Transform-Domain In-Loop Filter with Block Similarity for HEVC

Xinfeng Zhang†, Weisi Lin†, Ke Gu‡, Qiaohong Li‡, Shanshe Wang* and Siwei Ma*

† *Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore*

‡ *School of Computer Science and Engineering, Nanyang Technological University, Singapore*

* *Institute of Digital Media & Cooperative Medianet Innovation Center, Peking University, Beijing, China*

†{xfzhang, wslin}, ‡ {guke, qli013}@ntu.edu.sg, *{swma, sswang, wgao}@pku.edu.cn

***Abstract*—In-loop filtering is an important technique in modern video coding standards. In this paper, we propose a transform-domain in-loop filter to further improve the compression performance of high efficiency video coding (HEVC) standard. The proposed method estimates block transform coefficients by adaptively fusing two prediction sources according to their uncertainties respectively. The first prediction is the block transform coefficients of compressed video frames, the uncertainty of which is related to quantization parameters. The second prediction is the weighted average of transform blocks in a neighborhood, and the weights are designed according to block similarity. Its uncertainty is estimated based on the coefficient variance. To optimize the filtering performance, the parameters utilized in the proposed in-loop filter are learned from compressed videos for each quantization parameter offline, and frame level flags are utilized to switch the proposed in-loop filter according to rate-distortion cost. Extensive experimental results show that the proposed in-loop filter can further improves the compression efficiency of HEVC.**

***Index Terms*—In-loop filter, HEVC, block similarity, transform coefficient estimation**

I. Introduction

In-loop filtering is an important technique in state-of-the-art video coding standards, e.g., deblocking filter in H.264/AVC and sample adaptive offset (SAO) in HEVC. On one hand, the in-loop filters can improve the quality of current coded frame directly by reducing compression artifacts, e.g., blocking artifacts and ringing artifacts. On the other hand, the filtered frames can provide more accurate inter-prediction for subsequent frames, which leads to bit-rate savings further.

H.264/AVC first adopted the deblocking filter (DF) [1] into the coding loop, and applied different low pass filters to 4×4 block boundaries according to coding information, e.g., prediction modes, motion vectors and reference frame indices. Besides DF, HEVC further adopted a nonlinear in-loop filter, Sample Adaptive Offset (SAO) [2], to reduce compression artifacts by adding different offset values to each sample, and these offset values are derived by minimizing reconstruction errors at the encoder and transmitted to the decoder side. Adaptive Loop Filter (ALF) [3], [4] was another widely discussed in-loop filter in HEVC development, which derives Wiener filters by minimizing distortions between original samples and decoded samples and needs transmit filter parameters to the decoder side. However, these in-loop filters only take advantage of image local smoothness prior, which may be inefficient for edge and texture regions.

The well known nonlocal means (NLM) filter [5] achieves good denoising performance, which takes weighted average of nonlocal image patches as the noise-free estimation. The weights are determined by the similarity of image patches located at the source and target coordinates. Matsumura et al. [6] introduced the NLM filter into HEVC by elaborately designing the filter strength, template size and control flags to further improve its performance in video coding. Zhang *et al.* [7], [8] also utilized image nonlocal similar patches to construct a low-rank matrix, and then applied soft-thresholding operation to the singular values of the matrix to reduce compression artifacts. However, nonlocal image prior model is also not always valid for different image content.

In this paper, we propose a novel in-loop filter by fusing two prediction sources in transform domain adaptively to further improve video coding performance. In the proposed method, the reconstructed frame from deblocking filter is transformed with block discrete cosine transform (BDCT), and the transform coefficients are utilized as one prediction source, the uncertainty of which is determined by quantization parameters (QP). Then, we take the weighted average of non-local transform blocks as another prediction, the uncertainty of which is measured by the variance of coefficients in each band. In order to improve filtering performance, the filter parameters are optimized based on compressed videos offline, and frame level flags for different color components are added in picture header syntax structure to switch the proposed filter according to the distortion changes of the filtered frames. The proposed transform-domain adaptive in-loop filter (TALF) is integrated into HEVC reference software, HM7.0, and extensive experimental results show that the proposed TALF further improves the compression performance of HEVC significantly.

