ADAPTIVE OPTIMAL BIT-DEPTH ESTIMATION IN COMPRESSED VIDEO SENSING

Zhi-Jie He, Chun-Ling Yang, Rui-Dong Tang, Li-Hong Ma*

School of Electronic and Information Engineering, South China University of Technology Guangzhou 510640, China

ABSTRACT

Quantization bit-depth has an important effect on rate-distortion (RD) in compressed video sensing (CVS). This paper proposes an adaptive optimal bit-depth estimation (AOBE) model based on non-uniform quantization and frame-based DPCM. In the model, quantization bit-depth is calculated by the function of predicted residual features and sampling rate (SR). Experimental results demonstrate our proposed AOBE model has superior RD performance compared with the fixed bit-depth, while maintains almost the same complexity at sampling side.

Index Terms— compressed video sensing, encoding, non-uniform quantization, bit-depth estimation

1. INTRODUCTION

Compressed Video Sensing (CVS)[1] is the application of the theory and principles of Compressed Sensing (CS)[2] to video coding, which performs the sampling and compressing procedures simultaneously. The very simple sampling operation and efficient compression make CVS competitive in resource-limited environments, e.g., video surveillance systems and wireless multimedia sensor networks, etc. Much attention has been devoted to CVS reconstruction algorithms[1, 3, 4], while there are few researches on the quantization of CS measurements, especially for the optimization of the rate-distortion (RD) performance.

Uniform scalar quantization (SQ) to each CS measurement is straightforward, but proved inefficient for ignoring the characteristic of the CS measurements. Liangjun Wang *et al.* proposed progressive quantization[5] for exploiting hidden correlations between CS measurements in compressed image sensing (CIS), but it is difficult to extend it to CVS. Sungkwang Mun *et al.* proposed a quantization method by combining uniform SQ and Differential Pulse-Code Modulation (DPCM-plus-SQ)[6] for block-based CIS. And based on it, Jian Zhang *et al.* proposed the spatially directional predictive coding (SDPC)[7] scheme. However, the complexity at encoding side is increased, which violates the original intention of CS. To take advantage of temporal correlation for CVS at encoder, frame-based DPCM quantization[8] is proposed. The measurement residuals between adjacent frames instead of the CS measurements are quantified and transmitted to the decoder. Frame-based DPCM quantization provides a superior RD performance for CVS.

Uniform SQ is utilized in all the quantization algorithms above, but usually the CS measurements or the measurement residuals is not uniformly distributed. Literature [9] proposed non-uniform SQ (NSQ) incorporating with DPCM for block-based CIS, which effectively improved the quality of reconstructed images. Nevertheless, how to extend NSQ to CVS has not been studied in the literatures. On the other hand, all methods above use fixed bit-depth for the measurements or measurement residuals at different bit-rate for all sequences. However, in consideration of the fact that the values of the residual are small for slow moving sequences because of the high similarity between adjacent frames, while it is opposite for fast moving sequences. Therefore, the optimal quantization bit-depth should vary with the motion intensity of the sequence. In addition, at a low sampling rate (SR), in lack of measurements, high bit-depth fails to improve the reconstruction performance, and at a high SR, with more measurements, high bit-depth helps to reconstruct much more details. So the optimal quantization bit-depth should be considered at different SR and for different sequences. To our best knowledge, this issue was only discussed in [10], where the authors presented a sampling rate and bit-depth optimization method by minimizing the quantization distortion. However, the residual calculation method is not reasonable in [10], which brings a large cumulative error and results in poor video reconstruction performance.

In this paper, firstly, we propose a Non-uniform Scalar Quantization scheme based on frame-based DPCM (DPCM-NSQ) for the CVS measurements, which makes full use of the distribution features of the measurement residuals to reduce the quantization distortion. Secondly, an adaptive optimal bit-depth estimation model is proposed based on DPCM-NSQ. This model can adaptively obtain the optimal quantization bit-depth according to sampling rate and prediction residual characteristic.

