Shearlet-based Loop Filter

Johannes Erfurt, Wang-Q Lim, Heiko Schwarz, Detlev Marpe and Thomas Wiegand†

Video Coding & Analytics Department, Fraunhofer Heinrich Hertz Institute (HHI), Berlin, Germany
†Image Communication Chair, Technical University of Berlin, Germany
email: [johannes.erfurt – wang.lim – heiko.schwarz – detlev.marpe – thomas.wiegand]@hhi.fraunhofer.de

***Abstract*—In video coding, in-loop filtering has attracted attention due to its increasing coding performances. In this paper the shearlet-based loop filter is proposed using a sparsifying transform, the shearlet transform, which can identify the important structures of natural images such as edges in the sparse transform domain. This allows for separating efficiently the important information from noise components. Our novel approach for in-loop filtering is to apply a shearlet transform to the decoded image, separating important structures from noise and perform an inverse shearlet transform combined with Wiener filtering. This effectively removes compression artefacts due to quantization noise and keeps the important features of the original image. Simulation results show that our shearlet based loop filter can improve the state-of-the-art video coding standard HEVC through up to 10.5% bit rate reduction along with improved subjective visual quality.**

***Index Terms*—shearlets, sparsity, in-loop filtering, Wiener filtering, classification,**

I. Introduction

The latest video coding standard, HEVC [1], [2] comprised of two in-loop filters. First the deblocking filter (DBF) [3] to mainly reduce blocking artefacts and second the sample adaptive offset (SAO) [4] to further reduce coding artefact after DBF achieving high bit rate reductions. Finally a third in-loop filter, the adaptive loop filter (ALF) [5] was proposed within the standardization process but in the last step dismissed. All loop filters have in common that they take the decoded image as an input and alter the pixel values aiming to reduce the distance between the reconstructed and original samples. The improved image gets in the picture buffer (see Fig. 1) which typically leads to better subsequent predictions, i.e. further coding gain of frames predicted from the current one. The adaptive loop filter approaches this task by classifying each pixel location into different classes. All pixel locations with same gradient direction and intensity are grouped into the same class, which eventually gives a partition of the set of all pixel locations. After that a Wiener filter is applied for each of the classes leading to a de-noised reconstructed image. The filter information is signaled to the bitstream. The drawback of this approach is that through the classification process in image domain quantization noise is not identified and separated and compression artefacts are hardly removed through the following filtering. The main idea of our approach for in-loop filtering is to build classes corresponding to noise components and the important structures of the original image such as edges and textures, followed by applying a Wiener filter for each of those classes. This would effectively perform de-noising, attenuating noise while keeping the important features of the original image through Wiener filtering. The fundamental property of shearlets [6], [7] is that the essential information of the image governed by anisotropic features such as edges can be identified by only a few shearlet coefficients whose absolute values are large while most of its coefficients are nearly zero. This leads to sparse representations for a large class of natural images, which means those images can be sparsely approximated by taking a few significant coefficients. This property is beneficial not only for data compression but also for de-noising, feature extraction, classification and other higher level tasks. The underlying assumption for our approach is that shearlet coefficients with large absolute values of the original image are less sensitive to quantization noise so that they still correspond to the important structures of the original image even after they are corrupted by quantization noise while the rest of the shearlet coefficients essentially corresponds to noise components. This allows us to build two classes, one for important structures and one for noise by simply classifying each shearlet coefficient depending on its absolute value. For that reason the shearlet transform is applied to the decoded image to compute its shearlet coefficients and each of them is classified as *significant* corresponding to a large absolute value and *non-significant* corresponding to a small absolute value. After that classification we apply a Wiener filter for each class to de-noise the distorted image followed by the inverse shearlet transform. Ideally this provokes a suppression of coding artefacts and an amplifying of details in the image.

The paper is structured as follows. In section II we briefly review the basic concept of shearlets. In section III the main procedure of our proposed in-loop filter called shearlet-based loop filter (SLF) is explained. Section IV shows numerical results and analyses the coding performance and conclusions are drawn in section V.

