Local Entropy Minimization and Measurement Rate Allocation for Compressed Sensing of Depth Video

Shengwei Wang, Li Yu
School of Electron. Inf. & Commun.
Huazhong Univ. of Sci. & Tech.
Wuhan, Hubei, China
{kadinwang, hustlyu}@hust.edu.cn

***Abstract*— Compressed sensing (CS) provides a new method to encode depth videos, which utilizes the sparsity of depth maps to improve the coding efficiency. In this paper, we design a novel CS-based depth video codec. The codec adaptively decomposes blocks from 64×64 to 8×8 sub-blocks via wavelet transforming. The decomposition divides smooth regions and complex boundaries by frequencies, and minimizes the local entropy of sub-blocks. Moreover, a measurement rate allocation algorithm is also proposed, which utilizes the rate-distortion optimization (RDO) to allocate the measurement rate for each block. The experimental results demonstrate that, compared with H.264/AVC and H.265/HEVC, the proposed codec improves the quality of virtual views by 1-2 dB and 0.2-0.5 dB PSNR respectively. Meanwhile, the codec reduces coding complexity greatly.**

I. Introduction

In Recent years, three dimensional (3D) technology has achieved great advances. 3D applications such as 3D television (3DTV) [1] and free-viewpoint TV (FTV) [2] have also gradually developed form PC to mobile terminals. Especially, Google has released Google Cardboard [3], which makes it convenient for everyone to get access to virtual reality (VR). Above 3D applications all need to display the whole 3D scene for viewers, which consume a big amount of data. In order to reduce the data volume, depth maps are used to reconstruct the 3D scene via depth image based rendering (DIBR) [4].

Traditional video codecs such as H.264/AVC [5] and H.265/HEVC [6] are designed for encoding texture videos. They are not fit for encoding depth videos. Specifically, traditional codecs do not fully utilize the highly sparsity of depth maps. The theory of compressed sensing (CS) [7][8][9] provides a novel way to encode depth videos by utilizing the sparsity of depth maps. In the CS theory, through optimization, the sparsity of a discrete signal can be exploited to recover the signal from far fewer samples than required by the Shannon-Nyquist sampling theorem. Especially, while the depth maps are sparser, the compression efficiency can become higher. Moreover, due to the simplification of CS sampling, the CS algorithm can make the codec low power consumption, which meets the demand of the mobile environment.

Due to the advantages of CS and the sparsity of depth maps, researchers have proposed CS-based codec for depth videos [10][11][12][13]. Sarkis *et al.* [10] proposed a basic codec for compressed depth video sensing, which set a fixed measurement rate for each block in depth videos. The characteristics of depth maps, which contain smooth regions and sharp object boundaries, was not considered. Duan *et al.* [11] proposed a depth video coding algorithm. In the algorithm, the CS acquisition utilized sub-sampled 2-D discrete cosine transform (DCT) to obtain de-correlated samples for each size fixed block, leading to redundant sampling. Liu *et al.* [12] designed a quad-tree partition scheme for compressed depth video sensing, without considering the measurement rate allocation. In Vijayanagar *et al.* [13] depth video codec, the depth blocks are classified into U, CS and Skip blocks, and only CS blocks are compress sensed. In his codec, the characteristics of depth maps are utilized, and the classification of blocks can be more precious to improve the coding efficiency.

The CS-based depth video codec should consider and utilize the property of depth maps, which are consisted of smooth regions and sharp boundaries. Meanwhile, in the codec, the measurement rate of each block should be adaptively determined by different sparsity degree of blocks. In the proposed codec, the depth block is adaptively decomposed from 64×64 to 8×8 sub-blocks via wavelet transforming. The decomposition separates smooth regions and sharp boundaries of the block, and minimizes the local entropy of sub-blocks. Moreover, the codec applies a measurement rate allocation algorithm to adaptively determine the sample rate for each block. Due to the simplification of CS sampling, the coding complexity is also greatly reduced, compared with H.264/AVC and H.265/HEVC. The codec with low complexity adapts to mobile devices for low power consumption.

