QUAD-TREE PARTITIONED COMPRESSED SENSING FOR DEPTH MAP CODING

Ying Liu, Krishna Rao Vijayanagar, and Joohee Kim

Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616

ABSTRACT

We consider a variable block size compressed sensing (CS) framework for high efficiency depth map coding. In this context, quad-tree decomposition is performed on a depth image to differentiate irregular uniform and edge areas prior to CS acquisition. To exploit temporal correlation and enhance coding efficiency, such quad-tree based CS acquisition is further extended to inter-frame encoding, where block partitioning is performed independently on the I frame and each of the subsequent residual frames. At the decoder, pixel domain total-variation minimization is performed for high quality depth map reconstruction. Experiments presented herein illustrate and support these developments.

Index Terms— Quad-tree decomposition, compressed sensing, sub-Nyquist sampling, sparse signals, depth map, total-variation

1. INTRODUCTION

Recent advance in display and camera technologies has enabled three-dimensional (3-D) video applications such as 3-D TV and stereoscopic cinema. In order to provide the "look-around" effect that audiences expect from a realistic 3-D scene, a vast amount of multiview video data needs to be stored or transmitted, leading to the desire of efficient compression techniques. One proposed solution is to encode two views of the same scene captured from different viewpoints along with the corresponding depth (disparity) map. With texture video sequences and depth map sequences, an arbitrary number of intermediate views can be synthesized at the decoder side using depth image-based rendering (DIBR) techniques [1]. Depth maps, therefore, are considered as an essential coding target for 3-D video applications.

Typically, depth maps are characterized by irregular piecewise smooth areas separated by sharp object boundaries, with very limited texture. To efficiently compress such images, traditional methods focus on constructions of linear functions to effectively represent smooth areas [2] or transforms that are adapted to edges [3],[4]. Recently, new depth map coding paradigms have been developed under the compressed sensing (CS) framework. CS is an emerging body of work that deals with sub-Nyquist sampling of sparse signals of interest [5]-[7]. Rather than collecting an entire Nyquist ensemble of signal samples, CS can reconstruct sparse signals from a small number of (random [7] or deterministic [8]) linear measurements via convex optimization [9], linear regression [10],[11], or greedy recovery algorithms [12]. In CS based depth map coding, data compression can be achieved via sub-Nyquist sampling as well as subsequent entropy coding of the CS samples, while high quality depth image reconstruction is still available by sparsity-aware decoding.

An existing CS based depth map coding algorithm [13] performs CS acquisition with a sub-sampled 2-D discrete cosine transform (DCT) to obtain de-correlated CS samples which are further entropy encoded, and reconstruction is performed via enforcing gradient sparsity in the pixel domain. Such technique provides higher coding efficiency than random sensing matrix used in typical CS based video compression strategies, as well as better reconstruction quality than pure inverse 2-D DCT in standard video decoders. However, the algorithm is based on fixed block size CS acquisition, leading to redundant sampling of large irregular uniform areas. On the other hand, graph based transform (GBT) has also been proposed for CS based depth map coding [14]. In such scheme, partial Hadamard transform is used as the sensing matrix and GBT is constructed per depth image block for sparse representation. Although the GBT provides more effective sparse transform than other orthogonal basis such as 2-D DCT, the construction of block-adaptive GBT increases encoder complexity, and the side information needed to specify the GBT heavily reduces the overall coding efficiency.

In this paper, we aim at developing a CS based variable block size encoder for efficient depth map compression. To avoid redundant CS acquisition of large irregular uniform areas, a simple top-down quad-tree decomposition algorithm is proposed to partition a depth map into uniform blocks of variable sizes and small blocks containing edges. Lossless 8-bit compression is then applied to each of the uniform blocks and only the edge blocks are encoded by CS and subsequent entropy coding. Such variable block size encoder is then extended to inter-frame encoding, where the quad-tree decomposition is independently applied to the I frame and subsequent residual frames in a group of pictures (GOP) of I-P-P-P structure. At the decoder, pixel-domain total-variation minimization is applied to the de-quantized CS measurements (or

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

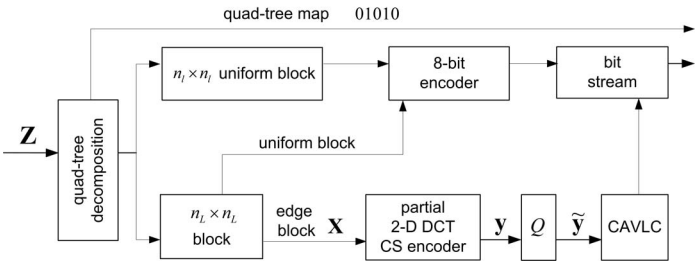

Fig. 1. Quad-tree partitioned CS encoder.

sub-sampled 2D-DCT coefficients) for edge block reconstruction.