The remainder of this paper is organized as follows. Section II introduces the proposed transform-domain adaptive in-loop filter in detail. Extensive experimental results are reported in Section III, and Section IV concludes the paper.

II. The transform-domain adaptive in-loop filter

The framework of the proposed transform-domain adaptive in-loop filter is illustrated in Fig.1. In the proposed method,

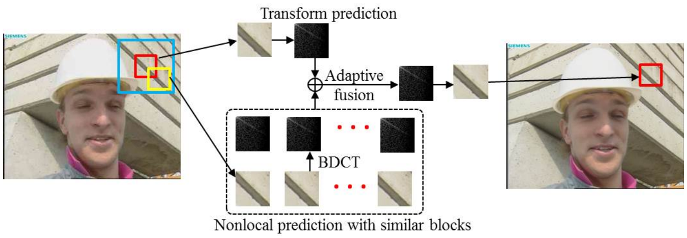
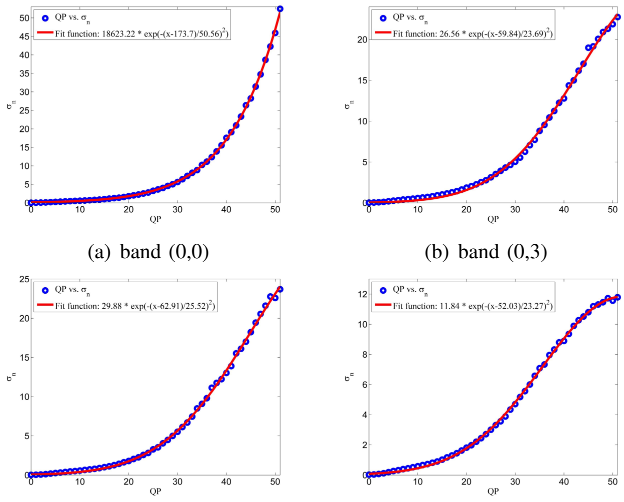


Fig. 1. The framework of the proposed transform-domain adaptive in-loop filter.

we first transform all the blocks in the frame, $\mathcal{I}_y$, reconstructed from deblocking filter with 8×8 BDCT. Then, we estimate the original DCT coefficients for every block by adaptively fusing two prediction sources, i.e., transform prediction and nonlocal prediction with similar transform blocks, according to their uncertainties. In order to improve the filtering performance, the processed blocks are overlapped, and the final reconstructed samples are weighted average of reconstructed samples from different block estimations. To ensure the filtering efficiency, frame level control flags are utilized to switch the filtering process for each color component, i.e., *TALF_Y_ON*, *TALF_U_ON*, *TALF_V_ON* corresponding to Y/U/V color components respectively. If the distortions of filtered color component decrease at the encoder side, the corresponding flag is signaled as *ture*, which indicates the decoder that the corresponding color component should be filtered. If the distortions of filtered color component increase at the encoder side, the corresponding flag is signaled as *false*, which indicates the decoder that the filtering process is skipped for the corresponding color component. For one frame, there are only three bits utilized to signal these frame level control flags.

A. Transform prediction

Although the reconstructed samples from deblocking filter contaminated by quantization noise, they are also an efficient prediction for original ones, the uncertainty of which is related with quantization noise levels. In this paper, we take the variance of quantization noise in each band, $\sigma_n^2(u,v)$, to measure the quantization noise level, and utilize its reciprocal to reflect the uncertainty of transform prediction. The larger the $\frac{1}{\sigma_n^2(u,v)}$ is, the uncertainty of transform prediction is less. Considering the quantization noise is directly related with quantization parameters (QP), we learn the relationship between standard deviation of quantization noise and QPs from three widely used WQVGA video sequences, i.e., *BasketballPass*, *RaceHorses* and *BlowingBubbles*, which are compressed by HEVC with different QPs. Fig.2 shows their relationship for different bands, which can be well approached by Gaussian function. Therefore, we estimate the standard deviations of quantization noise in each band of 8×8 transform blocks for different QPs and different frame types, i.e., I/P/B frames. Table I shows an example of the standard deviations of quantization noise, which are estimated for I frame compressed at QP=27. We can



