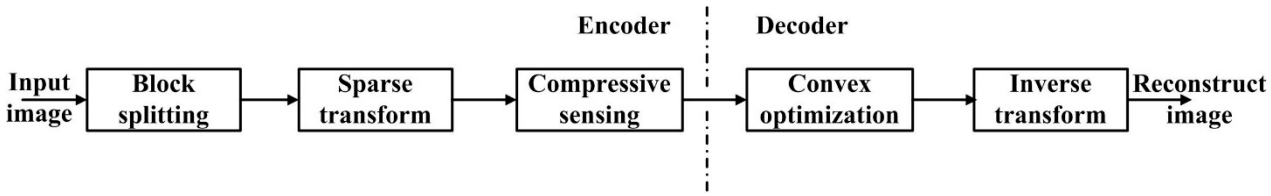
Depth images represent the distance information between the camera and the objects in the scene. The depth images are often treated as grey scale image sequences, which are similar to the luminance component of texture video. However, differently from the texture video, the depth image has its own special characteristics. Firstly, the depth image signal is much sparser than the texture video under certain transform basis, such as Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT), *etc*. It contains no texture but sharp object boundaries, since the gray levels are nearly the same in most regions within an object but change abruptly across the boundaries. Furthermore, the depth image is not directly used for display, but it plays an important role in the virtual view synthesis. The distortion of depth data, especially around the object boundaries, will seriously degrade the quality of the rendered virtual views [6]. Therefore, how to employ the depth image characteristics for efficient compression is an essential part in 3D systems.

In view of the sparsity characteristics of depth images, we attempt to apply compressive sensing (CS) [7] to represent depth information efficiently. CS is a new method to capture and represent compressible signals at a rate significantly below the conventional Shannon/Nyquist rate. In the conventional Shannon/Nyquist sampling theorem, when capturing a signal, one must sample at least two times faster than the signal bandwidth in order to avoid losing information. Due to the low sampling rate, CS can avoid the big burden of data storage and processing at the conventional encoder.

In recent years, CS is applied in image compression and the basic framework is shown in Figure 1. At the encoder, the input image can be processed block by block. For each block in the image, sparse transform, such as DCT or DWT, is used to produce the coefficients with sparse characteristics. Then compressive sensing is employed to encode the transform coefficients and generate the same amount of measurements for each block. At the decoder, a convex optimization method, such as the log-barrier or multiplier [8], can be adopted for the CS recovery. In the end, the corresponding inverse transform can be used for the image reconstruction. Block compressed sensing for natural images is proposed using the same measurement matrix, which is claimed that it can sufficiently capture the complicated geometric structures of natural images [9]. A new image/video coding approach is proposed, which can

combine the CS theory into the traditional DCT-based coding method to achieve better compression efficiency for spatially sparse signals [10]. Furthermore, the whole depth image can be processed by CS, and its performance is evaluated with rendered virtual view quality [11]. A novel compressed sensing framework is presented for depth image compression using adaptive graph-based transforms [12]. However, since the greedy algorithm is proposed to find the optimal edge image, which means higher complexity especially when the depth image block size increases.

Figure 1. Basic framework of image compression based on CS.



To address the above problems, in this paper, a novel depth image coding scheme is proposed based on a block compressive sensing method. The main improvements of the proposed scheme are as follows: (1) to ensure lower-complexity of the CS encoder, the entropy of pixels in each block is employed to represent the sparsity of depth signals; (2) in view of the different sparse characteristics of each block in the depth images, an adaptive measurement rate should be allocated for higher compression efficiency; (3) differently from the conventional CS, in this paper the measurements can be obtained directly in the pixel domain and the sparse transform is combined in the CS reconstruction, which can guarantee the lower-complexity of the CS encoder and the reconstructed image quality; (4) in order to better estimate the performance, objective and subjective quality of the rendered virtual views are taken into account.

The rest of this paper is organized as follows: in Section 2, the proposed scheme is presented step by step. In Section 3, the performance of the proposed scheme is examined. We conclude the paper in Section 4.

