ARTIFACT REDUCTION OF COMPRESSED VIDEO VIA THREE-DIMENSIONAL ADAPTIVE ESTIMATION OF TRANSFORM COEFFICIENTS

Xinfeng Zhang[1,2], *Ruiqin Xiong*[2], *Siwei Ma*[2] *and Wen Gao*[2]

[1]Key Lab of Intelligent Information Processing, Institute of Computing technology, Chinese Academy of Sciences, Beijing 100190, China
[2]Institute of Digital Media, Peking University, Beijing 100871, China
xfzhang@jdl.ac.cn, {rqxiong, swma, wgao}@pku.edu.cn

ABSTRACT

Block transform compressed videos usually suffer from annoying artifacts at low bit rates, caused by the coarse quantization of transform coefficients. The inter prediction utilized in video coding also induces block boundary artifacts when the neighboring blocks using different motion vectors from previous decoded frames. In this paper, a three-dimensional adaptive estimation method for video transform coefficients is proposed to reduce the artifact in compressed video. In the proposed method, transform coefficients of each block are estimated by adaptively fusing three prediction sources based on their reliabilities. One prediction source is the transform coefficients directly acquired from the decoded video, whose reliability is determined by the distribution of the quantization noise. The other prediction source is derived by a temporal autoregressive model of transform-blocks along motion trajectory among neighboring reference frames. Its reliability is estimated from prediction variance of local blocks. The last prediction source is derived from the nonlocal transform-blocks, whose reliability is estimated based on the distribution of the nonlocal coefficients and their similarity with the estimated block. Experimental results for compressed video sequences by HEVC show that, the proposed method can reduce the compression artifacts and improve both the objective and subjective quality.

Index Terms—post-processing, denoising, compression artifacts.

1. INTRODUCTION

Block discrete cosine transform (DCT) and intra/inter prediction technics are widely adopted in the existing video compression standards, e.g. MPEG-1/2/4, H.261/263/264, HEVC, to reduce the redundancy in video signals. In a typical video coding scheme, each frame is firstly divided into non-overlapped blocks, predicted from coded blocks in current frame or previously coded frames via motion estimation, transformed using DCT, quantized and entropy coded individually for each block. Due to coarse quantization of the transform coefficients, the decoded video images generally suffer from visually annoying artifacts at low bit rate. In addition, the motion compensated prediction also induces block boundary artifact, because the predicted blocks are generated by copying interpolated pixel data from different locations of possibly different reference frames [1].

In order to reduce the compression artifacts while maintaining compatibility with the existing coding standards, various post-processing techniques have been proposed in the literatures. Buades et al. [2] proposed the nonlocal means filter to predict each pixel by a weighted average of its surrounding pixels according to the similarity of image patches where the source and target pixel located. But this method does not take advantage of the compression information and it is difficult to determine the filtering strength. Zhai et al. [3] reduce the JPEG compression artifact by average the estimated block and its similar blocks, which are selected according to compression quality factors. In the works [4] and [5], the nonlocal estimation and the decoded value are adaptively fused according to their reliability to reduce compression artifacts further. Besides exploring the spatial correlation in the above methods, Dugad and Ahuja [6] design a spatial-temporal filter by combining a 1-D temporal Kalman filter and a 2-D spatial Wiener filter to remove compression noise while preserving important edges. However, the reliabilities of the two filters are difficult to decide.

In this paper, we propose a new approach to reduce compression artifacts by estimating the video image from transform domain. This is achieved by adaptively fusing estimated transform coefficients from three prediction sources based on their reliabilities, respectively. One prediction source is acquired by directly transforming decoded image blocks, whose reliability is determined by the variance of quantization error. The second prediction is acquired by representing the transform-blocks along motion trajectory with an autoregressive (AR) model and assuming that the local transform-blocks have similar model parameters. The prediction reliability is estimated based on the variance of prediction error for local blocks. The last

prediction utilizes nonlocal transform-blocks to infer the original coefficients, and these blocks are discriminatively utilized according to their similarity with the estimated block. The reliability is estimated according to the distribution of nonlocal transform-block coefficients and the quantization noise. Finally, we take the quantization steps to constrain the estimated coefficients to avoid over-smoothing.

