2015 IEEE International Symposium on Multimedia

Non-local Structure-based Filter for Video Coding

Jian Zhang†, Chuanmin Jia†, Siwei Ma†‡, and Wen Gao†‡

†Institute of Digital Media & Cooperative Medianet Innovation Center, Peking University, Beijing, China
‡Peking University Shenzhen Graduate School, Shenzhen, China
{jian.zhang, cmjia, swma, wgao}@pku.edu.cn

***Abstract*—The deblocking filtering (DF) in HEVC is only applied to the boundaries between the coding units, prediction units, or transform units, which actually exists two issues. On one hand, the simple DF in HEVC does not fully exploit structure information in video. On the other hand, DF in HEVC does not consider the inside areas, which often suffers from quantization distortion. To alleviate the above issues, in this paper, a non-local structure-based filter (NLSF) is proposed by simultaneously enforcing the intrinsic local sparsity and the non-local self-similarity of each frame in video. NLSF not only deals with the boundaries, but also deals with the inside areas, which is able to effectively reduce the artifacts while enhance the quality of the deblocking frames. Experimental results demonstrate that, compared with the original HEVC reference encoder implementation in AI configuration, the proposed NLSF can achieve up to 7.3% BD-rate saving by substituting for DF in HEVC.**

***Index Terms*—HEVC/H.265; deblocking fliter; loop filter; video coding**

I. Introduction

In April 2010, the ITU-T Video Coding Expert Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG) formed the Joint Collaborative Team on Video Coding (JCT-VC) to develop the new coding standard, which is known as high efficiency video coding (HEVC) or H.265, formally published in 2013 [1]. Many new coding tools and coding structures are adopted in HEVC, which enables a major advance in compression relative to its predecessors.

In the HEVC video coding standard, the in-loop filters play an important, which are applied in the encoding and decoding loops, after the inverse quantization but before saving the picture to the decoded picture buffer. HEVC standard specifies two in-loop filters, a deblocking filter, which is applied first, and a sample adaptive offset (SAO), which is applied to the output of the deblocking filter [2]. The purpose of the deblocking filter is to attenuate the discontinuities at prediction and transform block boundaries, while the SAO aims to further improve the quality of the decoded picture by reducing the ringing artifacts and changes in the sample intensity of areas of a reconstructed picture. Due to that the deblocking and SAO attenuate different artifacts, their benefits are usually additive when exploited together.

In HEVC, both the motion prediction and transform coding are block-based. The size of motion predicted blocks varies from 8×4 and 4×8, to 64×64 luma samples, while the size of block transforms and intra-predicted blocks varies from 4×4 to 32×32 samples. These blocks are coded relatively independently from the neighboring blocks and approximate the original signal with some degree of similarity. Since coded blocks only approximate the original signal, the difference between the approximations may cause discontinuities at the prediction and transform block boundaries [3] [4]. These discontinuities are attenuated by the deblocking filter.

In HEVC deblocking, the vertical boundaries in a picture are filtered first, followed by the horizontal boundaries. In a coding unit, the vertical boundaries between the coding blocks are processed starting from the left-most boundary towards the right-hand side. The horizontal boundaries are processed starting from the top-most boundary towards the bottom [4]. Since the deblocking filtering (DF) is only applied to the boundaries between the coding units, prediction units, or transform units, which actually exists two issues. On one hand, the simple DF in HEVC does not fully exploit structure information in video. On the other hand, DF in HEVC does not consider the inside areas, which often suffers from quantization distortion. These two issues degrade the quality of the deblocking frames.

In the past several years, non-local self-similarity has been emerging as one of the most properties of natural image and videos, which depicts the repetitiveness of higher level patterns (e.g., textures and structures) globally positioned in images and videos, and has achieved great success in various image/video restoration applications [5] [6] [7] [8] [9] [10]. In this paper, to enhance the quality of the deblocking frames, a non-local structure-based filter (NLSF) is proposed by simultaneously enforcing the intrinsic local sparsity and the non-local self-similarity of each frame in video. NLSF not only deals with the boundaries, but also deals with the inside areas, which is able to effectively reduce the artifacts while enhance quality of the deblocking frames. Experimental results demonstrate that, compared with the original HEVC reference encoder implementation in AI configuration, the proposed NLSF can achieve up to 7.3% BD-rate saving by substituting for DF in HEVC.