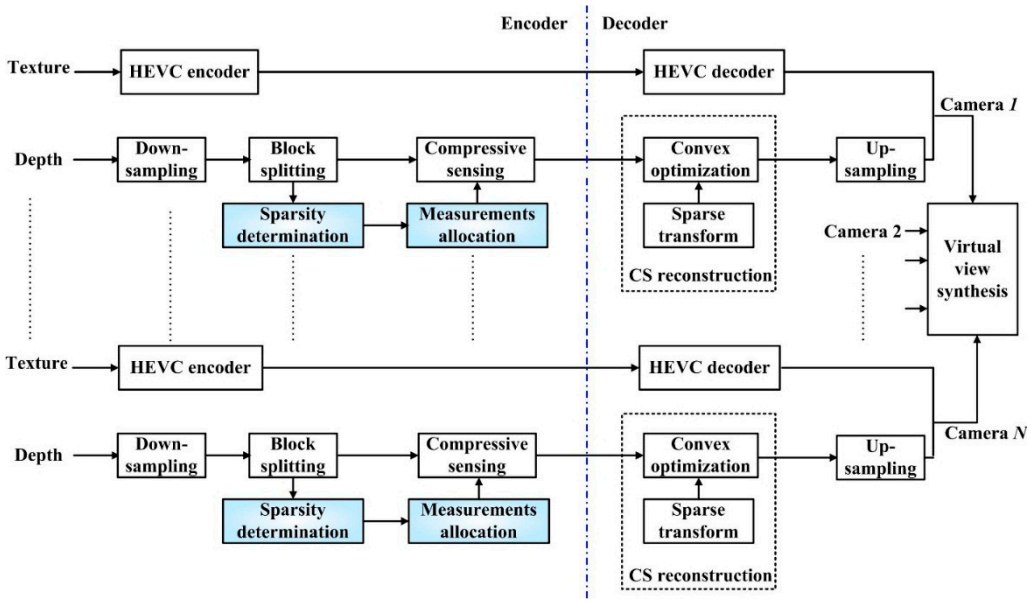
2. Proposed Scheme

2.1. Overview

Figure 2 illustrates the block diagram of the proposed scheme. *N* views from Cameras *1* to *N* can be processed independently and each view includes texture video and its corresponding depth image. Since texture video is very similar to the traditional two-dimensional (2-D) video, it can be compressed by a standard codec, such as High Efficiency Video Coding (HEVC), for high compression efficiency. In this paper, we focus on the compression of depth images. In view of the sparsity of depth images, a block compressive sensing method is applied to compress them. Firstly, in order to reduce the amount of computation, the original depth image can be down-sampled [13] and the sampling rate can be set as 0.5. Then the entropy of pixels in each block can be calculated to determine the sparsity in the pixel domain. According to the sparsity, adaptive measurements can be allocated to each block for better compression efficiency. It is noted that in order to reduce the complexity of the CS encoder, the sparse transform can

be shifted into the CS reconstruction. Therefore, at the decoder CS recovery can be obtained by solving a convex optimization problem combined with the sparse transform.

Figure 2. Block diagram of the proposed scheme.



2.2. Basic Idea of CS

Firstly, we will review the basics of the CS theory [7]. If $x \in R^n$ is a discrete signal and u is its coefficients in some orthonormal basis Ψ, then $x = \Psi^T u$. Here, x is said to be k-sparse with respect to Ψ if only k of n coefficients are non-zero. In CS theory, instead of encoding the k non-zero coefficients, the process of CS encoder is as follows:

$$y = \Phi x \tag{1}$$

where Φ is $m \times n$ matrix and $y \in R^m$. Since $m < n$, the original signal x can be compressed. At the CS decoder u can be reconstructed by solving the following optimization problem:

$$\begin{aligned} &\min \|u\|_1 \\ &\text{subject to } y = \Phi\Psi^T u \end{aligned} \tag{2}$$

Then according to $x = \Psi^T u$, the original signal x can be obtained.

In this paper, the CS encoder is utilized block by block for each frame to generate the CS frame. Each block can be organized to form a $n \times 1$ vector x. Here, the rows of the matrix Φ are samples of an independent identically distributed (i.i.d.) symmetric Bernoulli distribution. To be more specific, in the matrix Φ, the row consists of ± 1 and the probabilities of $+1$ and -1 are both 0.5. It is noted that for low complexity the matrix Φ is the same for all blocks. According to Equation (1), the measurement y can be produced directly in pixel domain, whose size is $m \times 1$. Then the measurement y can be encoded and transmitted to the channels. At the decoder we use a generic log-barrier algorithm to solve Equation (2). The corresponding matlab codes can be found in [14]. Furthermore, DCT basis is adopted as the orthonormal basis Ψ for simplicity. In this paper, DCT transform is not utilized at the encoder, but shifted into decoder. The corresponding details can be found in Section 2.4.