II. Compressed Sensing Theory

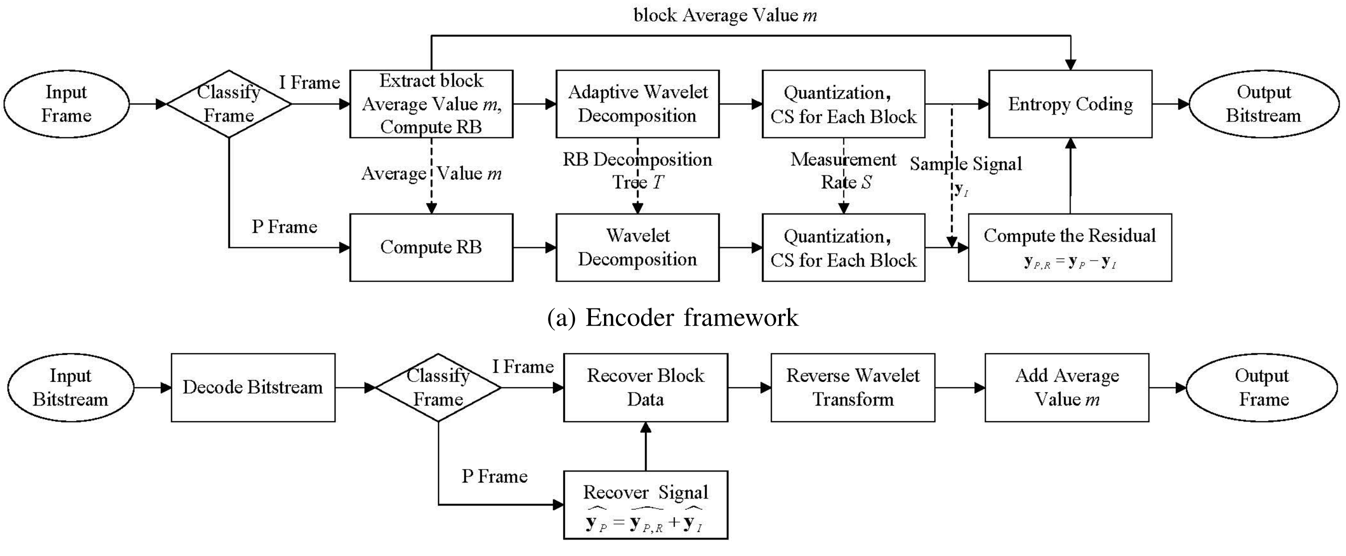
For a real-valued $N \times 1$ discrete signal $\mathbf{x}$, an orthonormal basis set $\mathbf{\Psi}$ is designed to represent $\mathbf{x}$ as:

$$\mathbf{x} = \mathbf{\Psi}\mathbf{s} \tag{1}$$

where $\mathbf{s}$ is the sparse coefficient vector. If the number of non-zero elements in $\mathbf{s}$ is k, $\mathbf{x}$ is k-sparse with respect to $\mathbf{\Psi}$ [8]. The CS theory proves that the signal $\mathbf{x}$ can be reconstructed from a very small measurement rate, which is less than the Nyquist sampling rate. Specifically, set M as the measurement number of $\mathbf{x}$, where $M \ll N$. Then, the measurement rate S can be calculated as follows:

$$S = \frac{M}{N} \tag{2}$$

978-1-5090-5966-9/16/$31.00 ©2016 IEEE


block Average Value m
Input Frame
Classify Frame
I Frame
Extract block Average Value m, Compute RB
Adaptive Wavelet Decomposition
Quantization, CS for Each Block
Entropy Coding
Output Bitstream
Average Value m
RB Decomposition Tree T
Measurement Rate S
Sample Signal $\mathbf{y}_I$
P Frame
Compute RB
Wavelet Decomposition
Quantization, CS for Each Block
Compute the Residual $\mathbf{y}_{P,R}=\mathbf{y}_P-\mathbf{y}_I$
Input Bitstream
Decode Bitstream
Classify Frame
I Frame
Recover Block Data
Reverse Wavelet Transform
Add Average Value m
Output Frame
P Frame
Recover Signal $\widehat{\mathbf{y}}_P=\widehat{\mathbf{y}_{P,R}}+\widehat{\mathbf{y}}_I$

(a) Encoder framework

(b) Decoder framework

Fig. 1. The framework of the proposed codec

Given an $M \times N$ measurement matrix $\mathbf{\Phi}$, combining the Eq. (1), the sample signal can be represented as:

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi \Psi s} \tag{3}$$

where $\mathbf{y}$ is the $M \times 1$ discrete sample signal.