The remainder of this paper is organized as follows. Section II elaborates the proposed non-local structure-based filter (NLSF). Extensive experimental results are reported in Section III. In Section IV, we concludes this paper.

978-1-5090-0379-2/15 $31.00 © 2015 IEEE
DOI 10.1109/ISM.2015.90

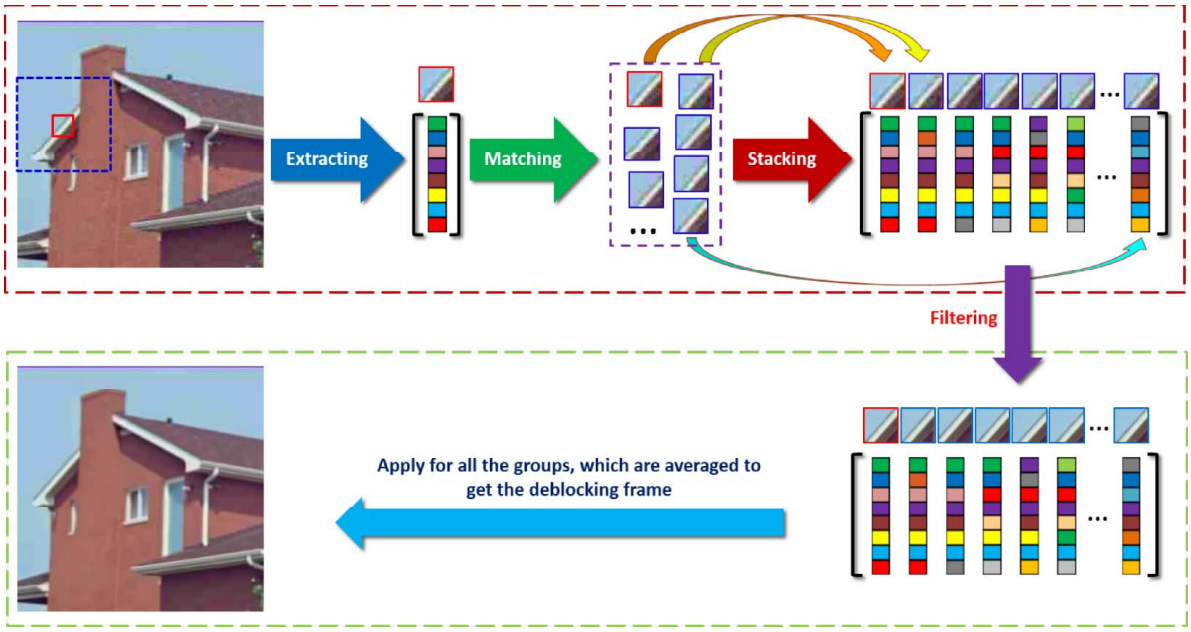

Fig. 1: Illustrations for the proposed non-local structured-based filter (NLSF).

II. Proposed Non-local Structure-based Filter

In our latest work [10], a new sparse representation model in the unit of group instead of block, named as group-based sparse representation (GSR), is proposed, which is able to exploit the local sparsity and the nonlocal self-similarity of natural images simultaneously in a unified framework. Motivated by the success of GSR in image inpainting, deblurring and compressive sensing recovery applications, in this paper, we propose to design the non-local structure-based filter (NLSF) on the basis of GSR model [10]. In this paper, different from GSR [10], the proposed NLSF has the advantages of being non-iterative and being parameter-free to the QP. The details are given below.

A. Group Construction

In this subsection, we will show how to exploit non-local structure within each frame in sequence to construct a group.