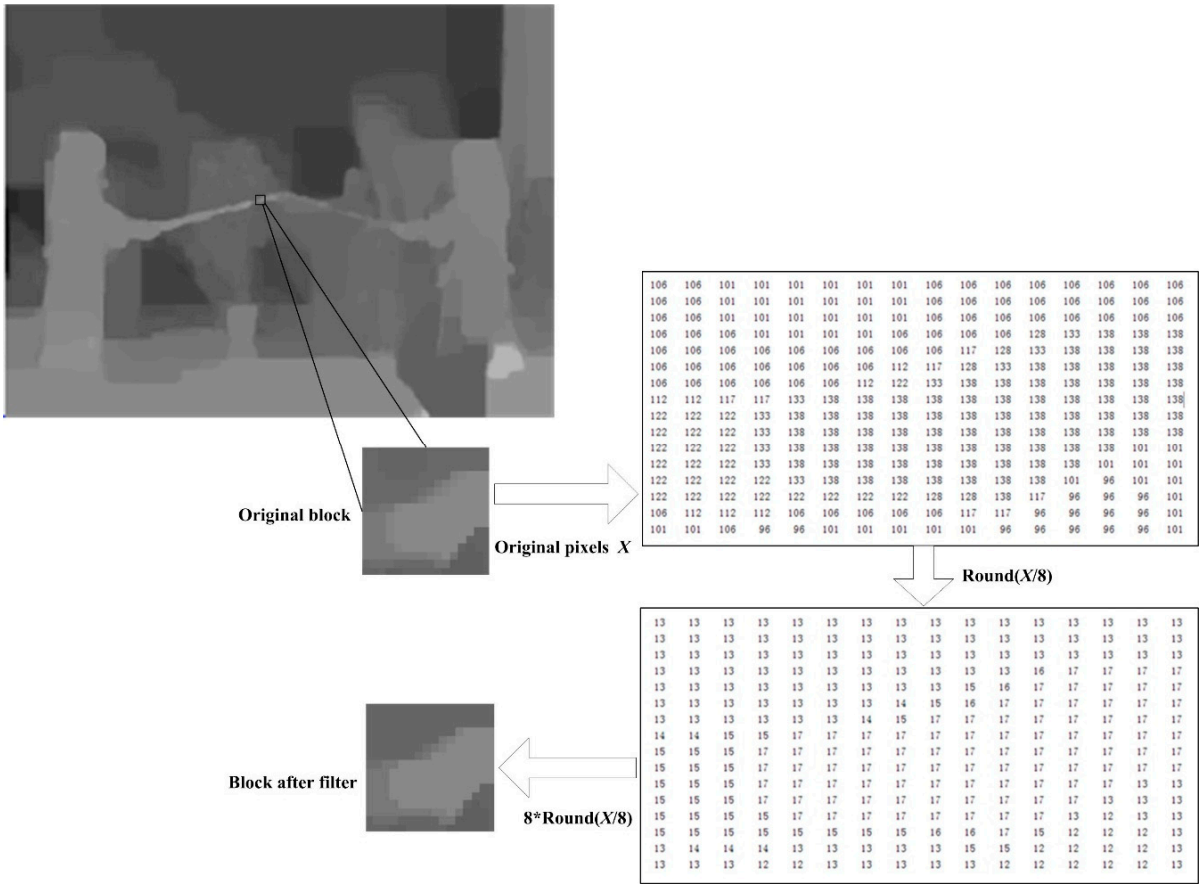
2.3. Entropy Calculation

In information theory [15], entropy is the average amount of information contained in the source. Therefore, for the image source, the entropy can represent the complexity of the image contents to a great extent. To be more specific, if the image content is very complex, the entropy can be larger while if the image content is very smooth, it can be smaller. According to the essence of the entropy, in this paper, it can be employed to measure the sparsity of the depth image. Generally, the entropy H of a discrete random variable X with possible values $\{x_1, x_2, ...x_i, ...x_n\}$ and probability mass function $P(X)$ can be defined as follows [15]:

$$H(X) = -\sum_i P(x_i) \log_2 P(x_i) \tag{3}$$

According to Equation (3), we can calculate the entropy of each block in the depth image. However, before the calculation, the background noise of the depth image should be removed first. It is noted that the removal of background noise aims to facilitate the accurate calculation of the entropy. Figure 3 shows an example of an anti-ground noise filter for depth images. Due to the background noise, the neighboring pixel values differ slightly from one another, which results in an inaccurate description of the information content using entropy. To remove the background noise without high computation complexity, an anti-ground noise filter can be utilized here. We adopt 8 as the stepper to quantize all the 256 pixel values of the original depth image. Finally, up to 32 (0–31) quantized values were left, which provided a good condition for the subsequent work.

Figure 3. Example of the anti-ground noise filter.



When we calculate the entropy of all blocks in the depth image, the probability of the appearance of each pixel can be counted in the calculation of the entropy, which is as follows:

