Nonlocal In-Loop Filter: The Way Toward Next-Generation Video Coding?

Ma, Siwei; Zhang, Xinfeng; Zhang, Jian; Jia, Chuanmin; Wang, Shiqi; Gao, Wen

2016

Ma, S., Zhang, X., Zhang, J., Jia, C., Wang, S., & Gao, W. (2016). Nonlocal In-Loop Filter: The Way Toward Next-Generation Video Coding? IEEE MultiMedia, 23(2), 16-26.

https://hdl.handle.net/10356/83405

https://doi.org/10.1109/MMUL.2016.16

Nonlocal In-Loop Filter: The Future Way Towards Next-Generation Video Coding?

Siwei Ma, Xinfeng Zhang, Jian Zhang, Chuanmin Jia, Shiqi Wang, and Wen Gao, *Fellow, IEEE*

Abstract—In-loop filtering has emerged as an essential coding tool since H.264/AVC due to the delicate design in reducing different kinds of compression artifacts. However, existing in-loop filters only rely on image local correlations, where the nonlocal similarities have been largely ignored. In this paper, we journey through the design philosophy of in-loop filters and discuss our vision for the future of in-loop filter research by exploring the potential of non-local similarities. Specifically, the group-based sparse representation, which jointly exploits image local and nonlocal self-similarities, lays a novel and meaningful groundwork to the in-loop filter design. Hard- and soft-thresholding filtering operations are further applied to derive the sparse parameters that are appropriate for the compression artifacts reduction. Experimental results show that such in-loop filter design can significantly improve the compression performance on top of the High Efficiency Video Coding (HEVC) standard, leading us a new direction to improve the compression efficiency in the future.

Index Terms—HEVC, In-loop filtering, nonlocal similarity, sparse representation.

1 Introduction

HIGH Efficiency Video Coding (HEVC) [1], which is the latest video coding standard jointly developed by ITU-T Video Coding Experts Group (VCEG) and Moving Picture Experts Group (MPEG), was claimed to achieve potentially more than 50% coding gain compared to H.264/AVC. During the development of HEVC, the performances of three kinds of in-loop filters have been intensively investigated, including deblocking filter (DF) [2], Sample Adaptive Offset (SAO) [3] and Adaptive Loop Filter (ALF) [4], and among them DF and SAO were finally adopted. However, these in-loop filters only take advantage of the image local correlations to reduce compression artifacts, the performance of which is limited.

Deblocking filter is the first adopted in-loop filter in video coding standard, i.e. H.264/AVC [5], to reduce the blocking artifacts caused by coarse quantization and motion compensated prediction. A typical example of the block boundary with blocking artifact is shown in Fig. 1. Specifically, H.264/AVC defines a set of low pass filters with different filtering strengths, which are applied to 4×4 block boundaries. There are five levels of filtering strength in H.264/AVC, and the filter strength for every block boundary is jointly determined by the quantization parameters (QP), correlations of samples on both side of block boundaries, and the prediction modes (intra/inter prediction). DF in HEVC is similar with that of H.264/AVC. However, it is only applied to 8×8 block boundaries when any of the criterions that the block boundary lies between coding units (CU), prediction units (PU) and transform units (TU) is satisfied. Due to the improvement of the prediction accuracy in HEVC, only three filtering strengths are utilized, leading to the complexity reduction compared to H.264/AVC.

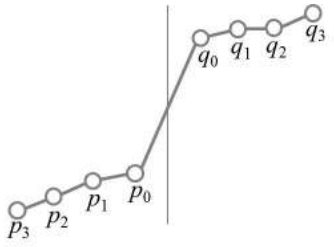


Fig. 1. 1-D example of block boundary with the blocking artifact, where $\{p_i\}$ and $\{q_i\}$ are pixels in neighboring blocks.

Sample Adaptive Offset (SAO) is a completely new in-loop filter adopted in HEVC. In contrast to the DF that only reconstructs the samples on block boundaries, all the samples are processed in SAO. As the sizes of CU, PU and TU have been largely extended compared with previous coding standards (i.e. CU: 8×8 to 64×64, PU: 4×4 to 64×64, TU: 4×4 to 32×32), the compression artifacts inside the coding blocks can no longer be compensated by DF. Therefore, SAO is applied to all samples reconstructed from DF by adding an offset to each sample to reduce the distortion. It has been proven to be a powerful tool to reduce ringing and contouring artifacts. In order to adapt the image content, SAO first divides an reconstructed picture into different regions, and then an optimal offset is derived for each region by minimizing the distortion between the original and reconstructed samples. It can use different offsets sample by sample in a region, depending on the sample classification strategy. In HEVC, two SAO types were adopted: edge offset (EO) and band offset (BO). For EO, the sample classification is based on comparison between the current and neighboring samples according to four 1-D neighboring patterns as shown in Fig. 2. For BO, the sample classification is based

- *S. Ma, J. Zhang, C. Jia, S. Wang and W. Gao are with the Institute of Digital Media, School of Electronics Engineering and Computer Science, Peking University, Beijing, China.*
  E-mail: {swma, jian.zhang, cmjia, sqwang, wgao}@pku.edu.cn
- *X. Zhang is with the Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore.*
  E-mail: xfzhang@ntu.edu.sg

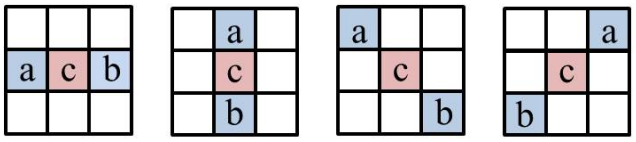


Fig. 2. Four 1-D directional patterns for EO sample calssification.

on sample values, i.e., the sample value range is equally divided into 32 bands. These offset values and region indices are signalled in bitstream, which may impose a relatively large overhead.