This work was supported by the National Natural Science Foundation of China (grant 61471173) and Guangdong Natural Science Foundation (grant 2016A030313455).

2. BACKGROUND

In Compressed Sensing, the original signal $x \in R^N$ is sampled by a linear projection operator $y = \Phi x$, where y is the CS measurements and Φ is an $M \times N$ ($M << N$) measurement matrix. According to CS theory, if x is sparse, it can be exactly reconstructed using the low dimensional measurements at decoder.

2.1 Block-based CS (BCS)

BCS[11] is widely used in CIS for its efficiency in image capturing and reconstruction. In BCS, an image is divided into $B \times B$ non-overlapping blocks and sampled by a measurement matrix:

$$y^{(j)} = \Phi_B x^{(j)} \tag{1}$$

Where, $x^{(j)}$ is a vector representation for block j of the input image x, $y^{(j)}$ is the corresponding measurements, and Φ_B is a $M_B \times B^2$ measurement matrix such that the sampling rate for the whole image is $SR = M_B / B^2$.

Due to the superiority of BCS, it is extensively used in CVS. Videos are divided into group of pictures (GOP), and each GOP consists of a key frame and several non-key frames. Then all frames are sampled by BCS separately, with high sampling rate (SR_K) for key frames and low rate (SR_{NK}) for non-key frames ($SR_K > SR_{NK}$).

2.2 Frame-based DPCM Quantization

In CVS, frame-based DPCM quantization[8] is helpful to reduce bit-rate, where the measurement residual between two frames instead of the original measurement is quantized. The measurement residual is obtained as follows:

$$d_t = y_t - y_{t-1} \tag{2}$$

Where, d_t is the prediction residual, y_t is the measurements of the current frame and y_{t-1} is the measurements of the reference frame. Note that if the reference frame y_{t-1} is key frames, the dimension of y_{t-1} is greater than y_t, so the prediction of y_t is a subset of y_{t-1}. Eq. (2) is rewritten as:

$$d_t = y_t - y_{t-1}(1 : M_{NK}) \tag{3}$$

Here, M_{NK} is the number of observations of non-key frames.

Since the value of prediction residuals d_t is usually smaller than the original measurements, frame-based DPCM used in CVS can reduce the quantization bitrate.

The measurement residuals can be quantized by uniform SQ or non-uniform SQ (NSQ). Due to the non-uniformity of the residual distribution, NSQ can better reduce quantization distortion than uniform SQ. Therefore, the DPCM-based NSQ scheme (DPCM-NSQ) is proposed in this paper, which will be described in the next section.

3. DPCM-NSQ SCHEME

When Gauss Random Measurement Matrix (GRMM) is used in CS, the measurements obey Gaussian distribution[12], so do the prediction residuals. Fig. 1 shows the residual distribution histogram of the 2nd frame of several videos. The residual obeys Gaussian distribution, obviously. This distribution is consistent with the NSQ characteristics. Hence the idea of using NSQ to quantify the residual of CS measurements for CVS is proposed in this paper.

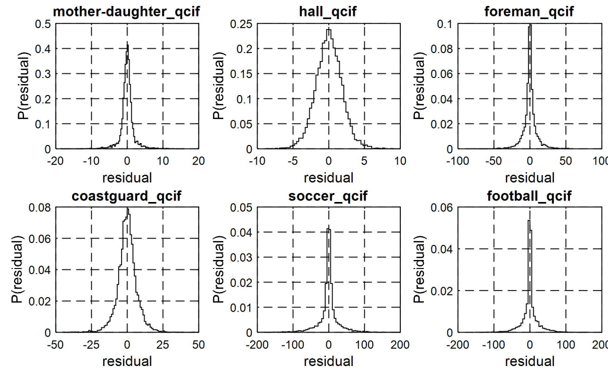

Fig. 1 Histogram of the residual for the 2nd frame.