II. Shearlets and Shearlet transform

Shearlets are defined by applying a parabolic scaling matrix A_j, a shear matrix S_k and translation for a fixed 2D function ψ, which is defined on $\mathbb{R}^2$, as follows:

$$\psi_{j,k,m} = |A_j|^{\frac{1}{2}}\, \psi(S_k A_j(\cdot - m)) \qquad (1)$$

where $|A_j|$ is the determinant of A_j and

$$A_j = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{\lceil \frac{j}{2} \rceil} \end{pmatrix}, \quad S_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \quad \text{for } j,k \in \mathbb{Z}, m \in \mathbb{Z}^2.$$

ISBN 978-90-827970-1-5 © EURASIP 2018

While the scale matrix A_j and the translation operator generate shearlets at different scales and positions in the image domain the shearing operator S_k establishes a rotation which changes the orientation of the generating function ψ. Now combining A_j with S_k provides localized anisotropic functions which are well adapted for anisotropic features such as the edge curves of images. This is illustrated in Fig. 3.

In context of image processing a *shearlet system* is a set of *shearlet filters* $(F_1, ..., F_K)$ which can be applied to an image X of dimension $N \times M$ and each filter F_k is a discrete version of a shearlet defined in (1). The number of filters K in the shearlet system is specified by two parameters, the number of *scales* L and the number of directions per scale D. Assume we have $L = 3$ scales and $D = 4$ directions for each scale (it is also possible to have a different number of directions for each scale). Then the shearlet system consists of $3 \cdot 4$ filters and one low-pass filter, in total 13 shearlet filters. Each filter $F_1, ..., F_K$ is applied separately to the image X through the 2D convolution operator, which leads to the *shearlet transform* S:

$$S(X) = ((X * F_k)(m))_{(k,m)\in\Lambda} \tag{2}$$

where Λ is a set of indices for the shearlet coefficients. The index set Λ is given as

$$\Lambda = \{(k, m) : k = 1, \dots, K, \text{ and } m \in \mathcal{D}\}$$

where

$$\mathcal{D} = \{(m_1, m_2) \in \mathbb{Z}^2 : 1 \le m_1 \le N, 1 \le m_2 \le M\}$$

is a set of all pixel locations of X. For a detailed introduction to shearlets we refer the reader to [8]. Fig. 4 illustrates the sparsity of the shearlet transform. Here the shearlet transform corresponding to a shearlet system with 3 scales and 4 direction per scale is performed for an input image resulting in 12 coefficient images (low-pass image omitted). Bright values correspond to high magnitude while dark blue values correspond to small magnitude of shearlet coefficients.

III. Shearlet-based Loop Filter

An essential step of the shearlet-based loop filter is that the de-noising process takes place in the transform domain after applying the shearlet transform. SLF is performed in addition to a coding/decoding procedure after DBF and SAO is performed as illustrated in Fig. 1 and 2. The filter operation takes place within the coding/prediction loop. The in-loop filtering process may reduce the coding noise by filtering the reconstructed image within the shearlet domain with Wiener filters created to minimize the distance from the original image as explained in next subsections. The filter coefficients are coded. The transmission and parsing of the filter information to/from the bitstream is shown in Fig. 1 and 2 as dotted lines.

A. *Shearlet transform and classification*

Assume the shearlet system consists of K shearlet filters $(F_1, ..., F_K)$, where K is specified through the number of scales and directions per scale. We want to point out that the number of shearlet filters is essential for the trade-off between complexity and coding performance as more shearlet filters increase the number of operations to perform the shearlet transform but on the other hand the more redundant the system the more beneficial for de-noising, which will be explained in detail in section V. Once the shearlet coefficients of a decoded image Y are calculated by (2), they are classified as *significant* and *non-significant* coefficients. Then we obtain the following two classes Λ_0 and Λ_1 defined by

$$\begin{aligned}\Lambda_0 &= \{(k, m) : |(Y * F_k)(m)| \le T \ , (k, m) \in \Lambda\}, \\ \Lambda_1 &= \{(k, m) : |(Y * F_k)(m)| > T \ , (k, m) \in \Lambda\} \end{aligned} \tag{3}$$

where a threshold T is either fixed or determined by the distortion of the decoded image Y. In that case it is calculated