By zig-zag scanning, the depth block data can be converted to an $N \times 1$ vector $\mathbf{x}$. The CS theory utilizes an $M \times 1$ vector $\mathbf{y}$ to sample $\mathbf{x}$ via a Gaussian random measurement matrix $\mathbf{\Phi}$. Due to $M \ll N$, the data volume of the block reduces greatly.

In order to recover $\mathbf{x}$ from $\mathbf{y}$, the total variational minimization (minTV) algorithm [14] is adopted. For depth blocks, minTV is applied in the gradient domain, as shown:

$$\min_{\widetilde{\mathbf{x}}} \|\widetilde{\mathbf{x}}\|_{TV} \quad s.t. \ \|\widehat{\mathbf{y}} - \Phi\widetilde{\mathbf{x}}\|_2 \le \varepsilon \tag{4}$$

where $\widetilde{\mathbf{x}}$ is the 1-D rasterization of the reconstruction CS block. The total variation $\|\widetilde{\mathbf{x}}\|_{TV}$ is defined as:

$$\|\widetilde{\mathbf{x}}\|_{TV} = \sum_{i,j} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2} \tag{5}$$

where $x_{i,j}$ is the pixel value at (i,j) in $\widetilde{\mathbf{x}}$.

III. Codec Architecture

The framework of the proposed encoder is shown in Fig. 1(a). The encoder sets the size of Group of Pictures (GOP) as 4 (IPPP), and the first frame of each GOP is intra (I) frame, the others are inter (P) frames. For the I frame, the encoder divides the frame into non-overlap 64×64 blocks, and process each block successively. For each block, the encoder firstly calculates the average pixel value m of the block, and writes m to the bitstream. Then, in order to reduce the amount of data preliminarily, the block minuses m, and the residual of the block (RB) remains. Next, an adaptive wavelet decomposition is conducted on the RB to divide high and low frequencies, since high and low frequency bands represent sharp boundaries and smooth regions respectively. Meanwhile, the process of the decomposition minimizes the local entropy of each wavelet sub-block, and generates a decomposition tree T of the RB. The detailed process of decomposition is described in Section. III.A. Next, the encoder quantizes the wavelet sub-block, and adaptively determines the measurement rate S of each block for CS. The process of rate-distortion optimization (RDO) guides the measurement rate allocation algorithm, for a well balance between coding quality and bit-rate. Section. III.B. shows the detailed algorithm. Finally, the block signal $\mathbf{X}_I$ is compress sensed as $\mathbf{y}_I$ by the Eq. (3), and $\mathbf{y}_I$ is entropy coded via the k^{th}-order Exp-Golomb code [15] to generate the bitstream.

The P frame is compressed in the similar way with I frame. The P frame obtains m, T and S from the co-located block of the I frame in the same GOP, which reduces both the temporal redundancy and the coding complexity. Moreover, the P frame dose not transmit the sample signal $\mathbf{y}_p$ of the block , but the residual signal $\mathbf{y}_{P,R}$, which are calculated as follows:

$$\mathbf{y}_{P,R} = \mathbf{y}_P - \mathbf{y}_I \tag{6}$$

where $\mathbf{y}_I$ is the sample signal of the co-located sub-block in the I frame.

The decoder framework is shown in Fig. 2(b). The decoder receives the bitstream and decodes it to the sample signals. For the I frame, the decoder reconstructs the wavelet blocks from the sample signals by minTV algorithm directly. Then, recover the original RBs by de-quantization and reverse wavelet transforming. Adding m, the RB becomes the original block of the depth map. The P frame is decoded in the same way, but the sample signal of the P frame $\mathbf{y}_P$ is the sum of the I frame sample signal $\mathbf{y}_I$ and the P frame residual signal $\mathbf{y}_{P,R}$. The recover algorithm is shown in Section. III.C. Moreover, m of the co-located block in the I frame is also added to the RB of

978-1-5090-5966-9/16/$31.00 ©2016 IEEE

the P frame to recover the original depth block.

A. Local Entropy based Adaptive Wavelet Decomposition

By calculating and extracting m of each block, the encoder only needs to compress RB. The RB consists of the smooth regions and the sharp object boundaries, which is the same as depth maps. After wavelet transforming, smooth regions and sharp boundaries preciously become low and high frequency bands respectively. Thus, different measurement rates can be carried out on different frequency bands to satisfy the error-tolerance of each band. Moreover, the wavelet basis is also an orthonormal basis, lying the foundation for following CS sampling.