The remainder of this paper is organized as follows. In section 2, we first review the basic scheme of video coding and then formulate the compression artifact reduction framework. Section 3 introduces the coefficient estimation method from three different prediction sources according to their reliabilities, respectively. Section 4 provides parameter estimation. Experimental results are reported in Section 5 and Section 6 concludes the paper.

2. THREE-DIMENSIONAL ADAPTIVE ESTIMATION OF TRANSFORM COEFFICIETS

In this section, we firstly briefly review a few concepts and notations relevant to hybrid video coding for convenience of later discussion, and then introduce the proposed three-dimensional adaptive estimation method for transformation coefficients.

2.1. Review of hybrid video coding

Suppose we have a video image I (a two dimensional grid) of size HxW, where $I(i,j,t)$ denotes a pixel and the induces i, j and t are the coordinates in the vertical, horizontal and temporal directions, respectively. We use $\mathrm{B}_{m,n}(i,j,t)$ to denote an image block of size NxN in I, with its top left pixel being $I(m,n,t)$. We use $\mathbf{X}_B$ to represent the transform coefficients of block B. The data in each block is first predicted from neighboring coded blocks in current image or previous coded images and the prediction residuals are transformed, quantized and entropy coded into the compressed bitstream.

The decoded video images are reconstructed by inverse transform and quantization. The reconstructed coefficients are,

$$\mathbf{Y}_B(u,v) = round\left(\frac{\mathbf{X}_B(u,v)}{Q(u,v)}\right) \square Q(u,v) \tag{1}$$

where $Q(u,v)$ is the quantization step. The original transform coefficients are constrained in $[\mathbf{Y}^{\min}(u,v),\ \mathbf{Y}^{\max}(u,v)]$. Therefore, in a few deblocking schemes, a projection onto convex set operation, $\mathbf{X}'=P_Q(\mathbf{X},\mathbf{Y})$, is defined to enforce the estimated coefficients,

$$\mathbf{X}'(u,v) = \begin{cases} \mathbf{Y}^{\min}(u,v), & if\ \mathbf{X}(u,v)<\mathbf{Y}^{\min}(u,v) \\ \mathbf{X}(u,v), & if\ \mathbf{Y}^{\min}(u,v)\le\mathbf{X}(u,v)\le\mathbf{Y}^{\max}(u,v) \\ \mathbf{Y}^{\max}(u,v), & if\ \mathbf{X}(u,v)>\mathbf{Y}^{\max}(u,v) \end{cases} \tag{2}$$

2.2. The framework of the three-dimensional adaptive estimation for transform coefficients

In a standard decoder, the coded image is reconstructed simply by inversely transforming the quantized coefficients for each coding block. To tackle the coding artifacts generated from coarse quantization and motion compensation, besides the reconstructed blocks from the decoded image, we also take advantage of the similarity of nonlocal blocks and the video signal continuity along motion trajectory in temporal domain. By introducing three distance metrics D_1, D_2 and D_3, the proposed compression artifact reduction method is formulate as the following optimization problem,

$$\begin{aligned} \underset{\mathrm{x}}{\operatorname{argmin}} \sum_{B\in\Omega} & D_1\left(\mathrm{X}_B,\mathrm{Y}_B\right) + D_2\left(\mathrm{X}_B,\{\mathrm{Y}_{B(t)}\}_{B(t)\in MC(B)}\right) \\ & + D_3\left(\mathrm{X}_B,\{\mathrm{Y}_{B'}\}_{B'\in N(B)}\right) \end{aligned}, \tag{3}$$

subject to the quantization constraint:

$$\mathrm{X}_B(u,v) \in \left[\mathrm{Y}_B^{\min}(u,v), \mathrm{Y}_B^{\max}(u,v)\right]. \tag{4}$$

Here, the $MC(B)$ is the block set in reference frames along motion trajectory, and the $N(B)$ is the nonlocal block set used to predict the target transform-block.

In Eq. (3), the first term D_1 measures the distance between the estimated coefficients and the reconstructed coefficients from decoded image by inverse transform and quantization, which can be regarded as the data fidelity. The second and third terms measure the conformance of the estimated coefficients with the temporal and nonlocal similarity models.

