# JND-Guided Perceptual Pre-filtering for HEVC Compression of UHDTV Video Contents

Eloïse Vidal[1], François-Xavier Coudoux[1(✉)], Patrick Corlay[1], and Christine Guillemot[2]

[1] Univ. Valenciennes, CNRS, Univ. Lille, YNCREA, Centrale Lille, UMR 8520 - IEMN, DOAE, 59313 Valenciennes, France
coudoux@univ-valenciennes.fr

[2] Inria, Campus Universitaire de Beaulieu, 35042 Rennes, France

**Abstract.** In this paper, two new perceptual filters are presented as pre-processing techniques to reduce the bitrate of HEVC compressed Ultra high-definition (UHD) video contents at constant visual quality. The proposed perceptual filters rely on two novel adaptive filters (called BilAWA and TBil) which combine the good properties of the bilateral and Adaptive Weighted Averaging (AWA) filters. Moreover, these adaptive filters are guided by a just-noticeable distortion (JND) model to adaptively control the strength of the filtering process, taking into account the properties of the human visual system. Extensive psychovisual evaluation tests conducted on several UHD-TV sequences are presented in detail. Results show that applying the proposed pre-filters prior to HEVC encoding of UHD video contents lead to bitrate savings up to 23% for the same perceived visual quality.

## 1 Introduction

Digital video formats have widely evolved in recent years and we have gone from the Standard Definition Television (SDTV) to High Definition Television (HDTV) then today the Ultra High Definition Television (UHDTV). This has led to a tremendous increase of visual data to be stored or transmitted and several digital video compression standards have been proposed by broadcast engineers to solve this problem. In particular, the newly developed High Efficiency Video Coding (HEVC) standard has come as the ideal solution to allow the wide deployment of UHDTV services for customers [1]. It is common to say that HEVC allows dividing by two the output compressed video bit rate compared to its H.264/AVC predecessor, for the same reconstructed video quality. Such performances can further be improved by optimizing the HEVC coding tools or by facilitating the encoding step by means of pre-filtering. The aim of any pre-filtering solution is to remove the spurious noise and insignificant details from the original video material in order to ease the compression. For a given visual quality, pre-filtering results in better coding efficiency. Hence, conventional denoising filters have been early used to prefilter original video contents prior to the encoding stage [7,28]. Such denoising filters can be controlled by the

encoder parameters like motion vectors and residual signal's energy to control the strength of a lowpass filter applied before compression [14,15]. The prefiltering process can be applied in the pixel domain [30], or in the frequency domain [16,17].

Recently, there has been a renewed interest in digital video pre-filtering and several pre-filtering algorithms have been proposed in the literature to help the encoding stage by reducing the high-frequency content prior to quantization [2,8,18]. In particular, several studies have proposed to control perceptually the low-pass prefilter hence reducing the visually insignificant information [9,19,25,29]. In [26,27], the authors proposed to control an anisotropic filter by a contrast sensitivity map. The proposed pre-processing filter is applied prior to H.264 encoding and has the particularity of depending on a number of display parameters. Recently, the authors in [9] proposed an original method for preprocessing the residual signal in the spatial domain, which integrates a HVS-based color contrast sensitivity model. Just-Noticeable Distortion (JND) models have also be employed to control a prefilter before compression [22,23], or to adapt the quantization stage to reduce the non visible frequency coefficients in H.264 [20] and HEVC [2,21,24].

