Modeling the perceptual distortion of dynamic textures and its application in HEVC

Conference Paper · September 2016

DOI: 10.1109/ICIP.2016.7533068

3 authors:

Karam Naser
University of Nantes
9 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Vincent Ricordel
Institut de Recherche en Communications et …
63 PUBLICATIONS **150** CITATIONS
SEE PROFILE

Patrick Le Callet
University of Nantes
370 PUBLICATIONS **3,050** CITATIONS
SEE PROFILE

MODELING THE PERCEPTUAL DISTORTION OF DYNAMIC TEXTURES AND ITS APPLICATION IN HEVC

Karam Naser, Vincent Ricordel and Patrick Le Callet

University of Nantes, IRCCyN UMR CNRS 6597
Polytech Nantes, Rue Christian Pauc BP 50609 44306 Nantes Cedex 3, France
karam.naser; vincent.ricordel; patrick.le-callet
@univ-nantes.fr

ABSTRACT

In the quest of perceptually optimized video coding, coding textures is representing a challenging case. While a large body of research was put into the perception of static textures, dynamic textures are still not sufficiently explored.

In this paper, we focus on short term consistent patches, known as dynamic textures, with a very limited spatial and temporal extent. We estimated the visual distortion due to HEVC compression via subjective testing. The estimated distortion profile per texture was used to optimize the HEVC coding process. Experimental results showed that this technique can offer a high bitrate saving for the same subjective quality.

***Index Terms*—** Dynamic Texture, Perceptual Optimization, Quality Assessment, HEVC

1. INTRODUCTION

HEVC [1], the latest video compression standard, is specifically designed to minimize the distortion for a fixed rate budget. It has been shown that it can achieve up to 50% bitrate saving, as compared to the previous standard (AVC), using both subjective and objective evaluations [2]. Despite this improvement in the video coding process, two main issues are still not yet taken into consideration. Firstly, HEVC relies mainly on pixel comparison for optimizing the rate and distortion, and secondly, it is dedicated to statistical redundancies and ignores many possible perceptual redundancies.

A long with the standard video compression, there has been a large effort towards the perceptual video compression [3][4]. The perceptual compression takes into account various aspects of the human visual system, such as sub/supra-threshold distortion characteristics, visual attention and visual sensitivity. Many approaches have been proposed to use this knowledge to enhance the coding efficiency and improve the overall quality of the decompressed videos.

This work was supported by the Marie Sktodowska-Curie under the PROVISION (PeRceptually Optimized VIdeo CompresSION) project bearing Grant Number 608231 and Call Identifier: FP7-PEOPLE-2013-ITN.

For textures, perceptual compression falls generally into 2 categories, low level and high level mechanisms of human visual system. Example of low level visual mechanism is the filter bank based visual similarity measure, that is used in [5] and [6]. On the other hand, high level mechanism relies on the fact that textures are not perceived in details, and modeling them as random process can lead to indistinguishable re-generation. Examples of this approach can be found in [7] and [8].

Estimating the perceptual quality is of a fundamental importance in the video compression scenario. Conventional quality measure, such as mean squared error (MSE), considers that the change in the pixel level is linearly proportional to the change in the perceived quality. This assumption, is valid only on a limited scope [9]. For this reason, there exists a bunch of research that aims at better estimation of the perceived quality.

In the context of estimating the visual quality of textures, there has been only limited studies. In particular, for static textures, 10 textures in [10] were subjectively assessed for various types of distortions, and it was observed that the traditional image quality metrics fail to predict that quality in the case of multiple distortions types. Another notable work was done in [11], where the authors showed that the quality of the synthesized texture images can be learned from the parameters of the synthesis algorithm. On the other hand, there has been no attempt to estimate the quality of dynamic textures, up to our knowledge.

The scope of this paper covers both visual quality assessment and perceptual compression of dynamic textures. Traditionally, video quality assessment focuses on estimating the quality for full sequences with temporal extent of 5-10 seconds. Such an approach, however, can not be directly used to drive a video coding system toward perceptual optimization. This is because the coding process makes all decisions in a block-wise manner, with limited memory to past and future blocks. In order to take the best decision, namely the perceptually most preferable, an immediate quality measure is needed. For this reason, this paper focuses on small short

time patches with homogeneous contents.

The paper proposes a systematic perceptual optimization algorithm, in which the change of the physical variable (namely MSE) on the perceived distortion is mapped directly to a perceptual scale, which is then used to lead the encoder to the best rate-perceptual distortion compromise. This approach can be considered as a generalized proof of concept for a content based perceptually optimized video compression.