Adaptive Loop Filter (ALF) is a Wiener-based adaptive filter and the coefficients of which are derived by minimizing the mean square errors between original and reconstructed samples. Numerous recent efforts have been dedicated in developing high efficiency and low complexity ALF approaches. In HEVC reference software HM7.0, the filter shape of ALF is a combination of 9×7-tap cross shape and 3×3-tap rectangular shape, as illustrated in Fig. 3. Therefore, only correlations within a local patch are utilized to reduce the compression artifacts. To adapt the properties of input frame, up to 16 filters are derived for different regions of luminance component. Such high adaptability also creates large overhead that should be signalled into bitstream. Therefore, these regions need to be merged at encoder side based on rate-distortion optimization (RDO), which makes neighboring regions share the same filters to achieve a good tradeoff between the filter performance and overheads. In [6], Zhang et al. proposed to reuse the filter coefficients and regions division in previous encoded frame to reduce overheads. In [7], Stephan et al. proposed to place the filter coefficient parameters in a picture-level header called Adaptation Parameter Set (APS), which makes in-loop filter parameters reuse more flexible with APS indices.

In this paper, we explore the performance of in-loop filters for HEVC by taking advantage of image local and nonlocal correlations. A nonlocal similarity based loop filter (NLSLF) is incorporated into the HEVC standard by simultaneously enforcing the intrinsic local sparsity and the nonlocal self-similarity of each frame in the video sequence. For a reconstructed video frame from previous stage, we firstly divide it into overlapped image patches, and subsequently classify them into different groups based on their similarities. Since these image patches in the same group are with similar structures, they can be represented sparsely in the unit of group instead of block [8]. The compression artifacts can be reduced by thresholding the singular values of image patches group-by-group based on the sparse property of similar image patches. Two kinds of thresholding methods, i.e., hard- and soft-thresholding, and their related adaptive threshold determination methods are also explored. Extensive experimental results are conducted on HEVC common test sequences, which demonstrate that the nonlocal similarity based in-loop filter significantly improves the compression performance of HEVC, and up to 8.1% bitrate savings can be achieved.

The remainder of this paper is organized as follows. In Section 2, we review related work in image denoising and in-loop filters based on image nonlocal correlations.

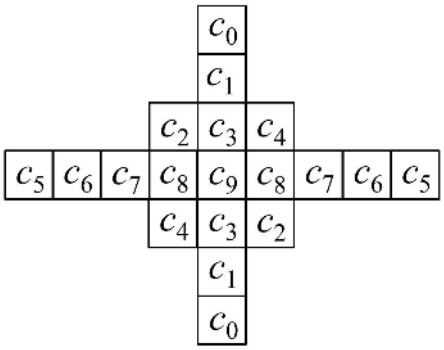


Fig. 3. ALF shape in HM7.0 (each square corresponds to a sample).

Section 3 presents the non-local in-loop filter for HEVC. Experimental results are reported in Section 4 and Section 5 concludes the paper.

2 Nonlocal Image Filter

In existing video coding standards, in-loop filters only focus on the local correlation within image patches, without fully consideration of the nonlocal similarities. However, in image restoration and denoising fields, many methods based on image nonlocal similarities have been proposed [9]–[13]. In [9], Buades et al. proposed the famous nonlocal means filter (NLM) to remove different kinds of noise by predicting each pixel with a weighted average of nonlocal pixels, where the weights are determined by the similarity of image patches located at the source and target coordinates. The well known denoising filter, BM3D [10], stacks nonlocal similar image patches into 3D matrices, and removes noise by shrinking coefficients of 3D transform of similar image patches based on image sparse prior model. Zhang et al. [11]–[13] utilized the nonlocal similar image patches to suppress compression artifacts, which are achieved by adaptively combining the pixels restored by the NLM filter and reconstructed pixels according to reliability of NLM prediction and quantization noise in transform domain. In [8], [14], [15], the authors utilize group of nonlocal similar image patches to construct image sparse representation, which can be further applied to image deblurring, denoising and inpainting. Although these nonlocal methods significantly improve the quality of restored images, all of them are treated as post-processing filters, such that the compression information has not been fully exploited.

In [16] and [17], Matsumura et al. firstly introduced the NLM filter to compensate the shortcomings of HEVC with only image local prior models, and delicately designed patch shapes, search window shapes and optimizing filter on/off control modules are utilized to improve the coding performance. In [18], Han et al. also employed the nonlocal similar image patches in a quadtree-based Kuan's filter to suppress compression artifacts, where the pixels restored by NLM filter and reconstructed pixels are adaptively combined together according to the variance of image signals and quantization noise. However, the weights in these filters are difficult to determine, leading to limited coding performance improvement.