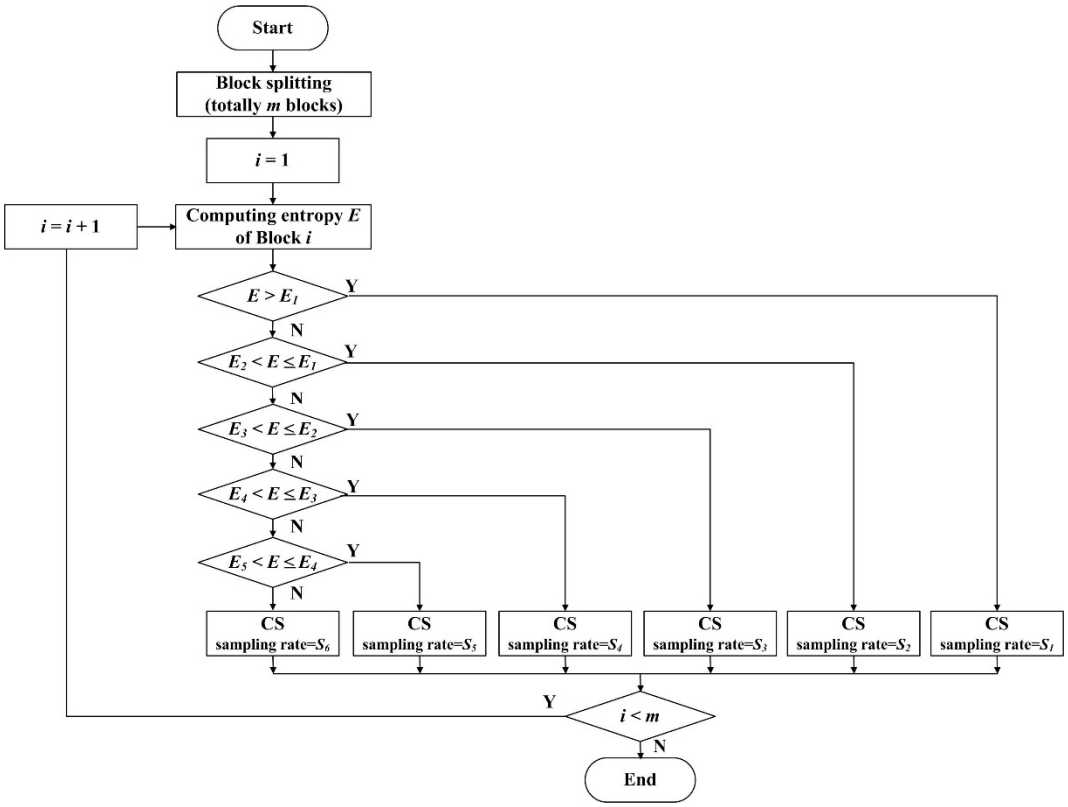
$$P(x_i) = \frac{n(x_i)}{N}, \quad x_i = 0,1,...,31 \tag{4}$$

Here, N is the total number of pixels in a block and $n(x_i)$ is the number of the quantized values x_i. As a result, the entropy of each block can be computing to measure the sparsity of the depth image.

2.4. Adaptive Measurement Allocation

In order to reconstruct a higher quality depth image at a lower sampling rate, we will allocate different sampling rates to different blocks according to their entropy, shown as the flowchart in Figure 4. For simplicity, the depth image can be divided into m blocks with size $n \times n$. Here, we can set $n = 16$. Assuming that $i = 1$, we can compute the entropy E of the first block of the depth image. According to the relationship between E and the threshold E_j ($j = 1,2,...,5$, and $E_5 < E_4 < E_3 < E_2 < E_1$), the corresponding sampling rate S_k ($k = 1,2,...,6$, and $S_6 < S_5 < S_4 < S_3 < S_2 < S_1$) can be allocated for each block until all the blocks have been processed. Here, $S_6 = 20\%$, $S_5 = 30\%$, $S_4 = 40\%$, $S_3 = 50\%$, $S_2 = 60\%$ and $S_1 = 70\%$. It is noted that due to the total six decisions, three bits are required for each block as the overhead of the proposed scheme.

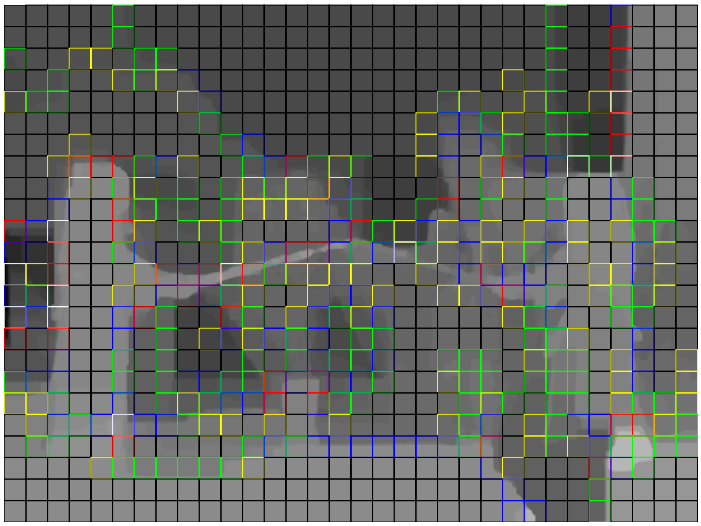
Figure 4. Flowchart of measurement allocation.



