Journal of Information Hiding and Multimedia Signal Processing ©2014 ISSN 2073-4212
Ubiquitous International Volume **5**, Number **3**, July **2014**

Depth Coding Based on Compressed Sensing with Optimized Measurement and Quantization

Mei Zhao[1], Anhong Wang[1], Bing Zeng[2,3], Lei Liu[4] and Huihui Bai[4]

[1]Institute of Digital Media and Communication
Taiyuan University of Science and Technology, China.
zhaosongzhaomei@163.com; wah_ty@163.com

[2]Institute of Image Processing
University of Electronic Science and Technology of China
eezeng@uestc.edu.cn

[3]The Hong Kong University of Science and Technology
Hong Kong SAR, China

[4]Institute of Information Science
Beijing Jiaotong University, Beijing , China
12112061@bjtu.edu.cn

Received December, 2013; revised January, 2014

Abstract. *Efficient coding of depth map is an essential part of 3-D video processing system due to the fact that the quality of each synthesized virtual view highly depends on the depth map. In this paper, we propose a compressed sensing (CS) based depth coding scheme. At the encoder side, the depth map is first pair-wisely measured in its Fourier domain, and quantization with a carefully-designed dead-zone is applied on all CS measurements (after considering the distribution of measurement values). An optimized trade-off between the measurement rate and quantization is employed to achieve the best-possible rate-distortion performance. At the decoder side, considering that the depth map usually consists of piece-wise constant areas and sharp edges, we solve a total variation (TV) minimization with constraints being put forward to preserve discontinuities at boundaries and at the same time enforce smoothness within the depth map. Experimental results show that our scheme achieves a significant improvement in rate distortion performance and a better synthesis quality as compared to the standard JPEG scheme.*

Keywords: 3-D video, depth map coding, compressed sensing, pair-wisely measurement, quantization with dead-zone. .

1. **Introduction.** Recently, the joint video team (JVT) proposed the structure of multi-view plus depth (MVD) [1] in which intermediate views can be synthesized by the neighboring views with their corresponding depth maps. The large amount of multi-view data in 3D video (3DV) applications can be efficiently compressed by the MVD format coding while the 3-D scene rendering at the decoder side is also very convenient. This makes MVD a very popular and promising solution for 3DV processing. The depth map represents 3-D scene information (i.e., the distance between capturing camera and object) in 3DV systems. In the view-synthesis techniques, depth information is very important and the quality of synthesized views highly depends on the depth map. Depth distortions will cause geometry changes and occlusion variations of the overlapping foreground and

background objects when pixels warped from an original view into a virtual view, both of which will yield texture distortions and thus degrade the quality of synthesized views [2]. Therefore, efficient depth map compression is crucial for 3DV system.

Depth map is typically considered as an 8-bit grayscale image. Therefore, a direct approach to process depth map is to treat it as a standard image (or image sequence) and compress it using the standard image or video compression tools such as JPEG or H.264/AVC. However, these standards have been designed to provide maximum perceived visual quality for texture/color images. Different from texture images, depth maps have unique characteristics that make the existing standard image/video coding techniques not suitable for depth map compression. First, depth map has discontinuous boundaries but smooth areas within these boundaries. Several literatures have already demonstrated that the discrete cosine transform (DCT) utilized in image and video coding is not efficient for blocks containing complex shaped edges [3-4], while edges and boundaries of depth map are very important for view synthesis. Second, the temporal consistency of depth video is much lower than the texture/color video as the depth capturing devices have not enough resolutions or the depth estimation algorithms are not satisfactory [5]. Temporal inconsistency will directly result in inefficient inter prediction which consequently leads to high bit-rates to encode the residual data. Moreover, conventional coding methods focus on the guarantee of perceived visual quality with an optimal rate-distortion (R-D) performance for viewers. However, the depth map is never displayed and it is solely used to assist virtual view synthesis at the decoder. Thus some techniques to enhance the visual quality such as smoothing filter make no sense for depth map.

Nowadays, how to preserve the fidelity of depth information with high efficiency has attracted a lot of attention. Various approaches have been proposed attempting to compress the depth map efficiently. Since the edge information significantly affects the rendering quality of synthesized view, it has been shown in [6] that the rendering quality can be increased by some special handling of object boundary regions of a depth map. Krishnamurthy et. al. [7] improved JPEG2000 standard and proposed a coding method based on the region of interest (ROI) which can avoid artifacts on the edges. In [8], regions where accurate depth is especially crucial are firstly identified. These methods generate a good rendering quality; however the compression ratio is not satisfactory. Therefore, a technique that can preserve boundary information well and achieve better compression efficiency is strongly required for 3DV systems.