The basic idea of GSR is to adaptively sparsify the natural image in the domain of group. Thus we first show how to construct a group. In fact, each group is represented by the form of matrix, which is in fact composed of nonlocal blocks with similar structures. To be concrete, as illustrated in Fig. 1, first, divide the frame $\mathbf{x} \in \mathbb{R}^{N^2}$ with size N^2 into K overlapped blocks of size $\sqrt{B_s} \times \sqrt{B_s}$, and each block is denoted by the vector $\mathbf{x}_k \in \mathbb{R}^{B_s}$, i.e. $k = 1,\ 2, \ldots, K$. Then, for each block $\mathbf{x}_k$, denoted by small red square in Fig. 1, within the $W_s \times W_s$ training window (big blue square), search its c best matched blocks, which comprise the set $\mathbf{S}_{\mathbf{x}_k}$. Here, Euclidean distance is selected as the similarity criterion between different blocks. Next, all the blocks in $\mathbf{S}_{\mathbf{x}_k}$ are stacked into a matrix of size $B_s \times c$, denoted by $\mathbf{X}_{G_k}$, which includes every block in as its columns, i.e. $\mathbf{X}_{G_k}=[\mathbf{x}_{G_k \otimes 1}, \mathbf{x}_{G_k \otimes 2}, \ldots, \mathbf{x}_{G_k \otimes c}]$. The matrix $\mathbf{X}_{G_k}$ containing all the blocks with similar structures is named as a group. Note that, each block $\mathbf{x}_k$ is represented as a vector, while each group $\mathbf{X}_{G_k}$ is represented as a matrix, as shown in Fig. 1. It is obvious to observe that each block corresponds to a group.

B. Group-based Filtering

This subsection will give the details about how to conduct effective filtering based on the group.

Given one frame $\mathbf{x} \in \mathbb{R}^{N^2}$ in the sequence, which suffers from blocking artifacts, first construct K groups according to the above subsection. For each group $\mathbf{X}_{G_k}=[\mathbf{x}_{G_k \otimes 1}, \mathbf{x}_{G_k \otimes 2}, \ldots, \mathbf{x}_{G_k \otimes c}]$, $k = 1,\ 2, \ldots, K$, we apply singular value decomposition to it, yielding

$$\mathbf{X}_{G_k} = \mathbf{U}_{G_k} \mathbf{\Sigma}_{G_k} \mathbf{V}_{G_k}^T = \sum_{i=1}^{m} \mathbf{\Upsilon}_{\mathbf{x}_{G_k \otimes i}} \left(\mathbf{u}_{G_k \otimes i} \mathbf{v}_{G_k \otimes i}^T \right), \tag{1}$$

where $\mathbf{\Upsilon}_{\mathbf{x}_{G_k}}=[\mathbf{\Upsilon}_{\mathbf{x}_{G_k \otimes 1}}; \mathbf{\Upsilon}_{\mathbf{x}_{G_k \otimes 2}}; \ldots; \mathbf{\Upsilon}_{\mathbf{x}_{G_k \otimes m}}]$ is a column vector, $\mathbf{\Sigma}_{G_k}=diag(\mathbf{\Upsilon}_{\mathbf{x}_{G_k}})$ is a diagonal matrix with the elements of $\mathbf{\Upsilon}_{\mathbf{x}_{G_k}}$ on its main diagonal, and $\mathbf{u}_{G_k \otimes i}, \mathbf{v}_{G_k \otimes i}$ are the columns of $\mathbf{U}_{G_k}$ and $\mathbf{V}_{G_k}$, separately.

In order to reduce the artifacts and enhance the quality, the hard thresholding operation is applied to $\mathbf{\Upsilon}_{\mathbf{x}_{G_k}}$, i.e.,

$$\boldsymbol{\alpha}_{G_k} = hard(\mathbf{\Upsilon}_{\mathbf{x}_{G_k}}, \tau), \tag{2}$$

where $hard\,(\mathbf{x}, a) = \mathbf{x} \odot \mathbf{1}\,(abs\,(\mathbf{x}) - a)$ denotes the operator of hard thresholding and $\odot$ stands for the element-wise product of two vectors. τ denotes the threshold, whose setting is elaborated in the next subsection.