The remainder of this paper is organized as follows. In Section 2, we describe in detail the proposed quad-tree partitioned CS depth map encoder, and a total-variation minimization based depth image reconstruction algorithm are elaborated in Section 3. In Section 4, the proposed CS encoder/decoder is extended to inter-frame coding. Experimental results and comparison studies are presented in Section 5. Finally, a few conclusions are drawn in Section 6.

2. PROPOSED QUAD-TREE PARTITIONED CS ENCODER

In the compressive depth map encoding block diagram shown in Fig. 1, each frame is virtually partitioned into non-overlapping macro blocks of size $n \times n$. A simple L-level top-down quad-tree decomposition is then applied to each macro block $\mathbf{Z} \in \mathbb{R}^{n \times n}$ independently to partition it into uniform blocks of size $\frac{n}{2^{\ell-1}} \times \frac{n}{2^{\ell-1}}$, $\ell \in \{1, 2, ..., L\}$, and edge blocks of size $\frac{n}{2^{L-1}} \times \frac{n}{2^{L-1}}$. Then, each uniform block is losslessly encoded using 8-bit representation, and CS is performed on each edge block $\mathbf{X} \in \mathbb{R}^{\frac{n}{2^{L-1}} \times \frac{n}{2^{L-1}}}$ in the form of

$$\mathbf{y} = \mathbf{\Phi}(\mathbf{X}), \tag{2.1}$$

where the sensing operator $\mathbf{\Phi}(\cdot)$ is equivalent to sub-sampling the P 2D-DCT coefficients of the lowest frequency after zig-zag scan. Then, the resulting measurement vector $\mathbf{y} \in \mathbb{R}^P$ is processed by a scalar quantizer with a certain quantization parameter (QP), and the quantized indices $\widetilde{\mathbf{y}}$ are entropy encoded using context adaptive variable length coding (CAVLC) as implemented for [15] and transmitted to the decoder.

The bit-saving property of the proposed CS depth map encoder relies on the quad-tree decomposition illustrated in Fig. 2. For each macro block $\mathbf{Z}$, the proposed quad-tree decomposition is performed as follows. Let the level index $\ell = 1$ be the lowest level corresponding to the macro block of size $n \times n$, then the block size at level-ℓ can be simply calculated as $n_\ell = n \times 2^{1-\ell}$. Let us denote a level-$\ell$ block as $\mathbf{X}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$. If all the pixels in $\mathbf{X}_\ell$ have the same values, the encoder transmits a '0' to indicate $\mathbf{X}_\ell$ is a uniform block, and the quad-tree decomposition is thereby terminated. If $\mathbf{X}_\ell$ is not uniform and $\ell < L$, the encoder transmits a '1' to indicate that $\mathbf{X}_\ell$ will be further split into four non-overlapping sub-blocks $\mathbf{X}^i_{\ell+1} \in \mathbb{R}^{\frac{n_\ell}{2} \times \frac{n_\ell}{2}}$, $i = 1, 2, 3, 4$. Decomposition is performed recursively for each sub-block in a left-to-right, up-to-down order indicated by the arrows in Fig. 2 until the highest level $\ell = L$ is reached, or a uniform block is detected. The resulting bit stream is transmitted as the "quad-tree map" to inform the decoder of the decomposition structure for successful decoding.

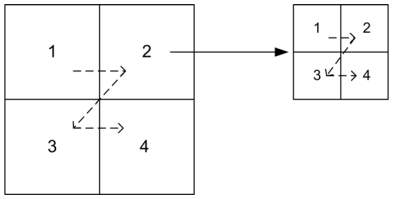

Fig. 2. Quad-tree decomposition.