Compressed sensing (CS) [9-11] is an emerging theory in signal processing which asserts that any sparse signal can be recovered from far fewer samples or measurements than the traditional Nyquist theory suggests. It has been used in several applications such as steganography[12] and image watermarking[13]. A distributed compressed video sensing algorithm is proposed in [14-15] using the adaptive sparse basis. For depth map compression, several CS-based methods have been recently proposed and it has been proved that the sharp discontinuities in depth can be accurately reconstructed. An adaptive CS framework for depth map compression using a family of graph-based transform was proposed in [16-17]. Using the variable density random sampling method [18] (as described in details in Section 2.2), a CS-based depth coding algorithm is proposed in [19], which firstly subsamples the depth map in the frequency domain and then reconstructs the image using a conjugate gradient minimization scheme. This method shows a better performance than JPEG and JPEG2000. However, since the quantization has not been considered for the measurements, the comparison seems unfair to some degree. Additionally, it does not take into account the characteristics of the Fourier coefficients which need further processing to compress the large volume of data.

In this paper, a new CS-based coding scheme is proposed, which considers the characteristics of depth map and measurement values as well as quantization with a carefully-designed dead-zone. An optimization between the measurement rate and quantization is discussed in our work. Compared to the work presented in [19], our contributions focus on the following aspects: First, we propose a pair-wise random sampling method in the Fourier domain and the sampling pattern is designed according to the characteristics of Fourier coefficients. Second, a uniform scalar quantizer with dead-zone, which considers the distribution of the Fourier coefficients, is applied on the measurements. Third, an optimization between the measurement rate and quan- tization is considered so as to achieve the best-possible R-D performance for the reconstructed depth map.

2. **Overview of compressed sensing theory.** Given a real value discrete signal $\mathrm{x} \in \mathrm{R}^N$ with length N, it is defined to be sparse if there exists a basis matrix $\Psi \in R^{N\times N}$ such that $\mathrm{x} = \Psi\alpha$ where $\|\alpha\|_0 = K \ll N$. The CS theory tells us that such a K-sparse signal can be reconstructed by the far fewer samples with certain accuracy:

$$\mathrm{y} = \Phi\mathrm{x} = \Phi\Psi\alpha = \Theta\alpha \tag{1}$$

where $y \in R^M$ denotes the measurement vector with length M, Φ is an $\mathrm{M} \times \mathrm{N}$ measurement matrix which is incoherent with Ψ and $\mathrm{M} = \mathrm{O}(\mathrm{Klog}(\mathrm{N}/\mathrm{K}))$, $\mathrm{K} < \mathrm{M} \ll \mathrm{N}$. The reconstruction problem can be formulated as an optimization problem by solving:

$$\min \frac{1}{2} \|\Phi\Psi\alpha - y\|_2^2 + \tau\|\alpha\|_1 \tag{2}$$

where the l_1 norm term enhances the sparsity of the solution in Ψ domain and the l_2 norm term ensures the fidelity between the solution and the measurements.

2.1. **Variable density sampling scheme.** A variable-density sampling scheme is proposed in [18] and the measurement matrix is generated based on the fact that the smaller the incoherence between Φ and Ψ, the smaller the interference produced by the sub-sampling would be [20]. To generate the measurement matrix, a probability density function (PDF) is firstly generated according to the distribution of the discrete Fourier transform (DFT) coefficients with the constraint that the sum of the PDF equals to the predefined number of the measurements. Then the binary sampling mask, where 1 at (m, n) indicates a sampling point and 0 means no measurement on that point, is selected as the measurement matrix according to the PDF, meanwhile making the incoherence between Φ and Ψ as small as possible. As demonstrated in [18-19], this measurement method is highly ef-ficient.

3. **The proposed framework.** The framework of our scheme is shown in Figure. 1. The depth image x is first sub-sampled by a variable-density sampling matrix Φ which is pair-wisely generated and the measurement vector y is obtained through Eq. (1). Quantization with a variable dead-zone is performed on the measurement values. Then, the quantized bit-stream $\mathrm{y_q}$ coupled with the sampling mask Φ is arithmetic encoded and transmitted to the decoder. At the decoder, the received bit-stream is firstly arithmetic decoded to get the quantized measurement $\hat{y}_\mathrm{q}$ and the sampling mask Φ, then $\hat{y}_\mathrm{q}$ is de-quantized. Finally, the output image is reconstructed by the l_1-norm conjugate gradient minimization scheme as that in [14].