3. THE PROPOSED METRICS AND RELIABILITIES FOR COEFFICIENT ESTIMATION

According to the discussion in Section 2, we choose the following distance metric for the data fidelity,

$$D_1\left(\mathrm{X}_B,\mathrm{Y}_B\right) = W_1\left\|\mathrm{X}_B - \mathrm{Y}_B\right\|_2^2, \tag{5}$$

$$W_1 \propto \frac{1}{\sigma_Q^2}. \tag{6}$$

Here W_1 is the weight inversely proportional to the variance of quantization error, σ_Q^2, reflecting the estimation reliability when predicting X_B from Y_B.

For the second distance metric, we take the temporal autoregressive model to estimate the target transform-block, assuming the continuity of video signal along motion trajectory in temporal domain.

$$D_2\left(\mathrm{X}_B,\{\mathrm{Y}_{B(t)}\}\right) = W_2\left\|\mathrm{X}_B - \sum_{i=-2,i\neq 0}^{2} \alpha_i \mathrm{Y}_{B_{t+i}}\right\|_2^2,\ B_{t+i}\in MC\left(B_t\right) \tag{7}$$

The derivation of AR parameters can be formulated into the following Least-Square (LS) problem,

$$\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\left\|\mathrm{Y}_{B'} - \sum_{i=-2,i\neq 0}^{2} \alpha_i \mathrm{Y}_{B'_{t+i}}\right\|_2^2,\ B'_{t+i}\in MC\left(B_t\right), B'\in L\left(B_t\right) \tag{8}$$

where $L(B_t)$ is the local neighborhood centered at block B_t. The prediction relationship is illustrated in Fig.1. The weight

W_2 in Eq. (7) should reflect the reliability of the prediction from temporal AR model and is inversely proportional to the variance of model error.

$$W_2 \propto \frac{1}{\sigma_{AR}^2} \tag{9}$$

$$E_{B'} = \mathrm{Y}_{B'_t} - \sum_{i=-2, i\neq 0}^{2} \alpha_i \mathrm{Y}_{B'_{t+i}} \tag{10}$$

$$M_{B_t} = \frac{1}{|L(B_t)|} \sum_{B' \in L(B_t)} E_{B'} \tag{11}$$

$$\sigma_{AR}^2 = \frac{1}{|L(B)|} \sum_{B' \in L(B)} \left(\mathrm{Y}_{B'}^t - M_{B'}\right)^2 \tag{12}$$

where $|L(B)|$ is the number of the blocks in this local neighborhood.

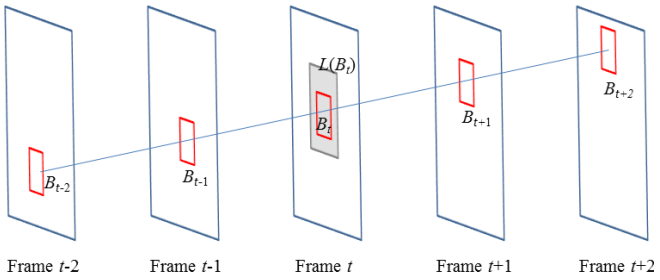


Fig. 1. The temporal AR prediction for target transform-block.

For the third distance metric, we take the three-dimensional nonlocal transform-blocks to infer the original coefficients. Due to image structure variation, different transform-blocks play different roles in the prediction process. Therefore, we use the weighted average of the coefficients in nonlocal blocks as the expectation of the original coefficient distribution and employ the variance of nonlocal coefficients to reflect the reliability of the prediction. Based on the above discuss, the third distance metric can be formulated as,

$$D_3\left(\mathrm{X}_B, \{\mathrm{Y}_{B'}\}\right) = W_3 \left\| \mathrm{X}_B - \sum_{B' \in N(B)} w_{B'} \mathrm{Y}_{B'} \right\|_2^2, \; B' \in N(B) \tag{13}$$

The weight W_3 is inversely proportional to the variance of nonlocal coefficients,

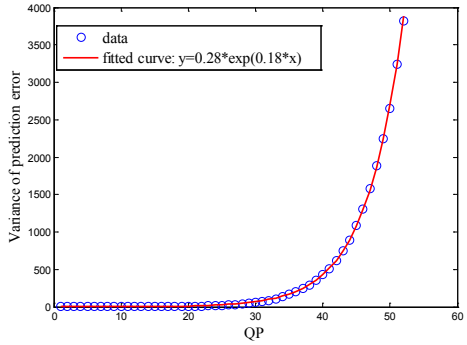
$$W_3 \propto \frac{1}{\sigma_{N(B)}^2}, \quad \sigma_{N(B)}^2 = \sum_{B' \in N(B)} w_{B'} \left(\mathrm{Y}_{B'} - \sum_{B' \in N(B)} w_{B'} \mathrm{Y}_{B'} \right)^2 \tag{14}$$

The weights used to differentiate nonlocal blocks should efficiently depress the negative effect of dissimilar blocks in the prediction process. Therefore, we take the difference of nonlocal blocks and estimated block to reflect their similarity and employ an exponent function to model the weight distribution of the nonlocal blocks.