In detail, the encoder designs a local entropy based adaptive algorithm to decompose block from 64×64 to 8×8, as Alg. 1 shows.

Algorithm 1 Adaptive wavelet decomposition for RB

```
Input:
    The RB in I frame, $B_{0,0}$.
Output:
    The decomposition tree of $B_{0,0}$, $T$;
    The decomposition result set, $\mathbf{B}$.
1: Initialization: Set the block size, $N = 64$, the decomposition level, $l = 0$, sub-block number $i = 0$;
2: Calculate the entropy $H_{l,i}$ of $N \times N$ block $B_{l,i}$ as follows:
       $H_{l,i} = E[-\log p_i] = -\sum_{i=1}^{n} p_i \log p_i$    (7)
   where $n$ is the number of values appeared in the $N \times N$ block, and $p_i$ is the appearance probability of the $i$-th value in the block. Eq. (7) is also the cost function.
3: Conduct wavelet transforming on $B_{l,i}$, and set $l = l + 1$, $N = N/2$. Thus, 4 $N \times N$ sub-blocks are obtained, indicating with $B_{l,j}$ $(j = 1, 2, 3, 4)$.
4: Calculate the local entropy of $B_{l,j}$ $(j = 1, 2, 3, 4)$ via the Eq. (7), indicating with $H_{l,j}$ $(j = 1, 2, 3, 4)$.
5: Compare $H_{l-1,i}$ with the sum of local entropy $\sum_{j=1}^{4} H_{l,j}$. If $H_{l-1,i}$ is larger, store the wavelet results and decomposition structure to $\mathbf{B}$ and $T$ respectively. Otherwise, abandon the results, and stop the algorithm.
6: If $l = 3$, stop the algorithm. Otherwise, repeat procedure 2-5 on 4 sub-blocks $B_{l,j}$ $(j = 1, 2, 3, 4)$ respectively.
```

In the algorithm, the RB in the I frame are decomposed, in order to distinguish smooth regions and sharp boundaries by the frequency bands. Meanwhile, the local entropy are imported as the cost function to optimize the wavelet decomposition tree. The cost function controls the level of decomposition to minimize the entropy of each leaf node. The node with smaller local entropy is also sparser, and can achieve higher compression rate. The process of decomposition is lossless, because the algorithm do not reduce the entropy of the whole block, but the local entropy of each wavelet sub-block.

For the P frame, the wavelet decomposition are conducted, according to the decomposition tree T of the co-located RB in the I frame at the same GOP.

B. Quantization and Measurement Rate Allocation

After adaptive wavelet transforming, each wavelet sub-block is converted from 2-D block to 1-D vector by zig-zag scanning. Then, in order to quantize the block, the encoder sets a dead-zone threshold Δ, and a quantization step q. The original sample vector $\mathbf{y}$ can be quantized by the following equation:

$$\mathbf{y}_q = \begin{cases} 0, & if\ |\mathbf{y}| \le \Delta \\ ceil(\frac{\mathbf{y}-\Delta}{q}), & if\ \mathbf{y} > +\Delta \\ floor(\frac{\mathbf{y}+\Delta}{q}), & if\ \mathbf{y} < -\Delta \end{cases} \tag{8}$$

where $\mathbf{y}_q$ is the quantized vector of $\mathbf{y}$. For simplification, the paper assumes $\Delta = q$. Meanwhile, the quantization step size of the low frequency block is fixed at 5, and that of the high frequency ranges from 2 to 20.

Since each block has different error-tolerance, different measurement rate should be set for each block. In the CS theory, the measurement number M is determined by the equation in [9] as follows:

$$M = cK \log \frac{N}{K} \tag{9}$$

where K is the sparse degree of the signal, and c is the adjustment coefficient.