In Figure 5, a typical example for the standard test depth image Kendo is shown to explain the adaptive measurement allocation. Here, we use different colors to represent different sampling rates, such as white for S_1, red for S_2, blue for S_3, green for S_4, yellow for S_5 and black for S_6. In view of

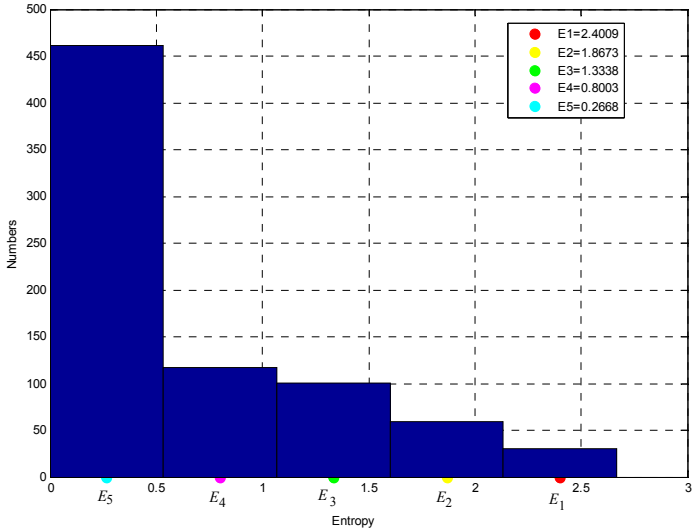
the characteristics of depth images, the most smooth block marked by black can be allocated the lowest sampling rate while the complex texture block marked by white can be allocated the highest sampling rate. As shown in Figure 5, since the smooth blocks are actually a larger percentage of all the blocks, higher compression efficiency may be achieved using unequal sampling rates than with a uniform sampling rate.

Figure 5. A typical example of measurement allocation.



It is noted that the threshold E_j can be computed by statistical methods. Firstly, since five thresholds should be taken into account, we can divide the entropy values of all blocks into five equal intervals. Here, also take the standard test depth image Kendo as an example, as shown in Figure 6. Furthermore, we can obtain the central values of each bin which are noted by colored circles in Figure 6. These central values can be considered as thresholds. We have to compute the entropy for all blocks, and decide the thresholds. Then for a different image, the entropy thresholds have to be computed again. Currently, we consider six levels of thresholding. More levels means better reconstructed image quality, but it also increases the computing complexity.

Figure 6. A typical example of threshold determination.



2.5. Improved CS Reconstruction

Here, the sparse transform can be shifted to the decoder to reduce the complexity of the encoder. Here, the log-barrier algorithm can be designed to solve quadratically constrained L_1 minimization:

$$\begin{aligned} &\min \|u\|_1 \\ &\text{subject to } \|Au-b\|_2 \le \varepsilon \end{aligned} \tag{5}$$

Here, $A=\Phi\Psi^T$, u is the coefficient of original pixel x in some orthonormal basis Ψ, and b is the vector of observation. It is noted that according to the log-barrier algorithm, some parameters should be updated due to the combination of sparse transform. Next the derivation is shown as follows:

$$y=\Phi x=\Phi\Psi^T u=Au \tag{6}$$

Then we will introduce the singular value decomposition (SVD) of A^T:

$$A=(A^T)^T=(USV^T)^T=VSU^T \tag{7}$$

Since A is an $m\times n$ matrix, U is an $m\times m$ unitary matrix, S is an $m\times n$ diagonal matrix and the $n\times n$ unitary matrix V^T denotes the conjugate transpose of the $n\times n$ unitary matrix V. Furthermore, according to Equation (7), the Equation (6) can be rewritten by:

$$y=VSU^T u \tag{8}$$

It also can be changed as follows:

$$S^{-1}V^T y=U^T u \tag{9}$$

By the comparison between Equation (5) and Equation (9), the parameters can be updated as follows: firstly, b in Equation (5) can be updated by $S^{-1}V^T y$. Secondly, A in Equation (5) is updated by U^T. Finally, the initial u can be replaced by Ub.

3. Experimental Results

In this paper, the standard test sequences shown in Table 1 are selected to validate the proposed scheme. The input for each view is the first color image frame with the corresponding depth image. In the practical application, the camera can process the multiviews image by image, which is like the intra-coding in the traditional method. Here, the experimental results are tested on a PC with a 2.67 GHz Intel CoreTMi5 CPU and the main scheme is implemented using MATLAB R2010a. The virtual viewpoint synthesis software with the version VSRS3.5 is adopted as the experimental platform.