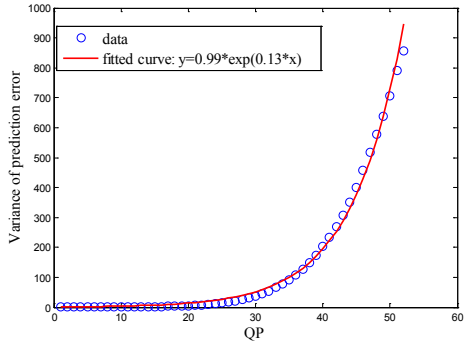
$$w_B = \frac{1}{Z} exp\left(- \frac{\left\| \mathrm{Y}_B - \mathrm{Y}_{B'} \right\|_2}{h} \right), \; Z = \sum_{B' \in N(B)} exp\left(- \frac{\left\| \mathrm{Y}_B - \mathrm{Y}_{B'} \right\|_2}{h} \right) \tag{15}$$

4. PARAMETER ESTIMATION

In order to solve the problem in Eq. (3), the variance of quantization error needs to be estimated from compression stream. We take the quantization parameters (QP) acquired from stream to estimate the error variance. Firstly, we use the latest coding standards, HEVC to compress video sequence with different QPs and then fit the relationship between the variance of quantization error and QP illustrated in Fig. 2. The size of transform-block is 8x8. We can see that exponent function can well fit their relationship. Therefore, we derive the variance of the quantization error for each band from QP based on this function.



(a) band (0,0)



(b) band (1,1)

Fig. 2. The relationship between the variance of quantization error and QP.

For the other two variances, we add the variance of quantization error to them considering that the AR parameters and the weights of nonlocal blocks are calculated according to the noisy blocks. Then, they are rewritten as,

$$\sigma_{AR}^2 = \frac{1}{|L(B)|} \sum_{B' \in L(B)} \left(\mathrm{Y}_{B'} - E_{B'}\right)^2 + \sigma_Q^2 \tag{16}$$

$$\sigma_{N(B)}^2 = \sum_{B' \in N(B)} w_{B'} \left(\mathrm{Y}_{B'} - \sum_{B' \in N(B)} w_{B'} \mathrm{Y}_{B'} \right)^2 + \sigma_Q^2 \tag{17}$$

5. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method. The common test video sequence with resolution 352x288 and 416x240 as listed in Table 1 and 2. These sequences are compressed using intra prediction by HEVC with and without in-loop filter (denoted as HEVC w/ and

HEVC w/o), respectively. We compare our method with the well-known nonlocal means filter (NLM) [3] and the in-loop filter in HEVC (including deblocking filter [8] and sample adaptive offset (SAO) [9]). The PSNR results are summarized in Table 1 and 2. From the tables, we can see that our proposed method obviously outperforms the other compared methods, up to 1.1 dB gain over HEVC w/o for *Mobile* at QP=37. The NLM filter cannot well adapt to different compression video artifacts even if the parameters are selected according to its performance. When the nonlocal similar blocks are not enough for some image blocks, the NLM is not efficient. The proposed method also outperforms the other two in-loop filter, deblocking filter and SAO, achieving about 0.47dB and 0.26dB gain.

In Fig.3, we show the subjective quality of reconstructed images. We can see that lots of compression artifacts exist in the video image acquired directly from standard decoder in Fig.3 (a), e.g. blocking artifact on the face. The proposed method can significantly reduce these compression artifacts.