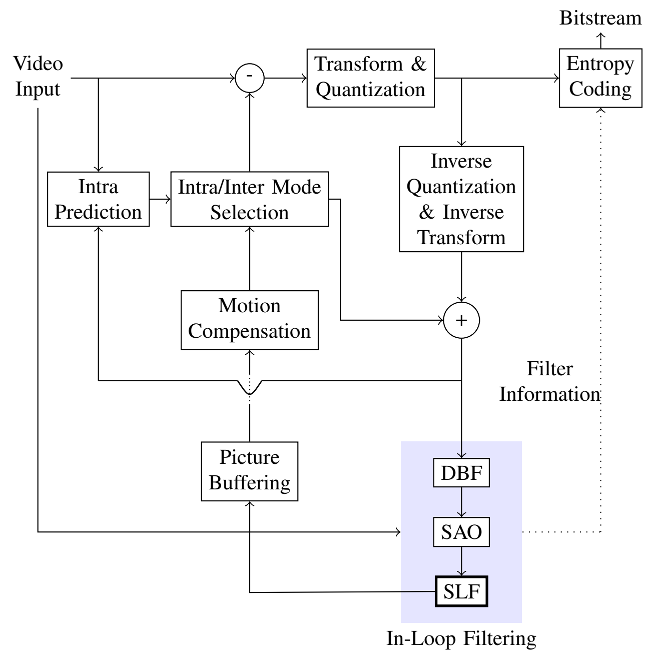

Fig. 1: Video Encoder Block Diagram

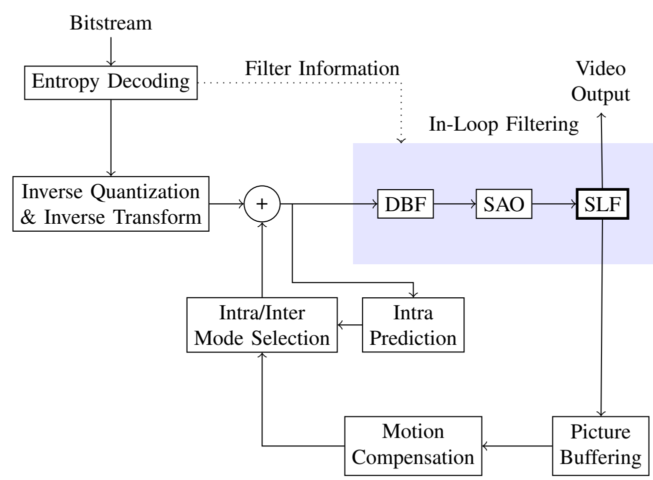

Fig. 2: Video Decoder Block Diagram

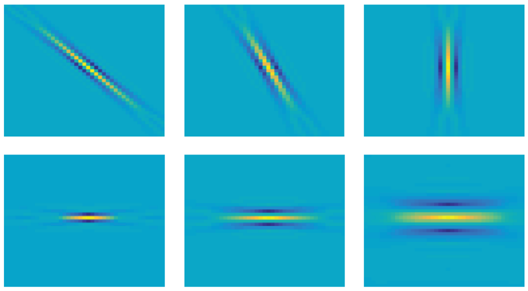

Fig. 3: Example of shearlet functions for different directions at same scale (top row) and different scales with same direction (bottom row).

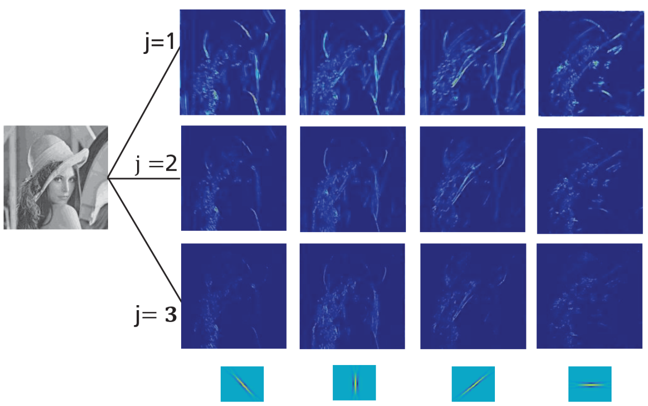

Fig. 4: Shearlet transform performed on an input image resulting in coefficient images for different scales (rows) and different directions (columns).