Non-uniform quantization is usually implemented by a non-linear companding function followed by a uniform SQ process. DPCM-NSQ scheme for CVS is shown in Fig. 2.

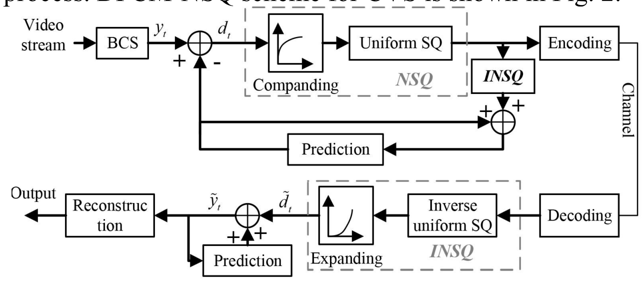

Fig. 2 DPCM-NSQ scheme for CVS. *NSQ* is non-uniform SQ; *INSQ* is inverse non-uniform SQ.

4. ADAPTIVE BIT-DEPTH ESTIMATION MODEL

The optimal quantization bit-depth for each frame based on DPCM-NSQ is related to the characteristic of the prediction residual and SR, which is also mentioned in [10]. To further improve RD performance of DPCM-NSQ, this paper proposes an adaptive optimal bit-depth estimation model. In the following subsection, we first give the measurement method of the residual characteristic, and then describe the optimal bit-depth model based on residual characteristic and SR detailedly.

4.1 The Measure of Residual Characteristic

The values of the prediction residual are usually small for slow moving sequences, and is opposite for fast moving sequences. From Fig. 1, hall and mother-daughter are slow moving sequences, the residual values are small. However, for the fast moving sequence of football, the residual values are extremely large. In addition, the residual roughly has the same distribution at different sampling rates for each sequence, and it follows Gaussian distribution. Therefore, the residual characteristic can be measured by:

$$Sc = (\mu + 2.58\sigma) - (\mu - 2.58\sigma) \tag{4}$$

Where, μ is the mean and σ is the standard deviation of the residual respectively. Generally, large Sc means fast moving sequences, vice versa.

4.2 Optimal Quantization Bit-depth Model

In order to obtain the optimal quantization bit-depth model, many simulations for a lot of sequences at different SR are implemented. Fig. 3 presents the scatterplot about *B* and *SR* when the best reconstruction quality (PSNR) is obtained at several sampling rates for several sequences. The optimal bit-depth almost increases linearly with the increasing of SR for each sequence. In addition, the best bit-depth varies for different sequences at the same SR. The prediction residual of fast motion sequence needs large bit-depth, and that of the slow motion sequence needs small bit-depth. As we know from section 4.1, different motion sequences have different residual characteristic Sc. So the optimal bit-depth for different sequence at each SR can be expressed by the function of residual characteristic Sc. Therefore, the adaptive optimal bit-depth model based on the residuals' characteristic (*Sc*) and *SR* for each video frame is proposed, and it is expressed as:

$$B = f(Sc) + g(SR) \tag{5}$$

Where, B represents the optimal bit-depth for each frame.

In the following, we will analyze the relationship $B_{SR}=g(SR)$ between B_{SR} and *SR*, the relationship $B_{Sc}=f(Sc)$ between B_{Sc} and *Sc*, and finally attain the expression of *B* about *SR* and *Sc*.

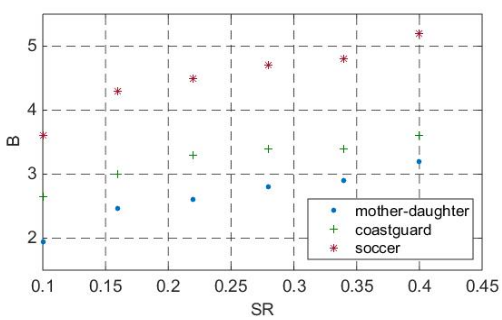

Fig. 3 The scatterplot of *B* and *SR* at best reconstruction quality for several video sequences (QCIF).