Table 1: the average PSNR results of different methods, video compressed by HEVC intra-coding, QP=37

Sequences	HEVC w/o	HEVC w/	NLM	Proposed
BlowingBubbles	30.10	30.24	30.10	30.67
BQSquare	29.44	29.59	28.02	30.47
Basketballpass	32.18	32.37	32.18	32.70
RaceHorses	30.49	30.66	30.45	30.93
Foreman	32.70	32.96	32.93	33.27
Bus	29.53	29.65	29.13	30.19
Footabll	29.80	29.96	29.51	30.25
Vectra	32.51	32.79	32.69	33.08
Carphone	33.21	33.56	33.50	33.93
Mobile	28.34	28.45	27.05	29.44
Average	30.83	31.02	30.56	31.49

Table 2: the average PSNR results of different methods, video compressed by HEVC intra-coding, QP=42

Sequences	HEVC w/o	HEVC w/	NLM	Proposed
BlowingBubbles	27.21	27.36	27.37	27.51
BQSquare	25.71	25.86	25.48	26.37
Basketballpass	29.40	29.55	29.50	29.68
RaceHorses	27.76	27.96	27.93	28.11
Foreman	29.88	30.19	30.23	30.46
Bus	26.21	26.31	26.25	26.54
Footabll	26.83	27.00	26.95	27.23
Vectra	29.35	29.63	29.63	29.76
Carphone	30.20	30.46	30.55	30.68
Mobile	24.49	24.60	24.32	25.13
Average	27.70	27.89	27.82	28.15

6. CONCLUSIONS

In this paper, we propose a new transform-domain approach for compression artifact reduction. In the proposed scheme, a compressed video image is restored by adaptive estimation of DCT coefficients from three prediction sources. The prediction sources include the decoded image directly from compression stream, the AR prediction based on the temporal continuity of video signal and the nonlocal similarity of image prior knowledge. In additional, we combine these predictions accord to their reliabilities which are derived from compression quantization parameters and the statistical characteristic of the prediction source. Experimental results demonstrated that the proposed approach can reduce the compression artifacts and improve the quality of compressed video.

Fig. 3. The reconstructed video image, *Foreman* compressed with QP=37, (a) HEVC decoded image without in-loop filter, (b) HEVC decoded image with in-loop filter, (c) image generated from NLM and (d) image generated from the proposed method

7. ACKNOWLEDGEMENT

This work was supported in part by National High-tech R&D program of China (863 Program, 2012AA010805), National Science Foundation of China (61322106, 61370114, 61121002), and Beijing Natural Science Foundation (4132039).

8. REFERENCES

[1] P. List, A. Joch, J. Lainema, G. Bjøntegaard and Marta Karczewicz, "Adaptive Deblocking Filter," IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, Jul. 2003.

[2] A. Buades, B. Coll , J.-M. Morel, "A Non-Local Algorithm for Image Denoising," IEEE international conference on Computer Vision Pattern Recognition, vol. 2, pp.60-65, Jun. 2005.

[3] G. Zhai, W. Zhang, X. Yang, W. Lin and Y. Xu, "Efficient Image Deblocking Based on Postfiltering in Shifted Windows," IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, no.1, pp.122-126, Jan. 2008.

[4] X. Zhang, R. Xiong, S. Ma and W. Gao, "Reducing Blocking Artifacts in Compressed Images via Transform-Domain Non-local

Coefficients Estimation," IEEE International Cnference on Multimedia and Expo (ICME), pp.836-841, Jul. 2012.
[5] X. Zhang, R. Xiong, X. Fan, S. Ma and W. Gao, "Compression Artifact Reduction by Overlapped-Block Transform Coefficient Estimation With Block Similarity," IEEE Transactions on Image Processing, vol.22, no.12, pp.4613,4626, Dec. 2013.
[6] R. Dugad and N. Ahuja, "Video denoising by combining Kalman and wiener estimates," in Proc. IEEE Int. Conf. on Image Process., pp.152–156, Oct.1999.
[7] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," IEEE Trans. on Image Processing, vol.16, no.8, pp. 2080-2095, Aug. 2007.
[8] A. Norkin, G. Bjøntegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson; M. Zhou, G. Van der Auwera, "HEVC Deblocking Filter," IEEE Transactions on Circuits and Systems for Video Technology, vol.22, no.12, pp.1746-1754, Dec. 2012.
[9] C.-M. Fu, C.-Y. Chen, Y.-W. Huang, and S. Lei, "Sample adaptive offset for HEVC," Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on , vol., no., pp.1,5, 17-19 Oct. 2011.