by

$$T = \kappa \cdot \sqrt{\frac{1}{N \cdot M} \sum_{m \in \mathcal{D}} |X(m) - Y(m)|^2} \quad (4)$$

with some fixed constant κ which can be found experimentally. T will be encoded into the bitstream. Class Λ_0 consists of all non-significant coefficients whose absolute values are smaller or equal to a threshold T. Class Λ_1 consists of all significant coefficients whose absolute values are larger than T.

B. Wiener filtering and Reconstruction process

After the classification of the shearlet coefficients is performed, we compute Φ_1 and Φ_2 by

$$\Phi_r = S^{-1}(\mathcal{X}_{\Lambda_r} \cdot S(Y)) \quad \text{for } r = 0, 1$$

where $\mathcal{X}_{\Lambda_r}$ is the characteristic function defined by

$$\mathcal{X}_{\Lambda_r}(k, m) = \begin{cases} 1, & \text{if } \ (k, m) \in \Lambda_r \\ 0, & \text{otherwise.} \end{cases}$$

c_0	c_4	c_8	c_{12}	c_8	c_4	c_0
c_1	c_5	c_9	c_{13}	c_9	c_5	c_1
c_2	c_6	c_{10}	c_{14}	c_{10}	c_6	c_2
c_3	c_7	c_{11}	c_{15}	c_{11}	c_7	c_3
c_2	c_6	c_{10}	c_{14}	c_{10}	c_6	c_2
c_1	c_5	c_9	c_{13}	c_9	c_5	c_1
c_0	c_4	c_8	c_{12}	c_8	c_4	c_0

			c_0			
		c_1	c_2	c_3		
	c_4	c_5	c_6	c_7	c_8	
c_9	c_{10}	c_{11}	c_{12}	c_{11}	c_{10}	c_9
	c_8	c_7	c_6	c_5	c_4	
		c_3	c_2	c_1		
			c_0			

Fig. 5: Square- and Diamond-shaped Filter

Φ_0 and Φ_1 are reconstructed images associated with the non-significant coefficients and the significant coefficients of a decoded image Y respectively and we have $Y = \Phi_0 + \Phi_1$. Note that important structures of the original image X essentially correspond to the significant shearlet coefficients of X, namely $(X * F_k)(m)$ with $|(X * F_k)(m)| > T$ for some fixed threshold T. Now assume that those significant coefficients are not sensitive to quantization noise so that the significant coefficients of Y essentially correspond to the important structures of X while most of non-significant coefficients of Y correspond to noise components. This implies that reconstructed images Φ_0 and Φ_1 associated with two classes Λ_0 and Λ_1 would essentially contain noise components and important structures of X respectively. Therefore one can effectively remove noise while keeping important structures by applying Wiener filters for Φ_0 and Φ_1 separately. This can be done by the following reconstruction using Wiener filters w_0 and w_1.

$$X_w = w_0 * \Phi_0 + w_1 * \Phi_1 \quad (5)$$

In (5) a pair of Wiener filters $w = (w_0, w_1)$ is chosen so that it minimizes the mean square error between the original image X and the reconstructed image $X_{\tilde{w}}$ with $\tilde{w} = (\tilde{w}_0, \tilde{w}_1)$ where $\tilde{w}_0, \tilde{w}_1$ are arbitrary $n \times n$ filters for some fixed integer n. In other words, $w = (w_0, w_1)$ is given as