A. Relationship between B_{SR} and *SR*

In the simulation, we found that when SR is small, the reconstruction quality would be quite poor even if the bit-depth is increased, for the reason that insufficient number of measurements restricts the reconstruction performance. However, when the number of measurements increases, the reconstruction performance largely depends on the accuracy of measurements, higher bit-depth bringing better reconstruction performance. As can be seen from Fig. 3, there appears to be a roughly linear relationship between B and SR. So the relationship between B_{SR} and SR is supposed as:

$$B_{SR} = k \cdot SR + d \tag{6}$$

Where, k and d are the model parameters.

B. Relationship between B_{Sc} and *Sc*

According to the characteristic of Gaussian distribution, 99% of the residual fall into the range of Sc. Suppose the corresponding bit-depth is B_{Sc}, the quantization step size can be expressed as:

$$\Delta = Sc / 2^{B_{Sc}} \tag{7}$$

Eq. (7) can be rewritten as $2^{B_{Sc}} = (1/\Delta)\cdot Sc$. If the quantization step size for each frame of different video sequences remains unchanged, the relationship between $2^{B_{Sc}}$ and Sc should be linear, which is expressed as:

$$2^{B_{Sc}} = a \cdot Sc + b \tag{8}$$

Here, b is a positive number greater than 1, ensuring that B_{Sc} is not less than zero.

Eq. (8) can be rewritten as:

$$B_{Sc} = \log_2\left(a \cdot Sc + b\right) \tag{9}$$

Where, a and b are the model parameters.

C. The adaptive optimal bit-depth model

From what has been discussed above, substitute Eq. (6) and (9) into Eq. (5), and the optimal bit-depth is expressed as:

$$B = f(Sc) + g(SR) = \log_2\left(a \cdot Sc + b\right) + k \cdot SR + d \tag{10}$$

The model parameters can be predicted using the (*B*,*SR,Sc*) groups at the best reconstruction quality by Levenberg–Marquardt algorithm[13].

Through a lot of experiments (including mother-daughter, suzie, coastguard, mobile, soccer and football), several groups of (*B*,*SR,Sc*) are gotten, he scatterplot about the optimal bit-depth *B* and (*SR,Sc*) from the experiment is shown in Fig. 4. Then Levenberg–Marquardt algorithm is used to obtain the best parameters in the proposed model (10), (*a*, *b*, *k*, *d*)=(0.054, 2.223, 4.24, 0.16). The adaptive optimal bit-depth model is expressed as:

$$B = log_2\left(0.054Sc + 2.223\right) + 4.24SR + 0.16 \tag{11}$$

Actually, 2^B must be an integer. So the practical quantization model between *B* and (*SR,Sc*) is gotten:

$$B = \log_2\left\lceil 2^{\log_2(0.054Sc+2.223)+4.24SR+0.16} \right\rceil \tag{12}$$

The fitting curve obtained by the optimal bit-depth model Eq. (12) is also shown in Fig. 4. It is obvious that our model provides a reasonable approximation of the *B*-(*SR,Sc*) relationship.

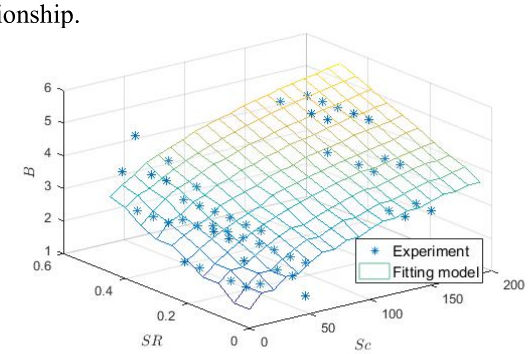

Fig. 4 Comparing the predicted (12) with experiment data.

