



New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding

Eloïse Vidal, Nicolas Sturm, Christine Guillemot, Patrick Corlay,
François-Xavier Coudoux

► To cite this version:

Eloïse Vidal, Nicolas Sturm, Christine Guillemot, Patrick Corlay, François-Xavier Coudoux. New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding. Signal Processing: Image Communication, 2017, 52, pp.124 - 137. 10.1016/j.image.2016.12.003 . hal-01591984

HAL Id: hal-01591984

<https://hal.science/hal-01591984>

Submitted on 22 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

New Adaptive Filters as Perceptual Preprocessing for Rate-Quality Performance Optimization of Video Coding

Eloïse Vidal, *Member IEEE*, Nicolas Sturm, Christine Guillemot, *Fellow IEEE*, Patrick Corlay and François-Xavier Coudoux, *Member IEEE*

Abstract—In this paper, we introduce two perceptual filters as pre-processing techniques to reduce the bitrate of compressed high-definition (HD) video sequences at constant visual quality. The goal of these perceptual filters is to remove spurious noise and insignificant details from the original video prior to encoding. The proposed perceptual filters rely on two novel adaptive filters (called BilAWA and TBil) which combine the good properties of the bilateral and Adaptive Weighting Average (AWA) filters. The bilateral and AWA filters being initially dedicated to denoising, the behavior of the proposed BilAWA and TBil adaptive filters is first analyzed in the context of noise removal on HD test images. The first set of experimental results demonstrates their effectiveness in terms of noise removal while preserving image sharpness. A just noticeable distortion (JND) model is then introduced in the novel BilAWA and TBil filters to adaptively control the strength of the filtering process, taking into account the human visual sensitivity to signal distortion. Visual details which cannot be perceived are smoothed, hence saving bitrate without compromising perceived quality. A thorough experimental analysis of the perceptual JND-guided filters is conducted when using these filters as a pre-processing step prior to MPEG-4/AVC encoding. Psychovisual evaluation tests show that the proposed BilAWA pre-processing filter leads to an average bitrate saving of about 19.3% (up to 28.7%) for the same perceived visual quality. The proposed new pre-filtering approach has been also tested with the new state-of-the-art HEVC standard and has given similar efficiency in terms of bitrate savings for constant visual quality.

Index Terms—Video Coding; Pre-processing; Image Filtering; Adaptive Weighting Average (AWA) Filter; Bilateral Filter; Just Noticeable Distortion (JND); Image Quality.

I. INTRODUCTION

IT is well-known that removing spurious noise or attenuating perceptually insignificant details by video filtering prior to encoding can improve the rate-quality performance of encoders [1]. Traditional noise filtering approaches using linear filters which compute the value of the filtered image as a weighted average of pixel values in the neighborhood are often employed [2], [3], [4], [5]. To cite a few examples, in conventional Gaussian low-pass filtering, the weights decrease with the distance from the filtered pixel. Nearby pixels generally share a similar value due to slow variations of luminance over space. Averaging them is a way of increasing the spatial correlation, hence compression efficiency, while introducing a negligible distortion.

However, in areas where the assumption of stationarity is not verified (e.g. corners, edges), the linear filtering will not only attenuate noise but will also lead to a strong attenuation of the high frequency structures and introduce blur. Therefore, there has been a remarkable effort to find nonlinear and adaptive operators which would smooth or increase correlation in smooth areas and at the same time better preserve image structures. Most adaptive filtering techniques use the standard deviation of those pixels within a local neighborhood to calculate a new pixel value. These methods include

anisotropic diffusion [6], bilateral filtering [7], [8] and adaptive weighted averaging [9]. Anisotropic diffusion uses the gradient of the image to guide the diffusion process, avoiding smoothing the edges [6]. Bilateral filtering first introduced in [7] is a non-linear filtering technique utilizing both the spatial and photometric distances to better preserve signal details. The link between anisotropic diffusion and bilateral filtering has been established in [10]. Bilateral filtering is actually the product of two local filters, one based on a measure of similarity between the pixel amplitudes - e.g. luminance channel of colored images - in a local neighborhood and the other one based on a geometric spatial distance. Both kernels are Gaussian kernels. An Adaptive Weighted Averaging (AWA) filter is proposed in [9] and used for motion-compensated filtering of noisy video sequences. Given its use in the temporal dimension, the dimension of the AWA filter support is in general small. It has also been successfully used in adaptive filtering [11].

This paper addresses the question of choosing a real time adaptive pre-processing filter prior to encoding which would maximize the bitrate saving while preserving the visual quality. The out-loop prefiltering approach applied prior to the encoding stage has been retained in the present work because it has the great advantage to be *universal*, i.e. it can easily be applied to any video encoder. Note here that we stress the term 'real time' as the chosen implementation is expected to meet high performance (30 or 60 fps high-definition video filtering) on a standard workstation. This is the reason why recently proposed denoising filter methods, such as the Non Local Means algorithm [12], will not be considered as they provide very high performance but at the cost of even higher computational time. Wavelet based methods will also not be accounted for [46], as they provide a complexity which is much higher than conventional spatial filtering. On top of that, any successful filtering done in the spatial domain can also be applied to the wavelet domain as it was successfully done in [13] with bilateral filtering.

The contributions of the paper are two-folds:

- 1) We first consider the well known AWA and bilateral filters and search for the best compromise between denoising performance (minimal absolute distance to the original) and lowest subjective visual distortion (lack of sharpness). The weights of the AWA filter start decreasing once the difference between the pixel luminance values exceeds a given threshold. Above this threshold, the decaying rate of the AWA filter is slow, which leads to a stronger smoothing effect when increasing the size of the filtering kernel. In contrast, the weights of the bilateral similarity kernel decay faster both in the similarity and geometric dimensions. The bilateral filter therefore better preserves edges and textures than the AWA filter does when the size of the support increases.

These observations naturally led us to introduce two novel adaptive filters designed around the bilateral paradigm (geometric kernel + similarity kernel) with however different approaches. These filters, called BilAWA and Thresholded Bilateral (TBil), improve the paradigm of bilateral filtering, and are fast enough for real-time computation. They combine the good properties of the AWA and bilateral filters and enable a larger filtering support, with the aim

E. Vidal and N. Sturm are with Digigram S.A., 38330 Montbonnot, France (eloise.vidal@gmail.com, sturm@digigram.com)

C. Guillemot is with INRIA, Campus Universitaire de Beaulieu, 35042 Rennes, France

P. Corlay and F.-X. Coudoux are with DOAE, IEMN (UMR CNRS 8520), University of Valenciennes, 59313 Valenciennes, France

to increase the noise and insignificant details removal performance while preserving the image structure and textures.

The proposed filters are first studied in a context of noise removal in High Definition (HD) images, using four different quality metrics: two of them express the distance in regards to the inverse problem of denoising, while the two others describe the overall sharpness quality of the filtered images. Experimental results show that the two proposed adaptive filters outperform the AWA and bilateral filters in terms of denoising while preserving the image texture and structure.

2) Perceptual pre-filtering is derived from the novel filters by integrating a Just Noticeable Distortion (JND) model to vary the filtering strength according to the visual significance of the image content. Indeed, the first set of experiments shows that the TBil and BilAWA filters offer the best compromise between noise removal and preservation of image sharpness compared to the original bilateral and AWA filters. This led us to retain these two filters to design the perceptual pre-filtering technique. To further remove details which are perceptually insignificant, a Just Noticeable Distortion model (JND) [14] is introduced to control the strength of the filters. Note that, as the filters themselves are not dependent on the JND model, other JND pixel-domain models, such as for example described in [16], could be used. The JND-guided BilAWA and TBil filters have been used as prefilters prior to MPEG-4/AVC encoding for increasing the rate/quality performances. Experimental results based on psychovisual evaluation tests show that a significant rate saving (up to 28.7% and 19.3% on average) can be achieved thanks to the pre-filters with equivalent perceived visual quality. The proposed new pre-filtering approach has been also tested with the new state-of-the-art HEVC standard and has given similar efficiency in terms of bitrate savings for constant visual quality. These average bitrate reduction ratios are comparable with some others obtained with methods recently described in the literature [52][49][48].

The paper is organized as follows. Section II gives the background in video pre-processing for compression as well as a brief overview of the original AWA and bilateral filters. Section III is dedicated to image quality discussion. Section IV presents the proposed adaptive filters based on the bilateral filtering paradigm and analyzes their behaviour for noise removal. Section V introduces the novel perceptual filters and analyzes their behaviour in light of the JND maps. Section VI presents the experimental protocol and discusses the results obtained with both perceptual filters under different MPEG-4/AVC and HEVC codec configurations (quantization parameter values and GOP structures). Section VII gives the conclusion and further work.

II. PRE-PROCESSING FOR VIDEO COMPRESSION: BACKGROUND

A. Pre-filtering for rate saving

Since the beginning of image and video compression codecs, conventional denoising filters have been applied prior to the encoding stage in order to reduce the undesirable high frequency content which degrades the encoder performances [4]. The external denoising filter can be controlled by the encoder parameters in a two-pass encoding process. For example, the authors of [17] exploit the motion vectors and the residual information energy to control the strength of a low-pass filter applied before a MPEG-2 coder. In contrast the authors of [18] use the motion vectors to reduce the complexity of a motion compensated 3-D filter. The denoising process can be embedded into the encoder by applying a spatial denoising filter on the residual information [19], or a frequential-domain filter on the wavelet [46] or DCT coefficients [20] [21]. Recently, the proposed denoising algorithm in [45] combines motion estimation using optical flow and patch processing.

Even when the video sequences are not altered by the presence of noise, which is a common situation in video professional applications,

adaptive low-pass filters are still useful to help the encoding stage at low bitrates by reducing the high-frequency content before the quantization stage [22] [23]. However, the adaptive filters suffer from smoothing effect which is annoying for high quality applications.

These observations have motivated several studies to control perceptually a low-pass filter hence reducing the visually insignificant information. Thus, a saliency map is employed in [24] to control both gaussian kernels of a bilateral filter to smooth the non saliency part of a video sequence prior to H.264 encoding. In [47], the authors present a video processing scheme based on a foveation filter which is guided by a sophisticated perceptual quality-oriented saliency model. In [53] [25] [52], an anisotropic filter controlled by a contrast sensitivity map is applied before x264 encoding. The proposed pre-processing filter has the particularity of depending on a number of display parameters, e.g. the viewing distance or the ambient luminous contrast. Recently, the authors in [51] proposed an original method for preprocessing the residual signal in the spatial domain prior to transform, quantization and entropy coding. In order to do that, the authors introduce a HVS-based color contrast sensitivity model which accounts for visual saliency properties. For video conferencing applications, a Region-of-Interest model [26] and a saliency map [27] were employed to apply low-pass filter on the background while preserving the human face which concentrates the visual attention. JND (Just-Noticeable distortion) models have also been employed to control a gaussian filter applied on superpixels before HEVC encoding [15], to reduce the amplitude of the residual information in MPEG-2 [14]. They have also been integrated to the quantization process to reduce the non visible frequential coefficients in H.264 [28] and HEVC [29] [48] [49]. Other video coding schemes as [50] use the structural similarity (SSIM) index instead of the JND criterion for rate-distortion optimization.

In our study, we consider the constraints of live streaming applications which require low latency and real time encoding. Consequently, we propose a low complexity external pixel-domain pre-filter controlled by an a priori model which does not need information from a first pass encoding. The reasons of developing an external pixel-domain pre-filter are two-folds: firstly, our solution is independent of the video codec used after the pre-processing step. Secondly, the main drawback of most current video compression standards is the presence of well-known blocking artifacts [38]: when the compression ratio increases, block boundaries tend to become visible. The various techniques applied on processed blocks at the encoder tend towards the same artifacts. By applying a pixel-domain filter on the entire frame prior the encoding process, our solutions will not induce blocking artifact. This is illustrated in Figure 1, where we compare the effect of increasing the Quantization Parameter value with respect to the perceptual filter for an equivalent bitrate reduction. Clearly, coarser quantization strengthens the blocking effect in the reconstructed image (see areas circled in red), while pre-filtering introduces a slight blurring effect which is much less annoying than blockiness for the end user.

