



HAL
open science

HVS based perceptual pre-processing for video coding

Madhukar Bhat, Jean-Marc Thiesse, Patrick Le Callet

► **To cite this version:**

Madhukar Bhat, Jean-Marc Thiesse, Patrick Le Callet. HVS based perceptual pre-processing for video coding. European Signal Processing Conference (EUSIPCO 2019), Sep 2019, Coruña, Spain. 10.23919/EUSIPCO.2019.8903172 . hal-02283750

HAL Id: hal-02283750

<https://hal.science/hal-02283750v1>

Submitted on 30 Mar 2021

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

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

HVS based perceptual pre-processing for video coding

* † Madhukar Bhat, † Jean-Marc Thiesse, * Patrick Le Callet

* *Group IPI Lab LS2N, University of Nantes, LUNAM University, 44306 Nantes, France*

* *patrick.lecallet@univ-nantes.fr*

† *VITEC, 99 rue Pierre Semard, 92320 Chatillon, France*

† *{madhukar.bhat, jean-marc.thiesse} @vitec.com*

Abstract—This paper proposes a perceptually optimized pre-processing technique using state of the art Human Visual System (HVS) model suitable for video compression to reduce bit-rate at the same quality. Visual masking models to accurately account HVS has been considered. Frequencies which are visually indistinguishable and need not be encoded are removed within visibility threshold. The scheme is optimized for multiple viewing distances to consider real-world scenarios. Extensive subjective and objective evaluation has been conducted to evaluate proposed pre-processing. Investigation shows significant bit-rate savings compared to a professional real time HEVC video encoder.

Index Terms—Pre-processing, Contrast sensitivity function, Subjective test, Human Visual system, Visual masking

I. INTRODUCTION

Perceptual optimization to remove perceptual redundancy and adaptation to real-world scenarios has been an active area in video compression [1]. Perceptual optimizations include modifications in encoding the video according to the scene presented and visual changes to the video outside the encoding loop [1]. These adaptations aim to introduce psycho-visual bit-rate savings and minimize resource allocation for the encoding process.

The coding loop is the most critical part of nowadays real-time encoders due to the amount of computations involved in coding decisions and reconstruction together. Thus, it is highly desirable to implement processing outside this loop using the source video to ease encoding. Using pipelining of the various modules finally allow designing the full encoder. Pre-processing is one relevant way to achieve perceptual optimization outside the video coding loop. It aims to remove visually redundant noise and high frequencies from the source video to ease compression and coding efficiency at a given quality level of the video.

Pre-processing can be carried out both in pixel domain or in frequency domain in many ways. There have been attempts of pre-processing as denoising filter [2] whose strengths were controlled by encoding parameters such as motion vectors and energy of the residual [3] [4]. These works mainly to reduce noisy components in the video introducing unnecessary complexity during encoding. In [5] [6] pre-processing is seen

as an adaptive low-pass filter to reduce high-frequency content which are visually not important but potentially challenging to encode. The adaptive low-pass filtering is applied using some more methods like diffusion [7], bilateral filtering [8] and adaptive weighted averaging [9]. Vidal et al. [10] introduced two filters combining good features of both bilateral and adaptive weighting average filters. In [11] [12] an anisotropic filter was proposed based on contrast sensitivity estimation for different real world video delivery scenarios such as viewing angle and distance. Just Noticeable Difference models(JND) have been employed in [13] [14] to control pre-filters. These JND models remove information within the noticeable threshold from a visual signal to avoid overestimation of removable frequencies. There is still room for improvements in accurate perceptual modeling of frequencies that could be removed combined with condition to be within the visibility threshold (JND) to ease the encoding process.

In contrast to above methods to model removable frequencies, Human Visual System (HVS) models are useful in accurately measuring distortions in content that are not visible to the human eye. In [15] [16] [17] [18] computational models of HVS intended to provide an estimate for the visibility difference between a distorted and original signal have been proposed. They include following elements: non-linear sensitivity to calculate absolute luminance, contrast sensitivity to visual frequencies, oblique effect and sub-band decomposition into visual channels, masking effect, Psychometric function and error pooling. These HVS tools can be carefully used to remove or add information from the source video within the visibility threshold [19]. Hence, HVS models theoretically perform better than other solutions in accurately measuring redundant information that could be removed from a source within JND to perform pre-processing algorithm.