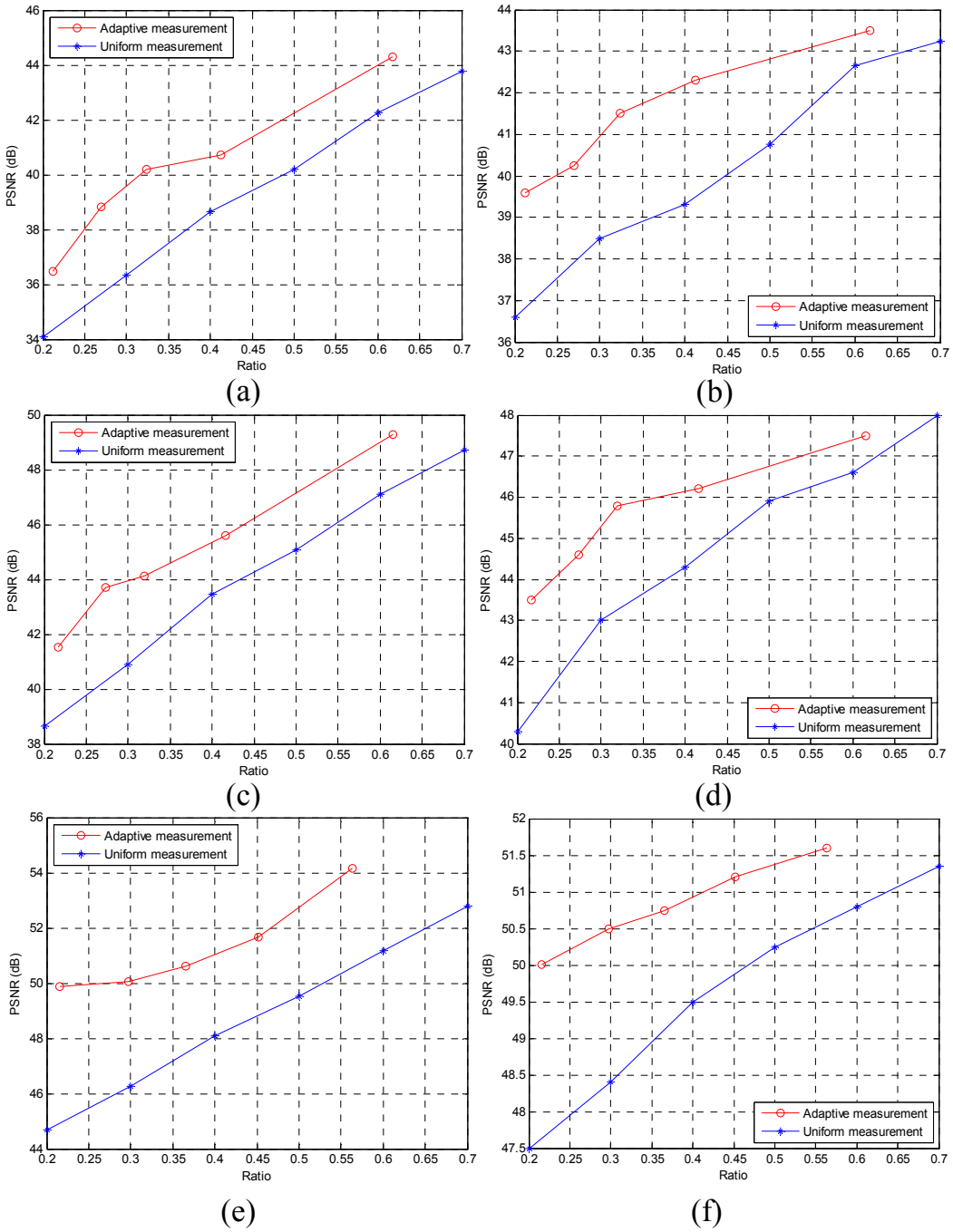
Table 1. Test sequences

Sequence	Resolution	View
Balloons	1024×768	1–3
Kendo	1024×768	1–3
Pantomime	1280×960	37–39

It can be seen from Figure 7a,c,e that the proposed scheme outperforms the uniform sampling scheme in PSNR values of depth map at the same ratio. Here, the ratio is the average ratio or average sampling

rate for adaptive measurements. In the three tested sequences, the PSNR values of the sequence Pantomime are higher than the two other sequences because this sequence has more smooth regions and better sparsity.

Figure 7. Objective quality comparison for Balloons, Kendo and Pantomime. (**a**), (**c**) and (**e**): for depth map; (**b**), (**d**) and (**f**): for synthesized virtual viewpoint.