B. Adaptive filtering

We focus on simple well-known spatial filters in order to obtain a low complexity implementation. In this Section we review the adaptive weighting average (AWA) [9] and bilateral [7] filters and then study the effect of their similarity kernels as well as of the geometric kernel present in the bilateral filter. While the first filter is very simple to implement, the second one is more elaborated and constitutes a reference in the field of noise reduction, as many elaborated filters have been based on it.

The AWA filter of parameters ϵ and a is defined by a so-called similarity kernel which computes the weights of the filter as

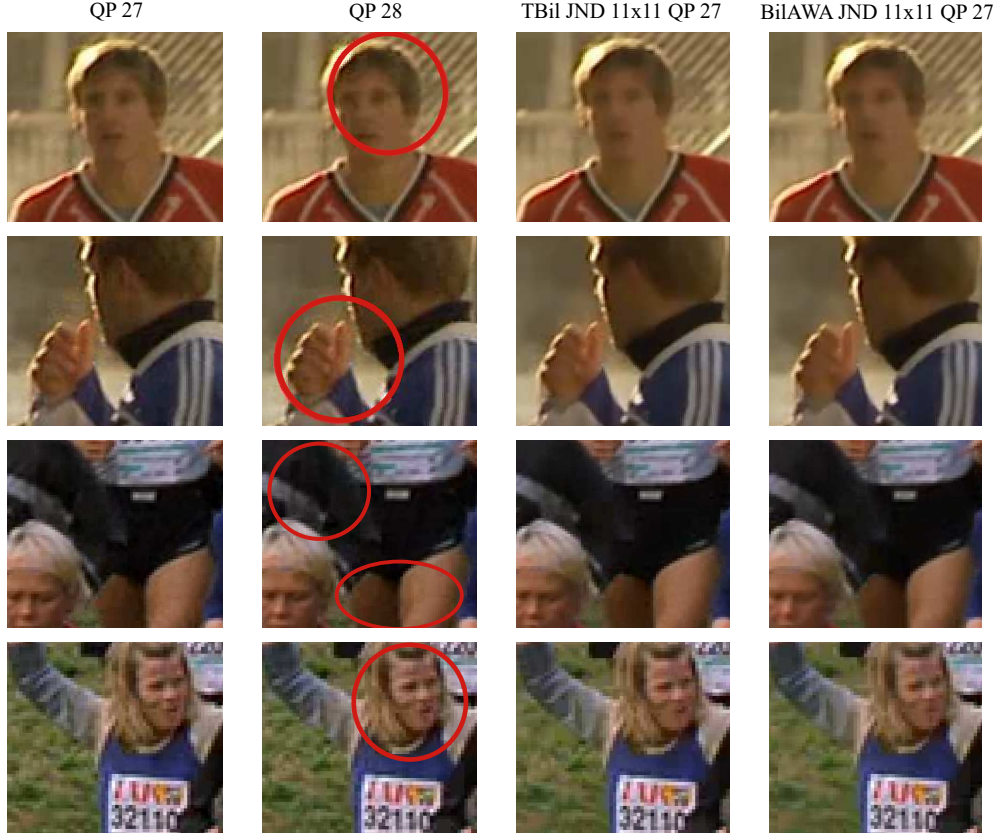


Figure 1. Comparison between coarser quantization and pre-filtering: the first and second lines show zoomed areas from the *Soccer* sequence, the third and fourth lines from the *CrowdRun* sequence. In the first column, the content has been encoded with H.264/AVC (QP=27, IBBP12). Encoding with coarser quantization QP=28 leads to a bitrate reduction of 13.5% for the *Soccer* sequence and of 12.9% for the *CrowdRun* sequence, but at the expense of visible blocking effect in the reconstructed image (second column, see areas circled in red). In comparison, our perceptual pre-filtering introduces a slight blur which is much less annoying than blockiness for the end user, for a similar bitrate reduction: $\Delta\text{Bitrate} = -9.2\%$ for the *Soccer* sequence and $\Delta\text{Bitrate} = -7.4\%$ for the *CrowdRun* sequence (third column). $\Delta\text{Bitrate} = -16.0\%$ for the *Soccer* sequence and $\Delta\text{Bitrate} = -14.2\%$ for the *CrowdRun* sequence (fourth column).

$$h_{i,AWA} = \frac{1}{1 + a * \max(\epsilon^2, \|I(\underline{x}) - I(\underline{x}_i)\|^2)}, \quad (1)$$

where $I(\underline{x})$ is the amplitudinal value of pixel at position $\underline{x} = (x, y)$, and $h_{i,AWA}$ is the filter coefficient at position $\underline{x}_i = (x_i, y_i)$, $i \in [1, N]$ within a $N \times N$ filtering mask centered at position \underline{x} . The threshold ϵ depends on the noise variance, and the differences in luminance below the threshold ϵ filter are considered as noise and are simply averaged to be reduced. The higher ϵ is, the more the filter tends to be an averaging filter.

The bilateral filter is instead defined as the product of a similarity kernel and a geometric kernel which are functions of the amplitudinal and spatial distances between the pixel at position \underline{x} being filtered and its neighboring pixels at positions \underline{x}_i respectively. The weights of the bilateral filter are computed as

$$h_{i,Bil} = h_{g,Bil}(\underline{x}, \underline{x}_i) h_{s,Bil}(\underline{x}, \underline{x}_i), \quad (2)$$

where $h_{g,Bil}$ denotes the geometric kernel of variance σ_g^2 which is function of the spatial distance given by

$$h_{g,Bil}(\underline{x}, \underline{x}_i) = \exp\left(-\frac{\|\underline{x} - \underline{x}_i\|^2}{2\sigma_g^2}\right), \quad (3)$$

and where $h_{s,Bil}$ denotes the similarity kernel of variance σ_s^2 which is function of the distance in amplitude between pixels, given by

$$h_{s,Bil}(\underline{x}, \underline{x}_i) = \exp\left(-\frac{\|I(\underline{x}) - I(\underline{x}_i)\|^2}{2\sigma_s^2}\right). \quad (4)$$

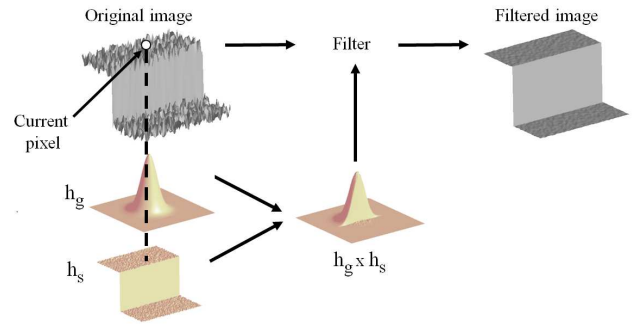


Figure 2. Illustration of the bilateral process with geometric $h_{g,Bil}$ and similarity $h_{s,Bil}$ kernels evaluated for the current pixel.

The bilateral filter is well known in order to be a non-iterative edge-preserving smoothing operator. Its luminance kernel prevents averaging across edges while still averaging within smooth regions. This feature is illustrated in Fig. 2. More sophisticated variants exist (e.g. [30]), but they need side information to guide the filtering process.

Fig. 3 illustrates the comparison between the similarity kernels for AWA and bilateral filters. The similarity kernels are different for the two filters. The AWA kernel filters more because the kernel is more extended. This is due to the fact that the function e^{-x} decreases faster than the function $\frac{1}{x}$.

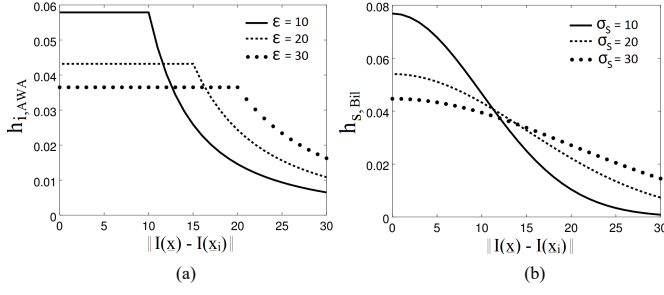


Figure 3. Comparison of AWA and bilateral similarity kernels: (a) evolution of weights for normalized AWA filter kernel as a function of squared luminance difference and for different threshold values; (b) evolution of weights for normalized bilateral filter similarity kernel as a function of squared luminance difference and for different variance values.

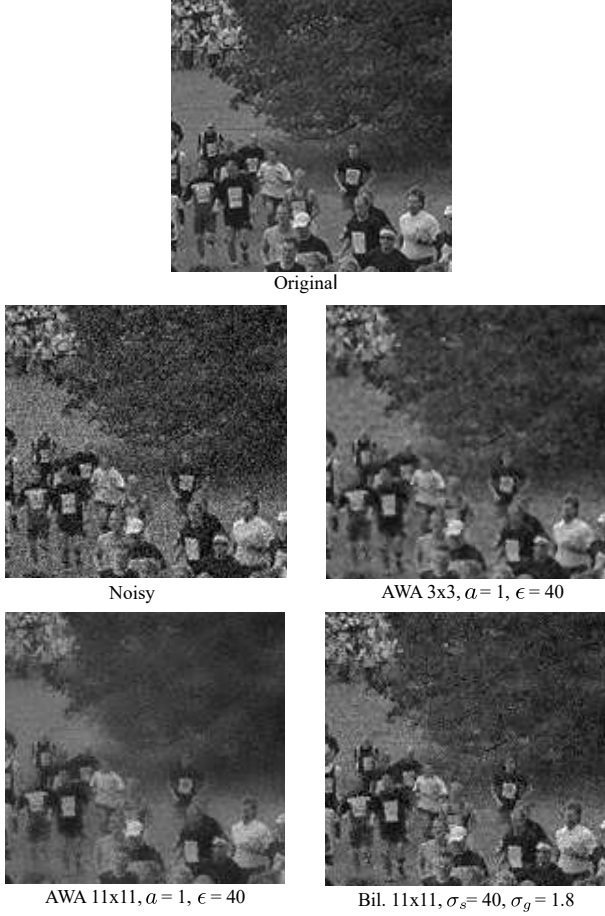


Figure 4. Illustrative example of the trade-off between noise reduction and smoothing performed by the AWA and bilateral filters.

The weights of the AWA filter decrease slowly and identically for any value of ϵ . As a consequence, the smoothing effect of the AWA filter is too strong especially when the size of its support increases. This is the reason why the filter is in general used with a small support (e.g., 3x3). In the case of the bilateral filter, the geometric kernel adds an other reduction of this averaging effect when filtering pixels far from the center.

III. QUALITY METRICS DISCUSSION

In what follows, we introduce new pre-processing filters derived from the AWA and bilateral filters described in the previous Section. It is essential to carefully select the quality metrics used to evaluate

the performances of the studied filters in a meaningful way. In this work, we mainly focus on the subjective quality metrics as psychovisual evaluation represent the ground truth for video quality. The subjective quality evaluation protocol will be detailed in Section VI. Psychovisual evaluation is also supported by the results of different objective metrics. Three kinds of objective metrics have been selected: first, the well-known PSNR and SSIM [32] objective metrics are used to quantify the similarity between each processed video frame and the corresponding original one: the more a processed frame is similar to a reference frame, the more the PSNR and the SSIM increase. These metrics are used to describe the ability of the filtering process to recover the original image and to solve the inverse problem of denoising. Although they do not reflect the perceived quality of the processed video, they constitute a good indicator of the strength of the filtering process. Secondly, two blurring effect measures are applied because blur is known to constitute the predominant artifact brought by low-pass filtering. Among the existing blur metrics, we use the Local Phase Coherence Sharpness Index or LPC-SI [33] (a no-reference blur metric based on the local measure of the phase coherence) as well as the Marziliano's blur metric [34] (based on the overall spread of the edges across the processed image) because these perceptual metrics exhibit a high correlation with subjective ratings of blurred images. Note that the less a frame is blurred, the higher the LPC-SI, while the opposite is true for the Marziliano's perceptual blur metric. Finally, a global perceptual full-reference quality measure has been incorporated as assessment index. Among the existing metrics, the so-called Feature-Similarity (FSIM) metric [39] was chosen because of its high consistency with subjective evaluations. The FSIM metric is a decreasing function of the global quality, with one means same quality as reference.