This paper introduces a new perceptual pre-processing filter named HVSPP which uses such HVS model for Luma and an adaptive low-pass filter for Chroma within JND threshold. HVSPP is built on visual tools such as contrast masking and luminance masking optimized for multiple viewing distance and screen brightness. The contrast sensitivity function (CSF)

is calculated in the complex multi-scale steerable pyramid [20] domain of the visual channel model of physical luminance of the source video. The complete model of masking effect is normalized to suit the pre-processing within the JND threshold to get the pre-processed video. Coding efficiency in-terms of bit-rate after application of HVSP is verified using extensive subjective and objective tests. The tests were conducted on a real time HEVC encoder developed by VITEC, a French company which expertises in real time video encoders.

The paper is organized as follows: Section II explains HVSP algorithm in detail. Experimental setup for the subjective assessment of the proposed solution is carefully explained in section III. Corresponding results are presented in section IV together with objective evaluation. Conclusions are drawn in section V.

II. FRAMEWORK FOR PROPOSED PERCEPTUAL PRE-PROCESSING FOR VIDEO CODING

In this section, the detail of the proposed HVS based perceptual pre-processing (HVSP) framework is given. Different HVS tools are applied in the pre-processing step to Luma component. Chroma component is applied with a filter based on morphological operations. The final HVSP signal is then encoded instead of source. Next subsections describe in Luma and Chroma Pre-processing in detail.

A. Luma pre-processing using HVS tools

As mentioned in the introduction, Luma pre-processing in the proposed algorithm uses HVS models to remove redundant information. The accuracy of the HVS model to account for masking effects in-terms of luminance and contrast with optimization for different real world use cases. This framework considers two viewing distances at 3H (height of the display) which is the recommended minimum distance of the observer from full HD display and 4.5H which is the reference viewing distance for consumers for optimization and evaluation. The luminance of the targeted display is 250cd/m2.

The proposed work-flow of the HVSP for Luma is shown in Fig 1. The HVSP algorithm is carried out in block-wise manner to target hardware implementation. As shown in the Fig.1 at first, Luma block is converted to physical luminance value. Then, the visual channel model is calculated. This is followed by, the visual channel model decomposed by multi-scale representation using the steerable pyramid. The contrast masking is then calculated for each scale and orientation. The calculated contrast masking at each scale and orientation is normalized and applied to multi-scale decomposition of luminance. Finally, the pre-processed block at each scale and orientation is reconstructed back to luminance and Luma. This Luma block is the HVSP block.

B. Luminance calculation

Generally in video representation and storage gamma encoded values are used for display purpose. The contrast sensitivity functions used in this paper work on the physical luminance in cd/m2. In the proposed model YUV videos are

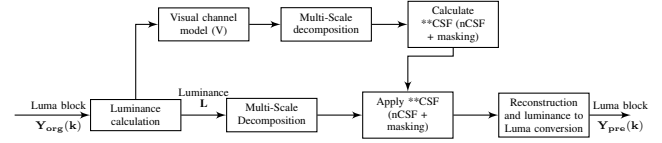


Fig. 1: Proposed work-flow: Perceptual Luma Pre-Processing for Luma component

dealt with, so a conversion of Luma component Y to physical luminance L in cd/m2 is necessary. The conversion is as follows:

$$L(\text{in cd/m}^2) = M * Y^\gamma + \text{black_level} \quad (1)$$

γ value here is 2.2. Y is the normalized Luma component and M is the maximum brightness of the screen in cd/m2. *black_level* is the minimum brightness of the screen.