Since the depth map is not directly used for display, the objective and subjective quality of the rendered virtual views should be taken into account. In the objective aspect, the synthesized virtual viewpoint image can be achieved by two original camera images. For example, for the tested sequences Balloons and Kendo, the depth and texture from the 1st and 3rd views can be used to synthesize the texture of 2nd view while for the sequence Pantomime, the depth and texture from the 37th and 39th views

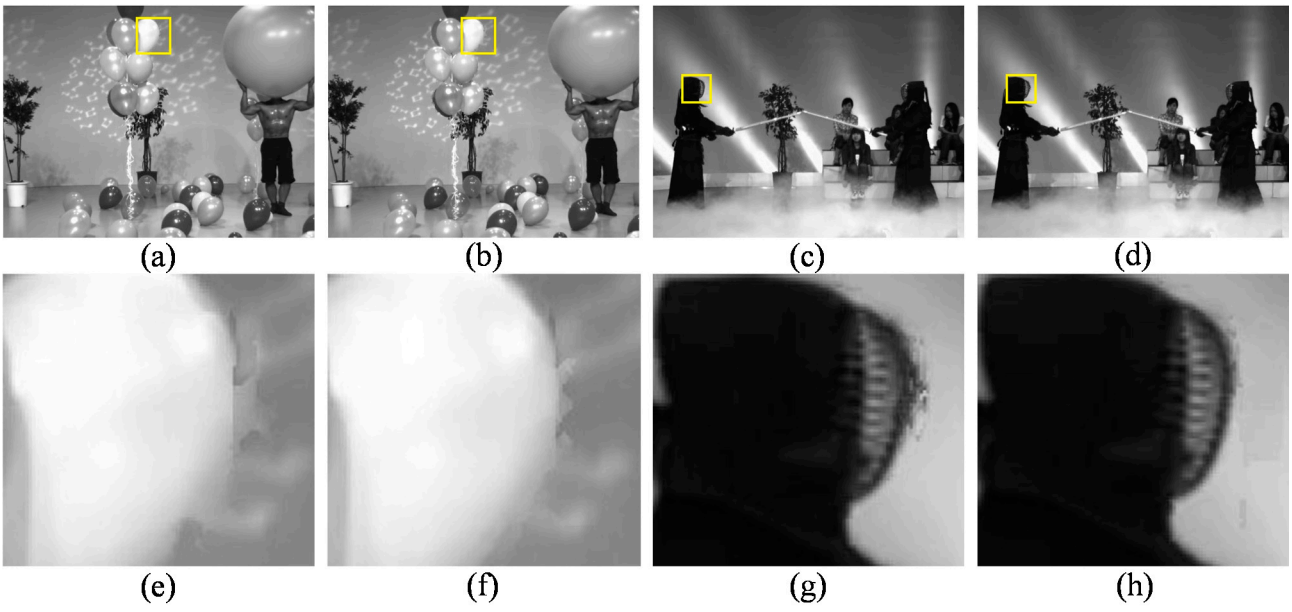
can generate the texture of 38th view. Then we make a comparison between the uniform sampling scheme and the proposed one by observing the quality of the synthesized image. In Figure 7b,d,f, it can be seen at the same average sampling rate, the synthesized image using the proposed scheme outperforms the uniform sampling scheme in PSNR values. Furthermore, the encoding and decoding times of the proposed scheme and the uniform one have also been shown in Table 2. From Table 2, we can find that the proposed scheme needs more time than the uniform one due to the increasing complexity.

Table 2. The depth image encoding and decoding time of uniform and adaptive methods.

Schemes	Balloons			Kendo			Pantomime		
	Ratio (%)	Encoding Time	Decoding Time	Ratio (%)	Encoding Time	Decoding Time	Ratio (%)	Encoding Time	Decoding Time
Uniform	20	0.0619	96.5081	20	0.0622	94.2078	20	0.0928	98.8772
	30	0.0709	99.7691	30	0.0709	98.1991	30	0.1072	101.4328
	40	0.0785	101.9015	40	0.0775	102.3725	40	0.1220	103.7480
	50	0.0873	105.1227	50	0.0849	107.5151	50	0.1353	106.4147
	60	0.0946	108.7454	60	0.0937	112.0063	60	0.1497	108.3903
	70	0.0980	110.1820	70	0.0965	115.0535	70	0.1513	111.1087
Adaptive	21.24	0.1709	98.2591	21.62	0.1710	97.6990	21.54	0.2687	100.1813
	26.94	0.1739	100.7861	27.36	0.1737	99.9364	29.76	0.2754	103.1046
	32.41	0.1772	103.5128	31.94	0.1761	103.0040	36.49	0.2849	104.7951
	41.20	0.1798	107.2502	41.60	0.1785	108.5315	45.2	0.2854	107.9946
	61.80	0.1844	109.7256	61.57	0.1837	112.2063	56.36	0.2923	109.8677

Next, we further discuss the reconstruction quality of synthesized images. From Figure 8, it can be seen that the better visual quality of synthesized images has been observed with the proposed scheme than uniform sampling scheme, especially in some parts denoted by a yellow rectangle.

Figure 8. Subjective quality comparison of synthesized virtual viewpoint for Balloons and Kendo. (**a**), (**c**), (**e**) and (**g**): uniform sampling; (**b**), (**d**), (**f**) and (**h**): proposed scheme.