IV. THE NEW ADAPTIVE FILTERS

A. New Bilateral Filters

The previous observations reveal two interesting features: (i) The AWA filter has a thresholded behaviour that exhibits the same filtering weight for values below ϵ and (ii) the bilateral filter has the fastest decay rate and preserves sharpness thanks to the use of a geometric kernel [31]. In Fig. 4, one can see how the bilateral filter is sharper than the AWA, and how the AWA induces much more blur with a large support such as 11x11 pixels. Examples are given for a noise standard deviation of $\sigma_n = 20$. This leads us to introduce two new filters which combine those two features.

The first proposed filter, called the BilAWA filter, is defined as the product of the similarity kernel of the AWA filter by a geometric kernel, as follows:

$$h_{i,BilAWA} = h_{g,BilAWA}(\underline{x}, \underline{x}_i) h_{s,BilAWA}(\underline{x}, \underline{x}_i), \quad (5)$$

where the kernel $h_{g,BilAWA}$ is the same as $h_{g,Bil}$ defined in Eq. 3 and the kernel $h_{s,BilAWA}$ is the same as $h_{i,AWA}$ (Eq. 1).

The second filter follows the AWA thresholding principle, but with a gaussian decay (see Fig. 3.b) and a geometric kernel. It is therefore called Thresholded Bilateral or TBil. It is described, using the same notation as Eq. 4, as

$$h_{i,TBil} = h_{g,TBil}(\underline{x}, \underline{x}_i) h_{s,TBil}(\underline{x}, \underline{x}_i), \quad (6)$$

$$h_{s,TBil}(\underline{x}, \underline{x}_i) = \min(e(-\frac{1}{2}), e^{-\frac{\|I(\underline{x}) - I(\underline{x}_i)\|^2}{2\sigma_s^2}}). \quad (7)$$

The threshold of the similarity kernel in Eq. 7 is chosen such that every value between 1 and σ_s has the same weight.

The size of the bi-dimensional filtering support is an important parameter of the pre-filtering algorithm. It has been chosen through simulations working on a large variety of video sequences. Filters

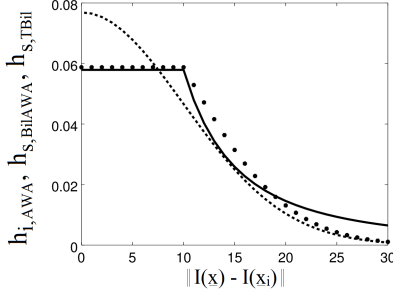


Figure 5. Comparison between the bilateral similarity kernel (dotted line, $\sigma_s=10$), the AWA kernel (BilAWA similarity kernel) (solid line, $\epsilon=10$) and the TBil similarity kernel (point, $\sigma_s=10$).

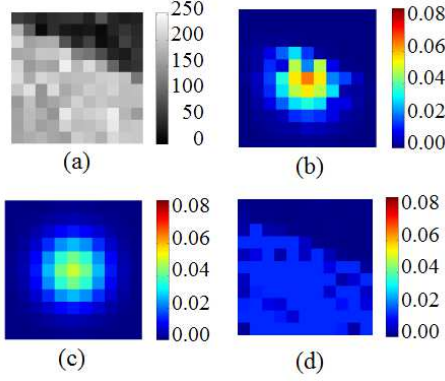


Figure 6. Illustration of the impact of the similarity kernel (c) and the geometric kernel (d) on the weights of the BilAWA filter (b) for a particular pixel and its 11x11 neighbors (a).

with squared kernel sizes of 3x3, 5x5, 7x7, 11x11 and 25x25 have been applied and the 11x11 kernel size has been retained as a satisfying compromise between introduced blur, bitrate saving and computational complexity. This 11x11 size is also consistent with the high-definition context of this work. SDTV filtering applications widely use bi-dimensional filtering support of 3x3 or 5x5 squared odd size [43] [44]. Comparing SDTV and HDTV, the 4:3 portion of a 1080 HDTV image is 1440x1080 or 1555200 pixels. If both SDTV and HDTV images are displayed at the same height, each 1080 pixel is about $\frac{1}{4}$ the size of a SDTV pixel [54]. This factor 4 of the resolution is consistent with the chosen size of 11x11 pixels.

Finally, Figure 5 gives a comparison between the bilateral similarity kernel, the AWA kernel (BilAWA similarity kernel) and the TBil similarity kernel, respectively. It can be noted that the similarity kernel of the new TBil filter offers an interesting compromise between the bilateral and AWA kernel decays. The combined effect of the geometric and similarity kernels as well as the adaptation to the local image content are illustrated in Figure 6.

B. Performance analysis in the context of noise reduction

We first analyze the behaviour of both the original and the novel filters in terms of the compromise they yield between removing spurious noise and preserving image sharpness. In the experiments we consider eight different images of size 1280 x 720 pixels extracted from a great variety of well-known high-definition video test sequences (*IntoTree*, *ParkJoy*, *MobCal*, *CrowdRun*, *Ducks*, *Ski*, *Soccer*, *Parkrun*). This dataset has been designed to cover a wide range of spatial activity values. These images are corrupted by additive white gaussian noise with standard deviation values of 10, 20 and 30, respectively. The

standard deviation of the bilateral geometric kernel σ_g has been fixed to a value of 1.8 as specified in [13]. Experiments confirmed that this corresponds to the best value in terms of denoising. In addition, we have chosen $\epsilon = \sqrt{2}\sigma_n$ for the AWA filter and $\sigma_s^2 = 2\sigma_n^2$ for the bilateral filter. Here, the performance analysis is limited to objective metrics evaluation. The distance and blur metrics discussed in Section III are used for performance assessment. We remind that:

- the PSNR values vary from 0 (opposite to the reference) to infinity (identical to the reference) while the SSIM values vary from 0 (opposite to the reference) to 1 (identical to the reference).
- the Marziliano's metric is an increasing function of the blur phenomenon while this is the opposite for the LPC-SI metric, with: LPC-SI=1 when no blur is present.

Results are given in Tables I and II, where one can first see that increasing the AWA filtering support from 3x3 pixels (column 1) to 11x11 pixels (column 2) logically leads to an increase of both distance from the original image and blur at low values of noise (standard deviation of 10 or 20). For higher value of noise (standard deviation of 30), the AWA 11x11 filter brings slightly improved results in terms of distance from the original image, compared to the 3x3 case, despite of an increase of blur phenomenon.

Table I
DENOISING PERFORMANCES: ABSOLUTE DISTANCE MEASUREMENT RESULTS. THE HIGHER THE NOTE IS, THE LOWER THE DISTANCE BETWEEN THE FILTERED IMAGE AND THE ORIGINAL ONE.

σ_n	Distance measures	AWA 3x3	AWA 11x11	Bilat. 11x11	BilAWA 11x11	Sim. 11x11	TBil 11x11
10	SSIM	.854	.836	.845	.860	.838	.862
	PSNR[dB]	31.7	30.3	31.4	31.6	31.1	31.7
20	SSIM	.683	.684	.668	.712	.659	.714
	PSNR[dB]	27.2	26.6	26.8	27.5	26.6	27.5
30	SSIM	.546	.558	.536	.585	.523	.598
	PSNR[dB]	24.4	24.5	24.2	25.0	24.2	25.1

Table II
DENOISING PERFORMANCES: BLURRING MEASUREMENT RESULTS. THE HIGHER THE MARZILIANO'S METRIC VALUE IS, THE MORE BLURRING ARTEFACT IS PERCEPTIBLE, WHILE THE OPPOSITE IS TRUE FOR LPC-SI METRIC.

σ_n	Blurring measures	AWA 3x3	AWA 11x11	Bilat. 11x11	BilAWA 11x11	Sim. 11x11	TBil 11x11
10	LPC-SI	.925	.918	.926	.923	.926	.927
	Marzil	3.72	3.80	3.34	3.77	3.22	3.41
20	LPC-SI	.916	.902	.922	.915	.921	.923
	Marzil	3.78	3.88	3.29	3.86	3.14	3.42
30	LPC-SI	.907	.884	.914	.903	.909	.914
	Marzil	3.80	3.92	3.26	3.89	3.09	3.44

When comparing the third and fifth columns of Tables I and II, the addition of the geometric kernel (Eq. 3) to the similarity kernel brings a significant improvement both in terms of blur limitation (in regards to LPC-SI) and distance from the original image (expressed here in terms of PSNR). Such improvements are also present when adding the geometric kernel to the AWA 11x11 filter (BilAWA filter, fourth column).

The BilAWA filter brings a significant improvement in PSNR compared to the bilateral filter (over 0.7dB for the higher value of σ_n) with slightly worse perceptual measurement results. Fig. 7 shows the differences on two images between the BilAWA and the bilateral filter for a noise value of 10 (first row) and 30 (second row). One can see clearly that fine details are better preserved by bilateral filtering at the expense of remaining noise.

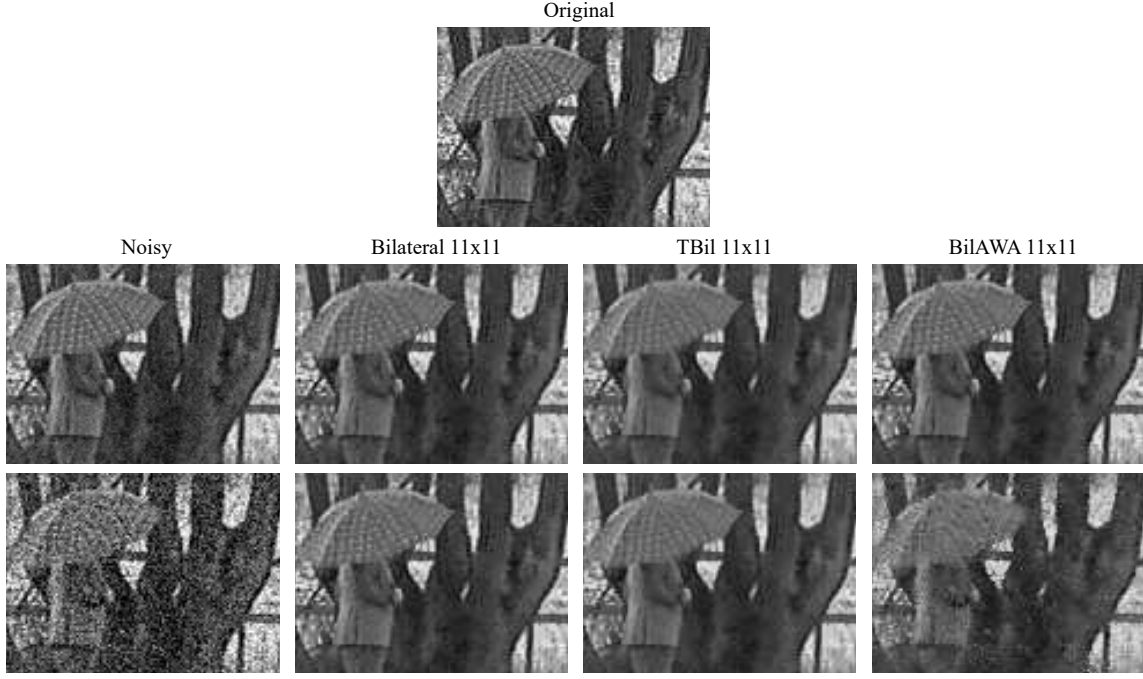


Figure 7. Denoising results obtained with different filters for $\sigma_n = 10$ (second row) and $\sigma_n = 30$ (third row) - *Parkrun* sequence, enlarged part.