(a) band (0,0)

(b) band (0,3)

(c) band (3,0)

(d) band (3,3)

Fig. 2. The relationship between standard deviation of quantization noise in 8×8 BDCT domain and QPs for intra coding frames

see that the standard deviations of quantization noise in high frequency bands are less than that in low frequency bands, which corresponds to less uncertainty.

TABLE I
THE STANDARD DEVIATION, $\boldsymbol{\sigma}_n(u,v)$, OF QUANTIZATION NOISE AT QP=27 IN 8×8 BDCT DOMAIN

4.03	4.01	3.94	3.92	3.89	3.56	3.45	3.18
3.96	3.94	4.02	3.72	3.55	3.34	3.26	2.86
4.11	3.92	3.98	3.72	3.78	3.36	3.20	2.87
4.05	3.82	3.74	3.57	3.57	3.29	3.09	2.67
3.97	3.76	3.61	3.44	3.52	3.13	3.26	2.62
3.84	3.58	3.48	3.27	3.13	3.05	2.88	2.47
3.82	3.56	3.31	3.09	2.94	2.77	2.72	2.22
3.36	3.07	2.86	2.62	2.46	2.35	2.10	1.91

B. Nonlocal prediction with similar transform blocks

Besides the transform prediction, we also utilize the weighted average of the nonlocal similar transform blocks to form another prediction for the target transform block, and the weights are determined based on the block similarity as follows,

$$w_{t,r} = \frac{1}{S} e^{-\frac{\|\mathbf{Y}_t - \mathbf{Y}_r\|_2}{h}}. \tag{1}$$

where S is a normalization constant, and h is a smoothness factor; $w_{t,r}$ represent the weight of referenced similar block $\mathbf{Y}_r$ relative to target transform block $\mathbf{Y}_t$. Then, the nonlocal prediction for band (u,v) is calculated as,

$$\widehat{\mathbf{X}}_t^{(c)}(u,v) = \sum_{r=1}^{R} w_{t,r} \mathbf{Y}_r(u,v). \tag{2}$$

Since the nonlocal prediction is based on the assumption of block similarity, when these blocks are more similar, the prediction will be more accurate. Therefore, we take the reciprocal of the variance of coefficients in the same band to reflect the uncertainty of the nonlocal prediction. The variance

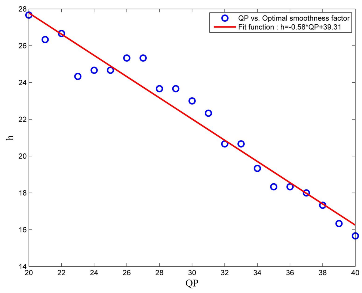
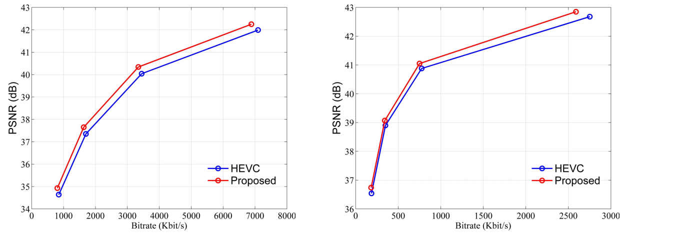


Fig. 3. The optimal smoothness factors for luminance component of intra coding frame.



(a) *Kimono* (b) *Johnny*

Fig. 4. Luminance PSNR curve for different sequences compressed by HEVC with and without the proposed TALF under LDP coding configurations with ALF off

of nonlocal transform coefficients for band (u, v) is calculated with,

$$\boldsymbol{\sigma}_c(u,v) = \frac{s}{R}\sum_{r=1}^{R}(\mathbf{Y}_r(u,v) - \boldsymbol{\mu}(u,v))^2, \quad (3)$$

$$\boldsymbol{\mu}(u,v) = \frac{1}{R}\sum_{r=1}^{R}(\mathbf{Y}_r(u,v)). \quad (4)$$

where s is a scale factor, which is used to compensate the influence on coefficient variance due to compression.