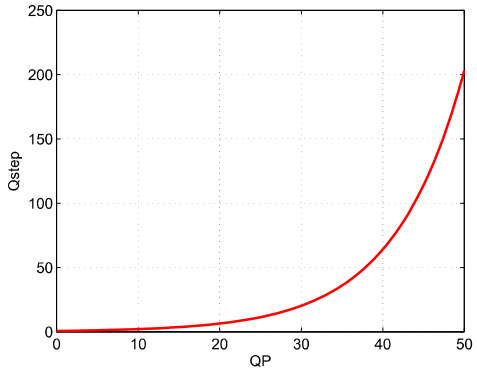

Fig. 2: Relationship between quantization step (Qstep) and quantization parameter (QP).

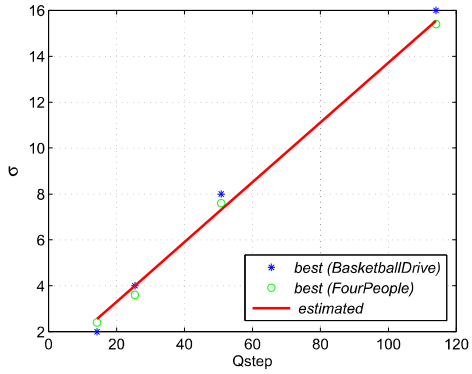

Fig. 3: Agreement between the values of σ estimated by Eq. (5) and the optimal ones (found experimentally), which give the highest PSNR for the deblocking *BasketballDrive* and *FourPeople* sequences using the proposed NLSF.

After achieving $\boldsymbol{\alpha}_{G_k}$, then we obtain the deblocking group $\hat{\mathbf{X}}_{G_k}$, expressed by

$$\hat{\mathbf{X}}_{G_k} = \sum_{i=1}^{m} \boldsymbol{\alpha}_{G_k} \otimes_i \left(\mathbf{u}_{G_k} \otimes_i \mathbf{v}_{G_k}^T \otimes_i \right). \tag{3}$$

This process is applied for all K groups to achieve $\hat{\mathbf{X}}_{G_k}$, $k = 1, 2, \ldots, K$. At last, all $\hat{\mathbf{X}}_{G_k}$ are averaged to get the deblocking frame $\hat{\mathbf{x}}$.

C. Parameter τ Estimation

From above description, the key for NLSF is the setting of parameter τ in Eq. (2). This subsection will present the details to give an adaptive and robust estimate for τ according to quantization parameter (QP) in video coding.

In this paper, we adopt Gaussian model to characterize the quantization noise between the original frame and the compressed frame due to its simplicity and effectiveness. Therefore, the estimate for τ is transformed to estimating the variance σ^2 of the quantization noise. Here, it is also worth emphasizing that the estimated noise variance σ^2 is not the real estimate of the variance of the difference between the original and the compressed images. Under assumption of Gaussian noise model, it is just the variance of the hypothetical Gaussian noise, which determines the level of adaptive smoothing that is able to reduce the artifacts generated by the quantization process [11].

As we know, similar to H.264/AVC, a QP is used to determine the quantization step size in HEVC. QP can take 52 values from 0 to 51 for 8-bit video sequences. An increase of 1 in QP means an increase of the quantization step size (Qstep) by approximately 12%. The resulting relationship between QP and Qstep for an orthonormal transform is formuated by [1]:

$$Qstep = 2^{\frac{(QP-4)}{6}}, \tag{4}$$

Fig. 2 also shows how Qstep increases non-linearly with QP. Since quantization consists of division by a quantization step size and subsequent rounding while inverse quantization consists of multiplication by Qstep, we directly study the relationship between σ and Qstep.