From those results, we can conclude that there is still a sharpness advantage to the bilateral filter. It seems that the decay of the filter for large luminance difference values observed in Fig. 3 is too slow for the AWA kernel. This may explain the decrease in sharpness obtained with the AWA 11x11 kernel and the BilAWA filter. This raises the need for the second new filter, called the TBil filter, which uses only the thresholded feature of the AWA filter while keeping a gaussian decay.

Results presented in Tables I and II show that all quality metrics indicate an image quality improvement after denoising using the TBil filter, except the Marziliano blur metric. However, this latter metric is clearly in favor of the TBil filter compared to the BilAWA one. The increase in PSNR compared to the bilateral filter goes up to 0.9dB for $\sigma_n = 30$. This shows that the TBil filter yields the best compromise between the bilateral and the AWA filter for noise removal.

V. JUST NOTICEABLE DISTORTION (JND) DRIVEN ADAPTIVE FILTER

A. Perceptually-Guided Filtering Process

In Section IV, two adaptive filters have been proposed and validated in terms of the trade-off between noise removal and preservation of image sharpness using different metrics. Both filters are modified versions of the bilateral filter, where the strength of the filter is controlled by a thresholding operation applied to the similarity term. The choice of the threshold is an important aspect of the filter design: in our case, we want to guide the video low-pass filtering to remove visual information hardly perceptible from the original video signal before compression, so as to increase the coding efficiency. By lowering typically the high frequency content of each image of the pre-processed video sequence, the perceptual pre-filtering process will lower the image entropy resulting in a reduced amount of data to be encoded.

For this, we propose to use a just noticeable distortion (JND) model. Generally speaking, the JND refers to the visibility threshold below which no changes can be detected in an image by most human observers. This JND is dependent on properties of the human visual

system (HVS) such as the contrast sensitivity function (CSF) of the eye or masking properties [35], and is based on local image characteristics. Several JND models have been proposed in recent years in the literature and applied to the digital video processing and coding areas. A good overview of these models is provided in [35], [14], [36]. These models mainly differ in the considered computational domain, respectively the pixel domain and the frequency domain. Recently proposed JND models [40], [41], [42] are highly complex and include some elaborated human visual properties like foveated masking effects. The JND frequency models are usually block-based and defined in the DCT domain, since the DCT is widely used in video processing applications. A JND threshold value is then obtained for each block from the modelling of spatial and temporal properties of the HVS in the frequency domain. The other possibility is to define the JND model in the pixel domain. Among the different models in the literature, the one defined by X. Yang et al. [14] is one of the most widely used because it provides a good compromise between accuracy and computational complexity. In what follows, we retain the pixel-based JND model proposed by Yang for several reasons: firstly, it doesn't rely on DCT block-based computation. It is consequently more suited to the proposed pixel-based pre-filtering process and avoids to introduce spurious blocking effect. Moreover, the JND models described in pixel domain are less complex and are better suited to real-time video processing, which constitutes one of the aims of our work as explained in Section I. In a first approach, we choose to use a 2D spatial filtering for real-time implementation purposes because a 3D spatio-temporal pre-filtering implies the use of motion compensation which is computationally expensive. Hence, the temporal masking model proposed by Yang is not taken into account here. Finally, it is important to note that our proposed perceptual filter can be applied with other JND pixel based models.

The spatial-only JND model denoted JND_S can be expressed as

$$JND_S = JND_{lum} + JND_{tex} - C * \min(JND_{lum}, JND_{tex}). \quad (8)$$

This model considers two phenomena: luminance masking noted JND_{lum} and texture masking, noted JND_{tex} . The third term

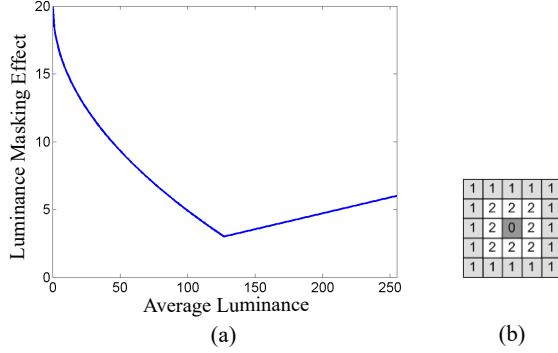


Figure 8. Luminance masking effect: (a) Approximation of the Weber-Fechner law, (b) Weighting windows for luminance background calculation.

accounts for the overlapping effect when both spatial and texture masking are present, with C a constant fixed experimentally to 0.3 [14].

1) *Luminance Masking*: Luminance masking noted JND_{lum} reflects the difference in sensitivity of the HVS to changes in an image based on the background luminance level. This visual phenomenon is usually modelled by the well-known Weber-Fechner law [35] reflecting the low sensitivity of the human eye to differences involved in dark image areas as illustrated in Figure 8.a. In [14], the authors propose the following approximation:

$$JND_{lum} = \begin{cases} 17 \left(1 - \sqrt{\frac{\overline{I(x,y)}}{127}} \right) + 3 & \text{if } \overline{I(x,y)} \leq 127 \\ \frac{3}{128} (\overline{I(x,y)} - 127) + 3 & \text{otherwise.} \end{cases} \quad (9)$$

where $\overline{I(x,y)}$ is the local average luminance value computed on a 5×5 neighborhood of each pixel using the weighting window illustrated in Figure 8.b).

2) *Texture Masking*: Texture masking, also called contrast masking, reflects the fact that high spatial activity within an image reduces the sensitivity of the eye to a visual stimulus. Moreover, it accounts for the high sensitivity of the human visual system to the contour information and the homogeneous areas. Indeed, the HVS is more highly sensitive to edges than textures. In order to separate the two high frequency contents, a gradient map $G(x, y)$ is first calculated to detect both edges and textures using four convolution masks g_k described in Figure 9, as follows:

$$G(x, y) = \max_{k=1,2,3,4} |grad_k(x, y)|$$

$$grad_k(x, y) = \frac{1}{16} \sum_{i=1}^5 \sum_{j=1}^5 p(x-3+i, y-3+j) * g_k(i, j). \quad (10)$$

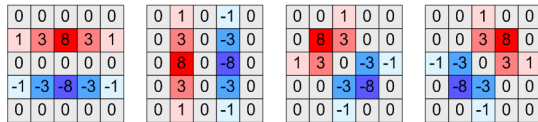


Figure 9. Four gradient directions used for $G(x,y)$ calculation.

Then an edge map $W_e(x, y)$ is calculated using the Canny operator combined with morphological processing. The resulting edge map is binarised and serves as a mask to remove strong edges from the gradient map, hence keeping only the textures with higher JND thresholds.

B. Using the model in the filters

The spatial-only JND mode given by Eq. (8) is used in the proposed perceptually-guided filtering approach to estimate sensitivity thresh-

olds pixel-by-pixel for each image of the processed sequence. Hence, the smoothing operation is adapted to the local sensitivity threshold, given by the JND, to selectively remove irrelevant high frequency details. Such low-pass filtering for perceptual image coding makes perfect sense as long as the bandwidth of the filter is driven by the JND value.

To do that, Equations (1) and (4) are revisited by simply substituting, in a first approach, the JND value to ϵ and σ_s parameters, respectively. Consequently, the similarity kernels of the two perceptual bilateral filters are expressed by the equations (11) and (12) for the BilAWA and the TBil filter respectively:

$$h_{s,BilAWA}^{JND}(\underline{x}, \underline{x}_i) = \frac{1}{1 + a * \max(JND^2(I(\underline{x})), \|I(\underline{x}) - I(\underline{x}_i)\|^2)}, \quad (11)$$

$$h_{s,TBil}^{JND}(\underline{x}, \underline{x}_i) = \exp\left(-\frac{\|I(\underline{x}) - I(\underline{x}_i)\|^2}{2 * JND^2(I(\underline{x}))}\right). \quad (12)$$

where $JND(I(\underline{x}))$ is the JND value of the pixel at location $\underline{x} = (x, y)$.

It is expected that the smoothing operation should be adaptively increased when the JND sensitivity threshold value increases. In this case, more distortion is supposed to be visually acceptable for the human observer. Similarly to noise filtering, we propose to act on σ_s , the similarity kernel variance of the bilateral filter, such that the filtering bandwidth is inversely proportional to the JND value. Hence, the value of the similarity variance σ_s is chosen in order to process strong edges and textures of the image in a separate way. In the case of the BilAWA filter, the JND parameter is similarly taken into account using the threshold ϵ : the higher the JND, the wider the photometric range used and the larger the weights of the filter kernel.

The distortion is directly related to the amount of information removed by filtering which depends on the image content. Therefore a feedback control loop would be further required to reach the exact JND threshold value. In the following experiments, the purpose of the proposed filters is to remove perceptually insignificant high frequency content. In most cases, the actual distortion due to filtering will be under the JND threshold. Preliminary results on the use of the JND model to control the pre-filtering have been reported in [11] considering the AWA algorithm as a basis filter operation.

C. Quality evaluation

In order to validate the proposed JND-driven filtering process, several simulations have been conducted on high-definition video sequences. Figure 10 gives for the *CrowdRun* sequence an illustration of the difference between two filtering approaches: the first one is based on the TBil filter defined in Section IV-A using a fixed threshold for the similarity kernel. The second filtering version is also obtained with the TBil filter: in this case, however, the TBil filter is controlled by the JND model. In the case of the non-perceptual TBil filtered version, the threshold value was chosen experimentally in order to have the same PSNR for the two filtered versions. In that way, we compare two frames with the same distance from the original frame. Figure 10 a) to c) correspond to an enlarged part of the original frame, the fixed TBil filtered version, and the JND-driven TBil filtered version, respectively. We can see that the perceptual filter (c) preserves the homogeneous parts of the frame (i.e. the grass) in comparison to the fixed non-perceptual filter (b). In addition, we present the JND map computed from the original frame (d), as well as the differences introduced by the non-perceptual filter (e) and the perceptual filter (f), respectively. For visualization purposes, Figures 10 e) and f) correspond to the logarithm of the absolute difference. We verify that the result of JND-driven filtering correlates

well with the JND map: the reduction of imperceptible details is mainly concentrated on areas of the processed image where the JND values are the most significant. Consequently, the processed image is sharper and of better perceived video quality without excessive blur.

D. Complexity evaluation

Performing the filtering on a support of dimension 11x11 pixels and computing the JND values are both time consuming operations. For instance, once the weights of the filter have been computed, a 11x11 filtering implies more than 250 Mega Multiplication Accumulations (MAC) per HD image, so more than 5 Giga MAC per second for a 25 fps. On top of that, since the filter is adaptive, 250 Mega filter weights have to be computed per image. This latter operation can be more or less complex given the filter kernel. For instance, AWA weights can be easily computed using a look up-table. The complexity of the whole JND computing process has been estimated at around 30 Giga MAC per second.

Real time processing is therefore not achievable on conventional computer without using aggressive acceleration techniques such as general-purpose processing on graphics processing units (GPGPU). We achieved real time computation of a modified version of the original spatial JND model given by Eq. (8) (with faster but less accurate edge detection) using OpenCL on a standard consumer GPU (ATI FirePro v4900) with the resort to various parallelisation techniques such as two-pass convolution and edge detection optimisation. The JND is computed at a framerate of 64 fps.