3. TOTAL-VARIATION MINIMIZATION RECONSTRUCTION

At the decoder, the reconstruction of each macro-block is performed independently. As described in Fig. 3, the decoder first reads the bit stream along with the binary quad-tree map to identify uniform and edge blocks. For uniform blocks, a simple 8-bit decoding is carried out. For edge blocks, the decoder performs entropy decoding to obtain the quantized partial 2D-DCT coefficients (or CS measurements) $\widetilde{\mathbf{y}}$. The elements of $\widetilde{\mathbf{y}}$ are then de-quantized to form the vector $\widehat{\mathbf{y}}$. Since depth map blocks containing edges have sparse spatial gradients, they can be reconstructed via pixel-domain 2-D total-variation[1] (TV) minimization in the form of

$$\widehat{\mathbf{X}} = \arg\min_{\mathbf{X}} \mathrm{TV}_{2\mathrm{D}}(\mathbf{X}) \tag{3.1}$$

$$\text{subject} \quad \text{to} \quad ||\widehat{\mathbf{y}} - \mathbf{\Phi}(\mathbf{X})||_{\ell_2} \leq \epsilon.$$

The reconstructed uniform blocks and edge blocks are regrouped thereafter to form the decoded macro block $\widehat{\mathbf{Z}}$.

4. EXTENSION TO INTER-FRAME CODING

So far, we have carried out the quad-tree based CS encoding for only intra frames. To exploit temporal correlation among successive frames, we now extend the proposed algorithm to

[1]The mathematical expression of pixel-domain 2-D total-variation is defined as in [16],[17].

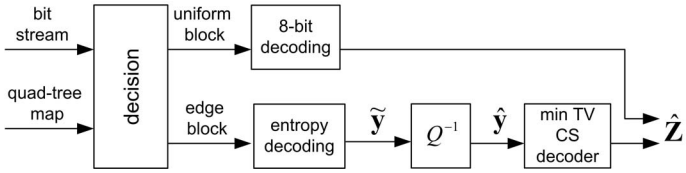

Fig. 3. Decoding system.

inter-frame encoding. At the encoder, the sequence of depth images is divided into groups of pictures with I-P-P-P structure. For each GOP, the I frame $\mathbf{F}_t$ at time slot t is encoded as described in Section 2, and the reconstructed image $\widehat{\mathbf{F}}_t$ is taken as the reference frame for encoding the subsequent three P frames. Hence, the k^{th} residual frame can be computed as

$$\mathbf{F}^r_{t+k} = \mathbf{F}_{t+k} - \widehat{\mathbf{F}}_t, \qquad (4.1)$$

$$k = 1, 2, 3.$$

Denote a residual macro block in $\mathbf{F}^r_{t+k}$ as $\mathbf{Z}^r_{t+k} \in \mathbb{R}^{n\times n}$, then it can be considered as the input of the quad-tree partitioned CS encoder depicted in Fig. 1. Specifically, an edge residual block can be computed as $\mathbf{X}^r_{t+k} = \mathbf{X}_{t+k} - \widehat{\mathbf{X}}_t$ where $\mathbf{X}_{t+k}$ is the pixel block in the k^{th} P frame to be encoded and the predicted block $\widehat{\mathbf{X}}_t$ is the co-located block in the decoded I frame $\widehat{\mathbf{F}}_t$.

For reconstruction, the uniform residual blocks can be recovered via 8-bit decoding. For an edge residual block, the CS measurements, or the de-quantized sub-sampled 2D-DCT coefficients received can be represented as

$$\begin{aligned} \widehat{\mathbf{y}}^r_{t+k} &= \Phi(\mathbf{X}^r_{t+k}) + \mathbf{n}^r_{t+k} \\ &= \Phi(\mathbf{X}_{t+k}) - \Phi(\widehat{\mathbf{X}}_t) + \mathbf{n}^r_{t+k}, \end{aligned} \qquad (4.2)$$

where $\mathbf{n}^r_{t+k}$ is the noise due to residual quantization. Hence, a pixel-domain TV minimization algorithm can be applied to reconstruct the P frame pixel block $\mathbf{X}_{t+k}$ in the form of

$$\widehat{\mathbf{X}}_{t+k} = \arg\min_{\mathbf{X}} \mathrm{TV}_{2\mathrm{D}}(\mathbf{X}) \qquad (4.3)$$

$$\text{subject to} \quad ||\widehat{\mathbf{y}}^r_{t+k} + \Phi(\widehat{\mathbf{X}}_t) - \mathbf{\Phi}(\mathbf{X})||_{\ell_2} \le \epsilon.$$

5. EXPERIMENTAL STUDIES

In this section, we experimentally study the performance of the proposed CS depth map coding system by evaluating the R-D performance of the synthesized view. Two test video sequences, *Balloons* and *Kendo*, with a resolution of 1024×768 pixels are used. For both video sequences, 40 frames of the depth maps of view 1 and view 3 are compressed using the proposed quad-tree partitioned CS encoder, and the reconstructed depth maps at the decoder are used to synthesize the texture video sequence of view 2 with the View Synthesis Reference Software (VSRS) [18].