C. Frame reconstruction and parameter determination

Based on the two prediction sources, we can get a better prediction by adaptively fusing them according to their uncertainties,

$$\widehat{\mathbf{X}}_t(u,v) = \mathbf{Y}_t(u,v) + \frac{\boldsymbol{\sigma}_n^2(u,v)}{\boldsymbol{\sigma}_c^2(u,v) + \boldsymbol{\sigma}_n^2(u,v)}\widehat{\mathbf{X}}_t^{(c)}(u,v). \quad (5)$$

Since the blocks are overlapped, the final filtered frame is the average of multiple samples by inversely transforming these estimated block coefficients.

In the proposed TALF, there are two parameters, i.e., scale factor s in Eqn.(3) and smoothness factor h in Eqn.(1) to be determined. The first parameter s is determined empirically, $s = 2$ for intra coding and $s = 5$ for inter coding. The smoothness factor, h, is related with the quantization noise level, which can be reflected by QPs. We take two WQVGA video sequences, i.e., *RaceHorses*, and *BlowingBubbles* to learn the optimal smoothness factors for different QPs, which achieve largest amount of distortion reduction. Fig.3 shows the relationship between the QPs and the optimal smoothness factors. We fit the relationship with polynomial for different coding methods including intra coding and inter coding, and for luminance and chroma components respectively.

III. Experimental Results

The proposed TALF has been implemented into HEVC reference software, HM7.0, and is placed between DF and SAO. In this section, we verify the efficiency of the proposed TALF with and without ALF on widely used video sequences as shown in Tables II and III. The first 50 frames are encoded with four QPs, i.e., 22, 27, 32 and 37. Three coding configures are tested, which are all intra coding (AI), low delay P coding (LDP) and random access coding (RA). The anchor is HM7.0 with and without ALF.

Compared with HM7.0 without ALF in Table II , the proposed TALF further improves the coding performance, achieving 2.4%, 4.8% and 2.2% bit-rate savings on average for luminance component under AI, LDP and RA coding configurations, respectively. It also achieves up to 6.0% bit-rate savings on average for chroma components under LDP coding configuration. Especially for *Kimono* and *Johnny*, the proposed TALF achieves up to 11.6% and 9.8% bit-rate savings. Compared with HM7.0 with ALF, the proposed TALF also achieves 1.8%, 2.9% and 1.7% bit-rate savings for luminance component on average. The improvement is not significant as that when ALF is off, because the compression noise has been significantly reduced by the proposed TALF, which deteriorates the performance of ALF. Fig.4 shows the PSNR curves for sequences, *Kimono* and *Johnny*, and the proposed TALF achieves bit-rate savings over a very large bit rate range. Although the proposed method improves the compression performance obviously, it also introduces significant computational burdens. On average, the encoding time increases by 638%, 186%, 175% compared with that of HM7.0 for AI, LDP and RA respectively. Fast algorithms should be investigated for the proposed TALF to speed it up in our future work.

IV. Conclusion

In this paper, we have proposed a novel transform-domain adaptive in-loop filtering method by fusing transform coefficient and nonlocal transform coefficients in similar blocks according to their uncertainties. The transform-domain method can estimate prediction uncertainty band-by-band, which makes the multiple prediction fusion more efficient than that in spatial domain. Based on the experimental results on widely used video sequences, the proposed TALF can efficiently improve the compression performance further compared with HEVC, without introducing extra overhead, since three frame level flags are negligible.