5. EXPERIMENTAL RESULTS

5.1 Simulation Conditions

In the simulations, $A-law$ companding function is used in DPCM-NSQ scheme.

$$y=\begin{cases}\dfrac{A|r|}{1+\log(A)}\bullet\operatorname{sgn}(r), & 0\le|r|\le\dfrac{1}{A}\\[2ex] \dfrac{1+\log(A|r|)}{1+\log(A)}\bullet\operatorname{sgn}(r), & \dfrac{1}{A}\le|r|\le 1\end{cases}\tag{13}$$

Where, r represents the normalized residual, A is the parameter to control companding ratio and $sgn(\cdot)$ is a sign function.

The value of A has a great effect on the restoration precision of measurements. According to the characteristic of Gaussian distribution, 68% of the small residual values are included in the range of $|r\text{-}\mu|\le\sigma$ (μ usually equals to 0). Therefore, we propose the adaptive A as:

$$1/A=\sigma\Rightarrow A=1/\sigma\tag{14}$$

Different video sequences are used to test the proposed optimal bit-depth model, including suzie, football, foreman, and stefan (QCIF, 15Hz). The first 88 frames of each video sequence are used, and GOP size is 8. Each frame is divided into non-overlap blocks of size 16×16 and sampled by GRMM. Two-stage multi-hypothesis reconstruction algorithm (2sMHR)[14] is exploited to recover videos from the limited measurements. PSNR (dB) is used to evaluate the reconstruction quality.

5.2 Performance Analysis of the Fitting Model

The simulation results are shown in Fig. 5 and Fig. 6, AOBE means DPCM-NSQ combined with the proposed adaptive optimal bit-depth estimation model, 8-NSQ expresses our proposed DPCM-NSQ with 8-bits, and Ideal-NSQ is the ideal results by picking out the best performance from plenty pairs of (B,SR) under DPCM-NSQ. The results are also compared with the traditional video coding scheme, H.264[15]. Note that Ideal-NSQ uses the same bit-depth for the whole video sequence. In the AOBE model, bit-depth is estimated by Eq. (12) for each frame. So each sequence's bitstream is composed of three parts: the NSQ companding ratio control parameter A, the quantization bit-depth B, and the quantized prediction residual. Assuming that 16 bits are used to quantifying A and B, respectively, the impact on the bitstream for a 15Hz QCIF video sequence is $(16+16)\times15/1000=0.48\,kbps$, which is almost negligible.

From Fig. 5 and Fig. 6, it is easily observed that our adaptive optimal bit-depth estimation (AOBE) model has a great improvement compared with 8-NSQ and is close to Ideal-NSQ, owing to the adaptive estimation for the DPCM measurement residuals with different features at different sampling rates. H.264 has the best RD performance, which is at the expense of increasing the complexity of the encoding side.

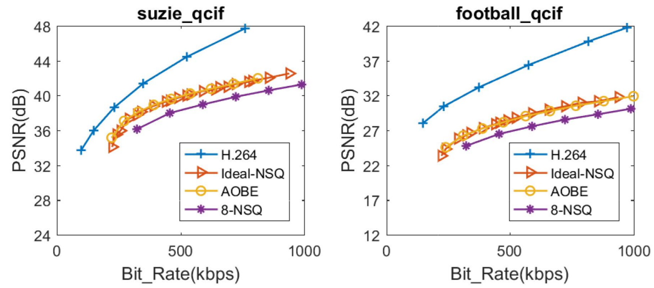

Fig. 5 RD performance for training set sequences.

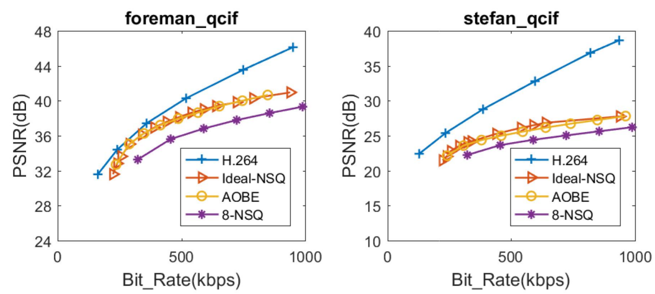

Fig. 6 RD performance for testing set sequences.