First, we observe the optimal values of σ found experimentally for the sequences *BasketballDrive* and *FourPeople* compressed with different QPs (QP = 27, 32, 38, 45) corresponding to different Qsteps calculated by Eq. (4), as illustrated in Fig. 3. It can be inferred that different sequences with the same Qp or Qstep have similar optimal values of σ, which means that σ is only related with QP or Qstep. Hence, in this paper, we propose to estimate the optimal value of σ directly from Qstep by curve fitting using the following empirical formulation:

$$\sigma = 0.13 * Qstep + 0.71. \tag{5}$$

The red curve in Fig. 3 expresses the relationship between the estimated σ and Qstep, which is very obvious and clear. The value τ is also has a linear relationship with σ, expressed as

$$\tau = \sigma * (B_s + \sqrt{c}). \tag{6}$$

Finally, combining Eqs. (4) (5) (6), we further obtain the relationship between the estimated τ and QP below

$$\tau = (0.13 * 2^{\frac{(QP-4)}{6}} + 0.71) * (B_s + \sqrt{c}). \tag{7}$$

Extensive experiments in Section III will verify the robustness and effectiveness of the proposed model Eq. (7).

III. Experimental Results

In this section, we will demonstrate the effectiveness of our proposed non-local structure-based filter (NLSF) over the deblocking filter (DF) in HEVC. Our comparison scheme is conducted by substituting NLSF for DF, which are implemented and integrated on HM 12.0 reference software [12]. The experimental environment is Intel i7-3770 CPU-3.40 GHz, 8 GB of RAM with professional version of Windows 7 OS. During the implementation, no other third-party static or dynamic library was included. All source code fulfills the standards of C++ 11 and can be transplanted to other platforms such as Linux version of HM.