Given the computational complexity of the JND, the use of a complex non local filtering would not allow a real-time processing on a standard workstation as it is especially heavy in terms of random memory access (the weak point of parallel systems). Therefore, local filtering keeps the algorithm simple and easily deployable on parallel architectures.

VI. EXPERIMENTAL RESULTS

By removing imperceptible details and by using constant quantization, we expect to reduce the bitrate necessary to encode a sequence without compromising the perceived quality. To evaluate the performances of the proposed perceptual pre-filters, we have compared the bitrate of the sequences, encoded at constant quantization parameter (QP), with and without pre-filtering and the quality of the decoded sequences based on psychovisual evaluation and completed thereafter by objective metrics. The test conditions are summarized in Fig. 11 and all the results are presented in Tables III, IV, V, VI and VII. The rate gain brought by pre-filtering the videos with the proposed BilAWA and TBil perceptual filters has been evaluated using both MPEG-4/AVC and HEVC codecs. The MPEG-4/AVC video coding standard is today largely adopted for both consumer and professional market sectors and is likely to remain widely used in the years to come, while HEVC is the new state-of-the-art video coding standard developed by JVT. The MPEG-4/AVC and HEVC codecs used in the experiments are the x264 and x265 implementation, respectively. Three 1280x720@50p sequences from our video database (*CrowdRun*, *Ski* and *Soccer*) are used to evaluate the proposed filters. We initially evaluated the BilAWA filter performances on six sequences as shown by Figure 12. The *IntoTree* sequence was not retained because of its marginal bitrate saving. We selected three sequences which all present high spatial information index but different temporal information indexes. The reason for choosing video contents with a high spatial index are two-folds: firstly, such contents allow to highlight the blurring artifact which is the main artifact introduced by pixel-domain low-pass filters. Secondly, it also validates the effectiveness of using

a spatial only JND map: if the proposed pre-processing solutions give satisfying results in this particular case, one can expect that the results could only be improved if the JND model is extended to a spatio-temporal one. Both proposed filters have been applied with a 11x11 mask. Results are also given when using the AWA filter with a 3x3 mask. The three filters under test are controlled by the original luminance spatial JND model given by Eq. (8) (without edge detection simplification). All the filters have been applied only on the luminance component, the chrominance components are copied from the original sequence.

A. Encoders set-up

For MPEG-4/AVC, the High Profile configuration of the codec has been used with CABAC encoding. Evaluation was performed with and without the deblocking filter. Note that the proposed perceptual pre-filtering technique can be used with or without the deblocking filter which is in the coding loop, and with any encoder. The reasons for performing a test without the deblocking filter are two-folds: firstly, the deblocking filter is useful at low bitrates but is less relevant for high rate and high quality encoding as targeted here for High Definition sequences. Secondly, it can introduce blurring effects that are not acceptable for professional applications. In addition, removing this filter allows us to more specifically analyze the behaviour of the proposed perceptual filters. These results are presented in Table IV. For comparison, the results with the deblocking filter are presented in Table VI.

For HEVC, the Main Profile configuration of the codec has been used, with both deblocking and SAO loop filters applied and CTU up to 64x64 pixels.

For both encoders, we analyze the gain brought by the proposed perceptual filters for two QP and GOP configurations. Our work is focused on high quality, so we chose low QP values of 22 and 27, corresponding to bitrates between 5.5Mbps and 53 Mbps for the three HD test sequences. We used two GOP structures: IBBP using one Intra picture every 12 frames (GOP IBBP(12)) and Intra-only GOP.

B. Subjective and objective quality protocol

Subjective evaluation has been conducted on a 47 inch monitor of resolution 1920x1080, in a dedicated room with white walls and a controlled light following the Recommendation ITU-R BT 500-13 [37]. Sixteen observers, five females and eleven males, have participated in the subjective tests. They were all non-expert with (or corrected-to-) normal visual acuity. To evaluate the difference of perceived quality between a sequence encoded with and without pre-filtering, we used the simple stimulus paired comparison protocol, as in [37], with a 7-level comparison scale presented in Fig.13. The Comparison Mean Opinion Scores (CMOS) are presented in Table IV.

In order to comfort the subjective evaluation tests, these ones are completed by the computation of the PSNR and LPC-SI objective metrics [33], in the same manner as in Section IV-B. We also add the Feature-Similarity (FSIM) full-reference objective quality metric [39], selected because it can achieve very high consistency with subjective quality scores. In Table IV, VI and VII, we present the difference of PSNR, LPC-SI and FSIM scores between the sequences encoded with and without pre-processing, called Δ PSNR, Δ LPC-SI and Δ FSIM respectively. LPC-SI and FSIM values have been multiplied by 100 for readability. We remind that:

- the PSNR values vary from 0 (opposite to the reference) to infinity (identical to the reference).
- the LPC-SI is a no reference metric. It is a decreasing function of the blur phenomenon. LPC-SI=1 when no blur is present.

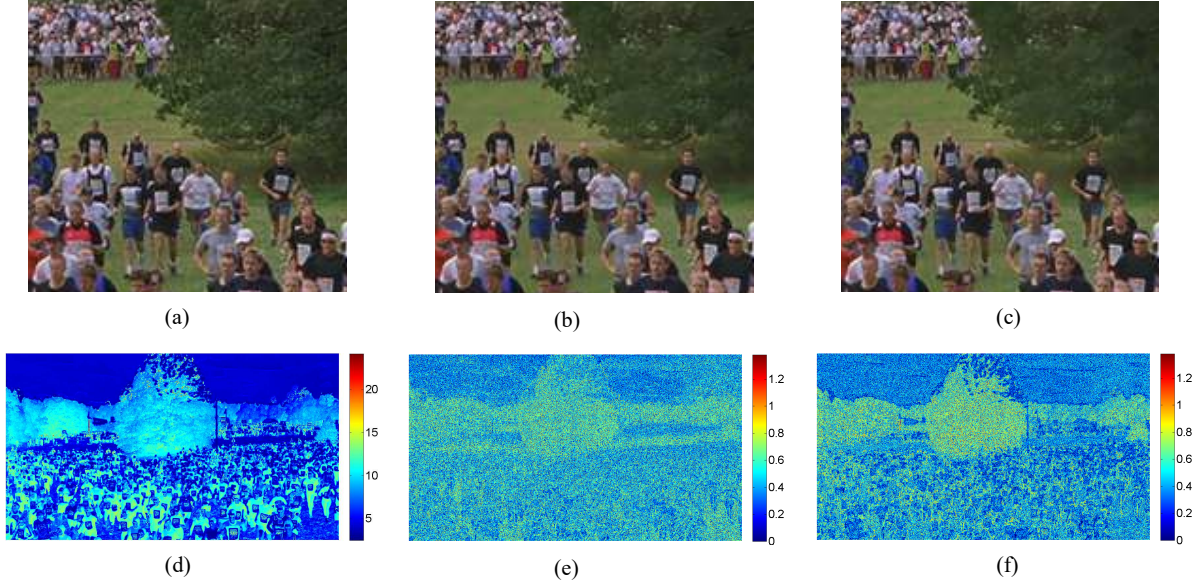


Figure 10. Comparison between perceptual filter and fixed threshold filter at same PSNR. (a), (b), (c) represent a portion of the original frame, the filtered one with the fixed TBil filter (TBil(11x11, 7.8)) and the perceptual TBil filter (TBil(11x11, JND)) respectively. The two filtered versions have the same PSNR = 39.4 dB. The figure also presents the JND map of the original frame (d), and the difference map of the fixed filtered frame (e) and the perceptually filtered one (f) compared to the original frame.

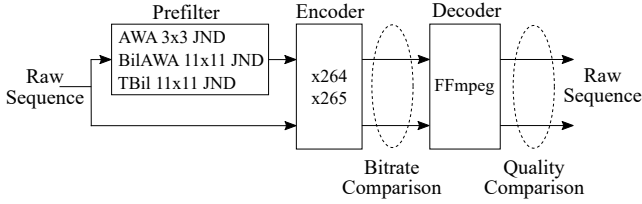


Figure 11. Experimental protocol.

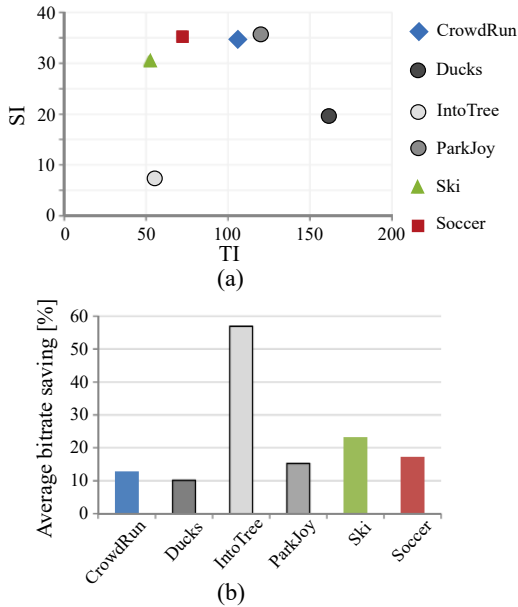


Figure 12. Video test material: (a) spatial (SI) and temporal (TI) information indexes ; (b) average bitrate savings obtained with BiLAWA JND 11x11 filter applied prior to x264 encoder with QP22, GOP IBBP12 and no Deblocking filter. Grey: excluded videos - Color: selected video for the rest of the evaluation.

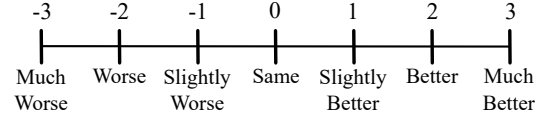


Figure 13. 7 level- comparison scale.

- the FSIM metric is in the $[0, 1]$ range with “1” meaning excellent quality compared to the reference.

C. Performance analysis of proposed TBil and BiLAWA filters

Before studying of perceptual prefilters, we analyse the behaviour of non-perceptual BiLAWA and TBil filters compared to the bilateral filter. The filters are driven by a constant variance which was chosen as follows: the mean JND value of the sequence was used to control the BiLAWA filter, then the variance of the bilateral and TBil filters were fixed experimentally in order to obtain almost the same PSNR for all sequences, i.e. the same distance from the original video sequence. Table III presents the results with the x264 codec without deblocking filter, at QP 27 and IBBP GOP structure with a GOP length of 12 frames. One can notice that the BiLAWA filter allows a higher bitrate saving than the bilateral filter. In the following, we will analyze the perceptual pre-filters BiLAWA and TBil compared to the perceptual AWA filter initially proposed by the authors [11].