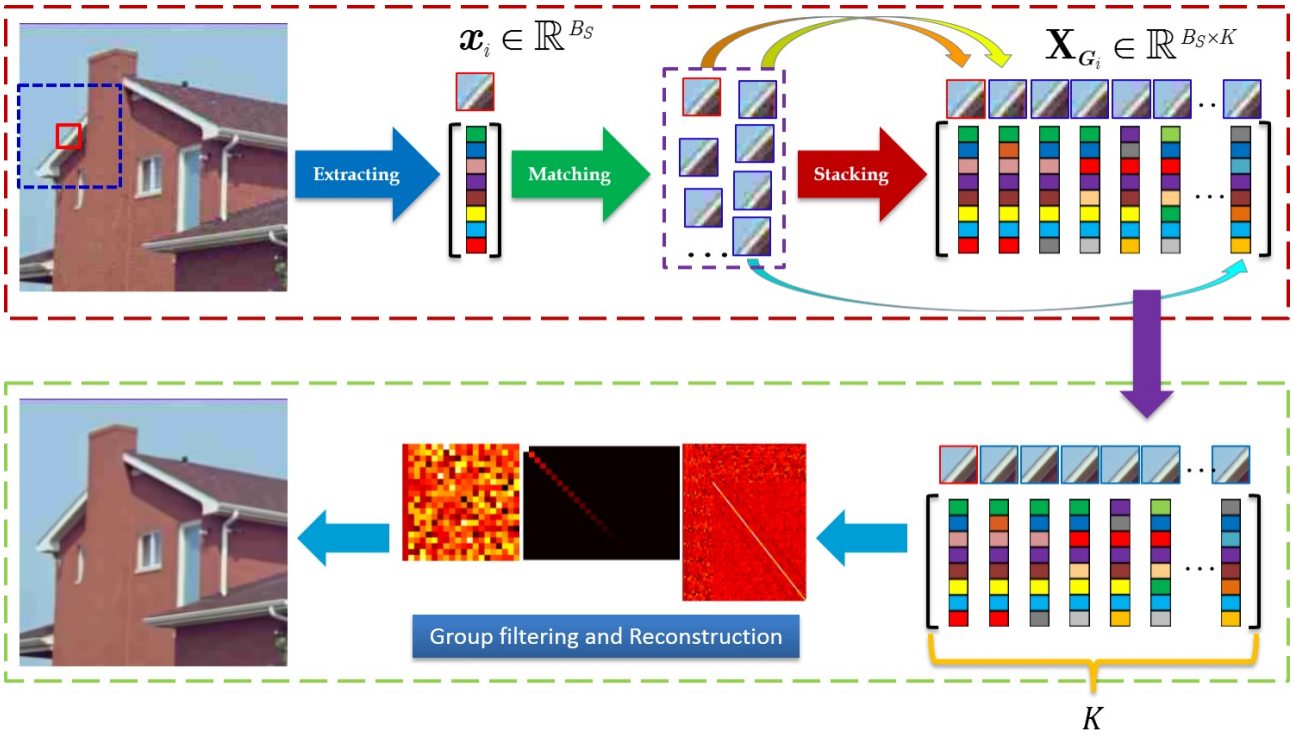


Fig. 4. Framework of the nonlocal similarity based loop filter (NLSLF).

3 The Nonlocal Similarity Based In-Loop Filter

In our previous work [8], a new sparse representation model is formulated in terms of a group of similar image patches, named as group-based sparse representation (GSR), which is able to exploit the local sparsity and the nonlocal self-similarity of natural images simultaneously in a unified framework. In this section, we describe how the nonlocal similarity based loop filter (NLSLF) is designed based on the GSR model, which can be divided into the following stages.

3.1 Patch Grouping

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 composed of nonlocal patches with similar structures. For a video frame, $\mathcal{I}$, we first divide it into S overlapped image patches with size of $\sqrt{B_s} \times \sqrt{B_s}$, and each patch is reorganized into a vector, $\boldsymbol{x}_k$, $k = 1, 2, ..., S$, as illustrated in Fig. 4. For every image patch, we find K nearest neighbors according to the Euclidean distance between different image patches,

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2. \tag{1}$$

These K similar image patches are stacked into a matrix of size $B_s \times K$,

$$\mathbf{X}_{G_i} = [\boldsymbol{x}_{G_i,1}, \boldsymbol{x}_{G_i,2}, ..., \boldsymbol{x}_{G_i,K}]. \tag{2}$$

Here $\mathbf{X}_{G_i}$ contains all the image patches with similar structures, which is termed as a group.

3.2 Group Filtering and Reconstruction

Since the image patches in the same group are very similar, they are able to be represented sparsely. For each group, we apply singular value decomposition to it and get image sparse representation,

$$\mathbf{X}_{G_i} = \mathbf{U}_{G_i} \mathbf{\Sigma}_{G_i} \mathbf{V}_{G_i}^T = \sum_{k=1}^{M} \Upsilon_{G_i,k} \left(\boldsymbol{u}_{G_i,k} \boldsymbol{v}_{G_i,k}^T \right), \tag{3}$$

where $\Upsilon_{G_i} = [\Upsilon_{G_i,1}; \Upsilon_{G_i,2}; ...; \Upsilon_{G_i,M}]$ is a column vector, $\mathbf{\Sigma}_{G_i} = diag(\Upsilon_{G_i})$ is a diagonal matrix with the elements of Υ_{G_i} as its main diagonal, and $\boldsymbol{u}_{G_i,k}$, $\boldsymbol{v}_{G_i,k}$ are the columns of $\mathbf{U}_{G_i}$ and $\mathbf{V}_{G_i}$, respectively. M is the maximum dimension of matrix $\mathbf{X}_{G_i}$

The matrix composed of corresponding compressed video frame is formulated as,

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{4}$$

where $\mathbf{N}$ is the compression noise, $\mathbf{X}$ and $\mathbf{Y}$ without any subscript represent the original frame and reconstructed frame, respectively. To derive the sparse representation parameters, we apply the thresholding, which is a widely used operation for coefficients with sparse property in image denoising problems. We apply two kinds of the thresholding methods, i.e., hard- and soft-threholding, to the singular values in Υ_{G_i}, which is composed of singular values of matrix $\mathbf{Y}$,

$$\boldsymbol{\alpha}_{G_i}^{(h)} = hard(\Upsilon_{G_i}, \tau) \tag{5}$$

$$\boldsymbol{\alpha}_{G_i}^{(s)} = soft(\Upsilon_{G_i}, \tau), \tag{6}$$

where the hard- and soft-thresholding are defined as,

$$hard(\boldsymbol{x}, \tau) = sign(\boldsymbol{x}) \odot (abs(\boldsymbol{x}) - \tau \mathbf{1}), \tag{7}$$

$$soft(\boldsymbol{x}, \tau) = sign(\boldsymbol{x}) \odot max(abs(\boldsymbol{x}) - \tau \mathbf{1}, \mathbf{0}). \quad (8)$$