In this paper, we present new adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding. These new pre-filters are derived from the well-known Adaptive Weighted Averaging (AWA) [4] and bilateral [5] filters and are implemented in the pixel domain. They are guided by means of a just-noticeable distortion (JND) model in order to be visually not perceptible (no excessive blur). A detail study of these prefilters with different experimental results obtained with HDTV contents has been given in [3]. Bit rate savings of about 20% are presented for the same perceived video quality. We extend the study of the performances of the pre-filtering algorithm to the case of HEVC compression of UHDTV contents. Since the visual quality constitutes the primordial criterion to validate any pre-filtering technique, we chose to validate our study in a privileged way on the basis of extensive subjective evaluation tests. This is an originality of this work because, to the best of our knowledge, there are very few papers that implement thorough subjective assessment of the quality of pre-processing techniques while it represents the ground truth. The paper is organized as follows: first, the pre-filtering solutions are described in detail. Then the results of applying the pre-filters on UHDTV sequences are presented. In particular, we have based the evaluation of the performances on a large set of subjective evaluation tests performed with respect to the methodology recommended by international standards. The test protocol is presented in detail and the subjective assessment results are given and analyzed. Results show that the pre-filter makes it possible to obtain significant bit rate reductions of the order up to 23% for the same perceived quality. Such results are comparable to those recently described in the literature [6]. Concluding remarks are drawn in Sect. 4.

## 2   Description of the Pre-filtering Algorithm

We propose a low complexity external pixel-domain pre-filtering approach controlled by an a priori model which does not need information from a first pass encoding. The pre-filtering step consists of removing from each image of the video sequence the non-essential data that will not be perceived by the human visual system (HVS). This is achieved by applying a low-pass filter whose response varies locally according to the perceptual relevance of the image content. To do that, the algorithm first computes in the pixel domain a just noticeable distortion (JND) map for the luminance component of each image. Then, the JND value at each pixel allows determining the strength of the smoothing operation. Imperceptible details and fine textures are consequently smoothed, hence saving bitrate without compromising video quality. We proposed two novel adaptive filters (called BilAWA and TBil) which combine the good properties of the well-known bilateral [5] and AWA [4] filters.

The JND-guided BilAWA filter is therefore defined as

$$h_{i,BilAWA}^{JND} = h_{g,BilAWA}(\underline{x}, \underline{x}_i)h_{s,BilAWA}^{JND}(\underline{x}, \underline{x}_i), \tag{1}$$

$$h_{g,BilAWA}(\underline{x}, \underline{x}_i) = exp(-\frac{||\underline{x} - \underline{x}_i||^2}{2\sigma_g^2}), \tag{2}$$

$$h_{s,BilAWA}^{JND}(\underline{x}, \underline{x}_i) = \frac{1}{1 + a\max(JND^2(I(\underline{x})), ||I(\underline{x}) - I(\underline{x}_i)||^2)}, \tag{3}$$

where $h_{g,BilAWA}$ denotes the geometric kernel of variance $\sigma_g^2$ which is a function of the spatial distance. $h_{s,BilAWA}^{JND}$ is the similarity kernel guided by the JND value $JND(I(\underline{x}))$. $I(\underline{x})$ is the amplitudinal value of pixel at position $\underline{x} = (x, y)$.

The JND-guided TBil filter is described as

$$h_{i,TBil}^{JND} = h_{g,TBil}(\underline{x}, \underline{x}_i)h_{s,TBil}^{JND}(\underline{x}, \underline{x}_i), \tag{4}$$

$$h_{s,TBil}^{JND}(\underline{x}, \underline{x}_i) = exp(-\frac{||I(\underline{x}) - I(\underline{x}_i)||^2}{2 * JND^2(I(\underline{x}))}), \tag{5}$$

where the geometric kernel is the same as $h_{g,BilAWA}$ described in Eq. 2. The threshold of the similarity kernel in Eq. 5 is chosen such that every value between 1 and $JND^2$ has the same weight.

Figure 1(a) illustrates the similarity kernel evolution of the BilAWA and TBil filters. Figure 1(b) illustrates the impact of the geometric and similarity kernel on the BilAWA filter (similar results are obtained for TBil filter).