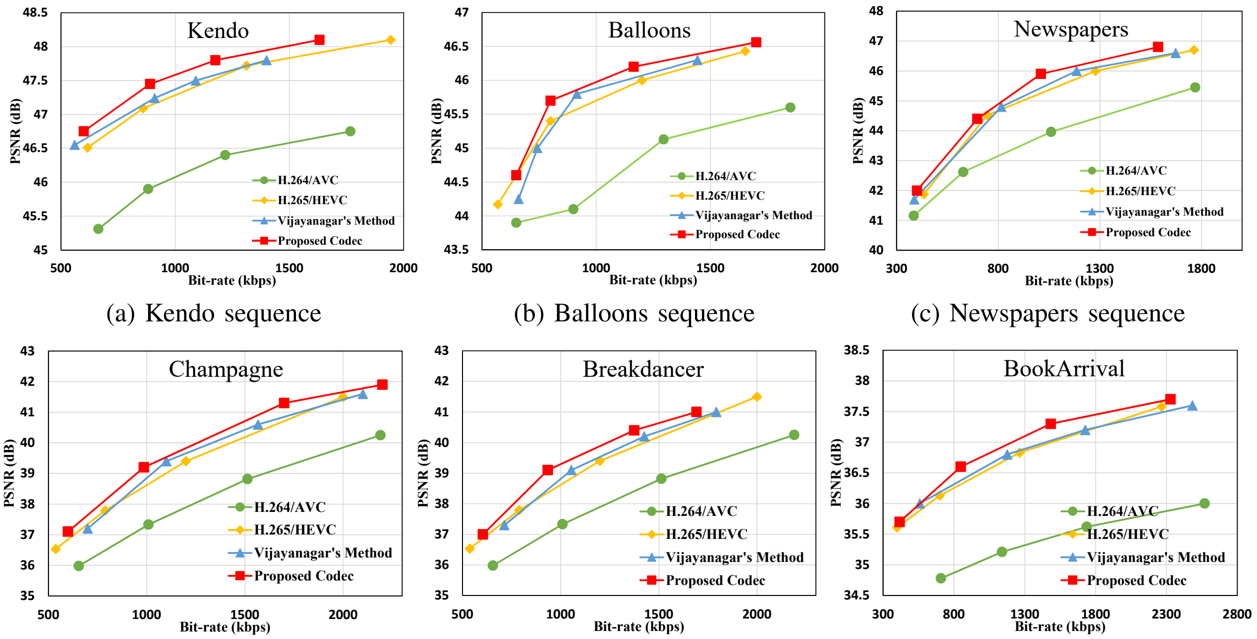
In order to determine the measurement rate for each block, the encoder applies rate-distortion optimization (RDO) algorithm to adjust c for each block. In detail, $\mathrm{X} = \{\chi_1, \chi_2 \cdots \chi_p\}$ represents the blocks in the frame, where p is the total number of blocks. Assume the set of adjustment coefficients for all the blocks is $C = \{c_1, c_2, \cdots c_p\}$. Meanwhile, $M = \{m_1, m_2 \cdots m_p\}$ shows the measurement number of blocks. The relationship between m_i and c_i is described as the Eq. (9). In order to trade off between the bit-rate and the video quality, C should satisfy the following RDO relationship:

$$\begin{aligned} C^* &= \arg\min_{C^* \in C} D(C, \mathrm{X}) \\ R(C, \mathrm{X}) &\le R_T \end{aligned} \tag{10}$$

where R_T is the target bit-rate. $D(C, \mathrm{X})$ and $R(C, \mathrm{X})$ is the total distortion and the total bit-rate. By a specified set of adjustment coefficients C, all the blocks in the frame can be sampled via CS. In order to calculate the $R(C, \mathrm{X})$, the sample signal $\mathbf{y}$ of the block is binarized and coded by Exp-Golomb. In order to calculate the $D(C, \mathrm{X})$, the algorithm reconstructs the wavelet sub-block from the sample signal $\mathbf{y}$. After this, the sum of mean square error (MSE) between the original block and the reconstructed block is set as the $D(C, \mathrm{X})$. By assuming that the rate and the distortion are additive, the Eq. (10) is re-written as following with optimal Lagrange multiplier:

$$C^* = \arg\min_{C^* \in C} \sum_{i=1}^{p} \left(D(c_i, \chi_i) + \lambda_{opt} R(c_i, \chi_i)\right) \tag{11}$$

978-1-5090-5966-9/16/$31.00 ©2016 IEEE



(a) Kendo sequence

(b) Balloons sequence

(c) Newspapers sequence

(d) Champagne sequence

(e) Breakdancer sequence

(f) BookArrival sequence

Fig. 2. R-D performance compared with the proposed codec, Vijayanagar's method, H.265/HEVC and H.264/AVC.