Acknowledgment

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the National

TABLE II
Bit-rate saving of the proposed TALF, anchor is HM7.0 with ALF off

Sequences		AI			LDP			RA		
		Y	U	V	Y	U	V	Y	U	V
1920 × 1080	Kimono	-4.4%	-2.2%	-2.0%	-11.6%	-8.7%	-6.1%	-4.0%	-4.5%	-2.3%
	ParkScene	-1.1%	-1.1%	-1.0%	-1.7%	-0.9%	-2.3%	-0.3%	0.0%	-0.5%
	Cactus	-2.2%	-1.6%	-3.1%	-5.1%	-4.7%	-3.7%	-2.5%	-3.1%	-3.0%
	BasketballDrive	-3.4%	-4.0%	-4.2%	-6.9%	-10.7%	-9.3%	-3.1%	-7.0%	-6.4%
	BQTerrace	-1.2%	-1.4%	-1.9%	-6.8%	-4.7%	-6.2%	-2.1%	-2.3%	-3.4%
1280 × 720	vidyo1	-4.9%	-4.2%	-4.6%	-7.0%	-9.9%	-9.8%	-4.7%	-8.9%	-8.5%
	vidyo3	-3.8%	-2.5%	-2.2%	-7.3%	-9.4%	-4.1%	-3.1%	-6.4%	-2.2%
	vidyo4	-3.6%	-4.7%	-4.0%	-6.3%	-13.4%	-11.6%	-3.3%	-10.3%	-8.7%
	FourPeople	-4.2%	-3.5%	-3.8%	-7.0%	-8.5%	-8.4%	-5.0%	-7.7%	-8.5%
	Johnny	-4.1%	-5.0%	-4.5%	-9.8%	-15.2%	-12.4%	-3.7%	-10.1%	-9.6%
	KristenAndSara	-4.5%	-4.4%	-4.7%	-8.3%	-11.9%	-12.7%	-5.1%	-7.8%	-9.9%
832 × 480	BasketballDrill	-1.3%	-3.3%	-4.2%	-1.5%	-3.8%	-4.7%	0.1%	-3.3%	-4.1%
	BQMall	-1.7%	-1.8%	-2.1%	-2.4%	-2.1%	-2.0%	-0.5%	-1.1%	-1.7%
	PartyScene	-0.2%	-0.8%	-1.0%	-0.4%	-0.2%	0.3%	-0.1%	0.7%	0.1%
	RaceHorsesC	-1.6%	-2.1%	-3.7%	-4.0%	-3.7%	-6.7%	-2.7%	-4.5%	-6.5%
	BasketballPass	-1.2%	-2.1%	-2.2%	-1.6%	-1.9%	-0.7%	-0.5%	-1.0%	0.4%
416 × 240	BQSquare	0.0%	-1.6%	-2.2%	0.1%	0.0%	-0.3%	0.4%	-0.3%	-1.2%
	BlowingBubbles	-1.0%	-2.0%	-2.1%	-0.9%	-1.8%	-0.2%	-0.3%	-1.3%	-0.4%
	RaceHorses	-1.9%	-1.8%	-2.5%	-2.4%	-1.5%	-1.6%	-1.8%	-2.7%	-3.0%
Overall		-2.4%	-2.6%	-3.0%	-4.8%	-6.0%	-5.4%	-2.2%	-4.3%	-4.2%