C. Visual channel model

Visual channel model mimics the human visual system closely when luminance L is fallen on the eye [18]. The model accounts for Intra-ocular light scatter which occurs when entering eye, photo receptor sensitivity which are the probability that photo receptors senses a photon, luminance masking which are responses of the eye for light which are usually non-linear and finally achromatic response which is joint cone and rod response. This transformation is represented as P.

D. Multi scale decomposition

The visual cortex of the human visual system presumably goes through decomposition of signals. To mimic that multi-scale decomposition is applied to visual channel model P. CSF model are better adapted at localized frequency in steerable pyramid compared to pixel domain, wavelet and cortex transform [18]. The multi-scale decomposition of the visual channel model P is achieved using a complex steerable pyramid in the current framework. To avoid reconstruction error introduced in the real-valued pyramid, a complex representation of the steerable pyramid is adapted in this framework. As shown in Fig. 1 there are two decompositions one is for P for calculating contrast masking and other is for L for applying contrast masking.

E. Contrast sensitivity calculation and application

In the proposed framework, the main step is to remove information from a Luma block within the noticeable threshold. Overall information that could be removed in each band is modeled as a sum of signal dependent noise (neural noise {nMask}) and signal independent noise (visual masking {CSF}). Signal independent noise which is CSF, try to remove information such as noise in the block within visibility threshold. The CSF $N_{CSF}(\rho)$ is modeled according to Barten's CSF model [21],

$$N_{CSF}(\rho) = \frac{1}{\sqrt{(1 + (p_1\rho)^{p_2}) \cdot (1 - e^{-(\rho/\tau)^2})^{p_3}}}, \quad (2)$$

where ρ is the spatial frequency in cycles per degree and p_1 to p_3 are fitted parameters adapted to 250cd/m².

The signal dependent noise N_{nMask} try to remove information such as oscillations in texture and structure which are within visibility threshold [18]. For f-th spatial frequency band and o-th orientation of the complex steerable pyramid, overall contrast masking(CM) is expressed as,

$$CM(f, o) = \frac{1}{\sqrt{N_{nMask}^2 + N_{nCSF}^{2p}}} \quad (3)$$

where, p is the gain control function which controls the shape of the masking function. The typical range of p is between 1.3 (contrast discrimination) up to 4 (contrast detection). In case of pre-processing the objective is to remove as much information as possible from the scene which are invisible i.e within JND. To make sure contrast masking (CM) is within JND threshold at scale f and orientation o of steerable pyramid, psychometric function can be calculated. This gives probability of detection at that band. It is expressed as,

$$\Psi(f, o) = 1 - e^{\log(0.5) \frac{|B_{PPP}(f,o) - B_{org}(f,o)|^p}{CM(f,o)}} \quad (4)$$

where, $B_{\{PPP\}}$ is the pre-processed band, $B_{\{org\}}$ is original band. The value of p is optimized in such a way that $\Psi(f, o)$ should be less than 0.5. Another objective attached for p is coding efficiency after encoding. Using these two conditions p has been set to 1.5 for high-pass and 3 for band-pass after careful analysis.

The total neural noise $CM(f,o)$ is in the domain of multi-scale decomposition of visual channel model(P). To have a reconstructible solution for application of neural noise to remove information from each band following normalization is performed.

$$B_{t_{normalized}}(f, o) = \frac{CM(f, o)}{\max(CM f, o)} \quad (5)$$

The information that can be removed from each band within probability of noticeable visual change is the normalized neural noise. This removal is performed as follows,

$$B_{PPP}(f, o) = B_{org}(f, o) - [B_{org}(f, o) * B_{t_{normalized}}(f, o)] \quad (6)$$

Then reconstruction of pre-processed steerable pyramid is carried out. Finally the physical luminance is converted back to Luma representation to give pre-processed block in the pixel domain. Fig. 2b shows the exaggerated absolute difference between source frame and pre-processed frame. HVSP removes un-necessary noise in the flat region like sky along with visually redundant oscillations in texture and edges.

F. Chroma pre-processing

Pre-processing of Chroma follows a morphological analysis based pixel domain filter process controlled by quantization level of video to be encoded. According to encoding QP a threshold of how much a pixel can be changed is set. Then 3x3 and 5x5 Gaussian filters are applied to image separately and also gradient of the image with 3x3 and 5x5 kernel is found.