Here $\odot$ stands for the element-wise product of two vectors, $sign(\cdot)$ is the function extracting the sign of every element of a vector, $\mathbf{1}$ is a all-ones vector and τ denotes the threshold. After achieving the shrunk singular values, the restored group of image patches $\widehat{\boldsymbol{x}}$ is given by,

$$\widehat{\boldsymbol{x}} = \sum_{k=1}^{M} \boldsymbol{\alpha}_{G_i,k} (\boldsymbol{u}_{G_i,k} \boldsymbol{v}_{G_i,k}^T). \quad (9)$$

Since these image patches are overlapped extracted, we simply take the average of the overlapped samples as the final filtered values.

3.3 Threshold Estimation

Based on the above discussion, the filtering strength is determined by the thresholding level parameter τ in Eqns.(5) and (6). However, in view of the various video content compressed with different quantization parameters, this is a non-trivial problem that has not been well resolved. In essence, the optimal threshold is closely related with the standard deviation of noise denoted as σ_n, and larger thresholds correspond to higher σ_n values.

In video coding, the compression noise is mainly caused by quantizing the transform coefficients. Therefore, quantization steps can be utilized to determine the standard deviation of the compression noise, and a scale factor is utilized to adapt different prediction modes, including intra and inter predictions.

For hard-thresholding, the optimal values of σ_n are derived experimentally based on the sequences *BasketballDrive* and *FourPeople* compressed with different QPs (QP = 27, 32, 38, 45), which are further converted to the quantization step sizes (Qsteps), as illustrated in Fig. 5. It can be inferred that different sequences with the same QP or Qstep have similar optimal values of σ_n, implying that σ_n is closely related with QP or Qstep. Inspired by this, we propose to estimate the optimal value of σ_n directly from Qstep by curve fitting using the following empirical formulation,

$$\sigma = a * Qstep + b. \quad (10)$$

where the $Qstep$ can be easily derived from quantization parameter based on the following relationship in HEVC,

$$Qstep = 2^{\frac{(QP-4)}{6}}. \quad (11)$$

The parameters (a, b) for different coding configurations are illustrated in Table 1.

Based on the filtering performance, we further use the size and number of similar image patches in one group as a scale factor,

$$\tau = \sigma_n * (B_s + \sqrt{K}). \quad (12)$$

where c is a scale factor according to prediction mode (intra/inter prediction) and σ_n is the standard deviation of compression noise for the whole image, which is estimated based on Eqn.(10).

For soft-thresholding, based on the filtering performance, we take the optimal threshold formulation for Generalized Gaussian signals,

$$\tau = \frac{c\sigma_n^2}{\sigma_{\boldsymbol{x}}}, \quad (13)$$

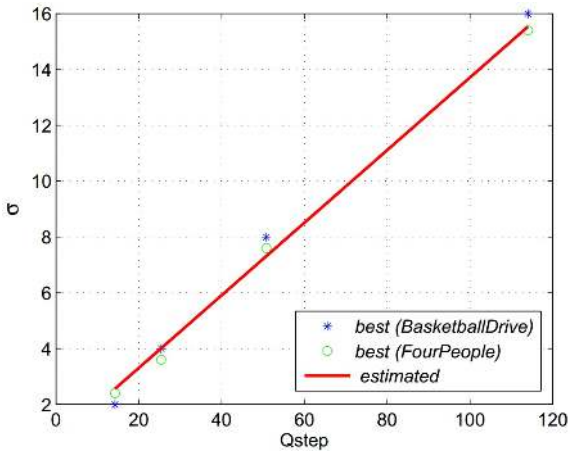


Fig. 5. Relationship between Qstep and standard deviation of compression noise.

where σ_x is the standard deviation of original signals that can be estimated by,

$$\sigma_{\boldsymbol{x}}^2 = \sigma_{\boldsymbol{y}}^2 - \sigma_n^2. \quad (14)$$

As the variance of compression noise, σ_n , is derived at the encoder side, we quantize it into the nearest integer range in [1,16], which are signalled with 4 bits and transmitted in the bitstream. Therefore, 12 bits are encoded in total for one frame with three colour components, e.g., YUV. The two thresholds for both hard- and soft-thresholding operations increase with the standard deviation of compression noise, which implies that the frames with more noise should be filtered with higher strength. Furthermore, the thresholds decrease with the standard deviation of signals, which can avoid over-smoothing for smooth areas.

3.4 Filtering On/Off Control

In order to ensure that the NLSLF consistently leads to distortion reduction, we introduce frame and LCU (Largest Coding Unit) levels on/off control flags that should be signalled in the bitstream. Specifically, regarding the frame level on/off control, three flags, *Filtered_Y*, *Filtered_U* and *Filtered_V*, are designed for the corresponding color component, respectively. When the distortions of the filtered image decrease, the corresponding flag is signalled as *true*, indicating that the image color component is finally filtered. For LCU level on/off control in luminance component, for each LCU a flag *Filterd_LCU[i]* is required to transmit. In picture header syntax structure, three bits are encoded to signal frame level control flags for each colour component, respectively. We place the syntax elements of LCU level control flags in coding tree unit parts, and only one bit is utilized for each LCU.