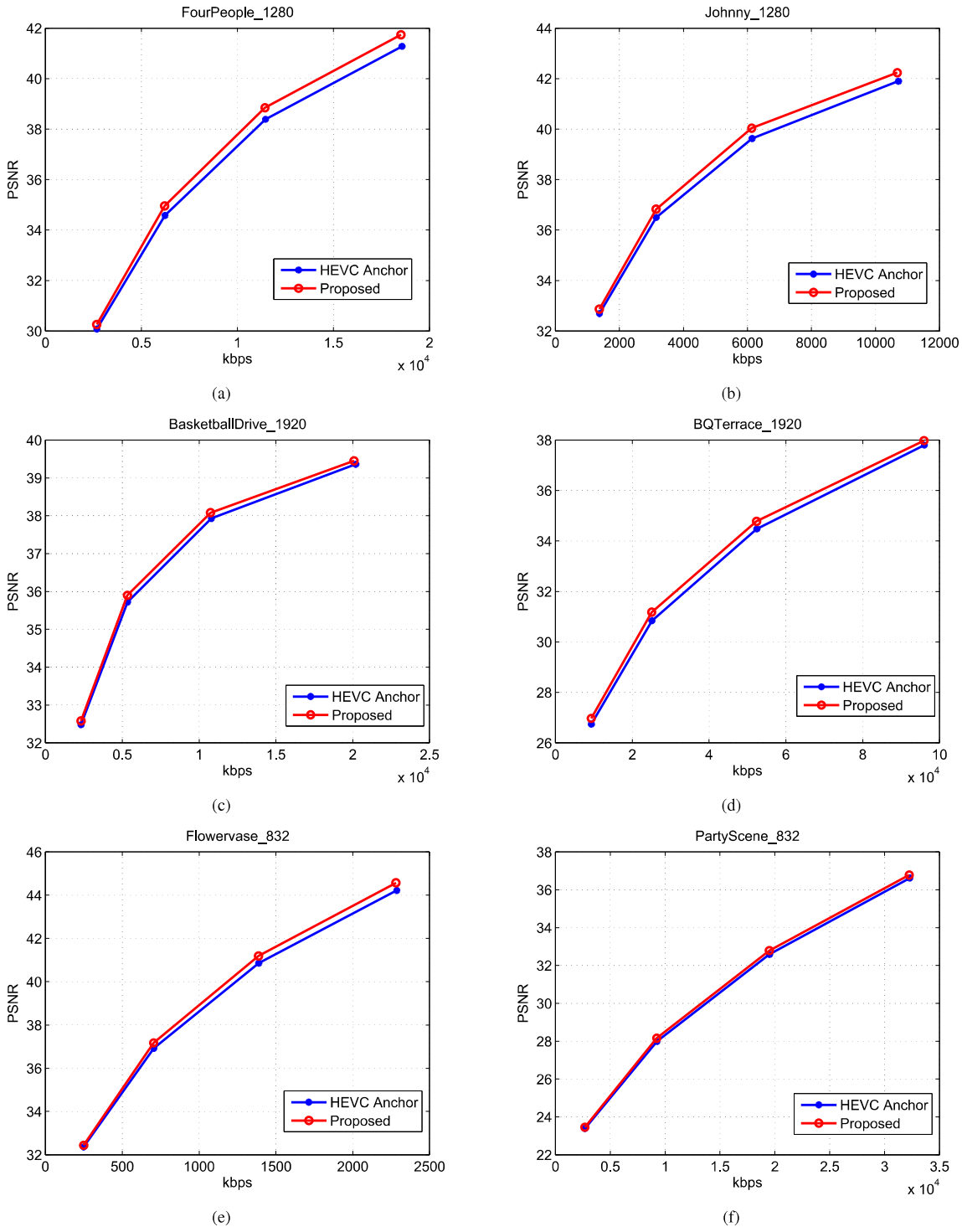

Fig. 4: The rate-distortion curves for all test sequences.

Test sequences with three different kinds of resolution has been chosen for evaluation procedure. Resolutions, frame rates, scene/character movements and other aspects have been taken into consideration when picking the test sequences (see Table I). HM is configured in All-Intra (AI) mode with quantization parameters (QP) of 27, 32, 38, 45. Table II shows experimental result using Bjontegaard's method [13] in terms of BD-rate (Y component).

(a)

(b)

(c)

(d)

(e)

(f)

Fig. 5: Subjective quality comparisons for the 1^{st} frame of Sequence Johnny by HEVC anchor and the proposed NLSF. (a) HEVC (bitrates = 6115.20 kbps, PSNR = 39.65 dB) (b) NLSF (bitrates = 6103.44 kbps, PSNR = 40.07 dB); (c) and (e) are the enlarged parts from (a); (d) and (f) are the enlarged parts from (b). Obviously, with even less bitrates, NLSF obtains much better quality than HEVC anchor.

Table II presents the PSNR performance of the proposed algorithm against HM reference software (Anchor). The table shows that the average BD-rate reduction is 5.4% for all six test sequences. Furthermore, it is obvious to see that the proposed algorithm achieves up to 7.3% BD-rate saving for Sequence Johnny, which is quite promising.

TABLE I: Properties of Test Sequences

Sequence	Format	Encoded Numbers	Frame Rate
BasketballDrive	1920×1080	10	50
BQTerrace	1920×1080	10	60
Flowervase	832×480	10	30
FourPeople	1280×720	10	60
Johnny	1280×720	10	60
PartyScene	832×480	10	50

TABLE II: Performance

Sequence	BD-Rate Saving (Y Component)
BasketballDrive	-5.0%
BQTerrace	-6.2%
Flowervase	-4.6%
FourPeople	-6.4%
Johnny	-7.3%
PartyScene	-2.9%
Average	$\mathbf{-5.4\%}$

More details performances can be seen in Fig. 4, in which we list all the rate-distortion performances for six sequences under four different QPs. All figures indicate that proposed method can achieve robust and outstanding performances in different categories of resolution, frame rates and content movement. Meanwhile, due to the parallelism feature of SVD algorithm, a accelerated version with parallelism can also be implemented for accelerating the whole routine. As for algorithm complexity analysis, experiment results show that the encoding time of the proposed method is about seven times as high as anchor HM reference software, which also means there exists spacious room for extension, acceleration and algorithm's optimization.