A thorough experimental analysis of these perceptual JND-guided pre-filters show that average bitrate savings of about 19% for H.264/AVC and 17% for HEVC, can be reached for the same perceived visual quality for HDTV contents. Further details about the design of the BilAWA and TBil filters can be found in [3]. Figure 2 illustrates the performances of the BilAWA prefilter on the Coast guard sequence encoded by means of H.264 (QP = 28). It can be noticed that the BilAWA prefilter preserves the sharpness of the image with no excessive blur. In
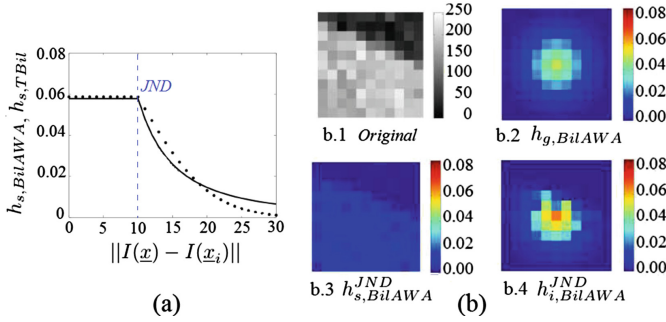
**Fig. 1.** (a) Comparison between the BilAWA similarity kernel (solid line) and the TBil similarity kernel (point) for a particular JND value of 10. (b) Illustration of the impact of the similarity kernel (b.3) and the geometric kernel (b.2) on the weights of the BilAWA filter (b.4) for a particular pixel and its $11 \times 11$ neighbors (b.1).

the present case, we verify that the prefiltering process allows a bit rate saving of about 30% after compression with the same visual quality.

As described previously, the strength of the prefiltering operation is controlled thanks to a JND metric. In cognitive psychophysiology of perception, discrimination threshold, also called Just-Noticeable Difference or Distortion (JND), refers to the smallest discernible discrepancy between two values of a stimulus. First used extensively in audio processing, the JND has been also used in digital video processing since the late 1990s [31]. A method of determining the JND for the video given by [32] is to present to a group of observers an original image and progressively increasing degraded versions. The image from which at least 75% of the observers perceive a difference gives us the threshold corresponding to one JND. The experiment can then be repeated by replacing the original image with the one-JND image to obtain the image corresponding to two JNDs and so on.

The method proposed by [32] is however difficult to implement and the results obtained strongly depend on the type of content used. This is why models of JND were developed. Whether in audio or video, the JND models always use the so called masking effects, specific to the human auditory and visual system. In our work, we choose the JND model defined by Yang in [22] because it is defined in the pixel domain and is therefore well suited to spatial domain filtering, especially in the context of real-time processing. This model is briefly described hereafter.

The spatio-temporal JND metric developed by Yang et al. is easy to implement and gives a satisfying precision. It is composed of two parts: the first part accounts for the spatial masking effect. Spatial masking takes into account the fact that the human eye is little sensitive to differences in the dark areas of an image according to the well-known Weber-Fechner curve (luminance masking), and also very sensitive to the edge information (texture masking). In practice, texture masking is derived from edge extraction using oriented Canny operators. The second part is related to the temporal masking effect. Temporal masking

refers to the fact that the system is more sensitive to differences in the stationary areas of a video. The temporal JND takes account of this characteristic and is evaluated from the frame difference. Finally a spatio-temporal JND value is obtained for each pixel. Hence, the higher the value, the stronger the smoothing operation applied to the pixel should be.



**Fig. 2.** Illustrative example of the performances of the new JND-guided pre-filters (Coast guard sequence, QP28): original and BilAWA pre-filtered version with a bit rate saving of about 30%.

## 3   Experimental Results

Extensive subjective visual assessment tests have been conducted in order to validate the performances of the pre-filtering algorithms in the context of HEVC compression of Ultra HD contents, as the visual judgment of human observers remains the ultimate criterion. These subjective assessment tests should respect to the methodology recommended by international standards in terms of video test material, viewing conditions or evaluation scale. The test protocol is detailed in the following paragraphs.