4 Experimental Results and Analysis

In this section, we implement the nonlocal similarity based in-loop filter in HEVC reference software, HM12.0. We denote the hard-thresholding filtering with threshold in Eqn.(12) as NLSLF-H, and the soft-thresholding filtering with threshold in Eqn.(13) as NLSLF-S. In order to better

TABLE 1
Coefficient for estimate σ for all configurations.

Component	AI		LDB		RA	
Type	*a*	*b*	*a*	*b*	*a*	*b*
Y	0.13000	0.7100	0.10450	0.4870	0.10450	0.4870
U	0.06623	0.8617	0.03771	0.8833	0.03771	0.8833
V	0.06623	0.8617	0.03771	0.8833	0.03771	0.8833

analyze the performance of the nonlocal similarity based in-loop filter, we further integrate the ALF of HM3.0 into HM12.0, in which the ALF tool has been removed, and compare the nonlocal similarity based in-loop filter with ALF.

The test video sequences in our experiments are widely used in HEVC common test conditions (CTC). There are 20 test sequences, which are classified into six categories, Class A$\sim$ class F. The resolution of class A is 2560×1600, class B is 1920×1080, class C is 832×480, class D is 416×240, and class E is 1280×720. Class F are not natural videos but screen content videos containing three different resolutions: 1280×720, 1024×768 and 832×480. Four typical quantization parameters are tested, i.e., 22, 27, 32 and 37. Three coding configures are tested respectively as that in CTC, i.e., all intra coding (AI), low delay B coding (LDB), and random access coding (RA). Along with the increase of K and B_s, the computational complexity increases rapidly, while the filtering performance may decrease for some sequences since dissimilar structures are more possibly to be included. Therefore, in our experiments, the size of image patches is set to $B_s = 6$ and the number of nearest neighbours for each image patch is set to $K = 30$ for all the sequences. For each frame, we extract image patches every five pixels according raster scanning order, which makes the image patches overlapped.

First, we treat the HM12.0 with and without ALF as anchors, respectively. The overall coding performance of NLSLF-S and NLSLF-H only with frame level control are illustrated in Table 2$\sim$ 5. Both of the two thresholding filters with nonlocal image patches achieve significant bitrate savings compared with that of HM12.0 without ALF. NLSLF-S achieves 3.2%, 3.1%, 4.0%, bitrate savings on average for AI, LDB and RA configurations, respectively. Moreover, NLSLF-H also achieves 4.1%, 3.3% and 4.4% bitrate savings on average for AI, LDB and RA configurations compared with HM12.0 without ALF. When the Nonlocal similarity structure based in-loop filters are combined with ALF, NLSLF-S achieves about 2.6%, 2.6% and 3.2% bitrate savings and NLSLF-H achieves about 3.1%, 2.8% and 3.4% bitrate savings compared with HM12.0 with ALF for AI, LDB and RA configurations, respectively. Although the improvements of the NLSLF are not so significant as that without ALF, they can still further improve the performance of HEVC with ALF. This verifies that the nonlocal similarity can further benefit compression artifact reduction compared with image local similarity. Since hard- and soft-thresholding operations are suitable for signals with different distributions, they show different coding gains on different sequences. Although NLSLF-H achieves better performance for most sequences than that of NLSLF-S in our experiments, soft-thresholding outperforms hard-thresholding for some sequences, e.g., Class E in LDB coding configuration and Class A in LDB and RA coding configurations.

Table 6 shows the detailed results of NLSLF-S with LCU level control for each sequence. Although LCU level control increases overheads, it can improve the coding efficiency as well by avoiding the over-smoothing case. It also shows that there is still room for improving the filtering efficiency by designing more reasonable thresholds for group-based sparse coefficients. Fig. 6 and Fig. 7 illustrate the rate-distortion curves of NLSF and HEVC without ALF for sequences, *Johnny*, *KristenAndSara* and *FourPeople*, respectively, which are compressed at different QP under RA configuration. We can see that coding performance is significantly improved in a wide bit range with the nonlocal similarity based in-loop filters.

We further compare the visual quality of the decoded video frames with different in-loop filters in Fig. 8. The deblocking filter only remove the blocking artifacts, and it is difficult to reduce other artifacts, e.g., ringing artifacts around the strips in the coat of image *Johnny*. Although SAO can process all the reconstructed samples, its performance is constrained by the large overheads, such that the blurring edges still exist. The nonlocal similarity based filters can efficiently remove different kinds of compression artifacts, and it also can recover destroyed structures by utilizing nonlocal similar image patches, e.g., most of the lines in coat being well recovered.

Although the NLSLF achieves significant improvement for video coding, it also introduces lots of computational burdens, especially due to SVD. Compared with HM12.0 encoding, the encoding time increase by NLSLF-H is 133%, 30% and 33% for AI, LDB and RA respectively. This also proposes new challenges to the loop filter research with image nonlocal correlations, which are also as our future work.

5 Conclusion

In this paper, we described our views on the in-loop filter design in the context of nonlocal similarities and chiseled a rough road toward the high efficiency in-loop artifacts removal for video compression. The novelty lies in adopting the non-local prior model in the in-loop filtering process, which leads to reconstructed frames with higher fidelity. To estimate the noise level, different kinds of thresholding operations have been examined, confirming that the nonlocal strategy can significantly improve the coding efficiency. This poses new chances not only to the in-loop filter research with non-local prior models, but also opens up new space for future exploration in nonlocal inspired high efficiency video compression.

TABLE 2
Performance of the NLSLF-S (Anchor: HM12.0 with ALF off).