In Table 3, the comparison with the traditional coding has been shown. Here, according to the main idea of JPEG or H.264 intra-coding, the traditional coding method is simulated based on a Discrete Cosine Transform (DCT). We decompose each block (16×16) of the original depth map by DCT and then perform the reconstruction using only the significant DCT coefficients. In Table 3, the traditional method has obtained higher PSNR values than the CS coding with more encoding time. Therefore, the CS method is suitable for real-time compression of high-speed camera images.

Table 3. Comparison with the traditional method.

Sequence	Ratio	PSNR (dB)		Encoding time (s)	
		Traditional	Proposed	Traditional	Proposed
Kendo	27.36%	45.5579	43.7066	0.9733	0.1736
Balloons	32.41%	43.0547	40.2082	0.9713	0.1772
Pantomime	21.54%	52.9541	49.8902	1.4823	0.2686

In the current stage, the CS method cannot compete with the traditional method in terms of compression efficiency. The main reason is that at the DCT encoder can nicely remove the correlation in the original image so that most information of the image can be recovered by a small amount of transform coefficients. In contrast, the CS encoder can realize compression mainly based on random sampling. At the decoder, it can apply sparse transform to gain performance. In the future, it is necessary for us to refer to the traditional method to improve the results of the CS method.

4. Conclusions

In this paper, we fully consider the sparse characteristics of depth images and propose a novel scheme based on block compressive sensing. Since the entropy can describe the sparsity of the depth image to some extent, adaptive measurement allocation is designed based on the entropy of each block. The simulation results show that compared with uniform sampling scheme, the proposed scheme has better rate distortion performance for both depth maps and synthesized virtual viewpoints.

Acknowledgments

This work was supported in part by Program for Changjiang Scholars and Innovative Research Team in University (No. IRT201206), 973 program (2012CB316400), National Natural Science Foundation of China (No. 61272051, No. 61210006, No. 61370111, No. 61272262, No. 61202240), Beijing Higher Education Young Elite Teacher Project (YETP0543) and the Fundamental Research Funds for the Central Universities of China (2014JBM028).

Author Contributions

Huihui Bai and Yao Zhao designed the research. Mengmeng Zhang and Anhong Wang performed the experiment. Meiqin Liu analysised the data. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Smolic, A.; Kauff, P.; Knorr, S.; Hornung, A.; Kunter, M.; Muller, M.; Lang, M. Three-dimensional video postproduction and processing. *Proc. IEEE* **2011**, *99*, 607–625.
2. Fehn, C.; de la Barré, R.; Pastoor, S. Interactive 3-DTV-Concepts and key technologies. *Proc. IEEE* **2006**, *94*, 524–538.
3. Tanimoto, M. Overview of free viewpoint television. *Signal Process. Image Commun.* **2006**, *21*, 454–461.
4. Vetro, A.; Wiegand, T.; Sullivan, G.J. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proc. IEEE* **2011**, *99*, 626–642.
5. Mori, Y.; Fukushima, N.; Yendo, T.; Fujii, T.; Tanimoto, M. View generation with 3D warping using depth information for FTV. *Signal Process. Image Commun.* **2009**, *24*, 65–72.
6. Merkle, P.; Morvan, Y.; Smolic, A.; Farin, D.; Müller, K.; de With, P.H.N.; Wiegand, T. The effect of multiview depth video on multiview rendering. *Signal Process. Image Commun.* **2009**, *24*, 73–88.
7. Candès, E.J.; Wakin, M.B. An introduction to compressive sampling. *Signal Process. Mag.* **2008**, *25*, 21–30.
8. Candès, E.J.; Romberg, J.; Tao, T. Stable signal recovery from incomplete and inaccurate Measurements. *Commun. Pure Appl. Math.* **2006**, *59*, 1207–1223.
9. Gan, L. Block compressed sensing of natural images. In Proceedings of the 15th International Conference on Digital Signal Processing, Cardiff, UK, 1–4 July 2007; pp. 403–406.
10. 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.
11. Sarkis, M.; Diepold, K. Depth map compression via compressed sensing. In Proceedings of IEEE International Conference on Image Processing, Cairo, Egypt, 7–10 November 2009; pp. 737–740.
12. Lee, S.; Ortega, A. Adaptive compressed sensing for depthmap compression using graph-based transform. In Proceedings of IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 929–932.
13. Ekmekcioglu, E.; Worrall, S.T.; Kondoz, A.M. Bit-rate adaptive downsampling for the coding of multi-view video with depth information. In Proceedings of the 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, Istanbul, Turkey, 28–30 May 2008; pp.137–140.
14. Compressive Sensing Resources. Available online: http://dsp.rice.edu/cs (accessed on 17 December 2014).
15. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.