### 3.1   Video Test Material

The UHDTV and 4K uncompressed test sequences used in the experiments are described in Fig. 3. Three sequences named Artic, Boat and Tahiti are issued from broadcast professional TV contents. The fourth sequence is an excerpt of the film Tears of Steel available on the Xiph.org web site [10]. In order to ensure that the sequences were representative of sufficiently varied situations, the spatial information (SI) and temporal information (TI) indexes [11] have been calculated for each sequence. Results given in Fig. 3 show that the selected sequences make it possible to largely cover the variety of spatial and temporal activities. All the sequences are of size $3840 \times 2160$ pixels and 420 chrominance format.

(a)                                              (b)

**Fig. 3.** (a) Test video sequences used in the experiments, clockwise from top left: Artic, Boat, Tears, Tahiti. (b) SI TI index.

## 3.2   Encoding Setup

The HEVC encoding of the UHD video sequences has been done using the x265 compression software. The Main Profile configuration of the codec has been used, with both deblocking and SAO loop filters applied and CTU up to $64 \times 64$ pixels. In order to analyze the gain brought by the proposed JND-guided pre-filters, different QP values have been considered varying from 27 to 41. The GOP structure was fixed to the IBBP one with a GOP length of 12 frames.

## 3.3   Subjective Assessment Test Protocol

The projection conditions of the subjective evaluation test sessions were made as possible consistent with the experimental conditions recommended by ITU Recommendation P.911 (Table 1). Among the different protocols preconized by the ITU P.911 Recommendation [12], the pairwise comparison (PC) methodology has been retained as the most relevant one in order to discriminate two systems with closed performances (Fig. 4).

**Table 1.** Experimental viewing conditions.

| Parameter | ITU Recommendation | Measured value |
|---|---|---|
| Size of the projection surface | Specify L × H | 4.45 m × 2.25 m |
| Viewing distance | 2H to 8H | 5H (row 1) or 6H (row2) |
| Screen luminance | 100 $cd/m^2$ | 48 $cd/m^2$ (CNC standard) |
| Ambient luminance | < 20 lux | 8 lux |
| Background chromaticity | D65 (x = 0.3127 y = 0.3290) | x = 0.3780 y = 0.3752 |

The video sequences have been encoded by means of HEVC with fixed QP value. Moreover, the PC methodology is based on a forced choice, which will make it possible to quickly check if the pre-filtering process induces a visible discomfort for the viewers.
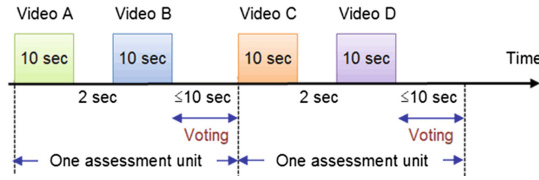
**Fig. 4.** Subjective evaluation protocol.

We choose a seven-level scoring scale (Fig. 5) that adds a nuance to the observer's judgment and limits also the recognition bias. Moreover, it introduces a notation at 0 which allows the observer to indicate if he/she perceives no difference. Finally, a panel of sixteen male and female observers has participated to the subjective assessment tests. A training session allows the participants to be familiar with the test protocol. After subjective evaluation tests, statistical analysis is applied on the raw data in order to compute the resulting mean opinion score (MOS) and associated 95% confidence interval.
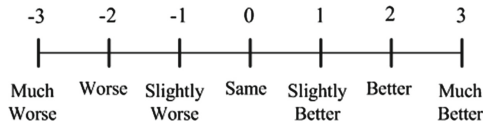


**Fig. 5.** Subjective evaluation scale.

### 3.4    Performances Analysis

Experimental results obtained thanks to both perceptual adaptive filters are presented in what follows. Figure 6 presents the bit rate saving results versus the CMOS for the BilAWA filter and the TBil filter. Table 2 presents the average results for all the sequences at each QP.