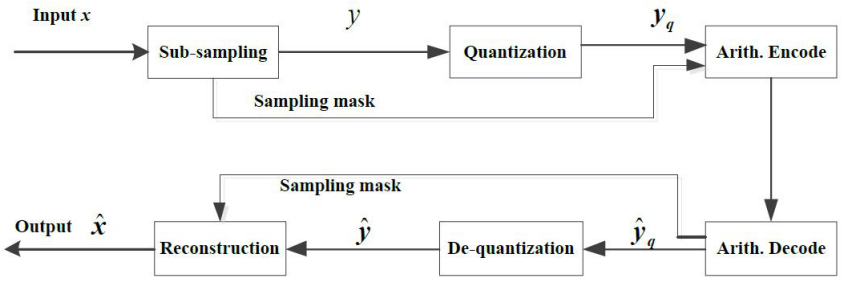

FIGURE 1. Block diagram of our proposed scheme

3.1. **Pair-wise random measurement.** The measuring method in [18] generates a sampling mask randomly without con-sidering the conjugate symmetry of the DFT coefficients. Actually, since the coeffi-cients of DFT are complex numbers with each coefficient composed of real and im-aginary parts, the quantization should be implemented to both of these two parts, which will nevertheless lead to a high bit-rate. Here, a pair-wise random measurement method which considers the conjugate symmetry of Fourier coefficients is put for-ward in our work.

Figure. 2 (a) shows the distribution of the Fourier coefficients. We can see that most energy of the image is concentrated in the k-space origin, which can be used as a prior information and thus a variable-density sampling scheme is preferable. Figure. 2 (b) shows the PDF under a specific measurement rate, from which we can see that the probability around the origin is larger, which is consistent with Figure. 2 (a). Then the pair-wise sampling mask is generated according to this PDF, which mainly consists of three steps:

(1) Determining the measurement matrix pair-wisely: First, half of the sampling mask is generated randomly based on the PDF, and then the other half is generated according to the symmetry. This ensures that two points symmetrical about the center be both sampled, generating a so-called pair-wise sampling. Note that this pair-wise sampling exploits the a prior information of DFT coefficients so that only a half of DFT coefficients and the sampling mask need to be transmitted, which will cut down the burden of quantization and entropy coding while without loss of recovery quality.

(2)Computing the incoherence between Φ and Ψ: When designing the pair-wise sampling matrix, i.e., the measurement matrix Φ, we should make sure that the incoherence between Φ and Ψ be minimal. Similar to [18], the point spread function (PSF) is used in our work to measure the incoherence:

$$\mathrm{PSF(i;j)} = e_j^* F_u^* F_u \mathrm{e_i} \tag{3}$$

(3)Selecting a sampling mask with minimal incoherence with Ψ: Because the generation of the sampling mask as mentioned above is random, we may accidentally choose a sampling pattern with a "bad" PSF. To prevent such situation, we repeat the procedure several times; record the corresponding PSF of the current sampling pattern each time; and finally the pattern with the lowest peak interference is selected as the sampling mask.

Figure. 2 (c) shows the output sampling mask matrix corresponding to the PDF in Fig. 2 (b). It is obvious that the sampling matrix follows the property the PDF while preserves symmetry to the origin.

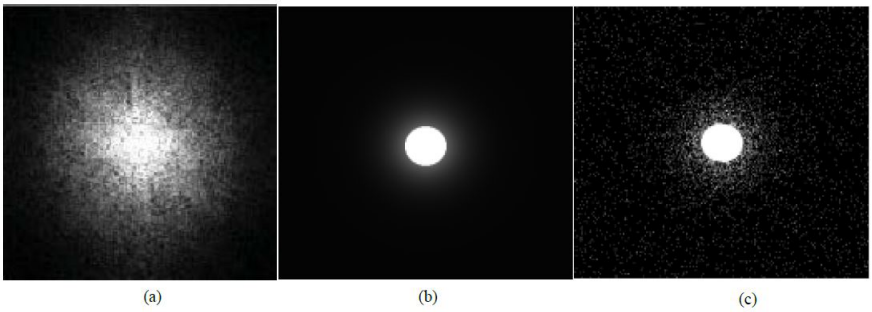

FIGURE 2. Illustration of the distribution of (a) DFT coefficients, (b) the PDF and (c) the sampling mask.