TABLE III
Bit-rate saving of the proposed TALF, anchor is HM7.0 with ALF on

Sequences		AI			LDP			RA		
		Y	U	V	Y	U	V	Y	U	V
1920 × 1080	Kimono	-3.8%	-1.6%	-2.0%	-5.5%	-3.8%	-4.8%	-3.7%	-3.8%	-3.7%
	ParkScene	-2.3%	-1.8%	-0.9%	-2.0%	-4.9%	-2.3%	-1.4%	-4.6%	-2.1%
	Cactus	-2.7%	-2.0%	-2.0%	-4.2%	-6.2%	-4.6%	-2.0%	-4.9%	-3.0%
	BasketballDrive	-2.5%	-1.2%	-1.4%	-3.5%	-1.4%	-1.8%	-2.3%	-1.5%	-1.2%
	BQTerrace	-0.2%	-0.3%	-0.4%	0.1%	-0.1%	-1.2%	0.1%	0.0%	-0.4%
1280 × 720	vidyo1	-1.4%	-1.1%	-1.9%	-3.0%	-2.4%	-1.2%	-1.6%	-2.8%	-2.2%
	vidyo3	-2.6%	-2.7%	-2.9%	-3.2%	-7.6%	-6.1%	-2.1%	-6.3%	-4.7%
	vidyo4	-0.9%	-1.0%	-1.2%	-4.2%	-4.0%	-5.3%	-1.7%	-1.6%	-3.1%
	FourPeople	-2.5%	-1.4%	-1.5%	-4.1%	-4.2%	-3.7%	-3.1%	-3.7%	-4.2%
	Johnny	-3.1%	-2.4%	-1.9%	-8.2%	-9.1%	-7.6%	-3.7%	-5.3%	-5.2%
	KristenAndSara	-3.1%	-2.3%	-2.3%	-6.2%	-7.1%	-5.8%	-4.2%	-3.7%	-4.9%
832 × 480	BasketballDrill	-1.2%	-2.7%	-3.6%	-0.8%	-4.5%	-5.3%	0.2%	-4.3%	-4.8%
	BQMall	-1.7%	-1.0%	-1.2%	-2.1%	-1.1%	0.0%	-0.9%	-0.3%	-0.7%
	PartyScene	-0.3%	-0.5%	-0.6%	-0.4%	-0.2%	0.0%	0.1%	0.3%	0.3%
	RaceHorsesC	-1.1%	-1.0%	-1.9%	-2.9%	-2.0%	-4.6%	-2.3%	-2.3%	-4.4%
416 × 240	BasketballPass	-1.4%	-1.9%	-1.8%	-1.4%	-2.1%	-2.1%	-1.0%	-1.7%	-1.4%
	BQSquare	-0.2%	0.3%	-0.6%	-0.1%	2.1%	0.8%	-0.1%	0.3%	1.1%
	BlowingBubbles	-1.1%	-1.2%	-0.8%	-1.0%	-1.2%	-0.8%	-0.6%	-1.8%	-0.4%
	RaceHorses	-1.2%	-0.3%	-1.1%	-2.1%	-1.7%	-3.0%	-1.7%	-1.3%	-2.3%
Overall		-1.8%	-1.4%	-1.6%	-2.9%	-3.2%	-3.1%	-1.7%	-2.6%	-2.5%

Research Foundation, Singapore, under its Interactive Digital Media (IDM) Strategic Research Programme. This work was also cooperated with Peking University, and partially supported by the National Basic Research Program of China (973 Program) under Grant 2015CB351800.

References

[1] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 614–619, Jul. 2003.

[2] C.-M. Fu, E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park, and W.-J. Han, "Sample Adaptive Offset in the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1755–1764, Dec. 2012.

[3] C.-Y. Tsai, C.-Y. Chen, T. Yamakage, I. S. Chong, Y.-W. Huang, C.-M. Fu, T. Itoh, T. Watanabe, T. Chujoh, M. Karczewicz, and S.-M. Lei, "Adaptive Loop Filtering for Video Coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 934–945, Dec. 2013.

[4] X. Zhang, R. Xiong, S. Ma, and W. Gao, "Adaptive loop filter with temporal prediction," in *Picture Coding Symposium (PCS), 2012*, May 2012, pp. 437–440.

[5] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, vol. 2, Jun. 2005, pp. 60–65.

[6] M. Matsumura, S. Takamura, and A. Shimizu, "Lcu-based framework with zero pixel line buffers for non-local means filter," *ITU-T SG16, JCTVC-J0165, Stockholm*, Jul. 2012.

[7] X. Zhang, W. Lin, R. Xiong, X. Liu, S. Ma, and W. Gao, "Low-rank decomposition-based restoration of compressed images via adaptive noise estimation," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4158–4171, Sept 2016.

[8] X. Zhang, R. Xiong, W. Lin, J. Zhang, S. Wang, S. Ma, and W. Gao, "Low-rank based nonlocal adaptive loop filter for high efficiency video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2016.