We can note that the two proposed pre-filters offer bit rate savings for the four UHDTV test sequences. These bit rate savings are very significant for low QP values, with a maximum of 23% for the Boat sequence processed by the BilAWA filter and then encoded with QP=27. In any case, the BilAWA filter is more efficient than the TBil one. The efficiency of the two pre-filters decreases as the QP value increases. For high QP values, the bit rate reduction is often moderate (around 5%). It might be due to the fact that stronger compression eliminates details anyway so that it removes the benefits of the pre-filtering step.

Considering the subjective visual assessment tests, the results demonstrate that both pre-filters do not affect the visual judgment of the observers. The MOS values are almost zero with a small 95% confidence interval. Hence, the proposed JND-guided adaptive filters allow reducing the bit rate while keeping
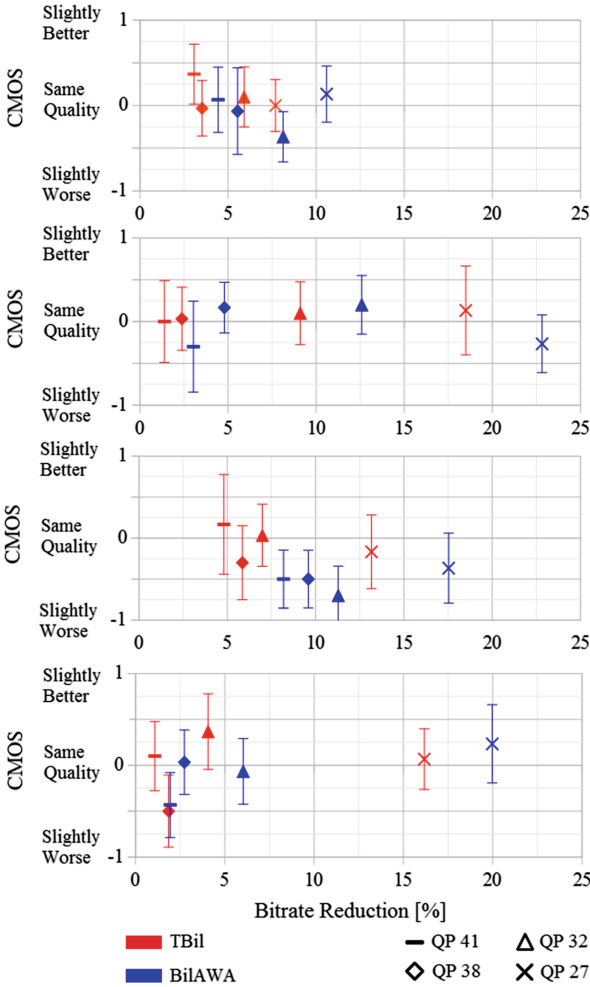
**Fig. 6.** Evaluation of proposed pre-filters. CMOS versus bitrate saving for BilAWA and TBil perceptual pre-filters in comparison with encoding scheme without pre-filters for ×265 encoding. From top to bottom: Artic, Boat, Tahiti and Tears test sequence.

the same perceived video quality. In addition to psychovisual evaluation tests, we have also considered some objective quality metrics. Among these, the Structural Similarity Image Measure (SSIM) has been retained because the SSIM is widely used in the digital video community [13]. When comparing the SSIM values of the compressed video sequences w/o pre-filtering, the average difference is between 0.0009 and 0.0041, showing that the two compressed sequences are perceptually undistinguishable (Table 2).

**Table 2.** Comparison of encoding performances of x265 codec with and without perceptual pre-filters. Analysis of bitrate reduction ($\Delta$Bitrate), subjective quality evaluation (CMOS and the confidence interval $\delta_{[95\%]}$) and associated objective measures variation($\Delta$SSIM). Average result on all sequences. x265 is used with the loop filters at GOP IBBP12.