3.2. **Quantization with dead-zone.** Some forms of quantization are necessary to produce a compressed bit-stream of the CS measurements. Uniform scalar quantization (USQ) is commonly used in the current coding schemes; however it is not suitable for our method since the measurements (obtained from the DFT coefficients) are not uniformly distributed. Figure. 3 shows the statistical distribution of the measurements of Aloe at a specific measurement rate (MR = 0.1, and note that when the MR varies, the general trend of the measurements distribution is stable). It is obvious that most of the measurements are around zero and only a few are significant which nevertheless make a great contribution to the reconstruction. Hence, it comes very naturally that adopting a quantization which deals with DFT coefficients according to their distribution. In this paper, a uniform scalar dead-zone quantization (USDZQ) is used, as shown in Figure. 4, where $2D_f$ is the size of dead-zone. The quantization with dead-zone can be described as:

$$y_q = Q(\text{Round}(k * y)) \tag{4}$$

where $Q(\cdot)$ denotes the uniform quantization, $k (k > 1)$ is a scale factor and $2D_f = \frac{1}{k}$.

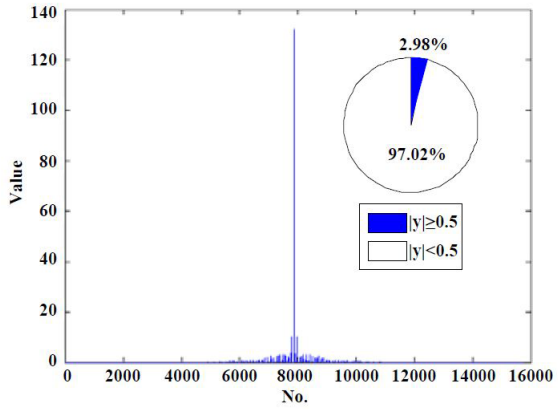

FIGURE 3. Statistical distribution of the measurements (in the case of MR=0.1 for Aloe).

3.3. **Optimization between measurement rate and quantization dead-zone.** Quantitation affects the R-D performance. Generally, when the dead-zone $2D_f$ decreases, the number of zeros created by the quantization will decrease accordingly. This will lead to an increased bit-rate R but lower distortion D. That is, the distortion D caused by quantization will be smaller when the dead-zone decreases but at the sacrifice of a higher

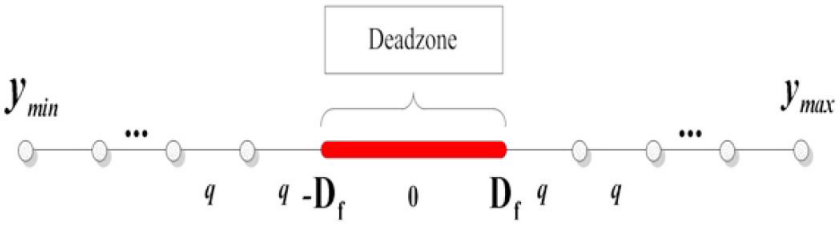

FIGURE 4. The scalar uniform quantizer with a dead-zone.

bit-rate R and vice versa. Meanwhile, MR is also a decisive parameter. In general, a larger MR often leads to a better reconstructed image with an increased bit-rate cost.

As illustrated in Section 3.2, most measurements of the depth map approach to zero, hence having little contribution toward the reconstruction. On the other hand, significant coefficients that influence reconstruction greatly make up only a small portion of the measurements (some 3% or less). Thus, the R-D optimization between the dead-zone and MR should be considered carefully. The basic principle of this optimization is to minimize D for a given R or vice versa, which can be expressed as:

$$\min\{D+\lambda R\} \tag{5}$$

where D and R denote the distortion of the depth map and the bit-rate required to transmit the depth map, respectively, and λ is the Lagrangian multiplier.

Usually, this R-D optimization problem can be solved by the Lagrangian multiplier method [21]. In this paper, we use a simple brute-force method to solve this problem. For a given target bit-rate, the method loops through different k with a step size 1. Then, for a determined k, MR is adjusted with a step size 0.05 until the bit-rate meets the target. The distortion of the depth map at each iteration is calculated, and finally the optimal tradeoff between D_f(or k) and MR can be found.