In the algorithm, λ_{opt} is acquired via the bisection algorithm. It should be noted that the proposed algorithm does not reconstructing the depth values while computing the distortion. Instead, to reduce the coding complexity, the algorithm uses the original wavelet transforming coefficients and the reconstructed coefficients to compute the MSE.

After obtaining λ_{opt}, the optimal C can be calculated. Then, for the i-th block, c_i is applied to determine the measurement rate for the blocks, and each block is sampled by CS via a Gaussian random measurement matrix $\mathbf{\Phi}$ with the Eq. (3). For the P frame, c_i is obtained from the co-located block of the I frame in the same GOP.

C. Sample Reconstruction

The decoder firstly recovers the sampling signal from the bitstream. Then, the minTV algorithm reconstructs the wavelet blocks. For the blocks in the I frame, the Eq. (4) describes the reconstruction process. Due to that the P frame transmits the residual of the sampling signals, the decoder should recover the sample signal $\widehat{y_P}$, as follows:

$$\widehat{y_P} = \widehat{y_I} + \widehat{y_{P,R}} \tag{12}$$

where $\widehat{y_P}$ is the complete sample signal of the P frame, and $\widehat{y_{P,R}}$ is the received residual sample signal. $\widehat{y_I}$ is the received I frame sample signal from the co-located sub-block in the same GOP with the P frame. Then, $\widehat{y_P}$ is used to reconstructed the original wavelet block by the Eq. (4).

IV. Experimental Results

The experiments are conducted on six test sequences. In each sequences, two views are selected as the reference views. Two reference views synthesize the virtual view. The Tab. 1 shows the detailed information of sequences and views in the experiments. 64 consecutive frames of each sequences are used to code. In every experiment, only the depth videos of reference views are compressed, and the quality of the virtual view is compared. The GOP of sequences is set as 4 (IPPP), and four different QPs are set to encode the sequences. View synthesis reference software (VSRS) 3.5 [16] is used to render the virtual views.

TABLE I
Experimental Sequences

Sequences	Left View	Right View	Virtual View
Kendo	View 1	View 3	View 2
Balloons	View 1	View 3	View 2
Newspapers	View 4	View 6	View 5
Champagne	View 37	View 39	View 38
Breakdancer	View 3	View 5	View 4
BookArrival	View 6	View 10	View 8

To verify the rate-distortion (RD) performance of the proposed codec. The proposed codec is compared with H.264/AVC, H.265/HEVC and Vijayanagar's method [13]. The reference software of H.264/AVC is JM 17. The reference software of H.265/HEVC is HM 12. The detailed RD performances of four methods are shown in Fig. 2. The bit-rate is the sum of the needed bit-rate to encode the left and the right reference depth videos. Peak signal noise ratio (PSNR) is computed with the luminance component of the virtual views, which are synthesized by the ground truth depth videos and encoded depth videos respectively.

It is clear to see that, the proposed codec has the best

978-1-5090-5966-9/16/$31.00 ©2016 IEEE

TABLE II
COMPLEXITY COMPARISON

Sequences	**H.264/AVC** **Average time(sec)**	**H.265/HEVC** **Average time(sec)**	**Vijayanagar's method** **Average time(sec)**	**Proposed codec** **Average time(sec)**
Kendo	1868.47	8135.51	107.59	140.93
Balloons	1764.38	7896.73	97.73	129.83
Newspapers	1859.37	8671.33	97.34	137.86
Champagne	1884.89	8734.88	91.43	148.17
Breakdancer	1870.35	9956.63	98.51	134.16
BookArrival	1785.11	8030.81	94.47	127.19

RD performance. Especially, compared with H.264/AVC and H.265/HEVC, the proposed codec improves the quality of virtual views by 1-2 dB and 0.2-0.5 dB PSNR respectively. As traditional video codecs, H.264/AVC and H.265/HEVC are designed to encode texture videos. They are not sensitive to high frequency signals, which represent sharp object boundaries in depth maps. Vijayanagar's method [13] does not fully consider the property of each depth block. The proposed codec distinguishes the sharp boundaries and smooth regions by dividing the frequencies via wavelet transforming. Meanwhile, different measurement rates are allocated to sample each block by the features of the block. The experimental results also demonstrate that the proposed codec improves the quality of depth video coding efficiently.