$$(w_0, w_1) = \arg\min_{\tilde{w}} \|X - X_{\tilde{w}}\|_2.$$

After the calculation of the Wiener filter coefficients, they are quantized in a similar way as the transform coefficients in standard video coding. These quantized filter coefficients are then coded losslessly, e.g. by entropy coding, and transmitted to the decoder [9]. The novelty is that Wiener filters are not applied to the decoded image Y itself, but rather in the shearlet domain to the functions Φ_0 and Φ_1. Now these functions are filtered by Wiener filters w_0 and w_1. X_w is now forwarded to the picture buffer and later used for subsequent predictions. For the Wiener filters different filter shapes are applicable. The filters could be e.g. square- or diamond-shaped, see Fig. 5. In that example of a 7×7 square-shaped filter only 16 or in case of a 7×7 diamond-shaped filter 13 coefficients for each class have to be sent to the bitstream. In the former case the position of the coefficient c_{15} or in the latter case c_{12} is at the centre position and is the position to be filtered. Both filter examples are rotational symmetric, i.e. if the pattern above is rotated 180 degrees, the coefficients are still in the same position.

ISBN 978-90-827970-1-5 © EURASIP 2018

IV. Simulation Results

For the future video coding standard the Fraunhofer Heinrich Hertz Institute developed a new reference module, the *NextSoftware* [10], [11]. It encodes all HM configurations with equal bit rate compared to HEVC test model reference software (HM version 16.6). We have tested Random Access (RA) for the quantization parameter values 27, 32, 37 and 42 for 2 GOBs (17 frames) and tested on mainly 4K sequences as can be seen in Table I. For the implementation of SLF we used a shearlet system with 65 shearlet filters with 4 scales and 16 directions per scale. For the implementation of the shearlet transform the shearLab [12] was utilized. Two different filter sizes are tested at the encoder, a 1×1 filter and a 7×7 square-shaped filter and then the one with best RD performance is taken. A filter on/off decision flag is signaled into bitstream to determine whether to take SLF or not depending on the RD performance.

Table I shows that 4.21% bit rate reduction can be achieved in comparison to HEVC on the selected set of test sequences, which is 0.8% more than ALF achieves. On top of coding gains there is primarily the subjective improvement to be noted as can be gathered from the test sequence *Johnny*. Here SLF and ALF (Fig. 6) were added to the standard HM configurations (on top of DBF and SAO). In case of SLF coding artefacts are removed especially in textured regions on the right where aliasing artefacts occur due to quantization. Even details in the image are improved (close to the collar). In case of ALF coding artefacts are still visible in many regions.

One drawback of SLF is that the shearlets transform is highly complex due to its redundancy. One possible approach to resolve this complexity issue is to use Graphic processing unit, which could already speed up its run-time significantly [12]. Also, note that the most time consuming part of our implementation of the shearlet transform is to compute a fast Fourier transform, whose complexity can be significantly reduced under certain sparsity assumptions [13]. Based on this approach, we expect a further speed-up is possible using the sparsity of the shearlet transform in the frequency domain. The redundancy of the system has also a significant impact on the coding gain and such a trade-off between complexity and coding gain has to be investigated. E.g. for an image with high variance a shearlet system with many filters are highly beneficial for achieving a good de-noising performance.

V. Future Research and Conclusion

In this paper, we have proposed a shearlet-based loop filter. The main idea of our approach is to decompose a decoded image Y into two components Φ_0 and Φ_1 corresponding to noise and the important structures of the original image X respectively using the shearlet transform. After this, we apply Wiener filtering for Φ_0 and Φ_1 separately followed by a inverse shearlet transform. Our simulation results show high bit rate reductions as well as improvement of subjective visual quality. However a major issue in SLF is that our assumption for the significant shearlet coefficients is not always fulfilled. In fact, the ideal classes for (3) should be given as

$$\tilde{\Lambda}_0 = \{(k,m) : |(X * F_k)(m)| \leq T \ , (k,m) \in \Lambda\},$$
$$\tilde{\Lambda}_1 = \{(k,m) : |(X * F_k)(m)| > T \ , (k,m) \in \Lambda\}. \quad (6)$$

We will explore more sophisticated classifications aiming at the ideal classification (6). In particular, the classification could be performed by taking different features to take the dependencies of neighbouring coefficients into account or trained ones using machine learning algorithms.

TABLE I: Coding Gains of SLF and ALF in JEM-7.0.

