IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 11, 2020, accepted March 29, 2020, date of publication April 1, 2020, date of current version April 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984994

Retinal Noise Emulation: A Novel Artistic Tool for Cinema That Also Improves Compression Efficiency

ITZIAR ZABALETA[1], MATEO CÁMARA[2], CÉSAR DÍAZ[2], TREVOR CANHAM[1], NARCISO GARCÍA[2], AND MARCELO BERTALMÍO[1]

[1]Image Processing for Enhanced Cinematography Group, Universitat Pompeu Fabra, 08035 Barcelona, Spain
[2]Grupo de Tratamiento de Imágenes, Information Processing and Telecommunications Center and ETSI Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Corresponding author: Itziar Zabaleta (itziar.zabaleta@upf.edu)

This work was supported in part by the European Union through the Horizon 2020 Research and Innovation Programme (Project HDR4EU) under Grant 761544 and (Project SAUCE) Grant 780470, and in part by the Spanish Government, Ministerio de Economía, Industria y Competitividad (AEI/FEDER), under Project TEC2016-75981 (IVME) and Project PGC2018-099651-B-I00.

ABSTRACT In cinema it is standard practice to improve the appearance of images by adding noise that simulates film grain. This is computationally very costly, so it is only done in post-production and not on the set. It is also limiting because the artists are not able to really experiment with the noise nor introduce novel looks. Furthermore, video compression requires a higher bit rate when the source material has film grain or any other type of high frequency texture. In this work, we introduce a method for adding texture to digital cinema that aims to solve these problems. The proposed algorithm is based on modeling retinal noise, with which the images processed by our method have a natural appearance. This ''retinal grain'' serves a double purpose. One is aesthetic, as it has parameters that allow to vary widely the resulting texture appearance, which make it an artistic tool for cinematographers. Results are validated through psychophysical experiments in which observers, including cinema professionals, prefer our method over film grain synthesis methods from academia and the industry. The other purpose of the retinal noise emulation method is to improve the quality of compressed video by masking compression artifacts, which allows to lower the encoding bit rate while preserving image quality, and to improve image quality while keeping the bit rate fixed. The effectiveness of our approach for improving coding efficiency, with average bit rate savings of 22.5%, has been validated through psychophysical experiments using professional cinema content shot in 4K, color-graded and where the amount of retinal noise was selected by a motion picture specialist based solely on aesthetic preference.

INDEX TERMS Image enhancement, noise, image perception, vision models, QoE, video compression, adaptive streaming.

I. INTRODUCTION

After the digital cinema revolution, many directors and cinematographers are becoming increasingly frustrated by some artistic limitations that the digital medium imposes. Since the beginning of cinema and for many decades, there was a wide variety of cameras, film stocks and film developing options that allowed cinematographers to experiment, find and test new possibilities for creative expression. But currently, most professional productions resort to the same digital cinema camera model, causing the default look to be quite homogeneous to begin with.

An established way to improve the appearance of digital images is to add to them a certain amount of fine-detail texture, and user studies have shown that observers indeed prefer images with some noise [1], [2]. In cinema and TV fiction the standard practice is to always add texture to digitally-shot content, and this texture invariably takes the form of film grain. This particular choice of texture aims to mimic the look of film, that is still considered as the gold standard by many cinematographers. In the movie industry the most popular methods for film grain emulation are based on assembling a

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Asikuzzaman.

database of scannings of different types of film stock with varying forms of grain, that are superimposed on the digital image that is processed. For these methods there is a compromise between speed and the realism of the result: the fastest algorithms overlay grain that is somewhat independent from the content of the digital image, which may produce noticeable artifacts especially when there is motion, and the methods producing the more visually pleasing results are computationally very intensive, requiring special hardware. In the academic literature the state of the art is the work of Newson *et al.* [3], who study the photographic process and analyze the distribution of grain in the film emulsion, and then use a computationally costly algorithm based on stochastic geometry in order to produce realistic-looking synthesized grain.

Due to their computational complexity, existing methods impose a restriction on the creative work of filmmakers, preventing them from having on the set an accurate representation of how the movie will look in postproduction after the synthetic texture is added. The emulation of film grain is also limiting because the artists are not able to really experiment with a wide diversity of texture options, nor to introduce novel looks. An added complexity is the following: image content with high frequency detail, like film grain, requires a higher bit rate in order to be compressed properly, i.e. for a given visual quality level, ''clean'' video requires less bits than video with film grain. For this reason there are several works in the field of video compression [4]–[7] that improve coding efficiency by a process consisting, first, of denoising the input video (removing the film grain), then modeling the noise with a set of parameter values which are transmitted alongside the denoised video, and finally at the decoder re-synthesizing the film grain noise and adding it back to the decoded denoised video. We are not aware of any of these methods actually being used in practice in video streaming, and they require practical solutions to very challenging problems like video denoising.

Regarding this latter point, the importance of coding efficiency cannot be emphasized enough. In the media industry there's a constant push in cinema, broadcast and streaming services towards ever higher resolution, framerate and dynamic range. The accompanying increase in data volume that these new formats bring imposes considerable demands on transmission bandwidth and memory, and as a result the problem of compressing video is as relevant as ever. Specifically, Ultra High-Definition (UHD)/4K video requires bit rates nearly 10 times higher than Standard Definition (SD) video and Full HD (FHD)/2K around four or five times higher [8], [9]. In addition, video traffic is expected to account for 82% of all IP traffic by 2022. Moreover, by that year, UHD/4K video will be around 22% of IP video traffic, and FHD/2K video around 57% [9]. Faced with these remarkable data, content and service providers are constantly searching for ways to provide increasingly higher quality video and quality of experience (QoE) at restrained bit rates.

In this respect, they can exploit the fact that most users are not able to perceive some objective quality drops under some conditions [10], [11]. In fact, the visibility of image distortions is reduced by the presence of another stimulus, a masking pattern. This phenomenon is called ''visual masking'' and it is a well known property of visual perception. Visual masking takes several forms, as it depends on different properties of the image stimuli: luminance, color, temporal variations, spatial patterns. Visual masking is a key perceptual phenomenon for the design of image and video compression algorithms [12], and in particular texture masking or pattern masking has been successfully applied for video coding [13].

In this work we propose a method for adding texture to digital cinema that overcomes all the abovementioned limitations.

The proposed algorithm emulates retinal noise, and the motivation for our using a retinal noise model is that the resulting images will have a more natural appearance, since noise is always present in retinal signals, in photopic (daytime) vision as well. Nonetheless, in our case the magnitude of emulated retinal grain that is added to day-like scenes is an artistic choice: it must be noticed that the method is proposed as an artistic tool and an aesthetic alternative to film grain emulation, rather than a physiologically accurate simulation of perceived noise. It has a very low computational complexity, so it is amenable for a real-time implementation that can be used on set. It has parameters that can be varied to achieve a wide range of texture appearances, allowing movie creators to try out new looks. Results are validated through psychophysical experiments in which observers, including cinema professionals, prefer our method over film grain emulation alternatives from academia and the industry.

Another contribution of our work is to show that the retinal noise emulation can also be used to improve the quality of compressed video by masking compression artifacts. Once the movie creator has taken the artistic decision to add a certain amount and type of this ''retinal grain'' to improve the look of the digital film, the movie distributor can use this fact to its advantage by encoding the ''clean'' content at a lower bit rate and adding the retinal grain after decoding, the exact same grain that the content creator decided was right for the movie for aesthetic reasons, thus masking the visual artifacts produced by the reduced bit rate and yielding the same QoE of a higher bit rate. The extra data that is required at reception to introduce the retinal noise is negligible, as it only consists of the values for the user parameters (up to 5 floating point numbers per frame). This is completely novel because in the literature, as mentioned above, the grain is roughly estimated via a denoising process (which is an open problem), parametric models of film grain provide coarse approximations, and those works have a limited application because they are intended just for films with grain, whereas our approach can be used with any kind of content. We performed psychophysical experiments using color-graded professional cinema content shot in 4K, where the amount of retinal noise was selected by a motion picture specialist based

solely on aesthetic preference. This content was encoded at different bit rates, and the retinal grain added after decoding. The participants rated the quality of the resulting videos, and the results show that when reducing the bit rate, the loss of perceived quality is consistently smaller when the video has had retinal grain added to it than when it has not. Our method is shown to yield remarkable savings in bit rate, of over 22.5% on average.

Additionally, we can mention some other application scenarios for our method. As the proposed scheme is able to provide better subjective quality to compressed video, it can also be applied to scalable video, thus opening its use in the adaptive bitrate scenarios considered in streaming applications. Even more, it can be applied in novel multicamera scenarios, like high quality free viewpoint video, where the synthesis artifacts due to occlusions and missing data could be hidden by the addition of retinal noise. Finally, our method can be applied in-camera to enhance photographs and video, specially in the case of acquisition devices with limited capabilities.

As a closing point, we want to briefly discuss the possibility of our proposed work being replaced by a deep neural network (DNN) procedure. In our opinion, given that the applications discussed in this paper are all based in the perceived (aesthetic) appearance of images and videos, the use of a DNN for these tasks would first require the ability of said DNN to represent aesthetic preference, and while there are some recent works in this regard, e.g. [14], [15], they have been shown to be unsuitable in the professional media production scenarios [16], [17] for which the method introduced in the current paper is intended.

II. SOME VISION FACTS AND MODELS

The retina is a thin sheet of neural tissue that lines the back of the eye and transforms light into electrical signals. It is composed of five cell types that are arranged in three cellular layers separated by two in-between layers, called plexiform layers (See Fig.1). The photoreceptor cells (rod and cones) in the outermost layer, absorb light and convert it into electrical signals. These signals are passed to bipolar cells, which in turn connect to retinal ganglion cells in the innermost layer. In addition to this vertical pathway the retinal circuit includes many lateral connections provided by horizontal cells in the outer plexiform layer and amacrine cells in the inner synaptic layer. Retinal ganglion cells are the output neurons of the retina and their axons form the optic nerve that transmits the visual signal from the retina to the brain.

Photoreceptors transform the light reaching the retina into electrical signals. The response of photoreceptors is nonlinear and, for a single cell without feedback, can be well approximated by the Naka-Rushton equation [19], which is a particular instance of a divisive normalization operation [20], i.e. a process that computes the ratio between the response of an individual neuron and some weighted average of the activity of its neighbors, and this in turns allows the photoreceptor response to adapt to the average light level therefore optimizing its operative range.

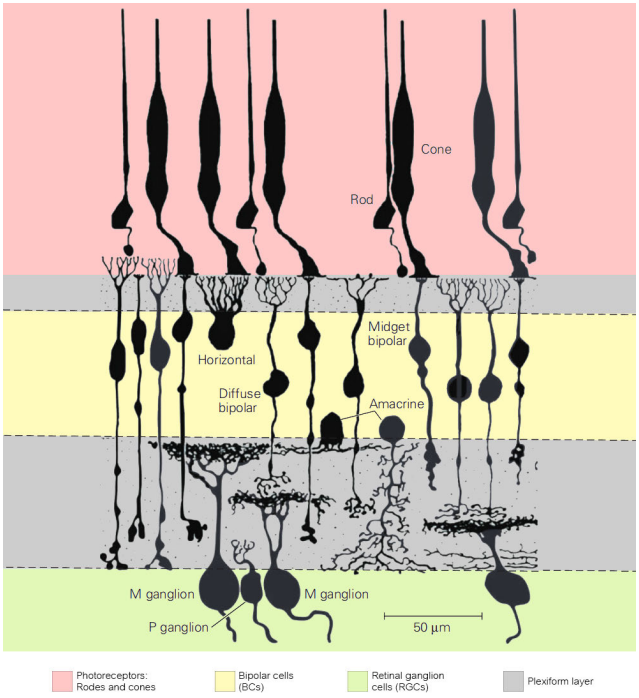


FIGURE 1. Neurons in the retina of the macaque monkey. Figure from [18].

The lateral inhibition or center-surround processing, in which a cell's response is modeled as the difference between the activity of the cell's closest neighbors and the activity of the cells in the near ring-shaped surround, allows to encode and enhance contrast therefore being key for efficient representation, and is present at every stage of visual processing from the retina to the cortex. Lateral inhibition is often modeled as a linear operation, a convolution with a kernel shaped as a difference of Gaussians (DoG).

In order to study noise in the retina, ganglion cells' responses have been measured with drifting gratings of various spatial and temporal frequencies and contrasts [21]. The precision of cell responses has been estimated in terms of a noise measure, defined as the variability of neuronal spike trains in response to this set of stimulus. It has been shown [21], [22] that noise (variability) remains constant for different levels of contrast, as it can be observed in Figure 2.

III. RETINAL NOISE EMULATION METHOD

The proposed method takes as input an image I, and creates an output image O with added texture emulating retinal noise. The transformations applied to the image are based on neurophysiological models of the visual system. The method can be summarized in the following main stages:

1) Transform the input image I with a model of retinal processes, producing an intermediate image R that emulates the retinal output.
2) Add "retinal" noise to R to obtain a noisy image R_n.

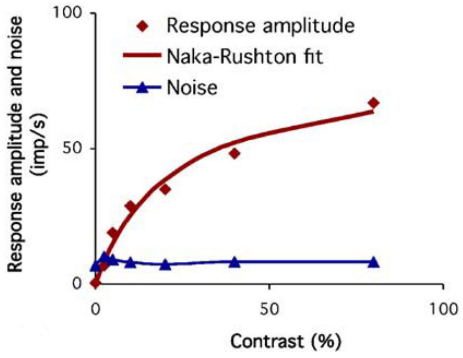


FIGURE 2. Noise is independent of the increase in response amplitude with contrast. Figure from [21].

3) Create the final output image O by applying to R_n the inverse of the previous transformations that emulate retinal processes.

A. THE ALGORITHM

The following transformations are applied separately in each RGB channel of the input image. In this section, the image I will represent each R, G, B channel of the input image in the range [0,1].

Given that our method is based on the emulation of retinal noise, we must start by ensuring that the input image I has values that are proportional to the intensity of light arriving at the retina. This is already the case if I is a RAW image, otherwise we assume that a nonlinear transform like gamma-correction has been applied to I with a standard exponent such as $1/2.2$ [23] and we undo it, obtaining the linear image I_L:

$$I_L = I^{2.2}. \tag{1}$$

After this, the photoreceptor response $P(I_L)$ to the light stimulus I_L is emulated via the Naka-Rushton equation [19], yielding I_P:

$$I_P = P(I_L) = \frac{{I_L}^n}{{I_L}^n + {I_s}^n}, \tag{2}$$

where I_s is the semi-saturation constant and n controls the slope of the Naka-Rushton curve.

The lateral inhibition or center-surround organization of both bipolar cells and retinal ganglion cells is modelled as a convolution between the photoreceptor response and a kernel K similar to a DoG. The resulting image R, which will be our proxy for the clean retinal image, is obtained as follows:

$$R = K * I_P. \tag{3}$$

Motivated by the work of [24], we choose for K the following form:

$$K = \mathcal{F}^{-1}\left(\frac{1}{0.81 + 0.2\mathcal{F}(G_K)}\right) \tag{4}$$

where $\mathcal{F}$ is the Fourier transform and G_K is a 2D Gaussian kernel with standard deviation equal to $1/3$ of the maximum of the image dimensions (height or width). A key advantage of this choice of kernel is that it is invertible, which will be very useful for us as we will see:

$$K^{-1} = \mathcal{F}^{-1}\left(0.81 + 0.2\mathcal{F}(G_K)\right). \tag{5}$$

We add to R a certain amount a of retinal noise n_r, that emulates the noise measured in the RGCs,

$$R_r = R + an_r, \tag{6}$$

and therefore the image R_r corresponds to the noisy image created in the retina.

For the noise signal n_r we use the same distribution as the noise observed in RGCs [21], [25], which has a constant standard deviation (that does not depend on the input contrast), and we impose as well a bandpass frequency spectrum as approximately given by the contrast sensitivity function of the visual system [26]:

$$n_r = (G_c - G_s) * I_{\mathcal{N}}, \tag{7}$$

where $I_{\mathcal{N}}$ is a Gaussian noise image with standard deviation $\sigma = 1$, and G_c and G_s are 2D Gaussian kernels. As it is mentioned in [26], contrast sensitivity depends on the orientation, and this effect could be modeled with the 2×2 covariance matrices Σ_c and Σ_s of the kernels G_c and G_s. In practice, in our experiments we will use symmetric kernels, and in this case the covariance matrices are not needed as the kernels can be described simply with the standard deviation of the Gaussians, σ_c for G_c and σ_s for G_s.

Recapping, the *noisy* retinal image R_r results from adding noise to

$$R = K * P(I^{2.2}), \tag{8}$$

so we can find the *noisy* light stimuli O that would directly produce R_r by undoing the previous chain of operations:

$$O = (P^{-1}(K^{-1} * R_r))^{\frac{1}{2.2}}. \tag{9}$$

The image O is the final output produced by our method, which as mentioned above is applied independently to each of the three color channels.

The linearization, the Naka-Rushton transform and their inverses can be encoded as 1D look-up tables (LUTs), so the method is essentially as fast as the time it takes to compute two convolutions, one with kernel K and the other with its inverse K^{-1}.

B. USER PARAMETERS

The full list of parameters for our method, by order of appearance, is: I_s, n, a, σ_c, σ_s. Default values are proposed for these parameters so the method can be used as fully automatic. Moreover, these parameters can be modified by the user to control the visual aspect of the noise in the resulting image.

For the Naka-Rushton equation, we have fixed both its parameters: $n = 0.74$ following [27], and $I_s = 0.18$ given that 18% is the reflectance value for mid-gray, taking 100%

FIGURE 3. Image with added retinal grain on the left. Close-up of the image with retinal noise in the center, and close-up of the original image on the right.

FIGURE 4. Resulting images with different parameter choices, from top to bottom, left to right: (a) $\sigma_c = 1$, $\sigma_s = 2$, $a = 0.05$, (b) $\sigma_c = 0.5$, $\sigma_s = 1$, $a = 0.05$, (c) $\sigma_c = 0.5$, $\sigma_s = 1$, $a = 0.1$, (d) Non-symmetrical kernels G_c and G_s, with covariance matrices $\Sigma_c = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.05 \end{pmatrix}$ and $\Sigma_s = \begin{pmatrix} 1 & 0 \\ 0 & 0.25 \end{pmatrix}$, $a = 0.05$ (e) $\sigma_c = 0.05$, $\sigma_s = 1$, $a = 0.05$, (f) Non-symmetrical kernels G_c and G_s, with covariance matrices $\Sigma_c = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.4 \end{pmatrix}$ and $\Sigma_s = \begin{pmatrix} 0.25 & 0 \\ 0 & 4 \end{pmatrix}$, $a = 0.05$.

as diffuse white. The intensity of the noise in the final output image O is controlled by the parameter a, that can vary in the range $[0, 1]$, and whose default value is set as $a = 0.015$. As a increases, the noise becomes more visible, as Fig. 4 shows. For the sizes of the Gaussians G_c and G_s used to generate n_r, we have chosen as default values $\sigma_c = 0.7$ and $\sigma_s = 1.5$. These parameters can be adjusted by the user allowing certain control in the size and distribution of the noise.

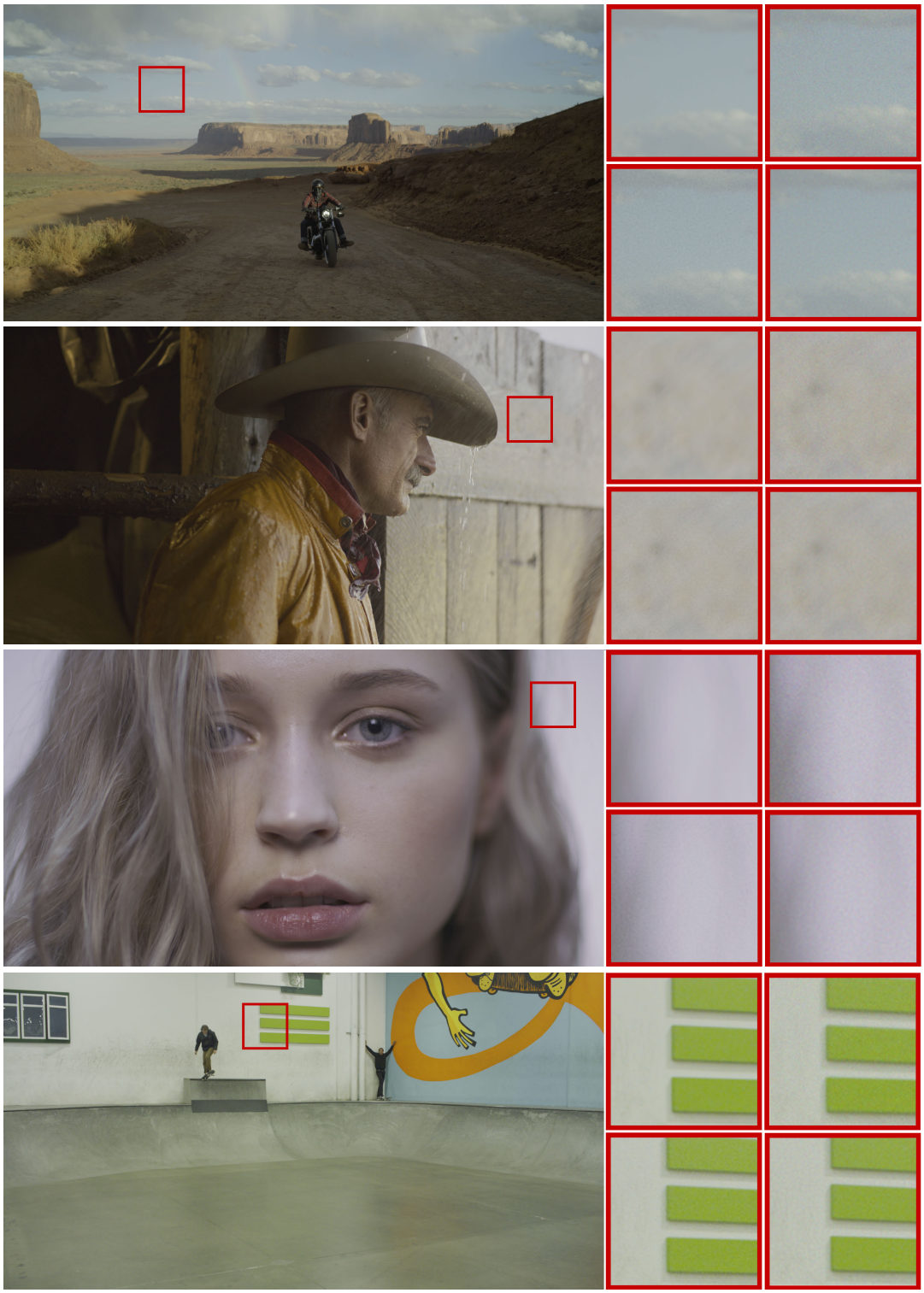


FIGURE 5. Left: frame without noise. At its right there is a zoomed-in detail in four versions, from left to right and top to bottom: original, film grain emulation result by Newson *et al.* [28], film grain emulation result by DaVinci Resolve 14, our retinal noise emulation result.

Higher values of σ result in bigger size noise. The effect of using non-symmetrical Gaussian filters, defined by their 2×2 covariance matrices Σ_c and Σ_s, is a non-symmetrically distributed noise as it can be observed in Fig. 4 (d) and (f). Modifying these values alters the power spectrum of the noise, as it can be observed in Fig. 6.

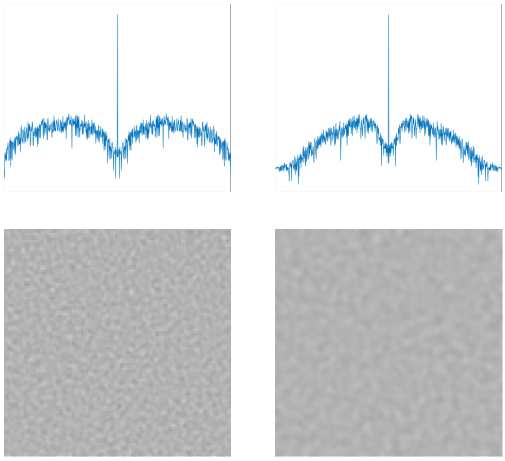


FIGURE 6. Proposed method applied to a flat grey image, with parameter values $\sigma_C = 0.7$, $\sigma_S = 1.5$ (left) and $\sigma_C = 1.2$, $\sigma_S = 2.6$ (right). Top: power spectrum. Bottom: result of proposed method.

C. PSYCHOPHYSICAL EVALUATION

The goal of adding noise to images with a creative intent is to produce results that are visually appealing for observers. Therefore, we conduct psychophysical experiments in order to validate our results and to compare them with methods from the state of the art in academia, like the algorithm of Newson *et al.* [3] in its implementation [28], and in the movie industry as well, like the film-grain emulation provided by the professional post-production software DaVinci Resolve 14.

For this study, the parameters of each method have been selected by a cinema expert to get the most appealing visual appearance of images according to his liking. For the method of Newson *et al.*, grain radius is set to $r = 1/200$, type of algorithm is pixel-wise and number of MonteCarlo iterations is set to $N = 1000$. For DaVinci ResolveFX texture film grain effect, the 35mm film settings have been used with an intensity of $I = 0.75$. For our method, the parameters used are $a = 0.015$, $\sigma_c = 0.7$ and $\sigma_s = 1.5$. Figure 5 shows samples of four video frames, corresponding to as many video sources, showing the four different versions used in the psychophysical experiments: a clean version, a version resulting from the addition of retinal noise, a version resulting from the addition of film-grain noise from the work by Newson *et al.*, and a version resulting from the addition of noise using the DaVinci ResolveFX texture film grain effect. All the methods are applied using the parameter values just specified. It is however worth noting that, in order to fully distinguish the difference between the methods, the temporal dimension is crucial.

For the evaluation, we used a room with dim ambient illuminance. Observers were instructed to sit approximately one meter away from the screen. A two-alternative forced-choice comparison (2AFC) technique was used: each observer was shown two consecutive videos on the screen, each of them obtained from one of the methods. The observers were asked to choose the most visually appealing video from the pair compared. Thirteen observers took part in the experiments,

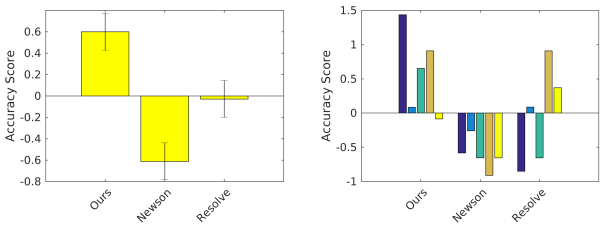


FIGURE 7. Accuracy scores of competing methods for adding texture: 13 observers took part in each experiment and 5 videos were used. Left: average. Right: scores per video.

one of them being a cinema professional from a major post-production house.

The analysis of the psychophysical experiment is presented in Fig. 7. To compute accuracy scores from the raw psychophysical data, we use the same approach as in [29] (Chapter 5), that is based on Thurstone's law of comparative judgment. As it can be seen in this figure, our method is preferred over the other two. This result is also consistent for each video separately in all the cases except one. The individual preference of the movie professional that took part in the experiment follows the same trend of the whole group of observers.

IV. RETINAL NOISE EMULATION FOR IMPROVING COMPRESSED VIDEO QUALITY

We test here the suitability of retinal noise emulation to improve the quality of compressed video, or, more precisely, to mask the image degradation inherent to the compression process (i.e. compression artifacts) in order to prevent users from perceiving it. To evaluate this masking effect, we have designed a set of experiments that simulates a video on demand (VoD) service where the content is provided via HTTP/TCP-based adaptive bit rate streaming (ABR) techniques [30]. Under this paradigm, content is encoded at different quality levels associated with unequal bit rates in accordance to a given quality ladder and segmented. These segments are stored in a server or set of servers (eg Content Delivery Network (CDN)) and are provided to the client upon request [31]. The client selects throughout the streaming session the segments that best suit the system state (channel available bandwidth, terminal capabilities, quality control policies…) to optimize the quality provided to the user [30].

The results of the experiments will indicate down to what point, if retinal grain is added, service providers will be able to decrease the encoding bit rates included in the quality ladders, and therefore the objective quality of the encoded sequences, without the users noticing. To be able to remove the effect of coding and so isolate that of retinal noise, sequences with and without noise are used in the experiments.

All the procedures and selections related to the subjective tests fulfill the guidelines included in Recommendations ITU-R BT.500-13 [32], ITU-T P.910 [33] and ITU-T P.913 [34].

TABLE 1. SRCs' spatial and temporal complexity.

	Balloon	Nature	Bugs	Closeup
SI	30.85	66.05	23.83	11.47
TI	28.64	44.08	43.21	43.21

A. TEST MATERIAL

The test material is made up of four 10-second-long 4K (4096×2160p) Source sequences (SRCs) acquired at 24 fps. They all use a 4:2:2 chroma subsampling at 12 bits. The SRCs were selected from a public dataset made available by Blackmagic including representative, varied, and habitual contents for users [35]. The number, duration, and characteristics of the source content were selected in accordance to Rec. ITU-T P.913 [34]. Table 1 includes the characteristics of the SRCs in terms of the average spatial and temporal complexity considering the spatial information (SI) and the temporal information (TI) indicators [33], [36].

Retinal noise was later on added to the 'clean' SRCs to obtain 'noisy' versions of them. The procedure is described next.

1) ADDITION OF RETINAL NOISE

The proofing observer, a motion picture specialist, was seated two picture heights away from a Sony PVM-250 25'' Full HD OLED display (Rec709, gamma 2.4 calibrated [23]). The display was driven by a Blackmagic 4K video card via DaVinci Resolve. A Tangent Element color correction panel was used as a control interface. The clips were shown in their native resolution (4096×2160), but were cropped to fit onto the Full HD display. The Tangent Element panel provided controls to shift the crop to different areas of the image. These controls could be operated with some small delay while the clip was playing.

First, proofing to select the proper noise shape parameters was conducted. Eight different noise shapes were each generated for a one-second excerpt of a single video clip at three different intensity levels.

Each of these shape sets were placed on the Resolve timeline and viewed through the OLED display in a dark surround. The pan and tilt functions were used to determine the areas of the clip where the noise was most prominent and observations were recorded about each of the clips' appearances. It was found that several of the generated noise shapes added textures to the image which appeared gritty and blocky, while others introduced a finer grain that was more pleasing to the eye. The most ideal grain shape of those generated in the first round was selected to be $\sigma_c = 0.5, \sigma_s = 1$, which differs from the values obtained in the experiment discussed in Section III-C because now the tests are performed on a higher resolution monitor.

This noise was then added at varying intensity levels (0.025, 0.05, 0.075) to all of the tested clips (1 second excerpts). For this test, the clips were arranged on the Resolve timeline along with the ''clean'' version of the clip with no noise added. These different intensities were then compared in the dark surround viewing environment. In many cases the clean versions themselves had a considerable amount of camera noise upon capture, which led to an unpleasant static on the clips. However, when some of the retinal noise was added, this camera noise was to some degree obscured and an overall more pleasing image was produced. In general, the lowest noise setting was selected for these images (0.025). In other cases of brighter images, it was found that a higher intensity of noise (0.05) seemed to improve the visual texture and appearance of the test clips, particularly in high frequency areas.

Figure 8 shows screenshots of the four sequences, with the optimal amount (in terms of image appearance) of retinal grain added to them.

TABLE 2. HRCs used to create the test sequences presented to the observers.

HRC	Resolution	bit rate (Mbps)	Retinal Noise
$HRC^{C}_{1,1}$	2160p	22.2	No
$HRC^{N}_{1,1}$	2160p	22.2	Yes
$HRC^{C}_{1,2}$	2160p	18.4	No
$HRC^{N}_{1,2}$	2160p	18.4	Yes
$HRC^{C}_{1,3}$	2160p	14.6	No
$HRC^{N}_{1,3}$	2160p	14.6	Yes
$HRC^{C}_{2,1}$	1440p	10.7	No
$HRC^{N}_{2,1}$	1440p	10.7	Yes
$HRC^{C}_{2,2}$	1440p	9.7	No
$HRC^{N}_{2,2}$	1440p	9.7	Yes
$HRC^{C}_{2,3}$	1440p	8.7	No
$HRC^{N}_{2,3}$	1440p	8.7	Yes
$HRC^{C}_{3,1}$	1080p	7.8	No
$HRC^{N}_{3,1}$	1080p	7.8	Yes
$HRC^{C}_{3,2}$	1080p	6.7	No
$HRC^{N}_{3,2}$	1080p	6.7	Yes
$HRC^{C}_{3,3}$	1080p	5.6	No
$HRC^{N}_{3,3}$	1080p	5.6	Yes
$HRC^{C}_{4,1}$	720p	4.5	No
$HRC^{N}_{4,1}$	720p	4.5	Yes
$HRC^{C}_{4,2}$	720p	3.7	No
$HRC^{N}_{4,2}$	720p	3.7	Yes
$HRC^{C}_{4,3}$	720p	2.9	No
$HRC^{N}_{4,3}$	720p	2.9	Yes
$HRC^{C}_{5,1}$	540p	2.0	No
$HRC^{N}_{5,1}$	540p	2.0	Yes
$HRC^{C}_{5,2}$	540p	1.7	No
$HRC^{N}_{5,2}$	540p	1.7	Yes
$HRC^{C}_{5,3}$	540p	1.4	No
$HRC^{N}_{5,3}$	540p	1.4	Yes

2) GENERATION OF TEST SEQUENCES

We have tested 30 different combinations of encoding and network parameter values. Each of these combinations, called Hypothetical Reference Circuit (HRC) [33], is applied to all source sequences, resulting in a set of Processed Video Sequences (PVSs), one per SRC and HRC, that are presented to the users for evaluation. The combinations considered in the tests are included in Table 2. HRCs are 4:2:0 and have a

FIGURE 8. Screenshots of the four sequences used in the subjective assessment, with optimal amount (in terms of image appearance) of retinal grain added. Each image shows a zoomed-in region (marked with a red square) in two versions, with and without the retinal grain. From left to right, top to bottom: "Balloon", "Nature", "Bugs", and "Closeup".

color depth of 10 bits to match common broadcast conditions. Furthermore, they preserve the framerate of the sources: 24 fps. Five of them (named $\mathrm{HRC}^{\mathrm{C}}_{i,1}$, where $i = \{1, \ldots, 5\}$) are anchor points derived directly from the quality ladders included in Apple' HLS technical note [8]. To decrease the gap between consecutive quality levels and so enable a finer-grained analysis, two additional HRCs (named $\mathrm{HRC}^{\mathrm{C}}_{i,j}$, where $j = \{2, 3\}$) were created from each anchor point $\mathrm{HRC}^{\mathrm{C}}_{i,1}$. The spatial resolutions of these 15 HRCs, marked with the superscript "C" for clean, nearly follow a geometric progression with ratio $1/\sqrt{2}$. So, each picture resolution is close to half the previous one. The other half of the set of HRCs, marked with the superscript "N" for noise, share the same characteristics as the original set of HRCs and, in addition, they include retinal noise.

All HRCs are H.264/AVC encoded. Bit rates for the anchor HRCs were obtained from the H.264/AVC ladder whenever the associated resolution was included there and extrapolated accordingly to the ladder rule for the remaining resolutions. The bit rates for the rest of the HRCs were obtained by linearly interpolating between the values of the anchor HRCs.

Finally, we generated a total of 120 Processed Video Sequences (PVS's). As mentioned before, the aspect ratios of the HRCs, and therefore of the PVS's, were 16:9 (∼1.78:1), following the SMPTE ST 2036-1 standard [37] and Recommendation ITU-R BT.2020 [38]. As the aspect ratio of the SRCs is 256:135 (∼1.9:1), according to the Digital Cinema System Specification of the Digital Cinema Initiatives (DCI) [39], the PVS's resolutions required a minor cinema-to-broadcast format adaptation that was conducted using a bicubic filter [40].

B. ENVIRONMENT AND EQUIPMENT

The test room was set to simulate home viewing conditions. Furthermore, the brightness was controlled according to recommended values [33]: 24.4 Lx in front of the subjects, 16.5 Lx to their left, 85.4 Lx to their right, 71.7 Lx above them, and 20.1 Lx behind them.

The device used in the tests was a TV set with a 43-inch screen and a 3840 × 2160 pixel resolution (Samsung UE43NU7475). The viewing distance was set to twice the height of the screen, in accordance to Rec. ITU-T P.913 [34].

C. METHODOLOGY

Before starting the experiments, the test designer read the guidelines of the tests to the observers. Next, subjects were trained by showing them examples of the best and worse quality levels they should expect for sequences with and without retinal noise (four extra PVS's created from a fifth content according to $\mathrm{HRC}^{\mathrm{C}}_{1,1}$, $\mathrm{HRC}^{\mathrm{N}}_{1,1}$, $\mathrm{HRC}^{\mathrm{C}}_{5,3}$ and $\mathrm{HRC}^{\mathrm{N}}_{5,3}$). In this way, subjects were more aware of the scale of qualities that they would encounter and rate the sequences accordingly.

During the experiments, all the PVS's, that is, every combinations of video sources -SRC's- and encoding and network conditions to be tested -HRC's-, including the reference sequence, were sequentially and randomly presented to the subjects. Each PVS was presented once to each subject. The order of presentation of the PVS's was different for each pair

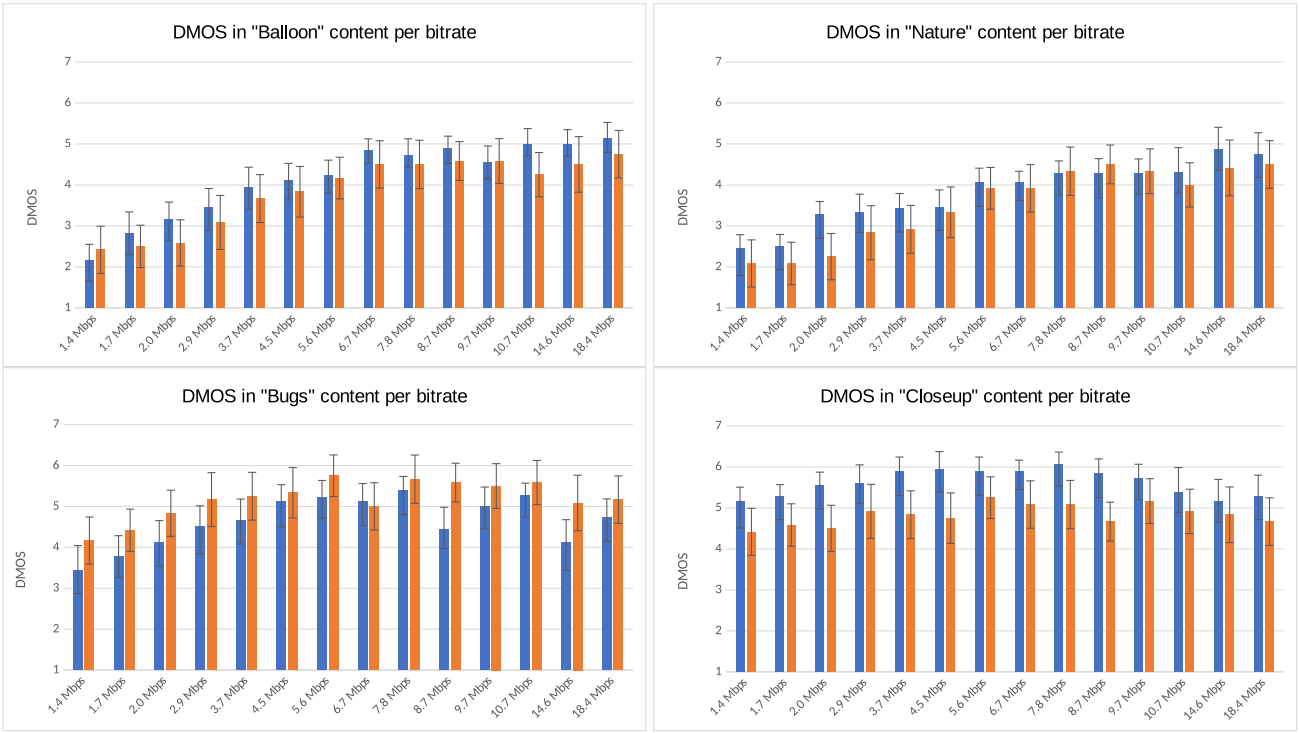


FIGURE 9. DMOS per content and bit rate. From left to right, top to bottom: "Balloon", "Nature", "Bugs", and "Closeup". Orange bars are used for 'clean' content and blue bars indicate content with retinal grain.

of observers and was set randomly, in accordance with Rec. ITU-T P.910 [33]. The whole session was slightly shorter than 30 minutes, as recommended by ITU-R BT.500-13. The test method followed in the tests is the Absolute Category Rating with Hidden Reference (ACR-HR) proposed in Rec. ITU-T P.910 [33], where subjects have five possible answers to choose from: ''Excellent'', ''Good'', ''Fair'', ''Poor'' and ''Bad''. The subjects were asked to assess each PVS right after its visualization. To help it, a four-second grey sequence was included between consecutive PVS's, also as stated in Rec. ITU-T P.910 [33].

There were 18 observers (6 women and 12 men) in the experiment, all of them having normal or corrected vision, aged between 20 and 30 years. The number of subjects is sufficiently significant, as stated in Rec. ITU-R BT.910 [33].The observers were rewarded for their participation in the tests, and a maximum of two observers were allowed in each test session. Due to the characteristics of the play-out system, the assessment was conducted sequentially on the two demi-sets of PVS's: first the PVS's including retinal noise (named $\mathrm{HRC}^{\mathrm{N}}_{i,j}$), called 'noisy' PVS's, and then the 'clean' PVS's (named $\mathrm{HRC}^{\mathrm{C}}_{i,j}$). No observers were rejected after the screening of the subjective results.

D. TEST RESULTS

Figure 9 depicts the results in terms of the evolution per content of the differential mean opinion score (DMOS) versus the encoding bit rate. The DMOS is defined as follows:

$$\mathrm{DMOS(PVS)} = \mathrm{MOS(PVS)} - \mathrm{MOS(REF)} + 5 \quad (10)$$

where MOS is the Mean Opinion Score computed for a given content (PVS or reference sequence). Therefore, the better the image quality of the sequence presented to the user (i.e. the more it looks like the reference one), the greater it will be in principle the MOS of that sequence, and so the resulting DMOS.

The DMOS values have been computed per user for each one of the 'clean' and 'noisy' sets of PVS's with respect to the scores given to their corresponding references, $\mathrm{HRC}^{\mathrm{C}}_{1,1}$ and $\mathrm{HRC}^{\mathrm{N}}_{1,1}$, as it is usually done in the literature. Each bar includes its 95% confidence intervals.

We can easily distinguish two trends in the figures per content: that of sequences ''Balloon'' and ''Nature'' and that of sequences ''Bugs'' and ''Closeu''. The results of the assessment of the first two sequences show a clear and steady decrease in the quality perceived by observers with the reduction of bit rate on average. However, the results for the other two sequences do not show any clear connection between the perceived quality and the bit rate. This is an interesting outcome of this exploratory analysis on the addition of retinal noise, as this discrepancy stems from the different nature of the video contents. On the one hand, all the elements in every picture in the sequences ''Balloon'' and ''Nature'' are in focus. On the other hand, a significant part of every picture in the sequences ''Bugs'' and ''Closeup'' are out of focus due to the limited depth of field used in their acquisition. As this experiment has been carried out on visual information in a bit rate limited scenario, it means that the last two sequences were always better treated by the compression and decompression system, as more bits could be devoted for the

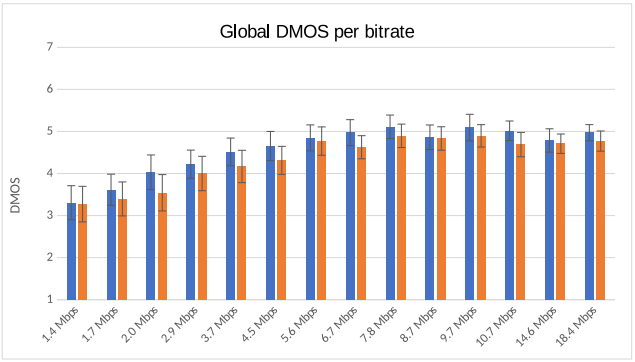


FIGURE 10. Global DMOS per bit rate. Orange bars are used for "clean" content and blue bars indicate content with retinal grain.

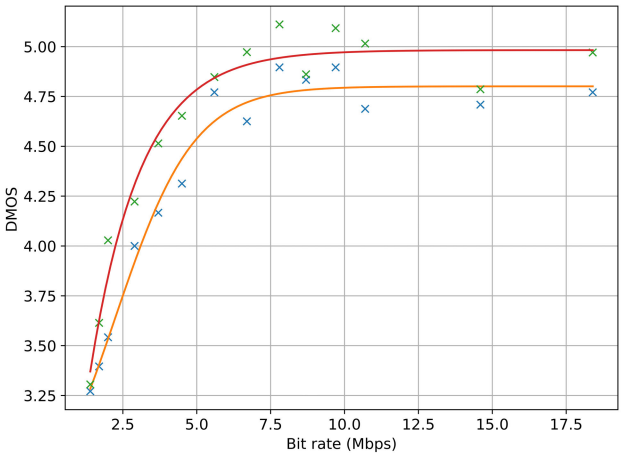


FIGURE 11. Regression on the bit rate-DMOS values. Green crosses represent "noise" scores, blue crosses are "clean" scores, the red line represents the regression on "noise" values and the orange line regression on "clean" values.

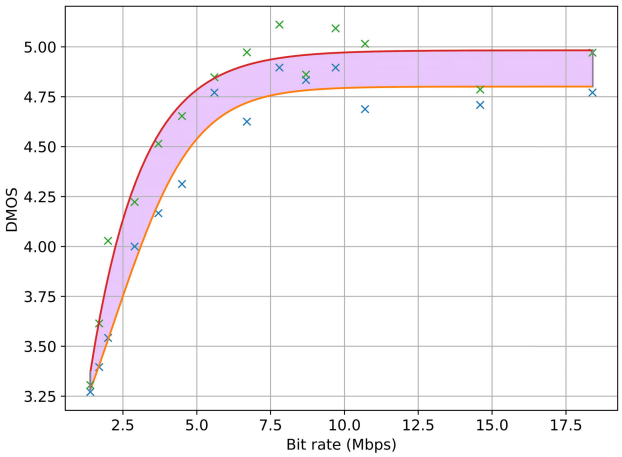


FIGURE 12. Regression on the bit rate-DMOS values including the area between the lines where the BD-DMOS is computed by vertical integration. This plot highlights the fact that, at any given bit rate, the addition of emulated retinal noise improves perceived image quality (the DMOS value is higher for the sequence with retinal noise).

encoding of the in-focus part of the picture. Therefore, the selection of test material in subsequent experiments should consider the depth of field information in addition to their spatial and temporal information. Moreover, it is important to note the type of content presented. Since the sequence "Bugs" did not show any benefit of being treated with noise, there is a possibility that documentary content may be less susceptible to being processed by adding noise. Thus, utmost care should be placed on content selection where texture noise could be added.

The analysis of the aggregated DMOS, presented in Figure 10, shows a reduced but significant advantage on the addition of retinal noise in the outcome of the assessment. Even if the confidence intervals overlap, the DMOS of each "noisy" PVS is always higher than the one of its "clean" counterpart, the advantage being more evident for resolutions up to high definition. We consider the results significant since there is a clear tendency for the results with noise to be better rated on average. From these results we conclude that, when the bit rate is reduced, the decrease in perceived image quality is smaller if the video has had retinal grain added to it.

In order to measure properly the benefits of applying our proposal to mask compression artifacts in terms of bit rate saving and quality improvement, we have applied a regression on the noise and clean sets of points. The regression has been performed using a sigmoid function [41] and the least squares method. Figure 11 shows the result of the process. The green and blue crosses represent the average user scores for noisy and clean contents, respectively. The red and yellow lines are their respective regression.

First, one can verify the conclusions drawn above: the addition of noise represents an enhancement of quality over the clean signal for all considered bit rates. Moreover, this result can also be seen from another point of view: for a given quality level, using the version that includes the emulation of retinal noise leads to significant savings in bandwidth. Both gains have been analyzed quantitatively. So, the former has been measured in terms of BD-DMOS by computing the area between the lines (represented in purple in Figure 12) by means of a vertical integration [42], [43]. Results point at a DMOS average improvement of 0.2. Regarding bit rate savings, they have been measured in terms of BD-Rate by computing the area between the lines (represented in purple in Figure 13) through a horizontal integration [42], [43]. Results indicate that the application of the retinal noise emulation method allows for a significant improvement in coding efficiency, with average bit rate savings of over 22.5%. Nevertheless, let us point out that our experiments suggest a potential weakness of our approach in that the applicability of the method may depend on the specific video sequence that is dealt with, because as remarked earlier there is the possibility that for some kind of content, like documentary footage, a cinema-like appearance is not preferable for the viewer. However, a mere previous analysis of the content

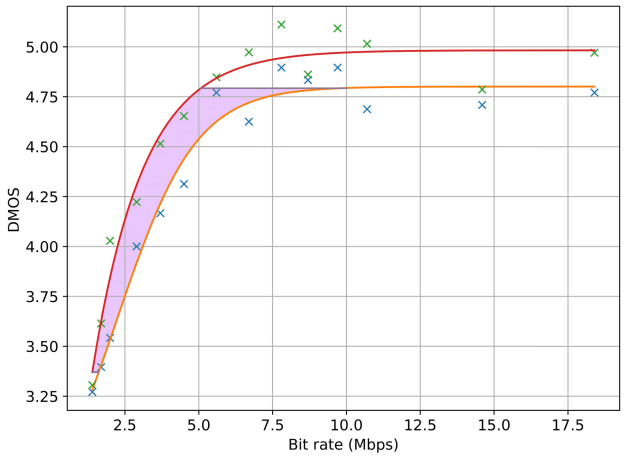


FIGURE 13. Regression on the bit rate-DMOS values including the area between the lines where the BD-Rate is computed by horizontal integration. This plot highlights the fact that, for any given perceived quality level (DMOS value), the encoding of the sequence with emulated retinal noise is more efficient (the required bit rate is lower than the one necessary to attain the same DMOS value with the clean sequence).

could determine the suitability of the inclusion of retinal noise.

V. CONCLUSIONS AND FUTURE WORK

We have presented a method for adding texture to digital cinema that is inspired by processes in the visual system and produces results that look natural and visually pleasing even for challenging scenes. The method has three parameters whose default values produce satisfactory results in a variety of scenarios, and that can be modified for artistic purposes, in order to achieve different looks. A psychophysical validation was conducted, showing that the proposed method outperforms algorithms from the state of the art in the academic literature and in the industry.

The retinal noise emulation method can also improve the quality of compressed video by masking compression artifacts. The aim was to help concealing distortions due to compression and thus allowing to maintain image quality while reducing the bit rate or improving image quality while maintaining the bit rate fixed. The proposed method has been validated through subjective assessment on 4K professional cinema sequences, where the amount of retinal noise was selected by a motion picture specialist based solely on aesthetic preference. The experiment has shown that the proposed method can yield very impressive savings in bit rate. A special effort has been made to maintain the rigorousness and reproducibility of the subjective tests carried out. As future work, we intend to explore the impact of the type of content in the usefulness of the addition of retinal grain noise to mask compression artifacts.

Our results point to a novel and, we believe, very promising avenue of research in computer vision which is the connection between vision models of retinal grain, perceived image quality (a very active area of interest in computer vision because the main challenges remain unsolved), image compression algorithms and image compression as performed by the visual system: the connection here is even more explicit since the classic work of Olshausen and Field [44], that allows to link convolutional neural netowrks (CNNs) trained for compression with the receptive fields that are actually measured in the human visual system. Ongoing work involves training CNNs for compression on natural images with and without retinal grain.

REFERENCES

[1] T. Kurihara, Y. Manabe, N. Aoki, and H. Kobayashi, "Digital image improvement by adding noise: An example by a professional photographer," *J. Imag. Sci. Technol.*, vol. 55, no. 3, 2011, Art. no. 030503.

[2] X. Wan, H. Kobayashi, and N. Aoki, "Improvement in perception of image sharpness through the addition of noise and its relationship with memory texture," *Proc. SPIE*, vol. 9394, 2015, Art. no. 93941B.

[3] A. Newson, J. Delon, and B. Galerne, "A stochastic film grain model for resolution-independent rendering," *Comput. Graph. Forum*, vol. 36, no. 8, pp. 684–699, Dec. 2017.

[4] B. T. Oh, C.-C.-J. Kuo, S. Sun, and S. Lei, "Film grain noise modeling in advanced video coding," *Proc. SPIE*, vol. 6508, Jan. 2007, Art. no. 650811.

[5] B. T. Oh, S.-M. Lei, and C.-C.-J. Kuo, "Advanced film grain noise extraction and synthesis for high-definition video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1717–1729, Dec. 2009.

[6] J. Dai, O. C. Au, C. Pang, W. Yang, and F. Zou, "Film grain noise removal and synthesis in video coding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 890–893.

[7] A. Norkin and N. Birkbeck, "Film grain synthesis for AV1 video codec," in *Proc. Data Compress. Conf.*, Mar. 2018, pp. 3–12.

[8] Apple. (2018). *HLS Authoring Specification for Apple Devices*. [Online]. Available: https://developer.apple.com/documentation/http_live_streaming/hls_authoring_specification_for_apple_devices

[9] Cisco. (2019). *Visual Networking Index: Forecast and Methodology, 2017–2022*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.pdf

[10] R. S. Allison, K. Brunnström, D. M. Chandler, H. R. Colett, P. J. Corriveau, S. Daly, J. Goel, J. Y. Long, L. M. Wilcox, Y. M. Yaacob, S.-N. Yang, and Y. Zhang, "Perspectives on the definition of visually lossless quality for mobile and large format displays," *J. Electron. Imag.*, vol. 27, no. 05, p. 1, Oct. 2018.

[11] S. Tavakoli, J. Gutierrez, and N. Garcia, "Subjective quality study of adaptive streaming of monoscopic and stereoscopic video," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 684–692, Apr. 2014.

[12] H. Sun and Y. Q. Shi, *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*. Boca Raton, FL, USA: CRC Press, 2008.

[13] M. Naccari and F. Pereira, "Advanced H.264/AVC-based perceptual video coding: Architecture, tools, and assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 766–782, Jun. 2011.

[14] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "PieAPP: Perceptual image-error assessment through pairwise preference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1808–1817.

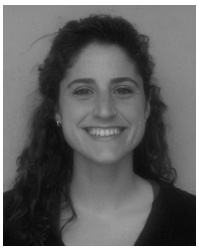
[15] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[16] M. Bertalmio, *Vision Models for High Dynamic Range and Wide Colour Gamut Imaging: Techniques and Applications*. New York, NY, USA: Academic, 2019.

[17] S. W. Zamir, J. Vazquez-Corral, and M. Bertalmio, "Vision models for wide color gamut imaging in cinema," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Nov. 14, 2019, doi: 10.1109/TPAMI.2019.2938499.

[18] E. R. Kandel, J. H. Schwartz, T. M. Jessell, M. B. T. Jessell, S. Siegelbaum, and A. Hudspeth, *Principles of Neural Science*, vol. 4. New York, NY, USA: McGraw-Hill, 2000.

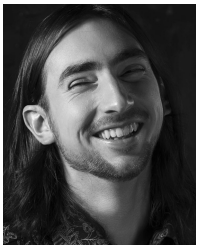
[19] R. Shapley and C. Enroth-Cugell, "Chapter 9 visual adaptation and retinal gain controls," *Prog. Retinal Res.*, vol. 3, pp. 263–346, Jan. 1984.
[20] M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation," *Nature Rev. Neurosci.*, vol. 13, no. 1, pp. 51–62, Jan. 2012.
[21] H. Sun, L. Rüttiger, and B. B. Lee, "The spatiotemporal precision of ganglion cell signals: A comparison of physiological and psychophysical performance with moving gratings," *Vis. Res.*, vol. 44, no. 1, pp. 19–33, Jan. 2004.
[22] L. J. Croner, K. Purpura, and E. Kaplan, "Response variability in retinal ganglion cells of primates.," *Proc. Nat. Acad. Sci. USA*, vol. 90, no. 17, pp. 8128–8130, Sep. 1993.
[23] M. Bertalmio, *Image Processing for Cinema* (Mathematical and Computational Imaging Sciences Series), 1st ed. Boca Raton, FL, USA: CRC Press, 2014.
[24] S. W. Zamir, J. Vazquez-Corral, and M. Bertalmio, "Automatic, fast and perceptually accurate gamut mapping based on vision science models," in *Proc. SMPTE Annu. Tech. Conf. Exhibit.*, Oct. 2017, pp. 1–14.
[25] C. L. Passaglia and J. B. Troy, "Impact of noise on retinal coding of visual signals," *J. Neurophysiol.*, vol. 92, no. 2, pp. 1023–1033, Aug. 2004.
[26] A. B. Watson and A. J. Ahumada, "A standard model for foveal detection of spatial contrast," *J. Vis.*, vol. 5, no. 9, p. 6, Oct. 2005.
[27] S. Ferradans, M. Bertalmio, E. Provenzi, and V. Caselles, "An analysis of visual adaptation and contrast perception for tone mapping," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2002–2012, Oct. 2011.
[28] A. Newson, N. Faraj, B. Galerne, and J. Delon, "Realistic film grain rendering," *Image Process. Line*, vol. 7, pp. 165–183, Jul. 2017.
[29] J. Morovic, "To develop a universal gamut mapping algorithm," Ph.D. dissertation, Colour Imag. Inst., Univ. Derby, Derby, U.K., 1998.
[30] A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer, and R. Zimmermann, "A survey on bitrate adaptation schemes for streaming media over HTTP," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 562–585, 1st Quart., 2019.
[31] Q. Fan, X. Li, S. Wang, S. Fu, X. Zhang, and Y. Wang, "NA-caching: An adaptive content management approach based on deep reinforcement learning," *IEEE Access*, vol. 7, pp. 152014–152022, 2019.
[32] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Standard Recommendation ITU-R BT.500, Jan. 2012.
[33] *Subjective video Quality Assessment Methods for Multimedia Applications*, Standard Recommendation ITU-R ITU-T P.910, Apr. 2008.
[34] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in any Environment*, Standard Recommendation ITU-R ITU-T P.913, Mar. 2016.
[35] *Blackmagic Pocket Cinema Camera 4K Gallery*. Accessed: Apr. 10, 2020. [Online]. Available: https://www.blackmagicdesign.com/products/blackmagicpocketcinemacamera/gallery
[36] L. Janowski and M. Pinson, "The accuracy of subjects in a quality experiment: A theoretical subject model," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2210–2224, Dec. 2015.
[37] *Ultra High Definition Television—Image Parameter Values for Program Production*, Standard ST 2036-1:2013, Oct. 2014.
[38] *Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange*, Standard Recommendation ITU-R BT.2020-2, Oct. 2015.
[39] "Digital cinema system specification," Digit. Cinema Initiatives, LLC, Hollywood, CA, USA, Tech. Rep. Version 1.3, Jun. 2018.
[40] M. Camara, C. Diaz, J. Casal, J. Ruano, and N. Garcia, "Perceptually equivalent resolution in handheld devices for streaming bandwidth saving," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 878–882, Jun. 2019.
[41] E. Upenik, M. Rerabek, and T. Ebrahimi, "On the performance of objective metrics for omnidirectional visual content," in *Proc. 9th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Erfurt, Germ, May 2017, pp. 1–6.
[42] G. Bjontegaad, *Calculation of Average PSNR Differences Between RD Curves*, document VCEG-M33, Apr. 2001.
[43] K. Rouis, M.-C. Larabi, and J. Belhadj Tahar, "Perceptually adaptive lagrangian multiplier for HEVC guided rate-distortion optimization," *IEEE Access*, vol. 6, pp. 33589–33603, 2018.
[44] B. A. Olshausen and D. J. Field, "Vision and the coding of natural images: The human brain may hold the secrets to the best image-compression algorithms," *Amer. Scientist*, vol. 88, no. 3, pp. 238–245, 2000.



ITZIAR ZABALETA received the Architecture degree from the Universidad de Navarra, in 2009, and the Mathematics degree from the Universitat de Barcelona, in 2017. She is currently pursuing the Ph.D. degree with the Image Processing for Enhanced Cinematography Group, Universitat Pompeu Fabra, Barcelona, Spain. In 2017, she did an internship on developing a segmentation algorithm for cell images using graph theory at the Bioinformatics Institute, A*STAR, Singapore, before joining her current studies at UPF. Her research interests include image and video processing algorithms, and color science for cinema.

MATEO CÁMARA was born in Madrid, in 1994. He received the B.Sc. degree in engineering (telecommunication technologies and services) and the M.S. degree in telecommunication engineering (accredited by ABET) from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2017 and 2019, respectively. He has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM, since 2017, where he has been actively involved in several research projects. His research interests are in the area of deep learning applied to video and video quality assessment.

CÉSAR DÍAZ received the Telecommunication Engineering degree (integrated B.Sc.-M.S.) and the Ph.D. degree in telecommunication engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2007 and 2017, respectively. Since 2008, he has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM, where he has been actively involved in Spanish and European projects. His research interests are in the areas of multimedia delivery, immersive communications, and quality of experience.



TREVOR CANHAM received the B.Sc. degree in motion picture science from the Rochester Institute of Technology, Rochester, NY, USA, in 2018. He held several positions at multiple post production facilities and delivered an undergraduate thesis on the effects of colored ambient illumination on image appearance on mobile displays with the Rochester Institute of Technology. This led to a seamless transition into his current position, where he continues to study methods for managing content appearance across different motion picture workflow stages and delivery destinations. He is currently working as a Research Engineer at the Image Processing for Enhanced Cinematography Group, Universitat Pompeu Fabra, Barcelona, Spain.

NARCISO GARCÍA received the Ingeniero de Telecomunicación degree (five years engineering program) and the Doctor Ingeniero de Telecomunicación degree (Ph.D. in communications) from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1976 and 1983, respectively. Since 1977, he has been a member of the Faculty of the UPM, where he is currently a Professor of signal theory and communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM. He was a Co-Writer of the EBU proposal, base of the ITU standard for digital transmission of TV at 34–45 Mb/s (ITU-T J.81). His current research interests include digital video compression, computer vision, and quality of experience. He has been actively involved in Spanish and European research projects, also serving as an evaluator, a reviewer, an auditor, and an observer of several research and development programs of the European Union. He was a recipient of the Junior and Senior Research Awards of the Universidad Politécnica de Madrid, in 1987 and 1994, respectively. He received the Spanish National Graduation Award and the Doctoral Graduation Award from UPM. He was an Area Coordinator of the Spanish Evaluation Agency (ANEP), from 1990 to 1992, and he was a General Coordinator of the Spanish Commission for the Evaluation of the Research Activity (CNEAI), from 2011 to 2014. He has been the Vice-Rector for International Relations of the Universidad Politécnica de Madrid, from 2014 to 2016.

MARCELO BERTALMÍO was born in Montevideo, in 1972. He received the B.Sc. and M.Sc. degrees in electrical engineering from the Universidad de la República, Uruguay, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, in 2001. He is currently a Full Professor at Universitat Pompeu Fabra, Spain. His current research interests are in developing image processing algorithms for cinema that mimic neural and perceptual processes in the visual systems, and to investigate new vision models based on the efficient representation principle. He was awarded the 2012 SIAG/IS Prize of the Society for Industrial and Applied Mathematics (SIAM), for co-authoring the most relevant image processing work published in the period 2008–2012. He has received the Femlab Prize, the Siemens Best Paper Award, the Ramón y Cajal Fellowship, and the ICREA Academia Award, among other honors. He has also received the ERC Starting Grant and the two ERC Proof of Concept Grants for vision-based image processing research with applications to the cinema industry; he has also written two books on the subject. He was an Associate Editor of *SIAM Journal on Imaging Sciences* and an elected Secretary of the SIAM's activity group on imaging.