D. Performance analysis of proposed JND-guided filters with H.264

Table IV presents the bitrate and the CMOS (Comparative Mean Opinion Score) results with confidence interval of 95% (read CMOS $\pm \delta_{95\%}$) for all sequences and encoding configurations of the x264 codec without deblocking filter. Figure 14 presents CMOS notes versus bitrate saving for the three evaluated pre-filters. The pre-filters yield very large rate savings (up to 28.7%) for same visual quality (very low CMOS values near to zero). We can note that none of the three filters causes a loss of perceived quality ($-0.29 < \text{CMOS} < -0.16$), and they all allow to reduce the bitrate (on average, 14.1%, 14.3% and 19.3% for AWA, TBil and BiLAWA filter respectively). We verify

Table III
COMPARISON OF ENCODING PERFORMANCES OF x264 CODEC WITH AND WITHOUT NON-PERCEPTUAL PRE-FILTERS. ANALYSIS OF BITRATE REDUCTION (Δ BITRATE), OBJECTIVE MEASURE VARIATION (Δ PSNR). x264 IS USED WITHOUT THE DEBLOCKING FILTER AT QP 27

		<i>Crowd Run</i>	<i>Ski</i>	<i>Soccer</i>	<i>Average</i>
x264	Bitrate [Mbit/s]	28.33	6.28	7.29	13.97
	PSNR [dB]	35.24	40.02	39.05	38.10
Bilateral (11x11)	Δ Bitrate [%]	11.88	14.31	14.62	13.61
	Δ PSNR [dB]	-2.03	-0.72	-1.16	-1.30
BilAWA (11x11)	Δ Bitrate [%]	14.42	15.30	17.00	15.58
	Δ PSNR [dB]	-2.03	-0.72	-1.18	-1.31
TBil (11x11)	Δ Bitrate [%]	11.06	13.97	13.70	12.91
	Δ PSNR [dB]	-2.00	-0.73	-1.14	-1.29

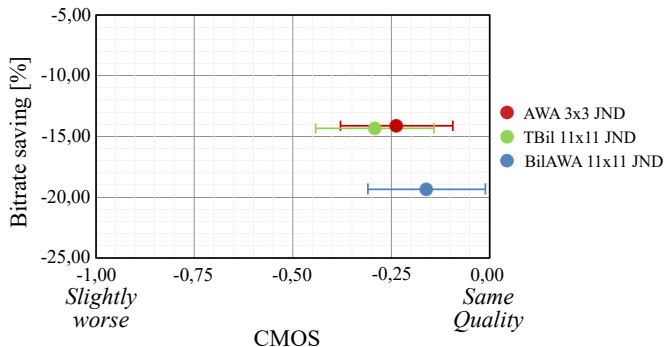


Figure 14. Evaluation of proposed pre-filters. CMOS versus bitrate saving for AWA, BilAWA and TBil perceptual pre-filters in comparison with encoding scheme without pre-filters. Global results for all sequences, GOP and QP configurations for x264 without deblocking filter.

that these average bit reduction ratios are comparable with the ones obtained with other methods recently described in the literature [25] [52]. In particular, the authors in [52] report bitrate savings between 9.6% and 30.4% (extreme case) for similar test data and coding conditions (four HD test videos, High-Profile H.264/AVC encoding using x264 encoder, original bitrates of 10Mbps at 15 Mbps). The bitrate reduction is directly correlated with the filtering strength and we have seen in Section IV-B that the AWA filter, applied with a 3x3 support, filters less details than the TBil filter with a 11x11 support, which filters less than a BilAWA filter with the same support.

In addition, Table IV presents the objective metric variations Δ LPC-SI, Δ FSIM and Δ PSNR. The Δ LPC-SI values (0.114 for AWA, 0.135 for BilAWA and 0.156 for TBil on average) are insignificant in comparison with the LPC-SI values of the encoded versions without pre-filters (91.316). The LPC-SI results are well correlated with the subjective evaluation, showing that no blurring effect is observed on the pre-processed sequences. Even if the Δ LPC-SI values are insignificant, they are always positive, meaning that the filtered encoded versions are less blurred than the encoded versions without pre-filter. That could be explained by the fact that LPC-SI is a no-reference metric and because perceptual pre-filters preserve edges of the sequences. Finally, it can be seen that the Δ FSIM values (-0.382 for BilAWA and -0.321 for TBil, on average) are small compared to the FSIM values of the compressed versions (99.726). This confirms the fact that the pre-filters do not introduce any degradation of perceived video quality.

a) PSNR analysis:

Table IV also presents Δ PSNR for the three tested filters. These results can be analyzed as follows: the perceptual pre-filter removes non-significant visual information to ease compression, and the

bitrate reduction will be even more significant than the amount of information removed is important. Consequently, the distance between the pre-filtered image and the original version is logically increased from a mathematical point of view, which results in a PSNR decrease. It can also be noted that the reduction of PSNR is directly correlated to the bitrate reduction. To further quantify the impact of the filters, Table V presents the PSNR results when the reference signal correspond either to the original sequence, or its pre-filtered version. We can note that the PSNR values of the pre-filtered then encoded sequences are logically higher when using the filtered version as a reference instead of the original one. In addition, these same PSNR values are also higher than the PSNR of the compressed sequences without pre-filtering. The reason is that the remaining information after filtering, having less details and being less noisy, is easier to encode.

Finally, this PSNR reduction occurs without loss of visual quality. This highlights the performances of the proposed JND-guided pre-filters as perceptually lossless processes since they can remove informations in the image (hence the decrease in PSNR) without affecting the visual quality. Indeed, the BilAWA filter brought the larger Δ PSNR (-2.90dB on average) and the larger Δ Bitrate (19.3%) without compromising the perceived quality (-0.16 CMOS) (Table IV). One can note a maximum PSNR reduction of 5.12 dB for a CMOS of 0.13, brought by the BilAWA filter on the *CrowdRun* sequence at QP 22 and GOP I only. One would expect that the differences are not perceived to a certain level of PSNR reduction, and then become visible. But there is no correlation between the Δ PSNR values and the CMOS scores. We take the example of the *CrowdRun* sequence pre-processed by the AWA filter: at QP=22 and GOP I only, there is no perceived differences (CMOS=0.00) for a PSNR reduction of 4.16 dB. In contrast, at QP=27 and GOP IBB12, the loss of details is slightly perceived (CMOS = -0.56) for a smaller PSNR reduction of -1.69 dB.

b) Comparison of the perceptual BilAWA and TBil filters:

Finally, the BilAWA filter yields the best performance, by reducing the bitrate by 19.3% with imperceptible quality difference (CMOS less than 0.16). We have seen that the TBil filter has the best denoising performances in terms of objective metrics. However, in a pre-processing context, bitrate saving and subjective evaluation tests show that the BilAWA is better suited. In fact, thanks to its gaussian decay, the TBil filter better preserves the details of a frame than the BilAWA filter which uses a $\frac{1}{x}$ decay. That can be observed in Figure 17 and confirmed by the PSNR values. But when the sequences are displayed, the subjective tests show that the observers cannot see any difference between the encoded sequence without pre-filter and with any of the three filters under test. It is important to notice that the results obtained by a frame analysis are not directly applicable to a video analysis. The BilAWA filter allows the highest bandwidth saving as it smooths more stronger the details.

c) QP impact:

One can compare the bitrate gain brought by the pre-filters with two different QP values with a IBBP(12) GOP. One can note that for all the filters, the rate saving is higher with a QP value of 22 compared with QP=27 (Figure 15). For example, the TBil filter brought a rate reduction of 17.3% at QP=22 and 10.2% at QP=27. This can be explained by the fact that the quantization process reduces the high-frequency content by itself when the QP increases.

d) GOP impact:

We can evaluate the pre-filters performances with different GOP configurations. For all the filters, on average for the three sequences, the rate reduction is larger when using a IBBP(12) GOP than a I only GOP (Figure 15). For example, the TBil filter leads to a rate reduction of 15.4% with a I only GOP and 17.3% with a IBBP(12) GOP. So the

Table IV

COMPARISON OF ENCODING PERFORMANCES OF x264 CODEC WITH AND WITHOUT PERCEPTUAL PRE-FILTERS. ANALYSIS OF BITRATE REDUCTION (Δ BITRATE), SUBJECTIVE QUALITY EVALUATION (CMOS AND THE CONFIDENCE INTERVAL $\delta_{95\%}$) AND ASSOCIATED OBJECTIVE MEASURES VARIATION (Δ LPC-SI, Δ FSIM AND Δ PSNR). x264 IS USED WITHOUT THE DEBLOCKING FILTER.

		GOP Intra-only				GOP IBBP(12)								Average
		QP 22				QP 22				QP 27				
		Crowd Run	Ski	Soccer	Average	Crowd Run	Ski	Soccer	Average	Crowd Run	Ski	Soccer	Average	
x264	Bitrate [Mbit/s]	110.48	27.47	37.11	58.36	53.41	13.24	16.28	27.64	28.33	6.28	7.29	13.97	33.32
	LPC-SI*100	92.273	89.330	92.403	91.335	92.275	89.331	92.400	91.335	92.264	89.222	92.343	91.277	91.316
	FSIM*100	99.908	99.819	99.852	99.860	99.879	99.764	99.794	99.812	99.673	99.383	99.465	99.507	99.726
	PSNR [dB]	40.39	43.97	43.36	42.57	39.22	42.86	42.09	41.39	35.24	40.02	39.05	38.10	40.69
AWA 3x3 JND +x264	Δ Bitrate [%]	-12.36	-17.24	-11.99	-13.87	-11.24	-22.68	-18.01	-17.31	-9.68	-12.72	-10.99	-11.13	-14.10
	CMOS	0.00	-0.50	0.00	-0.17	-0.44	0.06	-0.13	-0.17	-0.56	-0.31	-0.25	-0.38	-0.24
	$\pm\delta_{95\%}$	0.18	0.54	0.44	0.24	0.56	0.46	0.30	0.26	0.40	0.43	0.42	0.24	0.14
	Δ LPC-SI*100	0.136	0.156	0.072	0.121	0.139	0.135	0.077	0.117	0.133	0.122	0.056	0.104	0.114
	Δ FSIM*100	-0.088	-0.105	-0.095	-0.096	-0.090	-0.109	-0.099	-0.099	-0.109	-0.131	-0.100	-0.113	-0.103
	Δ PSNR [dB]	-4.16	-1.98	-2.09	-2.74	-3.27	-1.43	-1.45	-2.05	-1.69	-0.68	-0.77	-1.05	-1.95
BiLAWA 11x11 JND +x264	Δ Bitrate [%]	-15.79	-23.99	-17.16	-18.98	-15.66	-28.73	-24.33	-22.90	-14.20	-18.11	-15.95	-16.09	-19.32
	CMOS	-0.13	-0.31	-0.06	-0.17	0.19	-0.13	-0.38	-0.10	-0.38	-0.25	0.00	-0.21	-0.16
	$\pm\delta_{95\%}$	0.43	0.46	0.46	0.26	0.48	0.47	0.53	0.29	0.43	0.33	0.47	0.24	0.15
	Δ LPC-SI*100	0.187	0.162	0.080	0.143	0.188	0.144	0.086	0.139	0.187	0.117	0.062	0.122	0.135
	Δ FSIM*100	-0.343	-0.434	-0.336	-0.371	-0.345	-0.425	-0.333	-0.367	-0.391	-0.483	-0.347	-0.407	-0.382
	Δ PSNR [dB]	-5.32	-3.15	-3.17	-3.88	-4.42	-2.43	-2.38	-3.08	-2.51	-1.33	-1.39	-1.74	-2.90
TBil 11x11 JND +x264	Δ Bitrate [%]	-11.84	-21.27	-13.15	-15.42	-9.51	-24.76	-17.56	-17.28	-7.35	-14.16	-9.22	-10.24	-14.31
	CMOS	-0.06	-0.19	-0.50	-0.25	-0.50	-0.19	-0.50	-0.40	-0.19	-0.19	-0.31	-0.23	-0.29
	$\pm\delta_{95\%}$	0.52	0.41	0.36	0.25	0.40	0.41	0.31	0.22	0.45	0.37	0.50	0.25	0.15
	Δ LPC-SI*100	0.107	0.312	0.078	0.166	0.105	0.293	0.080	0.159	0.105	0.271	0.056	0.144	0.156
	Δ FSIM*100	-0.260	-0.400	-0.287	-0.316	-0.259	-0.389	-0.279	-0.309	-0.296	-0.438	-0.280	-0.338	-0.321
	Δ PSNR [dB]	-4.07	-2.77	-2.41	-3.08	-3.20	-2.08	-1.68	-2.32	-1.62	-1.04	-0.86	-1.17	-2.19

Table V

COMPARISON OF THE PSNR VALUES OF THE CODED SEQUENCES WHEN TAKING THE ORIGINAL OR THE FILTERED SEQUENCE AS THE REFERENCE. THE SEQUENCES ARE ENCODED WITH THE x264 CODEC AT QP 27, NO DEBLOCKING FILTER.