Sequences	AI			LDB			RA		
	Y	U	V	Y	U	V	Y	U	V
Class A	-4.3%	-4.0%	-3.9%	-3.5%	-3.3%	-2.3%	-4.8%	-6.1%	-5.7%
Class B	-2.9%	-3.3%	-4.0%	-3.0%	-4.2%	-4.2%	-4.3%	-5.5%	-4.7%
Class C	-2.8%	-4.6%	-6.2%	-1.6%	-3.4%	-5.4%	-2.1%	-5.1%	-6.5%
Class D	-2.0%	-4.5%	-5.5%	-1.3%	-2.4%	-2.5%	-1.6%	-3.5%	-4.4%
Class E	-5.8%	-5.3%	-4.4%	-7.9%	-10.0%	-9.5%	-9.8%	-9.4%	-8.6%
Class F	-2.5%	-3.1%	-3.4%	-1.7%	-2.8%	-3.3%	-2.2%	-4.4%	-4.7%
Overall	-3.4%	-4.1%	-4.6%	-3.2%	-4.4%	-4.5%	-4.1%	-5.6%	-5.8%

TABLE 3
Performance of the NLSLF-S (Anchor: HM12.0 with ALF on).

Sequences	AI			LDB			RA		
	Y	U	V	Y	U	V	Y	U	V
Class A	-1.8%	-2.3%	-2.4%	-1.0%	-3.9%	-2.5%	-2.2%	-5.2%	-5.0%
Class B	-1.8%	-2.1%	-3.0%	-1.8%	-3.9%	-4.7%	-2.6%	-5.0%	-5.2%
Class C	-2.7%	-3.5%	-4.5%	-1.7%	-4.4%	-5.9%	-2.2%	-5.6%	-6.4%
Class D	-1.9%	-2.8%	-3.7%	-1.7%	-2.2%	-3.2%	-1.8%	-3.7%	-4.6%
Class E	-3.9%	-2.8%	-2.1%	-6.1%	-7.5%	-6.0%	-7.4%	-7.3%	-6.2%
Class F	-2.4%	-2.9%	-3.2%	-1.9%	-3.6%	-3.9%	-2.0%	-4.2%	-4.5%
Overall	-2.4%	-2.7%	-3.2%	-2.4%	-4.2%	-4.4%	-3.0%	-5.1%	-5.3%

TABLE 4
Performance of the NLSLF-H (Anchor: HM12.0 with ALF off).

Sequences	AI			LDB			RA		
	Y	U	V	Y	U	V	Y	U	V
Class A	-4.9%	-3.0%	-3.5%	-3.1%	-1.2%	-1.4%	-4.2%	-3.1%	-2.8%
Class B	-3.2%	-2.2%	-3.9%	-3.2%	-3.5%	-3.7%	-4.3%	-3.9%	-3.8%
Class C	-3.6%	-4.9%	-6.9%	-1.9%	-3.4%	-4.8%	-2.5%	-4.2%	-5.9%
Class D	-3.1%	-4.4%	-5.9%	-1.5%	-2.5%	-2.8%	-2.1%	-3.4%	-3.4%
Class E	-7.1%	-8.5%	-8.9%	-7.4%	-9.5%	-10.5%	-10.0%	-11.4%	-12.1%
Class F	-3.5%	-4.4%	-5.0%	-2.4%	-2.8%	-3.6%	-3.0%	-5.0%	-5.4%
Overall	-4.2%	-4.6%	-5.7%	-3.3%	-3.8%	-4.5%	-4.3%	-5.2%	-5.6%

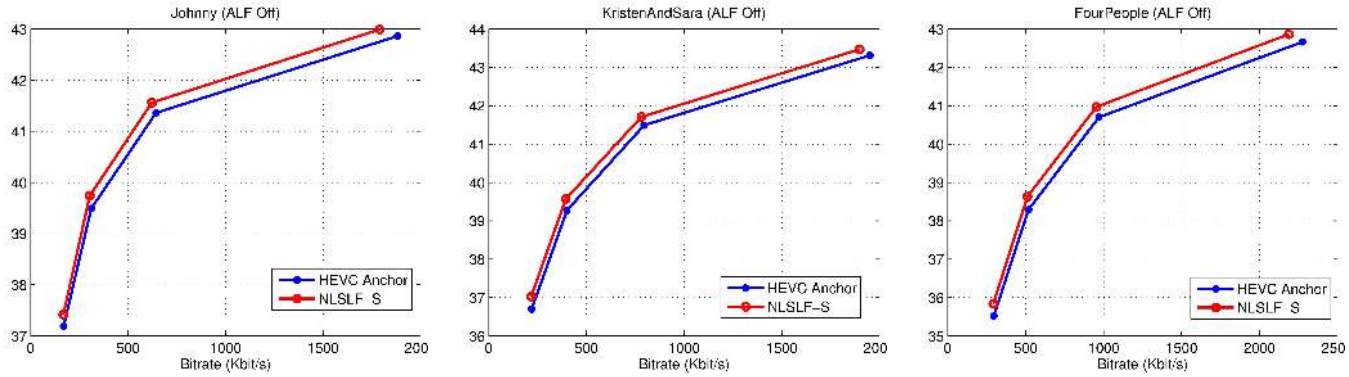
TABLE 5
Performance of the NLSLF-H (Anchor: HM12.0 with ALF on).