Fig. 2: (a) For crowdrun video sequence, Absolute difference (original-HVSP) for pre-processed frame(exaggerated by 100 times).

Then a particular pixel is changed according to its gradient. If a pixel is not a gradient pixel with 5x5 gradient filter then pixel value of Gaussian filtered image with 5x5 kernel is applied. Then the same step is repeated for 3x3 kernel. In this way strong edges are kept and flat areas are modified. 5x5 kernel is used see if the pixel is just a noise in flat area more robustly.

III. EXPERIMENTAL SETUP

The performance of the perceptual pre-processing framework is evaluated on four 1080p video sequences. Extensive paired comparison subjective tests and in addition objective tests have been conducted to validate performances of HVSP in the context of video compression using real time HEVC encoder developed by VITEC. Subjective tests incorporated to evaluate the proposed framework follows ITU-T P.910 [24]. The subjective test protocols is explained in the following paragraphs.

A. Content and Stimuli selection

Video sequences used for evaluating perceptual pre-processing algorithm is selected containing a wide range of real-world scenarios. Four 1080p videos at 50fps is selected for the subjective test. Two videos are JCT-VC common test condition videos [22]. Another two are from SVT test videos [23]. Selected videos are listed in TABLE I. Selection of stimuli for subjective test is based on HVSP and original video encoded at 4 QP's. TABLE I shows QP's selected for each sequence. The selection of QP is based on their RD-Curve.

Sequences	Selected QP's
Crowdrun	22,27,32,38
BasketBallDrive	20,27,35,40
BQTerrace	21,25,32,38
OldtownCross	22,27,32,37

TABLE I: Selected sequences and stimuli

B. Test environment

The testing environment incorporated here is based on Recommendation ITU-T P.910 [24]. Two viewing distances corresponding to the height of the display (H) at 3H and 4.5H are used. The luminance of the screen used for the test is 250cd/m².

C. Paired comparison test design

Paired comparison test conducted in this work uses squared design [25] with 9 Hypothetical Reference Circuits (HRCs) and 18 pairs to compare for each sequence. HRC includes source and eight videos encoded at different QPs from HVSP and original video. 30 naive observers took part in the test for each viewing distance.

IV. RESULTS

In this section, performance of HVSP is discussed for the setup explained in the last section for video compression. First, BD-rate is calculated using objective metrics and then subjective test data is carefully analyzed to measure performance of HVSP. Objective metrics used for the BD-rate measurement are PSNR, MS-SSIM, VMAF and HDR-VDP. TABLE II and TABLE III shows BD-Rate for HVSP against original at 3H and 4.5H respectively. The performance of HVSP is almost the same as original at 3H and worse at 4.5H for PSNR. However, this is expected as PSNR does not perform well to capture perceptual improvements [27]. Other metrics show BD-rate gain for HVSP compared to original at the same quality. At 4.5H, average bit-rate savings at the same quality increase compared to 3H for all perceptual metrics except PSNR. Objective metrics except HDR-VDP are not optimized for different viewing distances. Hence, the more suitable way of accurately quantifying improvements of perceptual algorithms like HVSP is subjective test [28].

Bit-rate saving /Sequence	PSNR	VMAF	MS-SSIM	HDR-VDP
Crowdrun	0.17%	-1.34%	-0.1%	-0.55%
Basketball	0.04%	-2.25%	-1.29%	-3.97%
BQ_terrace	0.4%	-0.64%	1.3%	-2.86%
oldtowncross	-0.54%	-4.13%	-3.63%	-7.5%
Overall	0.01%	-2.09%	-0.93%	-3.72%