	PSNR[dB] reference	<i>Crowd Run</i>	<i>Ski</i>	<i>Soccer</i>	<i>Average</i>
x264	Original	35.24	40.02	39.05	38.10
AWA 3x3 JND + x264	Original	33.55	38.95	38.16	36.89
	Filtered	36.03	40.24	39.10	38.46
BiLAWA 11x11 JND + x264	Original	32.73	38.45	37.80	36.33
	Filtered	36.23	40.66	39.44	38.78
TBil 11x11 JND + x264	Original	33.63	38.40	38.13	36.72
	Filtered	36.01	40.52	39.25	38.60

pre-filters have a more strongly impact on a GOP using inter-frames (P or B) than an I only GOP. This observation shows that the pre-filters tend to better improve the inter prediction (motion estimation and compensation) than the intra prediction. However, one can notice that for the sequence with the highest spatial activity (*CrowdRun*), the pre-filters save more bitrate for the intra prediction.

e) Content impact:

Given the higher initial rate of *CrowdRun* which contains more high frequency details, one could have expected a higher bitrate reduction compared to *Ski*. However, the adaptive filters smooth the texture around the edges: this may lead to amplifying the discontinuity (activity) around the edges, with in turn an increase of bitrates in these areas. As a result, the *CrowdRun* sequence with a lot of edges, obtains a lower bitrate saving (Figure 16).

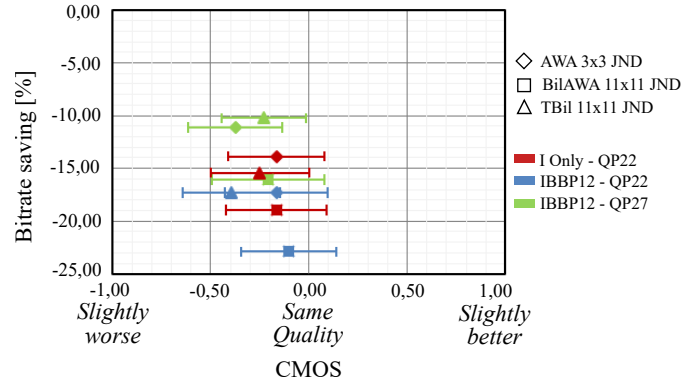


Figure 15. Evaluation of GOP and QP parameter impact on proposed pre-filters performances. CMOS versus bitrate saving for AWA, BiLAWA and TBil perceptual pre-filters in comparison with encoding scheme without pre-filters. Average results on *CrowdRun*, *Ski* and *Soccer* sequences, for x264 without deblocking filter.

f) Deblocking filter impact:

Table VI presents the results with the deblocking filter at QP 27 and IBBP GOP structure with a GOP length of 12 frames. One can notice that the use of the deblocking filter indeed increases the PSNR of the x264 codec (first row of the Table). But the results also show that the proposed filters bring gains when using the deblocking filter which are very similar to those we had with no deblocking filter, with similar variations in LPC-SI metrics, FSIM and PSNR.

E. Performance analysis of proposed JND-guided filters with HEVC

Finally, Table VII presents the results with x265 codec at QP 27 and IBBP GOP structure with a GOP length of 12 frames. These

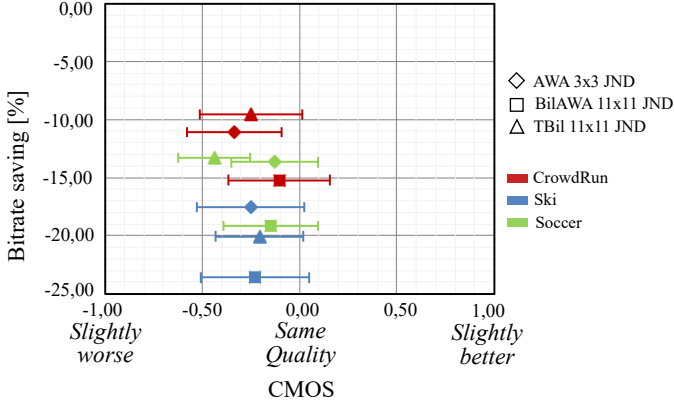


Figure 16. Evaluation of GOP and QP parameter impact on proposed pre-filters performances. CMOS versus bitrate saving for AWA, BilAWA and TBil perceptual pre-filters in comparison with encoding scheme without pre-filters. Average results on GOP and QP configuration for x264 without deblocking filter.

Table VI

COMPARISON OF ENCODING PERFORMANCES OF X264 CODEC WITH AND WITHOUT PERCEPTUAL PRE-FILTERS. ANALYSIS OF BITRATE REDUCTION (Δ Bitrate), OBJECTIVE MEASURES VARIATION (Δ LPC-SI, Δ FSIM AND Δ PSNR). X264 IS USED WITH THE DEBLOCKING FILTER AT QP 27 AND GOP IBBP12.

		Crowd Run	Ski	Soccer	Average
x264	Bitrate [Mbit/s]	31.36	7.02	8.43	15.60
	LPC-SI*100	92.262	89.297	92.360	91.306
	FSIM*100	99.670	99.360	99.460	99.497
	PSNR [dB]	35.74	40.54	39.61	38.63
AWA 3x3	Δ Bitrate [%]	-9.80	-12.13	-11.11	-11.02
	Δ LPC-SI*100	0.141	0.117	0.050	0.102
	Δ FSIM*100	-0.130	-0.180	-0.130	-0.147
	Δ PSNR [dB]	-1.81	-0.74	-0.82	-1.12
BilAWA 11x11	Δ Bitrate [%]	-14.31	-17.42	-16.04	-15.92
	Δ LPC-SI*100	0.187	0.103	0.074	0.121
	Δ FSIM*100	-0.440	-0.580	-0.410	-0.477
	Δ PSNR [dB]	-2.67	-1.46	-1.49	-1.87
TBil 11x11	Δ Bitrate [%]	-7.51	-13.83	-9.55	-10.30
	Δ LPC-SI*100	0.108	0.258	0.071	0.145
	Δ FSIM*100	-0.340	-0.520	-0.330	-0.397
	Δ PSNR [dB]	-1.76	-1.17	-0.94	-1.29

results show that the proposed filters bring gains when using HEVC which are very similar to those obtained with x264, with similar variations in LPC-SI, FSIM and PSNR metrics. One can observe that the rate gains brought by the proposed filters are even higher with HEVC than with MPEG-4/AVC. The reason is that the pre-filtering removes high frequency details, which favors the use of large CTU sizes in the quadtree partition of HEVC. This in turn induces a higher bitrate reduction. Our solutions bring bitrate reduction ratios varying from 11.1% to 17.4% depending on the selected filter. Such results are in accordance with the ones recently described in the literature, 9.6% [48] and 11.0% [49] on average. These results are obtained for similar test data and coding conditions (four and three HD test videos respectively, Main-Profile RA HEVC encoding using HM.11 encoder, QP 27, original bitrates of 2 Mbps to 5 Mbps). Finally, although formal subjective evaluation tests were not conducted in HEVC, our informal evaluation show that subjective quality loss between the HEVC compressed sequences with and without our perceptual pre-processing is imperceptible like in the H.264/AVC case. Thus we conclude that our pre-filtering methods can give significant bitrate

savings while keeping the same video quality.

Table VII

COMPARISON OF ENCODING PERFORMANCES OF X265 CODEC WITH AND WITHOUT PERCEPTUAL PRE-FILTERS. ANALYSIS OF BITRATE REDUCTION (Δ Bitrate), OBJECTIVE MEASURES VARIATION (Δ LPC-SI, Δ FSIM AND Δ PSNR). X265 IS USED WITH THE LOOP FILTERS AT QP 27 AND GOP IBBP12.

		Crowd Run	Ski	Soccer	Average
x265	Bitrate [Mbit/s]	28.13	5.49	6.29	13.30
	LPC-SI*100	92.244	89.157	92.260	91.220
	FSIM*100	99.790	99.780	99.830	99.800
	PSNR [dB]	35.22	40.56	39.77	38.52
AWA 3x3	Δ Bitrate [%]	-10.69	-13.32	-12.16	-12.06
	Δ LPC-SI*100	0.120	0.086	0.059	0.088
	Δ FSIM*100	-0.270	-0.600	-0.510	-0.460
	Δ PSNR [dB]	-1.87	-0.89	-0.97	-1.24
BilAWA 11x11	Δ Bitrate [%]	-15.44	-19.26	-17.35	-17.35
	Δ LPC-SI*100	0.164	0.115	0.068	0.116
	Δ FSIM*100	-0.460	-0.990	-0.790	-0.747
	Δ PSNR [dB]	-2.70	-1.67	-1.69	-2.02
TBil 11x11	Δ Bitrate [%]	-8.10	-15.01	-10.15	-11.08
	Δ LPC-SI*100	0.093	0.231	0.067	0.130
	Δ FSIM*100	-0.580	-0.940	-0.710	-0.743
	Δ PSNR [dB]	-1.77	-1.33	-1.11	-1.40

VII. CONCLUSION

In this paper, we have first described two novel adaptive filters (BilAWA and TBil) which efficiently exploit the features of the AWA and bilateral filters, while being amenable to the introduction of a control of the filtering process by a JND model. We have shown that these novel adaptive filters outperform the classical AWA and bilateral filters for noise removal. The introduction of the JND model then leads to perceptual adaptive filters which exhibit a strong interest as low complex real-time pre-processing techniques to improve HD real-time video compression efficiency by removing imperceptible details. Psychovisual evaluation tests have been conducted in order to validate the performances of the JND-guided adaptive pre-filters. Experimental results have exhibited significant bitrate saving without compromising the perceived visual quality. A maximum rate saving of 28.7% and an average rate saving of 19.3% have been obtained with the perceptual BilAWA filter applied before MPEG-4/AVC encoding at QP 22. Similarly, when using HEVC, a maximum bitrate saving of 19.3% and an average bitrate saving of 17.4% have been obtained at QP 27. Such performance is comparable to those recently described in the literature. The results of the different selected objective metrics corroborate the subjective quality rating as they do not reveal significant loss of quality or excessive blurring. Note that, although experimental results are given with one particular JND model, the proposed filters are independent of the model: the filtering parameters could as well be controlled by other JND pixel-domain models. Moreover the pre-filters act outside the encoding loop and can consequently be used with any image or video encoding scheme (JPEG XS, VP10). Further work will concern the improvement of the JND model including spatio-temporal and chrominance sensitivity. Moreover, the pre-processing algorithm could be implemented in the coding loop: in this case, it could benefit from the data computed during the pre-analysis step to weight the JND perceptual index or to adapt dynamically the filter kernel's support depending on the selected Coding Tree Unit (CTU) size, in the case of HEVC-compliant pre-processing.



Figure 17. Comparison of quality between a frame of the sequence *CrowdRun* encoded at QP22 without pre-filter (39.11dB) (a), with TBil(11x11, JND) ($PSNR_{frame} = 36.14\text{dB}$, $CMOS_{sequence} = -0.50$, $BitrateSaving_{sequence} = 9.51\%$) (b) and BilAWA(11x11, JND) ($PSNR_{frame} = 34.84\text{dB}$, $CMOS_{sequence} = 0.19$, $BitrateSaving_{sequence} = 15.66\%$) (c)

ACKNOWLEDGMENT

The Author would like to thank Marine Deneuve whom internship was the starting point for this work. This work has been supported by the French ANRT (Cifre #1098/2010).