Sequences	AI			LDB			RA		
	Y	U	V	Y	U	V	Y	U	V
Class A	-2.1%	-1.4%	-1.8%	-1.0%	-1.6%	-1.3%	-1.7%	-2.3%	-2.1%
Class B	-1.9%	-1.0%	-2.5%	-2.1%	-2.9%	-3.3%	-2.6%	-3.0%	-3.8%
Class C	-3.1%	-2.6%	-5.0%	-2.0%	-4.0%	-5.1%	-2.2%	-4.3%	-5.9%
Class D	-2.6%	-1.6%	-3.1%	-1.6%	-2.5%	-3.0%	-1.9%	-3.6%	-3.8%
Class E	-4.9%	-4.5%	-3.9%	-5.5%	-5.5%	-5.6%	-7.5%	-7.5%	-6.8%
Class F	-3.1%	-4.3%	-5.0%	-2.8%	-3.5%	-3.6%	-2.9%	-4.7%	-5.3%
Overall	-2.9%	-2.6%	-3.5%	-2.5%	-3.3%	-3.7%	-3.1%	-4.2%	-4.6%

Apart from in-loop filtering, the nonlocal information can motivate the design of other key modules in video compression as well. Traditional video coding technologies mainly focus on reducing the local redundancies by intra prediction with limited neighboring samples. The inter-prediction can be regarded as a simplified version of non-local prediction, which obtains predictions from a relatively large range compared with intra prediction, leading to significant performance improvement. However, to the maximum extent, only a unique pair of patches can be employed, e.g., one image patch in unidirection and two image patches in bidirection predictions. This significantly limits the potentials of the prediction technique, as the number of similar image patches can be further extended to fully exploit the spatial and temporal redundancies. With the new technological advances in hardware and software, we could have foreseen the arrival and maturity of these non-local based coding techniques. We also believe that the non-local based video coding technology described in this paper or similar technologies developed from this ground could play important roles in the future video standardization.

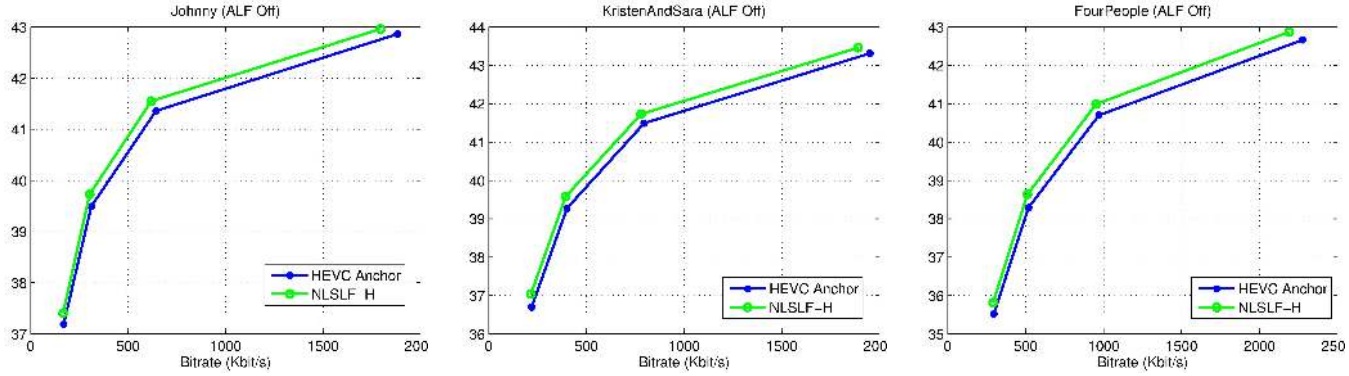
ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants 61322106, 61572047 and 61571017, and the National Basic Research Program of China (973 Program) under Grant 2015CB351800.



(a) (b) (c)

Fig. 6. The rate-distortion performance of NLSLF-S compared with HEVC (ALF OFF) for test sequences, (a) *Johnny*, (b)*KristenAndSara*, (c)*FourPeople* , which are compressed by HEVC RA coding.



(a) (b) (c)

Fig. 7. The rate-distortion performance of NLSLF-H compared with HEVC (ALF OFF) for test sequences, (a) *Johnny*, (b)*KristenAndSara*, (c)*FourPeople* , which are compressed by HEVC RA coding.

TABLE 6
Performance of the NLSLF-S with LCU level on/off control (Anchor: HM12.0 with ALF on).

Sequences		AI			LDB			RA		
		Y	U	V	Y	U	V	Y	U	V
Class A	Traffic	-2.0%	-2.0%	-2.4%	-2.3%	-1.9%	-1.5%	-2.9%	-3.9%	-3.2%
	PeopleOnStreet	-2.4%	-2.7%	-2.4%	-2.8%	-5.2%	-3.4%	-2.5%	-5.8%	-6.1%
Class B	Kimono	-1.9%	-1.0%	-1.8%	-3.0%	-4.3%	-4.4%	-1.5%	-2.8%	-4.1%
	ParkScene	-0.6%	-0.5%	-0.9%	-0.9%	1.4%	0.5%	-1.3%	-0.4%	-0.1%
	Cactus	-2.4%	-1.5%	-4.5%	-4.1%	-2.3%	-4.9%	-4.3%	-6.8%	-7.3%
	BasketballDrive	-1.9%	-4.7%	-5.2%	-2.5%	-9.1%	-8.5%	-2.3%	-8.0%	-6.9%
	BQTerrace	-2.8%	-2.5%	-2.7%	-4.6%	-2.5%	-4.9%	-7.2%	-4.4%	-5.6%
Class C	BasketballDrill	-4.3%	-7.0%	-8.6%	-3.1%	-10.2%	-11.9%	-3.3%	-11.8%	-13.0%
	BQMall	-4.2%	-3.8%	-4.0%	-4.7%	-4.3%	-4.5%	-4.4%	-5.4%	-5.0%
	PartyScene	-0.9%	-1.3%	-1.8%	-1.4%	0.9%	1.5%	-1.8%	-0.1%	-0.2%
	RaceHorsesC	-1.3%	-1.8%	-3.6%	-2.7%	-3.1%	-7.6%	-2.6%	-3.6%	-7.3%
Class D	BasketballPass	-3.4%	-4.5%	-4.7%	-2.4%	-4.0%	-3.6%	-2.0%	-5.2%	-4.6%
	BQSquare	-1.7%	-0.9%	-2.6%	-1.5%	1.0%	-0.4%	-2.4%	-0.8%	-1.9%
	BlowingBubbles	-1.1%	-2.9%	-3.6%	-1.9%	-2.7%	-0.5%	-2.2%	-3.7%	-4.1%
	RaceHorses	-2.1%	-3.3%	-4.4%	-3.3%	-1.0%	-5.6%	-2.7%	-4.6%	-7.2%
Class E	FourPeople	-3.2%	-2.5%	-1.7%	-4.8%	-5.6%	-4.5%	-5.6%	-5.2%	-4.7%
	Johnny	-4.9%	-3.0%	-1.7%	-6.7%	-7.7%	-5.3%	-8.1%	-6.8%	-5.8%
	KristenAndSara	-3.6%	-2.6%	-2.7%	-5.2%	-5.0%	-4.4%	-6.0%	-7.4%	-5.2%
Class F	BasketballDrillText	-4.4%	-6.7%	-7.8%	-3.3%	-8.2%	-8.5%	-3.7%	-10.3%	-10.8%
	ChinaSpeed	-1.7%	-2.5%	-2.5%	-2.9%	-2.1%	-3.1%	-2.3%	-4.6%	-4.4%
	SlideEditing	-1.9%	-0.5%	-0.8%	-2.1%	-0.2%	-0.4%	-2.1%	-0.5%	-0.8%
	SlideShow	-1.4%	-1.5%	-1.4%	-0.8%	-3.2%	-1.4%	0.0%	-0.7%	-0.9%
Overall		-2.5%	-2.7%	-3.1%	-3.1%	-3.7%	-3.9%	-3.3%	-4.8%	-5.0%