Fig. 5 shows the subjective quality comparisons for the 1^{st} frame of Sequence Johnny by HEVC anchor and the proposed NLSF. One can easily see that, even with less bitrates, the proposed NLSF achieves up to 0.42 dB PSNR gain over HEVC anchor. Besides, better visual quality results are obtained by NLSF (see Figs. 5(b), 5(d), 5(f)), which preserves more details and exhibits less artifacts than HEVC anchor (see Figs. 5(a), 5(c), 5(e)).

IV. Conclusion

In this paper, to improve the deblocking performance, a non-local structure-based filter (NLSF) is proposed by simultaneously enforcing the intrinsic local sparsity and the non-local self-similarity of each frame in video. NLSF not only deals with the boundaries, but also deals with the inside areas, which is able to effectively reduce the artifacts while enhance quality of the deblocking frames. Experimental results demonstrate that, compared with the original HEVC reference encoder implementation in AI configuration, the proposed NLSF can

achieve up to 7.3% BD-rate saving by substituting for DF in HEVC. Although with high complexity, the proposed NLSF shed light on the important role of deblocking filter. On going work is to address the algorithm parallelization and optimization for accelerating the proposed NLSF to make it practical.

V. Acknowledgement

This work was supported in part by the National Basic Research Program of China (973 Program, 2015CB351800), National Natural Science Foundation of China (61322106, 61421062), and Shenzhen Peacock Plan, which are gratefully acknowledged.

References

[1] Gary J Sullivan, Jens Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] Chih-Ming Fu, Elena Alshina, Alexander Alshin, Yu-Wen Huang, Ching-Yeh Chen, Chia-Yang Tsai, Chih-Wei Hsu, Shaw-Min Lei, Jeong-Hoon Park, and Woo-Jin Han, "Sample adaptive offset in the HEVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1755–1764, 2012.

[3] Peter List, Anthony Joch, Jani Lainema, Gisle Bjontegaard, and Marta Karczewicz, "Adaptive deblocking filter," *IEEE transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 614–619, 2003.

[4] Andrey Norkin, Gisle Bjontegaard, Arild Fuldseth, Matthias Narroschke, Makoto Ikeda, Karl Andersson, Minhua Zhou, and Geert Van der Auwera, "HEVC deblocking filter," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1746–1754, 2012.

[5] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel, "A non-local algorithm for image denoising," in *CVPR 2005*. IEEE, 2005, vol. 2, pp. 60–65.

[6] Jian Zhang, Ruiqin Xiong, Siwei Ma, and Debin Zhao, "High-quality image restoration from partial random samples in spatial domain," in *Visual Communications and Image Processing (VCIP), 2011 IEEE*. IEEE, 2011, pp. 1–4.

[7] Jian Zhang, Ruiqin Xiong, Chen Zhao, Siwei Ma, and Debin Zhao, "Exploiting image local and nonlocal consistency for mixed gaussian-impulse noise removal," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. IEEE, 2012, pp. 592–597.

[8] Jian Zhang, Debin Zhao, Chen Zhao, Ruiqin Xiong, Siwei Ma, and Wen Gao, "Image compressive sensing recovery via collaborative sparsity," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 2, no. 3, pp. 380–391, 2012.

[9] Jian Zhang, Debin Zhao, Ruiqin Xiong, Siwei Ma, and Wen Gao, "Image restoration using joint statistical modeling in a space-transform domain," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 6, pp. 915–928, 2014.

[10] Jian Zhang, Debin Zhao, and Wen Gao, "Group-based sparse representation for image restoration," *Image Processing, IEEE Transactions on*, vol. 23, no. 8, pp. 3336–3351, 2014.

[11] Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *Image Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 1395–1411, 2007.

[12] Frank Bossen, David Flynn, and Karsten Sühring, "HEVC reference software manual," *JCTVC-D404, Daegu, Korea*, 2011.

[13] Gisle Bjontegaard, "Calcuation of average PSNR differences between rd-curves," *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4 April 2001*, 2001.