3.4. **Recovery with minimal total variation.** Considering the smooth area in depth map, the optimization constraint of total variation (TV) is added to the reconstruction [19], and then Eq. (2) becomes:

$$\min \frac{1}{2} \|\Phi\Psi\alpha - y\|_2^2 + \tau\|\alpha\|_1 + \gamma TV(x) \tag{6}$$

where γ regularizes the weight of TV in the minimization.

The TV constraint aims at computing the finite differences of the depth map to measure the overall summation of variation in the image, which guarantees the smoothness of the depth map and preserves the edges well.

4. **Simulation results.** To validate the performance of our proposed scheme, three groups of comparison experiments are presented here: (1) the performance comparison of our scheme (when quantization is not performed) with [19], (2) the R-D performance comparison of depth map, and (3) the subjective and objective comparison of the synthesized virtual views. In (2) and (3), quantization is performed and we compare our method with JPEG and JPEG2000 standard under the same conditions. Four ground truth disparity images (Cones, Tsukuba, Art and Aloe) provided by Middlebury test bench [22] are tested in the experiments and the Peak Signal-to-Noise Ratio (PSNR) is used to evaluate the distortion.

As shown in Figure. 5, we first compare the performance of our proposed scheme with [15] under the same compression ratio (quantization is not performed here just as

[15] does). We can obviously see that our scheme has a great advantage over [15]. The contributing factors are: (1) pixel values are scaled within the interval [0,1], which is much useful, (2) a different reconstruction algorithm (borrowed from [14]) is used in our paper, and (3) we have chosen a suitable TV term to guarantee the re-construction quality.

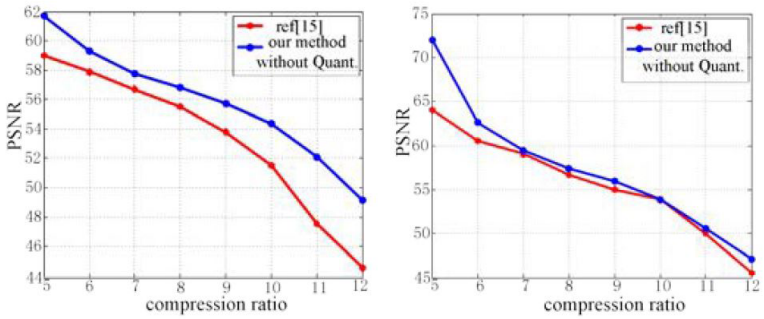

FIGURE 5. Comparison with Ref. [15].

When quantization is added to the scheme, the R-D performance comparison of depth map is presented in Figure. 6. It is obvious that the depth map compressed by our algorithm is better than JPEG under the same BPP and achieves a superiority of 1.1 dB, 1.56 dB, 2.3 dB, and 2.3 dB in PSNR on average over JPEG. Unfortunately, there still exists a gap with JPEG2000.

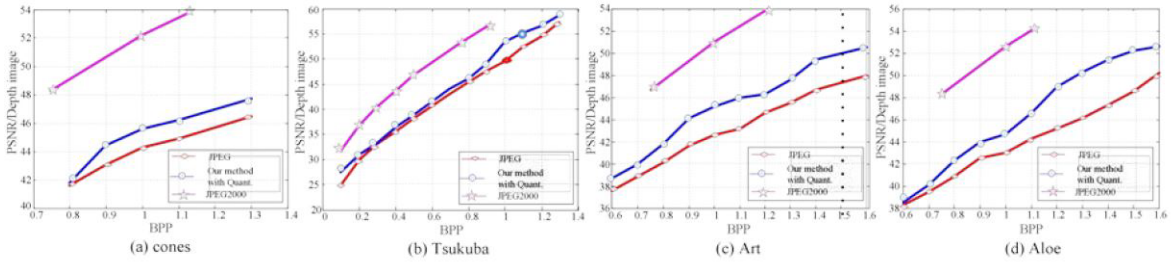

FIGURE 6. The rate-distortion comparison of the depth images.

Figure. 7 compares the R-D performance of the synthesized view among three schemes, where the synthesized right view of the Middlebury images is generated using the original left view and various compressed depth maps. Fig. 7 shows that our algorithm is more efficient than JPEG, yet a gap exists with JPEG2000.