|  |  | QP 27 | QP 32 | QP 38 | QP 41 |
|---|---|---|---|---|---|
|  | Bitrate [Mbs] | 7.92 | 3.44 | 1.46 | 1.03 |
|  | SSIM | 0.964 | 0.957 | 0.946 | 0.939 |
| BilAWA | **$\Delta$Bitrate [%]** | **−17.73** | **−9.51** | **−5.67** | **−4.40** |
| 11 × 11 | **CMOS** | **−0.07** | **−0.23** | **−0.09** | **−0.29** |
| JND | **± $\delta_{[95\%]}$** | **0.19** | **0.19** | **0.20** | **0.22** |
| ±x265 | $\Delta$SSIM * 100 | −0.410 | −0.272 | −0.201 | −0.167 |
| TBil | **$\Delta$Bitrate [%]** | **−13.88** | **−6.51** | **−3.21** | **−2.78** |
| 11 × 11 | **CMOS** | **0.01** | **0.15** | **−0.05** | **0.01** |
| JND | **± $\delta_{[95\%]}$** | **0.21** | **0.19** | **0.19** | **0.24** |
| ±x265 | $\Delta$SSIM * 100 | −0.290 | −0.168 | −0.109 | −0.085 |

We note a relation between the bitrate saving and the spatial and temporal index (Fig. 3(b)). The biggest bitrate reduction is obtained for the sequence Boat with the higher spatial and temporal activity.

## 4    Conclusion and Further Work

In this paper, we have analyzed the performances of two adaptive filters (BilAWA and TBil) guided by a JND model in the case of HEVC compression of UHDTV contents. The validation of the pre filtering techniques is based on extensive subjective evaluation tests. The introduction of a JND model leads to perceptually lossless adaptive filters which exhibit a strong interest to improve UHD real-time video compression efficiency by removing imperceptible details. The two proposed pre-filters offer significant bit rate savings for the UHDTV used test sequences. A maximum average bit rate saving of up to 23% has been obtained with the BilAWA filter for low QP values. We verify that the BilAWA filter is more efficient than the TBil one. Despite experimental results are given with the JND model developed by Yang et al., it should be noted that the proposed pre-filters are independent of the chosen model: the filtering parameters could as well be controlled by other more sophisticated JND pixel-domain models. Further work will concern the improvements of the JND model including chrominance sensitivity and visual saliency, as well as the refinement of the pre-filters parameters.

# References

1. Sullivan, G.J., Ohm, J.-R., Han, W., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Trans. Circu. Syst. Video Technol. **22**(12), 1649–1668 (2012)
2. Bae, S.H., Kim, J., Kim, M.: HEVC-based perceptually adaptive video coding using a DCT-based local distortion detection probability model. IEEE Trans. Image Process. **25**(7), 3343–3357 (2016)
3. Vidal, E., Sturmel, N., Guillemot, C., Corlay, P., Coudoux, F.: New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding. Sig. Process. Image Commun. **52**, 124–137 (2017)
4. Ozkan, M., Sezan, I., Tekalp, M.: Adaptive motion-compensated filtering of noisy image sequences. IEEE Trans. Circ. Syst. Video Technol. **3**(4), 277–290 (1993)
5. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV, pp. 836–846 (1998)
6. Vanam, R., Kerofsky, L., Reznik, Y.: Perceptual pre-processing filter for video on demand content delivery. In: IEEE ICIP, pp. 2537–2541, 27–30 October (2014)
7. Karunaratne, P., Segall, C., Katsaggelos, A.: A rate-distortion optimal video pre-processing algorithm. IEEE ICIP, vol. 1, pp. 481–484 (2001)
8. Buades, A., Lisani, J.L., Miladinovic, M.: Patch-based video denoising with optical flow estimation. IEEE Trans. Image Process. **25**(6), 2573–2586 (2016)
9. Shaw, M.Q., Allebach, J.P., Delp, E.J.: Color difference weighted adaptive residual preprocessing using perceptual modeling for video compression. Sig. Process. Image Commun. **39**, 355–368 (2015)
10. Test video sequences available at the Xiph web site. https://media.xiph.org/
11. Pinson, M.H., Wolf, S.: A new standardized method for objectively measuring video quality. IEEE Trans. Broadcast. **50**(3), 312–322 (2004)
12. ITU-T Recommendation P.911: Subjective audiovisual quality assessment methods for multimedia applications. Series P, December (1998)
13. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
14. Lee, J.: Automatic prefilter control by video encoder statistics. IET Electron. Lett. **38**, 503–505 (2002)
15. Jain, C., Sethuraman, S.: A low-complexity, motion-robust, spatio-temporally adaptive video de-noiser with inloop noise estimation. In: IEEE International Conference on Image Processing, pp. 557–560, October (2008)
16. Song, B., Chun, K.: Motion-compensated temporal filtering for denoising in video encoder. Electron. Lett. **40**, 802–804 (2004)
17. Varghese, G., Wang, Z.: Video denoising based on a spatiotemporal Gaussian mixture model. IEEE Trans. Syst. Circ. Video Technol. **20**(7), 1032–1040 (2010)
18. Segall, C.A., Karunaratne, P., Katsaggelos, A.K.: Pre-processing of compressed digital video. In: SPIE Image and Video Communication and Processing (2001)
19. Shao-Ping, L., Song-Hai, Z.: Saliency-based fidelity adaptation preprocessing for video coding. J. Comput. Sci. Technol. **26**, 195–202 (2011)