In our experiments, the macro block size $n = 128$, and a five-level ($L = 5$) quad-tree decomposition is implemented, resulting in uniform blocks of size $n_\ell \times n_\ell$, $n_\ell \in \{8, 16, 32, 64, 128\}$, and edge blocks of size 8×8. The number of CS measurements for each edge block is fixed at $P = 24$. For the intra-frame encoder, four different values of quantization parameters, QP=$\{$24,28,32,36$\}$ are utilized. For the inter-frame encoder, four pairs of quantization parameters are utilized for I and P frames as shown in Table 1. To reconstruct edge blocks from partial 2D-DCT CS measurements, TVAL3 software [19],[20] is employed to solve the TV minimization problems in (3.1) and (4.3).

Table 1. QP values for Inter-frame encoding.

	QP			
I frame	22	26	30	34
P frame	24	28	32	36

In our experimental studies, two proposed CS depth map encoders are examined for both sequences: the intra-frame and inter-fame quad-tree partitioned CS encoders. For comparison studies, we include three existing CS based depth map compression algorithms: intra- and inter-frame partial 2D-DCT CS encoder without quad-tree decomposition [13], and the intra-frame CS encoder based on partial Hadamard sensing matrix with GBT sparsifying basis [14]. For fair comparison, the same CAVLC scheme is used in the entropy coding of the CS samples for all five encoders. To solve the ℓ_1 minimization problem in the GBT-based algorithm, ℓ_1-magic software [21] is utilized.

Fig. 4 shows the rate-distortion characteristics of the *Balloons* sequence. The bit-rate indicates the average bits per pixel (bpp) of the compressed depth map sequences from view 1 and view 3, and the peak signal-to-noise ratio (PSNR) of the luminance component of synthesized view 2 is computed between the rendered view using compressed depth sequences and using the ground-truth depth sequences. Evidently, for a fixed PSNR, the proposed intra- and inter-frame quad-tree partitioned CS encoder provide as much as 0.075 bpp savings compared to their non-quad-tree counterparts. For GBT-based intra-frame CS encoder proposed in [14], the bit-rate shown in Fig. 4 includes only the bits that represent the partial Hadamard CS samples, while the side information needed for constructing the block-adaptive GBT sparsifying basis at the decoder is not included since its size highly depends on the entropy coding method. Nevertheless, it is sufficient to show that our proposed quad-tree based intra-frame CS encoder outperforms the GBT intra-frame CS encoder.

The same rate-distortion performance study is repeated in Fig. 5 for the *Kendo* sequence. For this fast motion sequence, our proposed intra- and inter-frame quad-tree partitioned CS

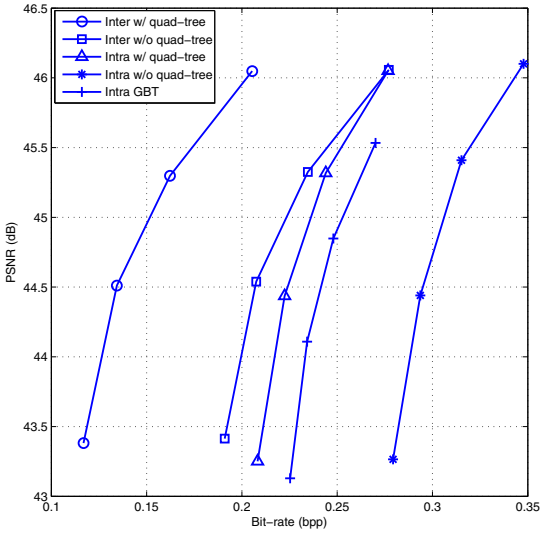

Fig. 4. Rate-distortion studies on the synthesized view 2 of the Balloons sequence.