Resolution	**Test Sequences**	**SLF**	**ALF**
4096 x 2160	Crosswalk	-5.64%	-1.32%
3840 x 2160	CampfireParty	-1.88%	-1.97%
	Drums100	-4.03%	-3.93%
	Tango	-4.17%	-1.66%
	ToddlerFountain	-2.13%	-2.61%
	CatRobot	-2.34%	-2.00%
	DaylightRoad	-3.63%	-2.47%
	Rollercoaster	-4.13%	-5.18%
	TrafficFlow	-4.24%	-2.80%
	BuildingHall	-3.42%	-3.60%
	FoodMarket	-6.42%	-2.54%
	ParkRunning	-3.59%	-3.80%
2560 x 1600	Traffic	-3.38%	-3.17%
	PeopleOnStreet	-5.04%	-4.82%
	NebutaFestival	-10.50%	-10.67%
	SteamLocomotive	-4.07%	-4.08%
1280 x 720	Johnny	-3.13%	-1.73%
	KristenAndSara	-4.13%	-2.94%
Average		**-4.21%**	**-3.41%**

References

[1] V. Sze, M. Budagavi, G. j. Sullivan, *High efficiency video coding (HEVC) algorithms and architectures*, Springer, 2014.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, *Overview of the high efficiency video coding (HEVC) standard*, IEEE Trans. Circuits and Syst. Video Technol. **22** (2012) No. 12.

[3] P. List, A. Joch, J. Lainema, G. Bjontegaard, M. Karczewicz, *Adaptive deblocking filter*, IEEE Trans. Circuits Syst. Video Technol. **13** (2003) No. 7.

[4] 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, W.-J. Han, *Sample adaptive offset in the HEVC standard*, IEEE Trans. Circuits Syst. Video Technol. **22** (2012) No. 12.

[5] C.-Y. Tsai, C.-Y. Chen, T. Yamakage, I. S. Chong, Y.-W. Huang, C.-M. Fu, T. Itoh, T. Watanabe, T. Chujoh, M. Karczweicz, S.-M. Lei, *Adaptive loop filtering for video coding*, IEEE J. Sel. Topics Signal Process. **7** (2013) No. 6.

[6] G. Kutyniok, J. Lemvig, W. Lim, *Shearlets and optimally sparse approximations, In: Shearlets: Multiscale analysis for multivariate data*, Birkhuser, Springer , 2012.

[7] G. Easley, D. Labate, W. Lim, *Sparse directional image representations using the discrete shearlet transform*, Appl. Comput. Harmon. Analy. **25** (2008) No. 1.

Fig. 6: Subjective image quality comparison: On the left side a reconstructed images with ALF and on the right side with SLF on top of all other in-loop filters (DBF and SAO).

[8] G. Kutyniok(Ed.), D. Labate(Ed.), *Shearlets: Multiscale analysis for multivariate data*, Birkhuser, Springer , 2012.

[9] D. Marpe, H. Schwarz, T. Wiegand, *Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard*, IEEE Trans. Circuits Syst. Video Technol. **13** (2003) No. 7.

[10] A. Wieckowski, T. Hinz, V. George, J. Brandenburg, J. Ma, S De-Luxn-Hernndez, B. Bross, H. Schwarz, D. Marpe, T. Wiegand, *NextSoftware: An alternative implementation the Joint Exploration Model (JEM)*, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JVET-H0084, 8th Meeting: Macao, CN, 18 Oct.-25 Oct. 2017.

[11] A. Wieckowski, T. Hinz, V. George, J. Brandenburg, J. Ma, S De-Luxn-Hernndez, B. Bross, H. Schwarz, D. Marpe, T. Wiegand, *AHG10: Updated NextSoftware as an alternative implementation the Joint Exploration Model (JEM)*, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JVET-I0034r2, 9th Meeting: Gwangju, KR, 20-26 Jan. 2018.

[12] G. Kutyniok, W. Lim, R. Reisenhofer, *Shearlab 3D: Faithful digital shearlet transform based on compactly supported shearlets*, ACM Trans. Math. Software **42** (2016), No. 5.

[13] Piotr Indyk, Michael Kapralov, and Eric Price *(Nearly) Sample-Optimal Sparse Fourier Transform*, SODA '14, Jan. 2014.