The rest of the paper is organized as follows. In Sec 2, the detail of subjective quality evaluation are described. Sec. 3 explains the perceptual optimization process used. The conclusion with future prospects is given in Sec. 4

2. ESTIMATING THE PERCEIVED DISTORTION

2.1. Method

Compared to the existing objective video quality metrics, the subjective evaluation is so far the most reliable method as it involves directly observers opinion about the visual quality.

Subjective quality assessment can be implemented using different methodologies, where the observers can respond to a continuous scale, discrete scale, and binary scale. It is generally agreed that the less the number of scales, the higher the precision in the observer response.

Taking into account the above discussion, we opted to use the Maximum Likelihood Difference Scaling (MLDS [12]) methodology, which is a suitable method for estimating the supra-threshold effects of a physical variable. In brief, each observer compares two pairs of distorted sequences, and selects the pair that shows a higher perceptual difference. A maximum likelihood approach is then used to estimate the perceptual difference at each value of the compression level.

Adapting MLDS in this work is straightforward. The observers were presented 4 sequences, that are horizontally 1 degree of visual angle apart, and 3 degrees vertically. Selecting a pair was done via the keyboard arrows, and by pressing "enter" to validate the selection. A screenshot of the used software is shown in Fig. 1.

The subjective test was conducted in a professional room specifically designed for subjective testing. It complies with the ITU recommendations regarding the room lighting and screen brightness [13]. The used screen was a TVLogic LVM401 with a resolution of 1920x1080 at 60Hz. The viewing distance was 3H, where H is the screen height.

2.2. Material

As being interested in short term patches of dynamic textures. We selected patches with minimal spatial and temporal extents. Spatially, the patches were chosen to be limited to foveal vision (2 degrees of visual angle), whereas temporally they were limited to 500 ms, which is more than the minimum fixation time (200 ms).

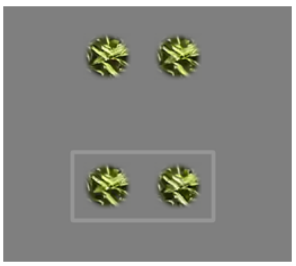

Fig. 1. Screen shot of the software used for MLDS

Two datasets were used in this work. The first one is DynTex dataset [14], which is a comprehensive dataset of 650 dynamic textures that has been extensively used for research. The other one is BVI textures [15], which is a new dataset of 20 high quality videos, designed specifically for subjective quality estimation.

Using these 2 datasets, we cropped homogeneous patches of spatial resolution of 2 degrees of visual angle (128x128 according to the viewing condition described in Sec. 2.1) of 500 ms temporal period (which corresponds to 13 and 30 frames for DynTex and BVI resp.). In total, we collected 37 sequences from DynTex and 6 from BVI.

As we have short time sequences, they were continuously repeated. To avoid temporal flickering artifact, they were temporally reversed upon each repetition. Beside this, a spatially circular window was applied to show the inner part of 64 pixels diameter, and the rest were faded using a gaussian filter (see Fig. 1).

In the subjective test, one can only afford limited number of sequences. Thus, we looked for sequences with distinguishable features. We considered the HEVC performance as an important feature for clustering the sequences. Using HEVC reference software (HM 16.2 [16]), the sequences where encoded to 10 levels of Quantization Parameter (QP), and the Bjontegaard delta PSNR (BD-PSNR [17]) was computed between all sequences. The sequence which has the minimum sum of BD-PSNR compared with all the other sequences is considered as the reference one, and the BD-PSNR with respect to this sequence is consider as the sequence feature. Accordingly, 8 sequences were retrieved using k-means clustering algorithm, which are shown in Fig. 2. For clarity, each video was assigned to a SeqId from 1 to 8, which follows the same order as shown in the figure (from left to right, and top to bottom).

2.3. Subjective Test Results

The raw data of selected pairs from the subjective experiment were converted to a perceptual scale using the software package provided by the authors in [18]. The resulting perceptual scale are shown in Fig. 3 for two sequences. The x-axes represents the overall average MSE of all the frames, whereas the

Fig. 2. Sequences used for subjective test

y-axes represents the perceived difference. The confidence intervals are computing by learning the observers probability and repeat 10000 simulations using a boot-strapping procedure as explained in [18].

The two curves shown in Fig. 3 represent two different trends in the MSE vs perceptual difference relationship. The first trend, as for SeqId 2, shows that there is a big deviation between the measure distortion (MSE) and the perceived one. On the other hand, the second trend, which is shown for SeqId 7, indicates that MSE is directly proportional to the perceived value of distortion.