REFERENCES

- [1] P. Karunaratne, C. Segall, and A. Katsaggelos, "A rate-distortion optimal video pre-processing algorithm," *IEEE International Conference on Image Processing*, vol. 1, pp. 481–484, Oct. 2001.
- [2] H. Kacem, F. Kammoun, and M. Bouhlef, "Improvement of the compression JPEG quality by a pre-processing algorithm based on denoising," *IEEE International Conference on Industrial Technology*, vol. 3, pp. 1319–1324, Dec. 2004.
- [3] P. van Roosmalen, R. Legendijk, and J. Biemond, "Embedded coring in MPEG video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 205–211, Mar. 2002.
- [4] O. Al-Shaykh and R. Mersereau, "Lossy compression of noisy images," *IEEE Transactions on Image Processing*, vol. 7, no. 12, pp. 1641–1652, Dec. 1998.
- [5] P. V. Roosmalen, A. C. Kokaram, and J. Biemond, "Noise reduction of image sequences as preprocessing for MPEG2 encoding," *European Signal Processing Conference*, no. 19, pp. 2253–2256, Sept. 1998.
- [6] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [7] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE International Conference on Computer Vision, ICCV*, 1998, pp. 836–846.
- [8] B. Shreyamsha Kumar, "Image denoising based on gaussian/bilateral filter and its method noise thresholding," *Signal, Image and Video Processing*, vol. 7, no. 6, pp. 1159–1172, 2013.
- [9] M. Ozkan, I. Sezan, and M. Tekalp, "Adaptive motion-compensated filtering of noisy image sequences," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 3, no. 4, pp. 277–290, Aug. 1993.
- [10] D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 844–877, June 2002.
- [11] E. Vidal, T. Hauser, P. Corlay, and F. Coudoux, "An adaptive video pre-processor based on just-noticeable distortion," in *Proc. 6th International Symposium on Signal, Image, Video and Communications, ISIVC 2012*, July 2012.
- [12] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, June 2005, pp. 60–65 vol. 2.
- [13] M. Zhang and B. Gunturk, "Multiresolution bilateral filtering for image denoising," *IEEE Transactions on Image Processing*, vol. 17, no. 12, pp. 2324–2333, Dec 2008.
- [14] X. Yang, W. Lin, Z. lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 6, pp. 742–752, June 2005.
- [15] L. Ding, G. Li, R. Wang, and W. Wang, "Video pre-processing with JND-based Gaussian filtering of superpixels," *Proc. SPIE 9410, Visual Information Processing and Communication VI*, 941004, March 2015.
- [16] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.
- [17] J. Lee, "Automatic prefilter control by video encoder statistics," *IET Electronics Letters*, vol. 38, pp. 503–505, May 2002.
- [18] C. Jain and S. Sethuraman, "A low-complexity, motion-robust, spatiotemporally adaptive video de-noiser with inloop noise estimation," *IEEE International Conference on Image Processing*, pp. 557–560, Oct. 2008.
- [19] L. Guo, O. Au, M. Ma, and P. Wong, "Integration of recursive temporal Immse denoising filter into video codec," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 2, pp. 236–249, Feb. 2010.
- [20] T. Chan, T.-C. Hsung, and D.-K. Lun, "Improved mpeg-4 still texture image coding under noisy environment," *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 500–508, May 2003.
- [21] B. Song and K. Chun, "Motion-compensated temporal filtering for denoising in video encoder," *Electronics Letters*, vol. 40, pp. 802–804, June 2004.
- [22] R. Kawada, A. Koike, and Y. Nakajima, "Prefilter control scheme for low bitrate tv distribution," *IEEE International Conference on Multimedia and Expo*, pp. 769–772, Jul. 2006.
- [23] C. A. Segall, P. Karunaratne, and A. K. Katsaggelos, "Pre-processing of compressed digital video," *SPIE Image and Video Communication And Processing*, 4310: 163–174, Jan. 2001.
- [24] L. Shao-Ping and Z. Song-Hai, "Saliency-based fidelity adaptation preprocessing for video coding," *Journal of Computer Science and Technology*, vol. 26, pp. 195–202, 2011.
- [25] R. Vanam, L. Kerofsky, and Y. Reznik, "Perceptual pre-processing filter for video on demand content delivery," *Proc. IEEE International Conference on Image Processing*, pp. 2537–2541, October 27–30 2014.
- [26] H. Kwon, H. Han, S. Lee, W. Choi, and B. Kang, "New video enhancement preprocessor using the region-of-interest for the videoconferencing," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2644–2651, Nov. 2010.
- [27] S.-P. Lu and S.-H. Zhang, "Saliency-based fidelity adaptation preprocessing for video coding," *Journal of Computer Science and Technology*, vol. 26, pp. 195–202, Jan. 2011.
- [28] M. Naccari and F. Pereira, "Advanced h,264/avc-based perceptual video coding: Architecture, tools, and assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 6, pp. 766–782, Jun. 2011.
- [29] —, "Integrating a spatial just noticeable distortion model in the under development hevcc codec," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 817–820, 22–27 May 2011.

- [30] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [31] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–73, Oct. 2009.
- [32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [33] R. Hassen, Z. Wang, and M. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. on Image Processing*, vol. 22, no. 7, pp. 2798–2810, July 2013.
- [34] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to jpeg2000," *Signal Processing: Image Communication*, vol. 19, pp. 163–172, Dec. 2004.
- [35] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, pp. 297–312, May 2011.
- [36] Z. Luo, L. Song, S. Zheng, and N. Ling, "H.264/advanced video control perceptual optimization coding based on jnd-directed coefficient suppression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 6, pp. 935 – 948, June 2013.
- [37] "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT.500-13*, Jan. 2012.
- [38] M. Yuen, "Coding artifacts and visual distortions," in *Digital Video Image Quality and Perceptual Coding*, Eds. H.R.Wu and K.R. Rao, CRC Press, pp. 87–122, May 2006.
- [39] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Trans. Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [40] Z. Wei and K.N. Ngan, "Spatio-Temporal Just Noticeable Distortion Profile for Grey Scale Image/Video in DCT Domain," *IEEE Transactions on Systems and Circuits for Video Technology*, vol. 19, no. 3, pp.337–346, Mar. 2009.
- [41] S.H. Bae and M. Kim, "A Novel Generalized DCT-Based JND Profile Based on an Elaborate CM-JND Model for Variable Block-Sized Transforms in Monochrome Images," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp.3227–3240, Aug. 2014.
- [42] S.H. Bae and M. Kim, "A DCT-Based Total JND Profile for Spatio-Temporal and Foveated Masking Effects," accepted for publication in *IEEE Transactions on Systems and Circuits for Video Technology*, DOI: 10.1109/TCSVT.2016.2539862, 2016.
- [43] G. De Haan, and E.B. Bellers, "Deinterlacing—An Overview," *Proc. of the IEEE*, vol. 86, no. 9, pp. 1839–1857, Sept. 1998.
- [44] T. Chen, Z. Yu and H.R. Wu, "Efficient edge line average interpolation algorithm for deinterlacing," *Proc. of SPIE*, 4067, pp. 1551–1558, May 2000.
- [45] A. Buades, J.L. Lisani and M. Miladinovic, "Patch-Based Video Denoising With Optical Flow Estimation," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2573–2586, June 2016.
- [46] G. Varghese and Z. Wang, "Video Denoising Based on a Spatiotemporal Gaussian Mixture Model," *IEEE Transactions on Systems and Circuits for Video Technology*, vol. 20, no. 7, pp. 1032–1040, Jul. 2010.
- [47] H. Oh and W. Kim, "Video Processing for Human Perceptual Visual Quality-Oriented Video Coding," *IEEE Transactions on Systems and Circuits for Video Technology*, vol.22, no. 4, pp. 1526–1535, 2016.
- [48] J. Kim, and M. Kim, "An HEVC-Compliant Perceptual Video Coding Scheme Based on JND Models for Variable Block-Sized Transform Kernels," *IEEE Transactions on Systems and Circuits for Video Technology*, vol. 25, no. 11, 1786–1800, Nov. 2015.
- [49] S.H. Bae, J.Kim, M. Kim, "HEVC-Based Perceptually Adaptive Video Coding Using a DCT-Based Local Distortion Detection Probability Model," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3543–3557, Jul. 2016.
- [50] S. Wang, A. Rehman, Z. Wang, S. Ma, W. Gao, "Perceptual Video Coding Based on SSIM-Inspired Divisive Normalization," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1418–1429, 2013.
- [51] M.Q. Shaw, J.P. Allebach, E.J. Delp, "Color difference weighted adaptive residual preprocessing using perceptual modeling for video compression," *Signal Processing: Image Communication*, vol 39, pp. 355–368, Apr. 2015.
- [52] L.J. Kerofsky, R. Vanam, and Y.A. Reznik, "Improved Adaptive Video Delivery System Using a Perceptual Pre-processing Filter," *Proc. GlobaSIP 2014*, pp. 1058–1062, 2014.
- [53] R. Vanam, and Y.A. Reznik, "Perceptual pre-processing filter for user-adaptive coding and delivery of visual information," *Proc. PCS 2013*, Dec. 2013.
- [54] "High Definition Television," *Sony Training Services*, Technical documentation Ref. STS/POSTER/HOGHDEF/V2, 2007.



Eloïse Vidal is Research Project Manager at Digigram S.A. (Montbonnot, France). She received the Ph.D. degree in 2014 from the University of Valenciennes, France. Since 2010, she has developed close collaborations with the Departement of Opto-Acousto-Electronics of the Institute of Electronics, Microelectronics, and Nanotechnologies, France (UMR 8520). Her research interests include video preprocessing, video quality, H.264/AVC and HEVC encoding and more generally video compression optimization.



Nicolas Sturm is head of research at Digigram S.A., Montbonnot, France. He holds a PhD degree in Signal Processing (2011) from University Paris XI-Orsay. From 2011 to 2012 he was a post-doctoral fellow with Institut Langevin - ESPCI, Paris, working on informed source separation and audio coding. His research interests include processing and delivery of multimedia signals, audio and video coding, source separation, audio recording and mixing.



Christine Guillemot is currently Director of Research at INRIA (Institut National de Recherche en Informatique et Automatique) in France. She holds a PhD degree from ENST (Ecole Nationale Supérieure des Telecommunications) Paris (1992). From 1985 to 1997, she has been with France Télécom in the areas of image and video compression for multimedia and digital television. From 1990 to mid 1991, she has worked as "visiting scientist" at Bellcore Bell Communication research) in the USA. Her research interests are signal and image processing, and in

particular 2D and 3D image and video coding, joint source and channel coding for video transmission over the Internet and over wireless networks, and distributed source coding.

She has served as Associate Editor for IEEE Trans. on Image Processing (from 2000 to 2003), for IEEE Trans. on Circuits and Systems for Video Technology (from 2004 to 2006), and for IEEE Trans. on Signal Processing (2007–2009). She is currently associate editor of the Eurasp journal on image communication (since 2010), for the IEEE Trans. on Image Processing (2014–2016), and for the IEEE journal on selected topics in signal processing (since 2013). She has been a member of the IEEE IMDSP (2002–2007) and IEEE MMSP (2005–2008) technical committees. She is currently a member of the IEEE IVMSIP - Image Video Multimedia Signal Processing - technical committee (since 2013). She is the co-inventor of 24 patents, she has co-authored 9 book chapters, 59 international journal publications and around 150 articles in peer-reviewed international conferences. She is IEEE fellow since January 2013.



Patrick Corlay received the Ph.D. degree in 1994 from the University of Valenciennes, France. He is an Assistant Professor at the Department of Opto-Acousto-Electronics of the Institute of Electronics, Microelectronics, and Nanotechnologies, France (UMR 8520). His current research interests are in telecommunications, digital transmission over wired networks (power line, ADSL), and optimal quality of service for video transmission



François-Xavier Coudoux (M'07) received the M.S. and Ph.D. degrees in electrical engineering from the University of Valenciennes, Valenciennes, France, in 1991 and 1994, respectively. Since 2004, he has been a Professor in the Department of Opto-Acousto-Electronics of the Institute of Electronics, Microelectronics, and Nanotechnologies, Valenciennes, France (UMR 8520). His research interests include telecommunications, multimedia delivery over wired and wireless networks, image quality, and adaptive video processing.