TABLE II: BD-rate at same quality for observer at 3H the height of screen

Bit-rate saving /Sequence	PSNR	VMAF	MS-SSIM	HDR-VDP
Crowdrun	1.87%	-2.52%	-0.27%	-1.19%
Basketball	4.67%	-3.42%	-1.92%	-5.02%
BQ_terrace	2.27%	-5.25%	-2.13%	-5.07%
oldtowncross	0.67%	-3.23%	-3.93%	-5.09%
Overall	2.37%	-3.61%	-2.06%	-4.09%

TABLE III: BD-rate at same quality for observer at 4.5H the height of screen

As explained in section III, an extensive subjective test is conducted to measure performance of HVSP. Paired comparison data obtained from subjective test is processed using Bradley-Terry model [29] to get relative score (D) for each sequence with 95% confidence intervals(CI). It is important to note for Bradley-Terry scores that, BD-quality at same rate for different sequences does not have any meaning [28]. This is because subjective scores are calculated relatively on per-sequence basis with 95% CI and are not comparable to other sequences. Hence, only BD-rate at relative subjective scores

can be calculated using subjective comparison of encoder based on fitted curve [30]. This model takes 95% CI attached to relative scores from Bradley-Terry scores (D) into account. Mean, maximum and minimum bit-rate gain for both the distances are given in TABLE IV. They are calculated as,

$$\Delta R = \psi(\hat{r}_1(D), \hat{r}_2(D), D_L, D_H) \quad (7)$$

$$\Delta R_{min} = \min\{\psi(\hat{r}_1^-(D), \hat{r}_2^+(D), D_L, D_H), \psi(\hat{r}_1(D), \hat{r}_2(D), D_L, D_H)\} \quad (8)$$

$$\Delta R_{max} = \max\{\psi(\hat{r}_1^-(D), \hat{r}_2^+(D), D_L, D_H), \psi(\hat{r}_1(D), \hat{r}_2(D), D_L, D_H)\} \quad (9)$$

Where, ΔR is the mean bit-rate difference between original and HVSP using mean relative subjective scores using ψ fitting function [30]. $\hat{r}_1(D)$ indicates rate at particular relative scores using mean scores (mean grade series from D_L to D_H). $\hat{r}_1^{+/-}(D)$ indicates rate at relative score including confidence interval ($D \pm CI$, max/min grade series). ΔR_{min} and ΔR_{max} indicates minimum and maximum bit-rate difference, hence indicating maximum and minimum bit-rate gain respectively [30]. Overall mean subjective bit-rate savings are 5.76% at 3H and 8.5% at 4.5H with maximum subjective bit-rate saving going up to 17.41% and 19.86% for respective distances.

Sequences	Mean Bit-rate gain (ΔR)		Minimum Bit-rate gain (ΔR_{max})		Maximum Bit-rate gain (ΔR_{min})	
	3H	4.5H	3H	4.5H	3H	4.5H
Crowdrun	-6.74%	-4.71%	4.94%	4.22%	-13.07%	-15.64%
Basketball	-5.02%	-9.01%	4.23%	4.41%	-15.64%	-21.82%
BQ_terrace	-6.05%	-2.53%	14.5%	14.09%	-20.59%	-14.93%
oldtowncross	-5.24%	-17.74%	4.06%	-5.00%	-20.32%	-27.08%
Overall	-5.76%	-8.5%	6.93%	4.43%	-17.41%	-19.86%

TABLE IV: BD-rate at same quality for observer at 3H and 4.5H. (A negative value indicates decrease in bit-rate (gain) for HVSP at same subjective score compared to original)

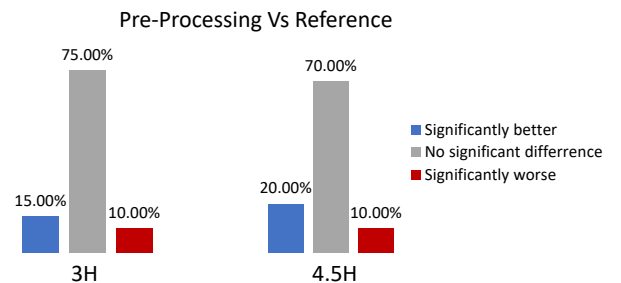


Fig. 3: Significance of difference between HVSP and original encoded at same QP

In another analysis of subjective data, paired comparison data of stimuli at same QP is considered. The significant subjective difference between HVSP (S1) and original (S2) video encoded at same QP is calculated. They are significantly different if,