Moreover, in order to verify the coding complexity, the encoding time of 4 methods is shown in Tab. 2. The average time means the average time of encoding 8 times for a sequence (4 QPs and 2 views) and is calculated as follows:

$$\mathrm{AvgTime} = \frac{\sum_{i=1}^{4} (T_{i,left} + T_{i,right})}{8} \qquad (13)$$

Where T means the time of encoding a single view of a sequence. The subscript i means the video is encoded with i-th QP, and the subscript $left$ and $right$ indicate the left and the right reference views respectively.

As the table shows, the proposed codec, compared with H.264/AVC and H.265/HEVC, reduces complexity greatly, making the codec low-power consumption. Due to the elimination of motion estimation, and mode selection in intra/inter prediction, compared with H.264/AVC and H.265/HEVC, the proposed codec can simplify the coding process greatly. The proposed codec has similar complexity with Vijayanagar's method, since both two methods adopt the CS-based codec framework. Moreover, the proposed codec adopts the adaptive wavelet transforming for a better encoding quality, which only costs slightly more time than Vijayanagar's method.

V. CONCLUSION

This paper proposed a novel CS-based codec for depth video coding. The codec adaptively decomposed blocks from 64×64 to 8×8. The process of decomposition divided smooth regions and complex boundaries by frequencies, and minimized the local entropy of each leaf node for better compression rate. Moreover, the codec utilized RDO algorithm to allocate the measurement rate for each block, which ensured that the codec balanced the video quality and the total bit-rate. The experimental results demonstrate that, compared with H.264/AVC and H.265/HEVC, the proposed codec improves the quality of virtual views by 1-2 dB and 0.2-0.5 dB PSNR respectively. Meanwhile, the codec reduces the coding complexity greatly.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (NSFC) (No. 61231010), National High Technology Research and Development Program (No. 2015AA015903, No. 2015AA015901), Research Fund for the Doctoral Program of MOE (No. 0120142110015).

REFERENCES

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

[2] M. Tanimoto, "Overview of FTV (free-viewpoint television)," *IEEE Int. Conf. Multimedia and Expo.*, pp. 1552–1553, Jun. 2009.

[3] (2016) Google VR and Google Developers. [Online]. Available: http://developers.google.com/vr/

[4] C. Zhu, Y. Zhao, L. Yu, and M. Tanimoto, *3D-TV System with Depth-Image-Based Rendering (DIBR)*. Springer, 2014.

[5] S. K. Kwon, A. Tamhankar, and K. R. Rao, "Overview of H.264/MPEG-f part 10," *J. Visual Commun. & Image Representation*, vol. 17, no. 2, pp. 186–216, 2006.

[6] G. J. Sullivan, J. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits & Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] E. J. Candè and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[9] E. J. Candès *et al.*, "Compressive sampling," in *Intl. Congr. of Math.*, vol. 3, 2006, pp. 1433–1452.

[10] M. Sarkis and K. Diepold, "Depth map compression via compressed sensing," in *IEEE Intl. Conf. Image Process.*, 2009, pp. 737–740.

[11] J. Duan, L. Zhang, Y. Liu, R. Pan, and Y. Sun, "An improved video coding scheme for depth map sequences based on compressed sensing," in *IEEE Intl. Conf. Multimedia Tech.*, 2011, pp. 3401–3404.

[12] Y. Liu, K. R. Vijayanagar, and J. Kim, "Quad-tree partitioned compressed sensing for depth map coding," in *IEEE Intl. Conf. Acoust., Speech and Signal Process.*, 2014, pp. 870–874.

[13] K. R. Vijayanagar, Y. Liu, and J. Kim, "Adaptive measurement rate allocation for block-based compressed sensing of depth maps," in *IEEE Intl. Conf. Image Process.*, 2014, pp. 1307–1311.

[14] S. Hawe, M. Kleinsteuber, and K. Diepold, "Dense disparity maps from sparse disparity measurements," in *IEEE Intl. Conf. Comput. Vision*, 2011, pp. 2126–2133.

[15] I. E. G. Richardson, *The H.264 Advanced Video Compression Standard*. Wiley John + Sons, 2010.

[16] "View synthesis reference software (VSRS 3.5)," *Tech. Rep. ISO/IEC JTC1/SC29/WG11*, Mar. 2010.