Next, comparisons with JPEG2000 are conducted from the objective aspect for images. The synthesized views are presented in Figure. 8, where bit-rates used to compress depth

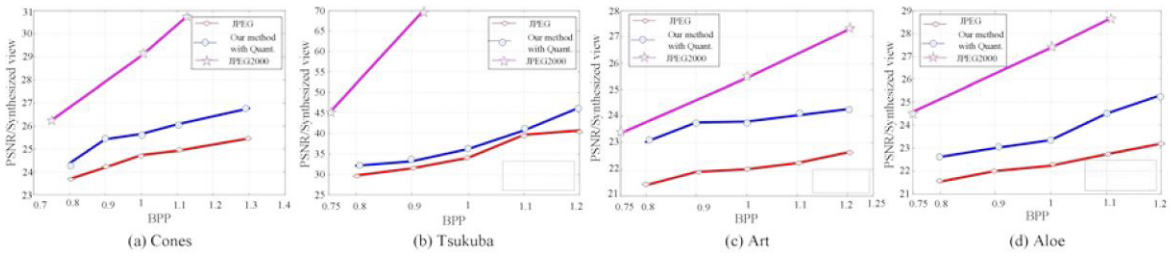

FIGURE 7. The rate-distortion comparison of the synthesized view.

maps are 1.0 BPP for Cones, Art and Aloe, and 0.5 BPP for Tsukuba, respectively. We can see that the synthesized right images with CS have superior (or at least the same) visual quality than those with JPEG and JPEG2000.

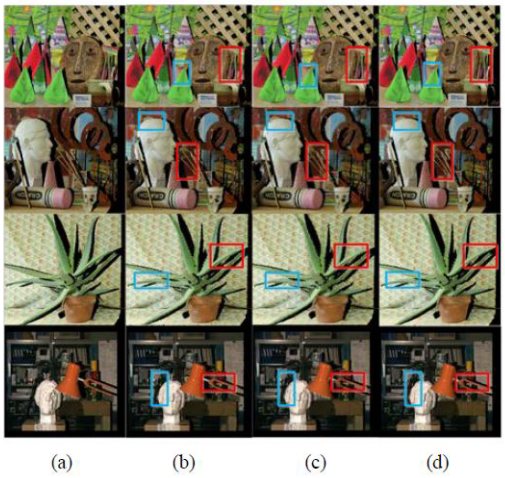

FIGURE 8. Resulting right frame synthesized using the original left frame of the Middlebury stereo images and various depth maps. From up till down:*Cones, Art, Aloe and Tsukuba.(a) Ground truth, (b) compressed with CS, (c) compressed with JPEG, (d) compressed with JPEG 2000.*

Although our method is not efficient than JPEG2000 in terms of ratedistortion performance, the perceived quality of the synthesized view performs better than it especially on edge areas as highlighted in Figure. 9, which demonstrates that the proposed scheme achieves less visual artifacts and preserves the edges better. This also more or less demonstrates that PSNR is not the exact quality evaluation for the reconstructed depth images. Another reason for the efficiency loss may be that the R-D optimization does not consider the synthesized views.

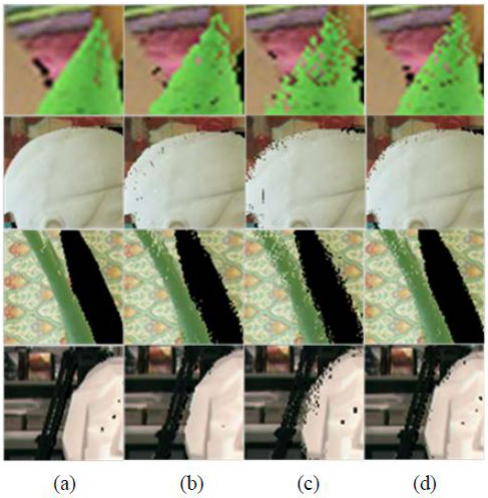

FIGURE 9. Highlighted parts of Fig. 8. From up till down: Cones, Art, Aloe and Tsukuba. (a0): original parts got from (a); (b0): original parts got from (b); (c0): original parts got from (c); (d0): original parts got from (d).

5. **Conclusion.** In this paper, a depth map coding based on compressed sensing (CS) is proposed. Different from the previous methods, quantization with different dead-zones is taken into account and the optimization between the measurement rate and quantization dead-zone is also considered. A new pair-wisely CS measurement is adopted in the Fourier domain and a minimal TV constraint is imposed in the recon- struction to preserve the edges while enhance the smoothness of the depth map. Experimental results show that our proposed scheme can preserve the arbitrary shaped edges well, and the synthesized view achieves better visual quality compared to the standard JPEG and JPEG2000. However, there are still some issues to be considered in the future, e.g., the rate-distortion optimization for the synthesized view should be taken into account. The relation between depth distortion and view synthesis distortion needs more investigation, and a more suitable metric should be used to measure the distortion of the depth map and synthesized view.