$$\|D(S1) - D(S2)\| > CI \quad (10)$$

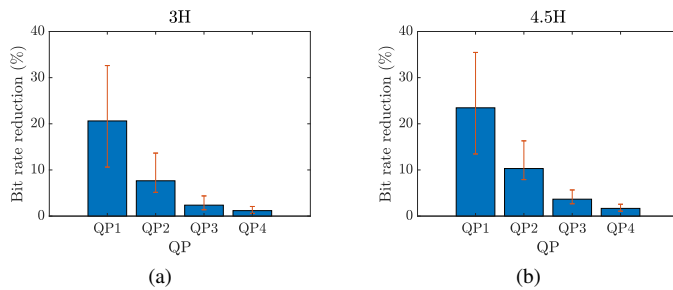


Fig. 4: Bit-rate reduction for HVSP at same QP

In most cases, the stimuli pair at the same QP are not significantly different, as shown in Fig. 3. This means, there is no significant subjective quality difference between HVSP and original video at same QP for most of the cases with 95% of CI. This implies that, statistically bit-rate reduction produced by HVSP at a particular QP can be considered as bit-rate saving at that QP for HVSP. The average bit-rate reduction at QP's from Section III for different sequences are shown in Fig. 4. Bit-rate reduction with HVSP compared to original reduces with an increase in QP, because overall bit-rate needed to encode the video also reduces. It can be concluded that, subjectively HVSP can produce a bit-rate gain of up-to 36% depending on QP with 95% CI.

V. CONCLUSION

This paper introduced perceptual pre-processing to improve video compression efficiency. In the proposed framework, various HVS tools are used to remove information from the source video within the JND threshold. The performance of HVSP has been carefully evaluated using objective metrics and a subjective test. The analysis of results demonstrates significant benefits of using this perceptual pre-processing. HVSP demonstrates very promising subjective BD-rate gains of 5.75% at 3H and 8.5% at 4.5H. This performance is captured by most of the tested objective metrics except PSNR. Above all, the subjective assessment proves that at same QP, the difference between HVSP and original videos are not significant in most of the cases with HVSP requiring significantly less bit-rate to encode.

REFERENCES

- [1] Chen Z, Lin W, Ngan KN. Perceptual video coding: Challenges and approaches. In *Multimedia and expo (ICME)*, 2010 IEEE international conference on 2010 Jul 19 (pp. 784-789). IEEE.
- [2] Karunaratne, P., Segall, C., Katsaggelos, A.: A rate-distortion optimal video preprocessing algorithm. *IEEE ICIP*, vol. 1, pp. 481484 (2001).
- [3] Lee, J.: Automatic prefilter control by video encoder statistics. *IET Electron. Lett.* 38, 503505 (2002)
- [4] Jain, C., Sethuraman, S.: A low-complexity, motion-robust, spatio-temporally adaptive video de-noiser with inloop noise estimation. In: *IEEE International Conference on Image Processing*, pp. 557560, October (2008)
- [5] Shaw, M.Q., Allebach, J.P., Delp, E.J.: Color difference weighted adaptive residual preprocessing using perceptual modeling for video compression. *Sig. Process. Image Commun.* 39, 355368 (2015)
- [6] Shao-Ping, L., Song-Hai, Z.: Saliency-based fidelity adaptation preprocessing for video coding. *J. Comput. Sci. Technol.* 26, 195202 (2011).