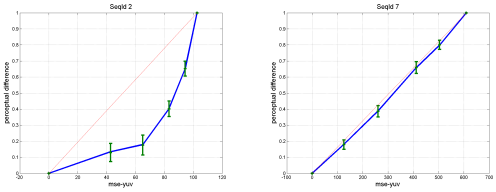

Fig. 3. Subjective test results of 2 sequences representing 2 different trends

3. PERCEPTUAL OPTIMIZATION OF HEVC

3.1. Optimization Process

Looking carefully at the curves in Fig. 3, we can see that sequences belonging to the first trend are suitable candidates for perceptual optimization. This is because the high deviation between the perceived distortion and the measured distortion can lead to wrong decisions inside the coding loop and thus would not lead to the optimal rate quality compromise. The SeqId's of the sequences belonging to this trend are given in the first rwo of Table 1.

In HEVC, the best prediction mode and block splitting are selected when they minimize the combined rate and distortion cost. The distortion is the Sum of Squared Differences (SSD), which is the MSE value multiplied by the number of pixels belonging to the block under consideration, and the combination of rate and distortion is done via a Lagrangian multiplier (λ). The optimum value of λ corresponds to the negative derivative of distortion over rate.

A straightforward way to utilize the subjective test result in the video compression scenario (HEVC) is to map the distortion measure in HEVC to its perceptual value. To achieve this, we used linear piece-wise mapping functions derived from the subjective test, was used to convert the measured MSE into a perceptual value (SSD_p) as follows:

$$\begin{aligned} SSD_p &= (\alpha MSE + \beta) \times N \\ &= \alpha SSD + \beta N \end{aligned} \tag{1}$$

Where N is the number of pixels belonging to the given block. The new lambda (λ_p) value can be also derived as follows:

$$\begin{aligned} \lambda_p &= -\frac{\partial SSD_p}{\partial r} \\ &= (\frac{\partial SSD_p}{\partial SSD}) \times (-\frac{\partial SSD}{\partial r}) \\ &= \alpha \times \lambda \end{aligned} \tag{2}$$

Thus, the λ_p is a scaled version of the previous λ.

According to this analysis, the perceptual optimization process, for each texture type, consists of piece-wise mapping function and scaling factor.

3.2. Estimating the Bitrate Saving

Estimating the bitrate saving between two compression algorithms, modified and original one, is curried out via comparing the bitrates of same quality points (iso-quality points). Finding these iso-quality points is equivalent to estimating the threshold where one cannot distinguish the difference between the two compression algorithms.

We designed a specific subjective test to estimate this threshold, namely a forced choice yes/no method. We fixed the reference encoder bitrate (HM 16.2), and used the optimized encoder to produce 7 bitrate values around the reference rate. Each pair of dynamic patches, obtained from reference and optimized encoder, was shown to the observers 6 times. The observers task is to select the patch that is better in quality, a screen shot of the used software is shown in Fig. 4.

Using the same subjective setup as in 2.1, the preference probability was computed. The preference probability is a psychometric function that can be generally fitted with an S shaped function. We used Weibull function (from the Matlab psychophysics toolbox [19]) as a fitting function using the maximum likelihood estimation. An example of the preference probability with its fitting function is shown Fig. 5. The fitted preference probability is then used to infer the iso-quality point, which corresponds to 50% value of probability of preference.

Fig. 4. Screen shot of the software used for 2AFC

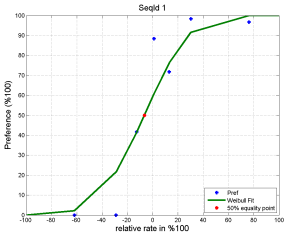

Fig. 5. Example of a psychometric preference function with Weibull fitting

3.3. Perceptual Optimization Results

The optimization process in Sec. 3.1 was used for the sequences shown in Table 1, as being sequences with possible perceptual optimization (see Sec. 3.1). To investigate the amount of possible bitrate saving, we considered the points where the maximum deviation between the MSE and perceptual difference is assumed to occur. This corresponds to the QP values shown in Table 1. The bitrate saving is computed as follows:

$$Saving = (R_d - R_p)/R_d \tag{3}$$

where R_d and R_p represent the rate of the default HEVC and the perceptually optimized version resp. We can clearly see in Table 1 that the proposed optimization process can highly reduce the bitrate (up to 29%).

3.4. Generalization of The Proposed Approach

The results obtained so far, are specific for each texture type, that was learned from the subjective experiment in Sec. 2. In order to practically deploy such an approach, it must be verified that it works consistently with other textures belonging to same texture types. To do so, we sampled 4 new sequences

SeqId	1	2	3	8
QP (default)	43	47	42	42
Bitrate saving (%)	6.5	12.3	**28.6**	7.5

Table 1. Relative bitrate saving

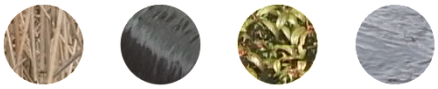

Fig. 6. Sequences used for validation test

SeqId	1'	2'	3'	8'
QP (default)	43	47	42	42
Bitrate saving (%)	9.2	5.4	10.5	**17.7**

Table 2. Relative bitrate saving

(shown in Fig. 6), which are the most similar to ones in Table 1. Once more, The used feature to assess the similarity is the same as in Sec. 2.2, namely HEVC rate-distortion behavior. These sequences were compressed using the perceptual optimization process used for the corresponding texture type. The corresponding bitrate saving, shown in Table 2, indicates clearly that the proposed approach is also valid for other sequences, sharing similar features.

4. CONCLUSION

In this paper, the need for short term quality measure was highlighted, as it is an important factor to steer the encoding process toward a perceptual direction, which can offer a bitrate reduction and/or quality improvement that can positively enhance the overall user experience.

A specialized subjective test methodology (MLDS) was used to estimate the perceptual distortion of HEVC compression, on a set of spatio-temporally homogeneous contents, known as dynamic textures. For a certain category of dynamic textures, a straightforward perceptual optimization was possible, achieving a bitrate saving up to 28%. The results were validated with other sequences and showed a consistency with the previous results.

The advantages of the proposed optimization algorithm are the simplicity and compatibility. That is, no need for a complicated quality metric and only linear mapping of the used distortion measure is needed. In terms of compatibility, there is no change in the reference decoder, so the sequences can be directly decoded by the HEVC standard.

The future work is in two directions. First is to learn more about the optimizable patches, and look insight into their discriminate features. Secondly, is to look for a better mapping between the measured and perceived distortion, such as differentiating the distortions in each prediction scheme (inter, intra), color components, and or temporal layer [1].

5. REFERENCES

[1] Gary J Sullivan, J-R Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] J-R Ohm, Gary J Sullivan, Holger Schwarz, Thiow Keng Tan, and Thomas Wiegand, "Comparison of the coding efficiency of video coding standardsincluding high efficiency video coding (HEVC)," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1669–1684, 2012.

[3] Jong-Seok Lee and Touradj Ebrahimi, "Perceptual video compression: A survey," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 684–697, 2012.

[4] Hong Ren Wu, Amy R Reibman, Weisi Lin, Fernando Pereira, and Sheila S Hemami, "Perceptual visual signal compression and transmission," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2025–2043, 2013.

[5] Guoxin Jin, Yuanhao Zhai, Thrasyvoulos N Pappas, and David L Neuhoff, "Matched-texture coding for structurally lossless compression," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 1065–1068.

[6] Karam Naser, Vincent Ricordel, and Patrick Le Callet, "Performance Analysis of Texture Similarity Metrics in HEVC Intra Prediction," in *Int. Workshop Video Processing and Quality Metrics for Consumer Electronics*. VPQM, 2015.

[7] Karam Naser, Vincent Ricordel, and Patrick Le Callet, "Local texture synthesis: A static texture coding algorithm fully compatible with HEVC," in *Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on*. IEEE, 2015, pp. 37–40.

[8] Johannes Ballé, Aleksandar Stojanovic, and Jens-Rainer Ohm, "Models for static and dynamic texture synthesis in image and video compression," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 7, pp. 1353–1365, 2011.

[9] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, June 2008.

[10] Milind S Gide and Lina J Karam, "On the assessment of the quality of textures in visual media," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE, 2010, pp. 1–5.

[11] Darshan Siddalinga Swamy, Kellen J Butler, Damon M Chandler, and Sheila S Hemami, "Parametric quality assessment of synthesized textures," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 78650B–78650B.

[12] Laurence T Maloney and Joong Nam Yang, "Maximum likelihood difference scaling," *Journal of Vision*, vol. 3, no. 8, pp. 5, 2003.

[13] ITUR Rec, "Bt. 500-11,," *Methodology for the subjective assessment of the quality of television pictures*, vol. 22, pp. 25–34, 2002.

[14] Renaud Péteri, Sándor Fazekas, and Mark J Huiskes, "DynTex: A comprehensive database of dynamic textures," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1627–1632, 2010.

[15] Miltiadis Alexios Papadopoulos, Fan Zhang, Dimitris Agrafiotis, and David Bull, "A video texture database for perceptual compression and quality assessment," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2781–2785.

[16] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG, " High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description, year = 2014," Tech. Rep.

[17] Gisle Bjontegaard, "Calculation of average PSNR differences between RD-curves," *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4 April 2001*, 2001.

[18] Kenneth Knoblauch, Laurence T Maloney, et al., "MLDS: Maximum likelihood difference scaling in R," *Journal of Statistical Software*, vol. 25, no. 2, pp. 1–26, 2008.

[19] David H Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, pp. 433–436, 1997.