6. CONCLUSION

In this paper, a non-uniform quantization scheme based on frame-based DPCM (DPCM-NSQ) is proposed to quantify the CVS measurements, in which an adaptive optimal bit-depth estimation model is proposed to further improve rate-distortion performance. The relationship between the bit-depth and the sampling rate, and the relationship between the bit-depth and the characteristic of the prediction residual, are both fully considered in the proposed quantization algorithm. Experimental results demonstrate that the proposed quantization method has a significant improvement in RD performance compared with the other quantization methods, and the gain is even up to 2dB for several sequences when compared with the 8-NSQ at the same bit-rate.

7. REFERENCES

[1] L. Ying, L. Ming, and D. A. Pados, “Motion-Aware Decoding of Compressed-Sensed Video,” *Circuits Syst. Video Technol. IEEE Trans.*, vol. 23, no. 3, pp. 438–444, 2013.

[2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] Y. Kuo, K. Wu, and J. Chen, “A scheme for distributed compressed video sensing based on hypothesis set optimization techniques,” *Multidimens. Syst. Signal Process.*, 2015.

[4] X. Gao, F. Jiang, S. Liu, W. Che, X. Fan, and D. Zhao, “Hierarchical frame based spatial–temporal recovery for video compressive sensing coding,” *Neurocomputing*, vol. 174, pp. 404–412, 2016.

[5] L. . Wang, X. . Wu, and G. . Shi, “Progressive quantization of compressive sensing measurements,” *Data Compression Conf. Proc.*, pp. 233–242, 2011.

[6] S. Mun and J. E. Fowler, “DPCM for quantized block-based compressed sensing of images,” *Eur. Signal Process. Conf.*, no. Eusipco, pp. 1424–1428, 2012.

[7] J. Zhang, D. Zhao, and F. Jiang, “Spatially directional predictive coding for block-based compressive sensing of natural images,” *2013 IEEE Int. Conf. Image Process. ICIP 2013 - Proc.*, pp. 1021–1025, 2013.

[8] Y. Chun-ling and O. Wei-feng, “Multi-Reference Frames-Based Optimal Multi-Hypothesis Prediction Algorithm for Compressed Video Sending,” *J. South China Univ. Technol. Sci. Ed.*, pp. 1–6, 2016.

[9] C. Qian, B. Zheng, and B. Lin, “Nonuniform quantization for block-based compressed sensing of images in differential pulse-code modulation framework,” in *The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014)*, 2014, no. 61271240, pp. 791–765.

[10] H. Liu, B. Song, F. Tian, and H. Qin, “Joint sampling rate and bit-depth optimization in compressive video sampling,” *IEEE Trans. Multimed.*, vol. 16, no. 6, pp. 1548–1562, 2014.

[11] S. Mun and J. E. . Fowler, “Block Compressed Sensing of Images Using Directional Transforms,” *Int. Conf. Image Process.*, no. November, pp. 3021–3024, 2009.

[12] Y. Baig, E. M. K. Lai, and J. P. Lewis, “Quantization effects on Compressed Sensing Video,” *ICT 2010 2010 17th Int. Conf. Telecommun.*, pp. 935–940, 2010.

[13] J. J. Moré, “The Levenberg-Marquardt algorithm: implementation and theory,” in *Numerical analysis*, Springer, 1978, pp. 105–116.

[14] W. Ou, C. Yang, W. Li, and L. H. Ma, “A two-stage multi-hypothesis reconstruction scheme in compressed video sensing,” in *IEEE international conference on image processing*, 2016.

[15] K. Sühring, “H.264/AVC Software JM 19.0.” http://iphome.hhi.de/suehring/tml/, 2015.