- [7] Perona, Pietro, and Jitendra Malik. "Scale-space and edge detection using anisotropic diffusion." *IEEE Transactions on pattern analysis and machine intelligence* 12.7 (1990): 629-639.
- [8] Tomasi, Carlo, and Roberto Manduchi. "Bilateral filtering for gray and color images." in *Proceedings IEEE International Conference on Computer Vision, ICCV*. IEEE, 1998, pp. 836846.
- [9] Ozkan, Mehmet K., M. Ibrahim Sezan, and A. Murat Tekalp. "Adaptive motion-compensated filtering of noisy image sequences." *IEEE transactions on circuits and systems for video technology* 3.4 (1993): 277-290.
- [10] Vidal E, Sturmel N, Guillemot C, Corlay P, Coudoux FX. New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding. *Signal Processing: Image Communication*. 2017 Mar 1;52:124-37.
- [11] Vanam R, Reznik YA. Perceptual pre-processing filter for user-adaptive coding and delivery of visual information. In *Picture Coding Symposium (PCS)*, 2013 2013 Dec 8 (pp. 426-429). IEEE.
- [12] Vanam R, Kerofsky LJ, Reznik YA. Perceptual pre-processing filter for adaptive video on demand content delivery. In *Image Processing (ICIP)*, 2014 IEEE International Conference on 2014 Oct 27 (pp. 2537-2541). IEEE.
- [13] Yang, X., Lin, W., Lu, Z., Ong, E., Yao, S.: Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile. *IEEE Trans. Circ. Syst. Video Technol.* 15(6), 742752 (2005)
- [14] Ding, L., Li, G., Wang, R., Wang, W.: Video pre-processing with JND-based Gaussian filtering of superpixels. In: *Proceedings of the SPIE 9410, Visual Information Processing and Communication VI*, vol. 941004, 4 March (2015)
- [15] P. Barten, Evaluation of subjective image quality with the square-root integral method, *JOSA A*, vol. 7, no. 10, pp. 20242031, 1990.
- [16] S. J. Daly, Visible differences predictor: an algorithm for the assessment of image fidelity, in *Digital Images and Human Vision*, A. B. Watson, Ed. The MIT Press, Aug. 1993, pp. 179206.
- [17] A. B. Watson and A. J. Ahumada, A standard model for foveal detection of spatial contrast ModelFest experiment, *Journal of vision*, vol. 5, pp. 717740, 2005.
- [18] Mantiuk R, Kim KJ, Rempel AG, Heidrich W. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on graphics (TOG)* 2011 Aug 7 (Vol. 30, No. 4, p. 40). ACM.
- [19] Urvoy, Matthieu, Dalila Goudia, and Florent Atrusseau. "Perceptual DFT watermarking with improved detection and robustness to geometrical distortions." *IEEE Transactions on Information Forensics and Security* 9.7 (2014): 1108-1119.
- [20] Portilla J, Simoncelli EP. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*. 2000 Oct 1;40(1):49-70.
- [21] Barten, Peter GJ. Contrast sensitivity of the human eye and its effects on image quality. Vol. 19. Bellingham, WA: Spie optical engineering press, 1999.
- [22] JCTVC-L1100, Common test conditions and software reference configurations, 2013.
- [23] L. Haglund, The SVT high definition multi format test set, 2006.
- [24] ITU-T. P.910, Subjective video quality assessment methods for multimedia applications, 2008.
- [25] O. Dykstra, Rank analysis of incomplete block designs: A method of paired comparisons employing unequal repetitions on pairs, *Biometrics*, vol. 16, no. 2, pp. 176188, 1960.
- [26] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [27] Huynh-Thu, Quan, and Mohammed Ghanbari. "Scope of validity of PSNR in image/video quality assessment." *Electronics letters* 44.13 (2008): 800-801.
- [28] P. Hanhart, L. Krasula, P. Le Callet, and T. Ebrahimi, How to benchmark objective quality metrics from paired comparison data? In *Quality of Multimedia Experience (QoMEX)*, 2016 Eighth International Conference on, IEEE, 2016, pp. 16.
- [29] Bradley, Ralph A. "14 paired comparisons: Some basic procedures and examples." *Handbook of statistics* 4 (1984): 299-326.
- [30] Hanhart, Philippe, and Touradj Ebrahimi. "Calculation of average coding efficiency based on subjective quality scores." *Journal of Visual Communication and Image Representation* 25.3 (2014): 555-564.