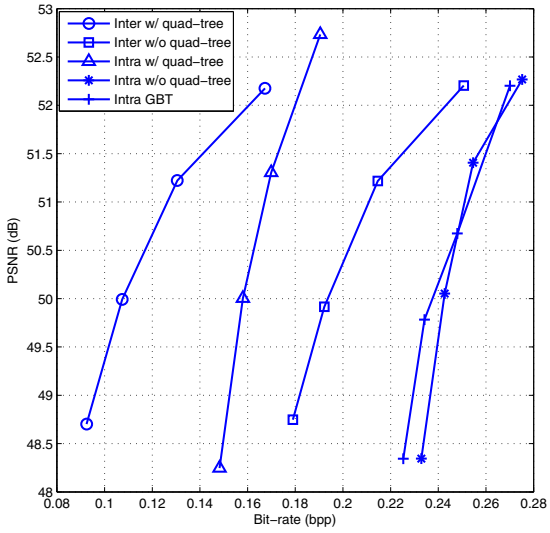

Fig. 5. Rate-distortion studies on the synthesized view 2 of the Kendo sequence.

encoder outperforms all the other three coding algorithms.

6. CONCLUSIONS

We proposed a variable block size CS coding system for depth map compression. To avoid redundant CS acquisition of large irregular uniform areas, a five-level top-down quad-tree decomposition is utilized to identify uniform blocks of variable sizes and small edge blocks. Each of the uniform blocks are encoded losslessly using 8-bit representation, and the edge blocks are encoded by CS with partial 2D-DCT sensing matrix. At the decoder side, edge blocks are reconstructed through pixel domain total-variation minimization. Since the proposed quad-tree decomposition algorithm is based on simple arithmetics, such CS encoder provides significant bit savings with negligible extra computational cost compared to pure CS-based depth map compression in literature. The proposed coding scheme can further enhance the rate-distortion performance when applied to an inter-frame coding structure.

7. REFERENCES

[1] A. Smolic, K. Müller, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems," in *Proc. IEEE Int. Conf. on Image Process. (ICIP)*, San Diego, CA, pp. 2448-2451, Oct. 2008.

[2] Y. Morvan, P. H. N. de With, and D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XIII*, vol. 6055, Jan. 2006.

[3] M. Maitre and M. N. Do, "Depth and depth-color coding using shape-adaptive wavelets," *J. Vis. Commun. Image R.*, vol. 21, no. 5-6, pp. 513-522, Mar. 2010.

[4] G. Shen, W.-S. Kim, S. K. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth-map coding," in *Proc. Picture Coding Symp. (PCS)*, Nagoya, Japan, pp. 566-569, Dec. 2010.

[5] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[7] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[8] K. Gao, S. N. Batalama, D. A. Pados, and B. W. Suter, "Compressive sampling with generalized polygons," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4759-4766, Oct. 2011.

[9] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure and Appl. Math.*, vol. 59, no. 8, pp. 1207-1223, Aug. 2006.

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267-288, 1996.

[11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, pp. 407-451, Apr. 2004.

[12] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[13] 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 *Proc. Int. conf. on Multimedia Technology (ICMT)*, Hangzhou, China, pp. 3401-3404, Jul. 2011.

[14] S. Lee and A. Ortega, "Adaptive compressed sensing for depthmap compression using graph-based transform," in *Proc. IEEE Int. Conf. on Image Process. (ICIP)*, Orlando, FL, pp. 929-932, Sept. 2012.

[15] A. A. Muhit, M. R. Pickering, M. R. Frater, and J. F. Arnold, "Video Coding using Elastic Motion Model and Larger Blocks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 5, pp. 661-672, May 2010.

[16] Y. Liu and D. A. Pados, "Decoding of framewise compressed-sensed video via interframe total variation minimization," *SPIE J. Electron. Imaging, Special Issue on Compressive Sensing for Imaging*, vol. 22, no. 2, Apr.-Jun. 2013.

[17] Y. Liu and D. A. Pados, "Rate-adaptive compressive video acquisition with sliding-window total-variation-minimization reconstruction," in *Proc. SPIE, Compressive Sensing Conf., SPIE Defense, Security, and Sensing*, Baltimore, MD, vol. 8717, May, 2013.

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

[19] C. Li, H. Jiang, P. Wilford, and Y. Zhang, "Video coding using compressive sensing for wireless communications," in *Proc. IEEE Wireless Communications & Networking Conf. (WCNC)* Cancun, Mexico, pp. 2077-2082, Mar. 2011.

[20] H. Jiang, C. Li, R. Haimi-Cohen, P. Wilford, and Y. Zhang, "Scalable video coding using compressive sensing," *Bell Labs Technical Journal*, vol. 16, pp. 149-169, Mar. 2012.

[21] E. Candès and J. Romberg, "ℓ1-magic: Recovery of sparse signals via convex programming," URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf.