From Contrast Sensitivity Function construction to Visual Weightings Computation for Digital Cinema

Chaker Larabi, Vincent Rosselli and Christine Fernandez-Maloigne
XLIM-SIC Lab., University of Poitiers, France

Abstract

Digital cinema is a very challenging field that will allow providing cinema material with a very high quality level. In the last years the key role of the Human Visual System (HVS) in the final perceived quality of the compressed images, becomes an undeniable reality. Therefore, it is natural to take advantage of the recent knowledge of human visual perception and models in an image compression system. Thus, in this paper we propose a reproducible technique for improving the perceptual JPEG 2000 image compression quality with a digital cinema profile. This technique consists of two main parts: a laboratory evaluation of a HVS model by the Contrast Sensitivity function (CSF), and the implementation of visual weightings for the JPEG 2000 scheme, using the evaluated HVS model, in the spatial Fourier domain of the wavelet decomposition sub-bands.

Introduction

Digital cinema is simply a new approach for making and showing movies. The basic idea is to use bits and bytes to record, transmit and replay images, rather than using chemicals on film [1, 2]. The main advantage of a digital technology is that it can store, transmit and retrieve a huge amount of information exactly as it was originally recorded. Analog technology loses information in transmission, and generally degrades with each viewing. Digital information is also a lot more flexible than analog information.

Digital cinema affects three major areas of movie-making:

- Production: how the movie is actually made;
- Distribution: how the movie gets from the production company to movie theaters;
- Projection: how the theater presents the movie.

In 2005, the DCI (Digital Cinema Initiatives) has adopted JPEG 2000 [5, 6] as the reference codec for digital cinema material [3]. Since this call, the ISO/JPEG 2000 working group has decided to request guidelines for using JPEG 2000 for digital cinema applications. An amendment has been added to the international standard in order to define two profiles specifically targeted to digital cinema. However, the working group still request to clarify issues concerning ICT and visual weightings for the cinema profile [4].

As all compression standards, JPEG 2000 can be improved by using a perceptual model. This latter allows to reproduce the behavior of the Human Visual System (HVS) to improve the compression results from a visual point of view. Two approaches are common in the JPEG 2000 literature: visual weightings and visual masking. Visual weightings are generated for standard viewing conditions: photopic, short viewing distance, 10 degrees of visual aperture. . .

In a cinema theater, the environment is very different. The viewing distance increases as the display is wide and the ambient lighting is very low (close to the mesopic vision). This allows to ask questions about the validity of the standard visual weightings for cinematic conditions and about the improvement (quality gain) that one can obtain by using appropriate weightings.

In this work, we focus on the construction of the contrast sensitivity function for cinematic conditions, on its application to JPEG 2000 - digital cinema profile and on the validation of the approach.

The remainder of this paper is organized as follows. The first Section describes the JPEG 2000 standard. Section 2 is dedicated to the construction of the Contrast Sensitivity Function in Cinematic conditions, while section 3 describes the computation of the visual weightings. After a validation, this paper ends by some concluding remarks

JPEG 2000 Overview

JPEG2000 is the image compression standard that provides a set of primordial tools for the emergent applications in the image field [5, 6]. JPEG2000 achieves strong compressions with an image quality higher than all the other standard techniques. Following the meetings of the JPEG 2000 committee of the International Standardization Organization at the New Orleans in December 2000, JPEG 2000 Part 1, has been officially declared international standard.

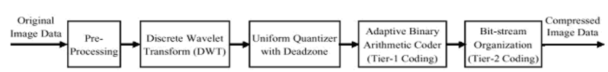


Figure 1. *Diagram of the JPEG2000 compression scheme*

As shown by Figure 1, JPEG 2000 is based on a Discrete Wavelet Transform (DWT), a scalar quantification, a contextual modeling, an arithmetic coder and finally a post-compression rate allocation. The DWT is dyadic and could be done either with the reversible filter The Gall (5,3), that provides a lossless coding, either with the bi-orthogonal irreversible filter Daubechies (9,7), that provides better results for the strong compressions, but without the possibility to do lossless compression. The quantifier follows a scalar approach based on a dead-zone, while being independent for each sub-band. This last is divided in rectangular blocks (named code blocks in JPEG 2000), generally of 64×64 size, undergoing entropic coding by the contextual modeling and the arithmetic coder.

The coded data are organized in layers representing levels of quality, using the post-compression rate allocation and then directed toward the code-stream in packets. The generated code-stream is divisible in different entities: a resolution, a layer (i.e. SNR- Signal to Noise Ratio), a position or a progressive component, or any possible combination.

The JPEG 2000 standard supports also a number of functions, among which many are inherent to the algorithm itself. An example consists in the random access that is possible thanks to the independent coding of the different code-blocks. The other supported functions are: the possibility to define a Region Of Interest (ROI), the resilience error to be robust to the random errors, the processing of multi-component images, the colors indexing, the simple rotation and the lossless compression, to mention some of them.

JPEG 2000 [4] is a coder based on the wavelet decomposition where the coefficients of each sub-band are divided in blocks of same size (named code-blocks) and each code-block is coded separately. The standard structure introduces the concept of quality layers that permits a process of post-compression optimization, where the different bitstreams parts of each code-block is collected in a given order (e.g., optimization according to the rate-distortion ratio), to construct the final bitstream.

This process of quality layers creation is an essential component in JPEG 2000. Thus, after the generation of one sub-bitstream for each code-block, it is to the encoder to determine the way to collect them in order to create the quality layers. Otherwise, the encoder must determine how to draw the lines of the quality layers or to choose the truncation points. This flexibility allows an adaptive rate allocation according to the code-blocks and makes usable a number of visual tools in JPEG 2000.

The Contrast Sensitivity Function

Concept

Human Visual System research offers mathematical models of how humans perceive the surrounding world. For example, models have been developed to characterize human sensibility to brightness and color [7, 8], over spatial frequencies. One of those models is the contrast sensitivity function (CSF). In order to describe this function, the concepts of contrast, contrast threshold and contrast sensitivity have to be developed first. Contrast is a key concept in vision science. The reason is that the information represented in the visual system is not the absolute light level but the contrast, which is the ratio of the local intensity and the average image intensity. The most representative definitions of contrast are those proposed by Michelson (equation 1) [9] and the Weber low for contrast . Those definitions were created from simple experiences on the luminance using sinusoidal gratings.

$$C = \frac{Y_{max} - Y_{min}}{Y_{max} + Y_{min}}, \tag{1}$$

where C is the contrast of Michelson, Y_{max} and Y_{min} respectively the maximum and minimal values of the grid brightness. The minimal contrast so that a sinusoidal grid, presented in to an observer, causes a neurons response, *i.e.* is visible, is called the sensitivity threshold to contrast. The reverse of this threshold is the sensitivity to contrast. This one varies in function of the spatial and temporal frequencies, from where the concept of CSF.

In literature, there are four methods to construct CSF.

- By appearance, the observers have to notify when they saw the pattern.
- By disappearance, the observers have to notify when they do not see anymore the pattern.
- By double-alternative, two images are displayed and the observers have to say where the pattern is.
- By fitting, the observers adjust the contrast level of the pattern to the boundary of visibility

Numbers of psychophysical experiments have been conducted following the previously mentioned approaches. These methods for the CSF construction are based on the concept of detection threshold. This threshold is the boundary beyond which the stimulus is perceived.

The standard description of color vision is in terms of the responses of an achromatic channel and two chromatic channels, one tuned to a red/green dimension and the other to a yellow/blue dimension. The achromatic channel, also named luminance channel, is well known as several study have been focused on it, but not in cinema conditions. The luminance CSF has a band-pass shape.

For the construction of chromatic CSF, monochromatic or opponent-color approaches can be used. In the first approach, the contrast is measured between a color and the black, and the obtained shape is band-pass. The second approach is based on chromatic opposition. As shown by Mullen, the chromatic CSF has a low-pass shape [10].

The DCI colorimetric specifications for digital cinema are very important. The color management chain is based on a reference projector. So, $X'Y'Z'$ has been adopted as the cinema color space where each component has a depth of 12 bits and a gamma correction of 2.6 is applied. The JPEG 2000 color conversion from RGB to YCC is allowed in the DCI recommendations.

Experimental conditions

According to *Wandell* [7], sine-waves stimuli are best perceived because only one frequency is contained in it. Then the grating must have a vertical orientation [12] even thought the diagonal sensitivity is different from the vertical/horizontal one.

The optimal cone of binocular vision is about 10 degrees of visual angle. In cinematic condition, the visual aperture is more than 20 degrees which include the peripheral vision. For our experiments we have chosen to construct our stimuli with an aperture of 20 degrees.

The quality of the stimuli must be optimal to obtain stable results. We assume that the HVS needs at least two periods to estimate the contrast. This restricts low frequencies. The projector and its inherent characteristics restrict the high frequencies. Cosine-wave gratings are preferred because they allow more measuring points. Figure 2 shows an example of used patterns.

Only few frequencies can be tested in a session to not exceed the fifteen to twenty minutes of assessment [11] and to avoid the visual fatigue of the observer. Nineteen frequencies were chosen from 0.139 cycle per degree (cpd) to 15.430 cpd for each channel. To ensure stable result, a repetition of the test is performed.

For the construction of the cinematic Contrast Sensitivity Functions, we constitute a panel of 40 observers between 18 and 51 years old. Gender balance has been respected. Before each test sequence, the observer vision is checked for its acuity and color

Figure 2. Example of pattern used in the construction of the CSF

Patch	x	y
White	0.315	0.338
Red	0.640	0.309
Green	0.255	0.695
Blue	0.148	0.070

Colorimetric Values for some standard patches

blindness.

The viewing distance was about 1.5 times the width of the theater screen. So, as our screen measures 572cm *times* 312cm, the observers were far from the screen for 858cm (cf. figure 3). We used a 2K projector (Barco DP90) with an XDC CineStore server. The maximum resolution is 2048x1080 pixels.

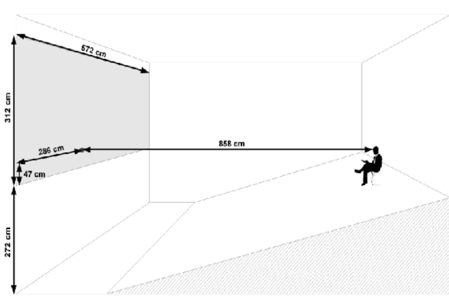


Figure 3. Description of the assessment theater and the selected positions for observers

The projector has to be calibrated to make our assessment compliant with the DCI specifications. The following table gathers the xy values for basic patches viewed at the standard observation distance.

Analytical Model

Several authors proposed analytical models of the achromatic CSF. The most known are *Mannos* and *Sakrison* whose proposed a model for brightness [8]. This model gives the relationship between the contrast sensitivity S and the spatial frequency f given in cycle per degree (cpd).

$$S(f) = b\left(a + \frac{f}{f_p}\right) \exp^{-\left(\frac{f}{f_p}\right)^c} \quad (2)$$

Where :

- a and c influence the curve slope respectively for low and high frequencies;
- f_p is the peak position of the CSF.

So, we used the previous model by tuning its parameters to fit with the results of the psychophysical experiment. As the chromatic CSF has a low-pass shape, we used a simple exponential function as the analytical model.

$$Sc(f) = a \exp^{(bf^c)} \quad (3)$$

Results

Before the modeling step, statistical tools have been run on the raw data in order to reject outliers and avoid to include them in the model.

The identification of the parameters of the model has been performed by the Least-Square minimization. The models fit quite well with the experimental data as the error for the achromatic channel is about 1.93%. For the Red-Green opposition, the error is 2.06%. And finaly, for the Blue-Yellow opposition, the error is 0.87%. Figure 4 shows the obtained CSF for the Blue-Yellow channel and the measured data points.

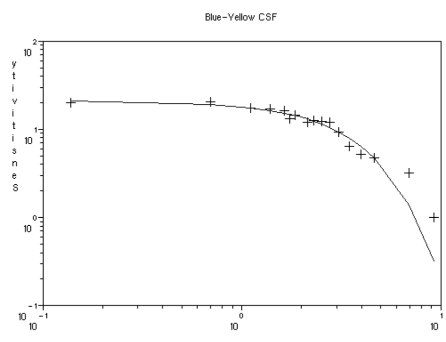


Figure 4. Blue-Yellow CSF with the originals data points

To be used in an imaging system, the Contrast Sensitivity Functions have to be normalized. Figure 5 gathers the CSF for both chromatic and achromatic channels.

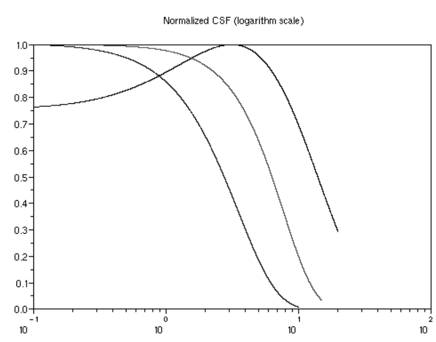


Figure 5. Normalized CSF: Luminance CSF in black, Red-Green CSF in red and Yellow-Blue CSF in blue.

These curves are drawn with the computed parameters integrated in the analytical models as indicated in the following equations.

$$S(f) = 5.54 \times \left(0.392 + \frac{f}{3.334}\right) \exp^{-\left(\frac{f}{3.334}\right)^{0.737}} \quad (4)$$

$$Sc_{red-green}(f) = 20.764 \times \exp^{(-0.023 \times f^{1.829})} \quad (5)$$

$$Sc_{blue-yellow}(f) = 21.083 \times \exp^{(-0.155 \times f^{1.476})} \quad (6)$$

Principle

As said in the introduction, JPEG2000 make use of the contrast sensitivity function (CSF) for visual optimization strategy in compression. Human eyes are less sensitive to high than to low frequencies. The CSF is used to determine what will be perceived in an image. The term weight is used to describe the sensitivity at a particular frequency. To use the CSF, which is described in visual frequencies of cycles/degree (cpd), it must be transfered to the compression domain which is in discrete frequencies (cycle/pixel). The visual weighting factors depend on the specific viewing conditions under which the image will be watched.

JPEG2000 accepts two methods for the integration of the CSF weights: Fixed frequency weighting and visual progressive weighting.

Fixed frequency weighting

The CSF curve is a continuous function of the spatial frequency but only one CSF weight is kept for each sub-band in a discrete wavelet transform. For example, depending on the specific viewing condition, the weight corresponding to the sensitivity of the mid-frequency of a sub-band could be chosen for that particular sub-band. This way of applying visual frequency weighting is referred to as fixed frequency weighting. The set of CSF weights can be incorporated in two ways in JPEG2000. In both cases, the decoder does not need to know the original CSF weights.

The first way is to modify the quantization step size. The quantization step size q_i of the transform coefficients of sub-band i is adjusted to be inversely proportional to the CSF weight w_i. The CSF normalized quantization values are then treated uniformly in the R-D optimization process.

The second way is to modify the embedded coding order. In this case, the quantization step sizes are not modified, but the distortion weights introduced into the R-D optimization are modified, with the CSF weight of each sub-band. This controls the relative significance of the bit-planes from the embedded bit-stream of each code-block.

Visual progressive weighting

JPEG2000 allows the implementation of visual progressive weighting, where different sets of visual weighting factors can be applied at different stages of the embedding. The motivation for visual progressive weighting is that the embedded bit-stream may be truncated later, so the viewing conditions may be very different. Visual progressive weighting allows the use of different sets of CSF weights that correspond to different viewing distances at different stages of the embedding.

Computation

In this work, we focused on the fixed frequency weighting that seems more adapted to the context of the study. To compute the Visual Weightings, we first take an X'Y'Z' image on which a color transform is applied towards the YCC color space. The aim is to have an opponent representation. The transformation of the image to the YCC color space is also the first step in the JPEG 2000 encoding process. Then the image is duplicated and Each component is transposed in the Fourier domain. Then, we use the modeled CSF as 2D filters.

Back to the spatial domain, a discrete wavelet transform is applied to both duplicated and original images. At this point each sub-band is processed independently to compute its particular weight. To achieve this, the following equation is used :

$$w_x = \sqrt{\frac{\sum (c_{i,j}^{csf})^2}{\sum c_{i,j}^2}} \quad (7)$$

where M is the set of coefficient of the computed sub-band, $c_{i,j}$ the original coefficients and $c_{i,j}^{csf}$ the coefficients after filtering.

Visual Weightings for Digital Cinema

The following table reports the obtained weighting factors for each component, each decomposition level and each sub-band for cinematic conditions.

	LH	HL	HH
1	0.641731	0.727859	0.472008
2	0.834714	0.852088	0.681370
3	0.975319	0.977757	0.922875
4	0.998426	0.998075	0.994100
5	1.001223	0.999697	0.997862

Y component

	LH	HL	HH
1	0.111436	0.109352	0.088143
2	0.149177	0.152355	0.094904
3	0.347601	0.362426	0.190024
4	0.636333	0.661777	0.472931
5	0.830503	0.865330	0.741372

Cb component

	LH	HL	HH
1	0.251607	0.262396	0.204351
2	0.420631	0.439335	0.228638
3	0.748305	0.754203	0.567974
4	0.913070	0.921399	0.841546
5	0.978377	0.976763	0.951529

Cr component

validation

The visual weightings obtained by the proposed approach are used with a JPEG 2000 encoder in order to compare them with visual weightings for standard conditions. The validation consists in objective and subjective evaluation. The first was performed using PSNR and SSIM while the second used a panel of observers. In order to ease the comparison for observers, we have encoded images at low bitrates i.e. 0.125 bpp, 0.250 bpp, 0.375 bpp and 0.5 bpp.

Figure 6 shows results for 0.125 bpp without VW, with standard VW and with cinematic VW and figure 7 for 0.250 bpp.

The following table gives PSNR and SSIM results for the bitrates shown in the previous figures. The results of SSIM are more correlated to human perception than PSNR. Hence, It demonstrates that the cinematic VW are suitable for low bitrates than the high bitrates.

PSNR and SSIM results for different bitrates and different configurations: without VW, with standard VW and with cinematic VW

SSIM			
Bitrate	No VW	St. VW	Cine VW
0.125	0.9626	0.9659	0.9693
0.250	0.9827	0.9853	0.9860
0.375	0.9893	0.9917	0.9912
0.500	0.9932	0.9945	0.9940
1.0	0.9976	0.9981	0.9959
PSNR			
0.125	31.7533	31.8770	31.9201
0.250	33.0704	33.1328	32.9119
0.375	34.4968	34.2977	33.5282
0.500	35.9227	35.6190	34.0539
1.0	42.3090	40.4013	38.4400

Figure 8 presents the results of subjective assessment for bitrates from 0.125 bpp to 0.5 bpp. Results show that the cinematic VW are relatively preferred to standard VW.

Conclusion

In this work, we have presented the construction of the Contrast Sensitivity Function for Cinema theater conditions. Even if the methodology is classical, the novelty lies in the conditions of assessment. This work can be considered as an input for the third amendment of JPEG 2000 standard and will be transmitted to SMPTE for an integration in the cinema specification.

In the same way, we also have presented the full process to compute visual weights. Results show that visual weighting are suitable for low bitrates when artifact are very visible. Images above the transparency threshold are difficult to assess since the quality is high enough.

Acknowledgments

This work has been funded by the European Community in the framework of the EDCine project (FP6/IP).

References

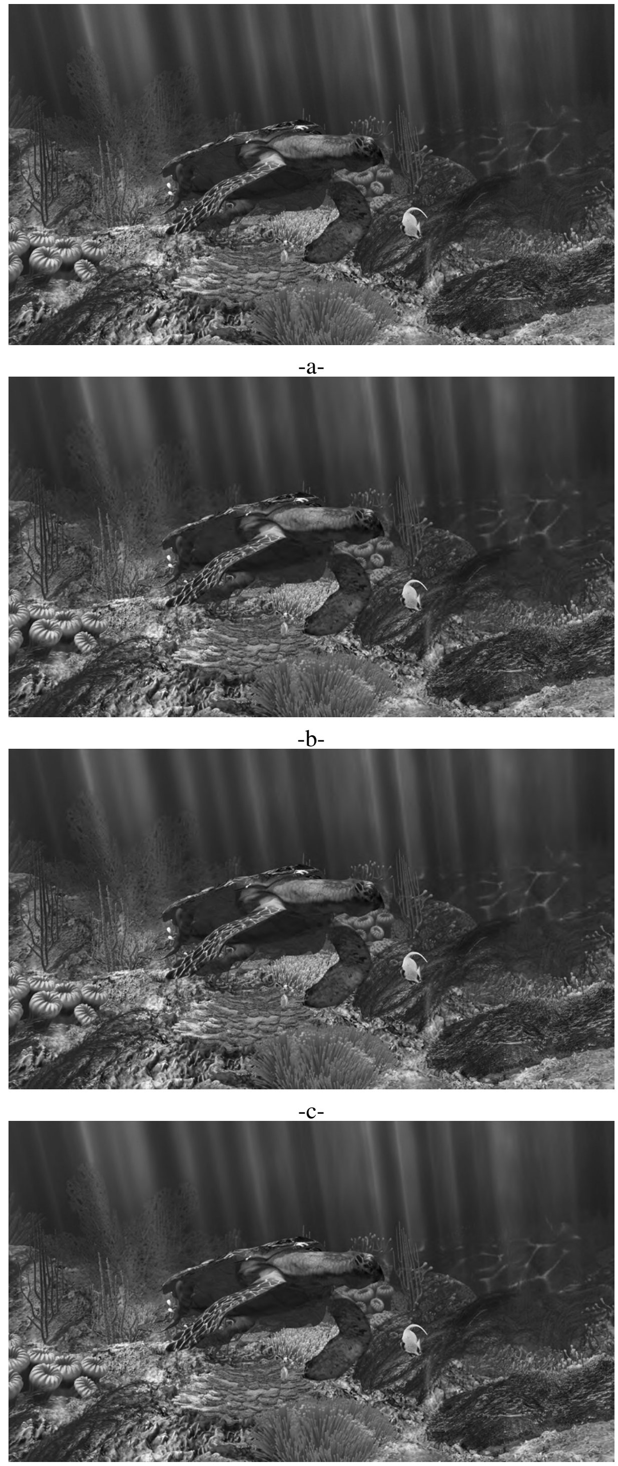
[1] C. S. Swartz, Understanding Digital Cinema: A Professional Handbook, Focal press, October 2004.

[2] Digdia, Digital Cinema - A Milestone Reached - transforming the entire industry, technical report, May 2005.

[3] DCI, Digital Cinema System Specification, technical report, july 2005.

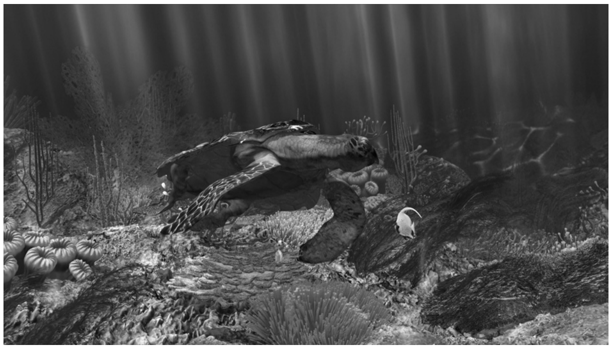
[4] ISO/SC29/WG1, Call for input to a technical report on JPEG 2000 DCinema applications. Technical report wg1n3810.

[5] A. Skodras, C. Christopoulos, and T. Ebrahimi, The JPEG 2000 still image compression standard, IEEE Signal. Processing Mag., vol. 18, pp 36-58, Sept. 2001.

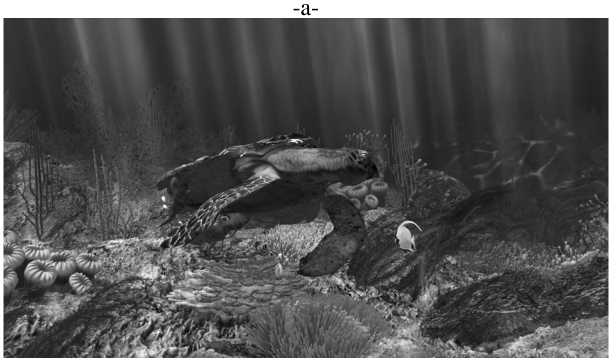


-a- -b- -c- -d-

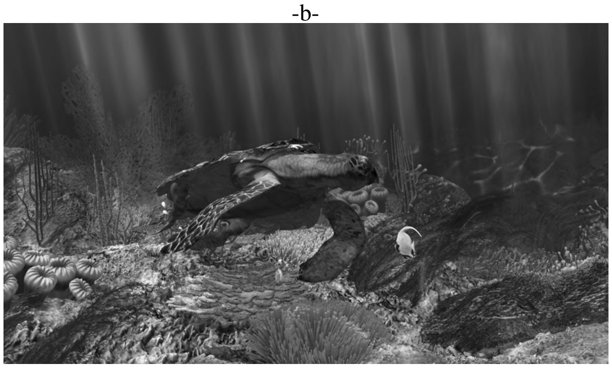
__Figure 6.__ Results at 0.125 bpp. a- original image, b- without VW, c- with standard VW and d-with cinematic VW.



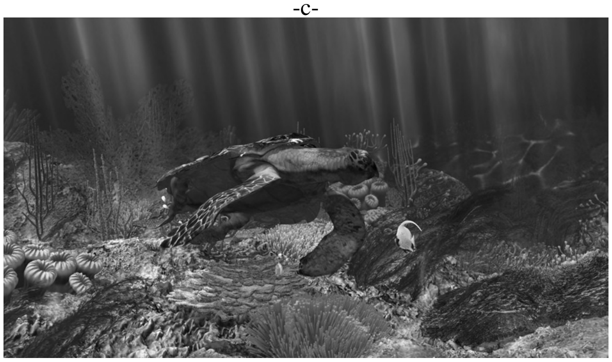
-a-



-b-



-c-



-d-

Figure 7. *Results at 0.250 bpp. a- original image, b- without VW, c- with standard VW and d-with cinematic VW.*

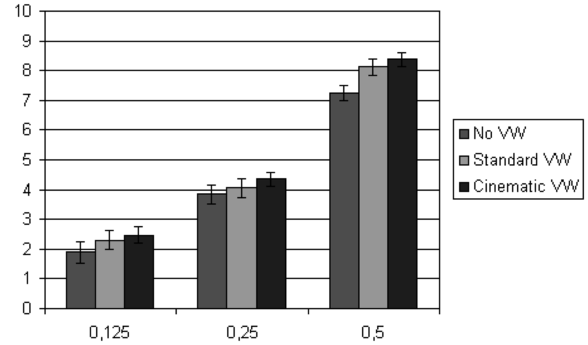


Figure 8. *subjective scores for 3 different bitrates*

[6] Information Technology JPEG 2000 Image Coding System, ISO/IEC FDIS15444-1: 2000 (Aug. 2000).

[7] Brian A. Wandell, Foundations of Vision, Sinauer Associates, Inc., Sunderland MA, first edition, 1995.

[8] J. L. Mannos and D. J. Sakrison, The effects of a visual fidelity criterion on the encoding of images, IEEE Trans. On Information Theor, vol. IT-20, no. 4, pp. 525-536, 1974.

[9] A. Michelson. Studies in Optics. PhD thesis, University of Chicago, 1927.

[10] K. Mullen, The contrast sensitivity of human colour visio to red-green and blue-yellow chromatic gratings, Journal of Physiology 359, pp. 381400, 1985.

[11] ITU-R, Methodology for the subjective assessment of the quality of television pictures, Tech. Rep. BT.500-10, ITU, Geneva, Switzerland, 2002.

[12] S. Daly, The visible differences predictor : an algorithm for the assessment of image Fidelity, in Digital images and human vision, pp. 179205, MIT Press, 1993.

Author Biography

Chaker Larabi received his master in automatics from the University of Lyon 1(1998) and his PhD in image processing from University of Poitiers (2002). Since then he is teaching and research assistant at the same university and manage the work on quality assessment methodologies, color appearance modeling, integrating the HVS models in compression applications. He is currently managing a work on digital cinema in the framework of a European project named EDCine. He is senior member IEEE, SMPTE member and member of IS&T. Currently, he is chair of AIC adhoc group of the JPEG committee and participates to CIE TC1-60.