6. **Acknowledgment.** This work is supported in part by National Natural Science Foundation of China (No. 61272262, 61210006), The Shanxi Provincial Foundation for Leaders of Disciplines in Science (20111022), Shanxi Province Talent Introduction and Development Fund (2011), Shanxi Provincial Natural Science Foundation (2012011014-3) and Program for New Century Excellent Talent in University (NCET-12-1037).

REFERENCES

[1] B. B. Chai, S. Sethuraman, H. S. Sawhney, and P. Hatrack, Depth map compression for real-time view-based rendering, *Pattern Recognition Letters*, vol. 25, No. 7, pp. 755-766, 2004.

[2] Y. Zhao, C. Zhu, Z. Chen, and L. Yu, Depth no-synthesis-error model for view synthesis in 3-D video, *IEEE Trans. Image Processing*, vol. 20, no. 8, pp. 2221-2228, 2011.

[3] B. Zeng, and J. J. Fu, Directional discrete cosine transforms–A new framework for image coding, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 3, pp. 305-313, 2008.

[4] W. S. Kim, S. K. Narang, and A. Ortage, Graph based transforms for depth video coding, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 813-816, 2012.

[5] M. Kang, and Y. Ho, Depth video coding using adaptive geometry based intra prediction for 3D video system, *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 121-128, 2012.

[6] 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 images, *Proc. of IEEE International Conference on Image Processing*, pp. 105-108, 2007.

[7] R. Krishnamurthy, B. B. Chai, H. Tao, and S. Sethuraman, Compression and transmission of depth maps for image-based rendering, *Proc. of International Conference on Image Processing*, pp. 828-831, 2001.

[8] D. Farin, R. Peerlings, and P. H. N. De, Depth-image representation employing meshes for intermediate-view rendering and coding, *Proc. of 3DTV Conference*, pp. 1-4, 2007.

[9] E. J. Candes, J. Romberg, and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489-509, 2006.

[10] D. L. Donoho, Compressed sensing, *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.

[11] E. Candes, and M. Wakin, An introduction to compressive sensing, *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21-30, 2008.

[12] C. Patsakis, and N. Aroukatos, LSB and DCT steganographic detection using compressive sensing, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 5, no. 1, pp. 20-32, 2014.

[13] H. C. Huang, and F. C. Chang, Robust image watermarking based on compressed sensing techniques, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 5, no. 2, pp. 327-332, 2014.

[14] X. Zhang, A. Wang, B. Zeng, and L. Liu, Adaptive distributed compressed video sensing, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 5, no. 1, pp. 98-106, 2014.

[15] L. Liu, A. Wang, Z. Li, and K. Zhu, An improved distributed compressive video sensing based on adaptive sparse basis, *Proc. of the IEEE International Conference on Robot, Vision and Signal Processing*, pp. 137-140, 2011.

[16] G. Shen, W. S. Kim, S. K. Narang, A. Ortega, J. Lee, and H. Wey, Edge-adaptive transforms for efficient depth-map coding, *Proc. of the 28th Picture Coding Symposium*, pp. 566-569, 2010.
[17] S. Lee, and A. Ortega, Adaptive compressed sensing for depth map compression using graph-based transform, *Proc. of the 19th IEEE International Conference on Image Processing*, pp. 929-932, 2012.
[18] M. Lustig, D. Donoho, and J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182-1195, 2007.
[19] M. Sarkis, and K. Diepold, Depth map compression via compressed sensing, *Proc. of the 16th IEEE international conference on Image processing*, pp. 737-740, 2009.
[20] E. Candes, and J. Romberg, Sparsity and incoherence in compressive sampling, *Inverse Problems*, vol. 23, no. 3, pp.969-985, 2007.
[21] A. Ortega, and K. Ramchandran, Rate-distortion methods for image and video compression: An overview, *IEEE Signal Processing Magazine*, vol. 15, no. 11, pp. 23-50, Nov.1998.
[22] Middlebury College, Middlebury stereo evaluation, available at http:// vision.middlebury.edu/stereo/eval..