REFERENCES

[1] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

Fig. 8. Visual quality comparison for sequence *Johnny* when ALF is off. Images in the first column are reconstructed with HEVC Anchor, images in the second column are reconstructed with NLSLF-S, and images in the third column are reconstructed with NLSLF-H.

[2] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] S. Wenger, J. Boyce, Y.-W. Huang, C.-Y. Tsai, P. Wu, and M. Li, "Adaptation Parameter Set (APS)," *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-F747, Torino*, Jul. 2011.

[8] J. Zhang, D. Zhao, and W. Gao, "Group-Based Sparse Representation for Image Restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.

[9] 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 vol. 2.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[11] X. Zhang, R. Xiong, S. Ma, and W. Gao, "Reducing Blocking Artifacts in Compressed Images via Transform-Domain Non-local Coefficients Estimation," in *2012 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2012, pp. 836–841.

[12] 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.

[13] X. Zhang, R. Xiong, S. Ma, and W. Gao, "Artifact reduction of

compressed video via three-dimensional adaptive estimation of transform coefficients," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct. 2014, pp. 4567–4571.
[14] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. 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.
[15] X. Zhang, W. Lin, J. Liu, and S. Ma, "Compression noise estimation and reduction via patch clustering," in *Proceedings of APSIPA Annual Summit and Conference*, vol. 16, no. 19, 2015.
[16] M. Matsumura, Y. Bandoh, S. Takamura, and H. Jozawa, "In-loop filter based on non-local means filter," *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-E206, Geneva*, Mar. 2011.
[17] M. Matsumura, S. Takamura, and A. Shimizu, "Largest coding unit based framework for non-local means filter," in *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, Dec. 2012, pp. 1–4.
[18] Q. Han, R. Zhang, W.-K. Cham, and Y. Liu, "Quadtree-based non-local kuans filtering in video compression," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 1044–1055, 2014.

Siwei Ma (M'12) received the B.S. degree from Shandong Normal University, Jinan, China, in 1999, and the Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005.

From 2005 to 2007, he was a Post-Doctorate with the University of Southern California. Then he joined the Institute of Digital Media, EECS, Peking University, where he is currently an Associate Professor. He has published over 100 technical articles in refereed journals and proceedings in the areas of image and video coding, video processing, video streaming, and transmission.

Xinfeng Zhang received the B.S. degree in computer science from Hebei University of Technology, Tianjin, China, in 2007, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014.

He is currently a Research Fellow in Nanyang Technological University, Singapore. His research interests include image and video processing, image and video compression.

Jian Zhang Jian Zhang received B.Sc. degree from Department of Mathematics, Harbin Institute of Technology (HIT), Harbin, China, in 2007, and received M.Eng. and Ph. D degrees from School of Computer Science and Technology, HIT, in 2009 and 2014, respectively. Currently, he is working as a postdoctoral fellow at National Engineering Laboratory for Video Technology (NELVT), Peking University (PKU), Beijing, China. His research interests include image/video coding and processing, compressive sensing, sparse representation, and dictionary learning. He was the recipient of the Best Paper Award at the 2011 IEEE Visual Communication and Image Processing.

Chuanmin Jia received the B.S. degree in computer science from Beijing University of Posts and Telecommunications, Beijing, China, in 2015, and he is currently working toward the Ph.D degree at Institute of Digital Media, EECS , Peking University. His research interests mainly focus on image processing and video compression.

Shiqi Wang (M?5) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in computer application technology from the Peking University, in 2014. He is currently a Postdoc Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. From Apr. 2011 to Aug. 2011, he was with Microsoft Research Asia, Beijing, as an Intern. His current research interests include video compression and image video quality assessment.

Wen Gao (M'92-SM'05-F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He is currently a Professor of computer science with the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing, China. Before joining Peking University, he was a Professor of computer science with the Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing. He has published extensively including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interfaces, and bioinformatics.