20. Naccari, M., Pereira, F.: Advanced H,264/AVC-based perceptual video coding: architecture, tools, and assessment. IEEE Trans. Circ. Syst. Video Technol. **21**(6), 766–782 (2011)
21. Naccari, M., Pereira, F.: Integrating a spatial just noticeable distortion model in the under development HEVC codec. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 817–820, 22–27 May (2011)
22. Yang, X., Lin, W., Lu, Z., Ong, E., Yao, S.: Motion-compensated residue pre-processing in video coding based on just-noticeable-distortion profile. IEEE Trans. Circ. Syst. Video Technol. **15**(6), 742–752 (2005)
23. Ding, L., Li, G., Wang, R., Wang, W.: Video pre-processing with JND-based Gaussian filtering of superpixels. In: Proceedings of the SPIE 9410, Visual Information Processing and Communication VI, vol. 941004, 4 March (2015)
24. Kim, J., Kim, M.: An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels. IEEE Trans. Syst. Circ. Video Technol. **25**(11), 1786–1800 (2015)
25. Wang, S., Rehman, A., Wang, Z., Ma, S., Gao, W.: Perceptual video coding based on SSIM-inspired divisive normalization. IEEE Trans. Image Process. **22**(4), 1418–1429 (2013)
26. Kerofsky, L.J., Vanam, R., Reznik, Y.A.: Improved adaptive video delivery system using a perceptual pre-processing filter. In: Proceedings of the GlobalSIP 2014, pp. 1058–1062 (2014)
27. Vanam, R., Reznik, Y.A.: Perceptual pre-processing filter for user-adaptive coding and delivery of visual information. In: Proceedings of the PCS 2013, December (2013)
28. Al-Shaykh, O., Mersereau, R.: Lossy compression of noisy images. IEEE Trans. on Image Process. **7**(12), 1641–1652 (1998)
29. Oh, H., Kim, W.: Video processing for human perceptual visual quality-oriented video coding. IEEE Trans. Syst. Circ. Video Technol. **22**(4), 1526–1535 (2016)
30. Guo, L., Au, O., Ma, M., Wong, P.: Integration of recursive temporal LMMSE denoising filter into video codec. IEEE Trans. Circ. Syst. Video Technol. **20**(2), 236–249 (2010)
31. Watson, A.B.: DCT quantization matrices visually optimized for individual images. In: Proceedings of the SPIE International Society for Optical Engineering, vol. 1913, pp. 202–216 (1993)
32. Watson, A.B.: Measurement of a JND scale for video quality. VQEG Final Report (2000)