RELIABILITY OF THE MOST COMMON OBJECTIVE METRICS FOR LIGHT FIELD QUALITY ASSESSMENT

Hadi Amirpour [1], *Antonio M. G. Pinheiro*[1], *Manuela Pereira* [1], *and Mohammad Ghanbari* [2,3]

[1] Instituto de Telecomunicações and Universidade da Beira Interior, Covilhã, Portugal
[2] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
[3] School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

ABSTRACT

Light field imaging is a promising technology for 3D computational photography. As Light Field images are represented for multiple views, their subjective evaluation is a very demanding task. Hence, identifying reliable objective quality assessment methodologies plays a very important role. In this paper six objective quality metrics; PSNR-Y, PSNR-YUV, SSIM-Y, MSSSIM-Y, FSIM-Y and HDRVDP2-Y are assessed for five state-of-the-art codecs at various bit-rates. Moreover, the metrics are computed in the linear, perceptually uniform and perceptual quantizer spaces. The results are compared against those of a subjective study and is concluded that the average FSIM-Y is the most reliable metric. The paper also introduces maps of the objective metrics to evaluate the quality dispersion among the different light field image views.

Index Terms— Light field, image quality assessment, objective metrics.

1. INTRODUCTION

Emerging technologies can produce richer information on the 3D world. Light field image is a very promising technology that has recently gained a greater attention with the introduction of commercial cameras by Lytro[1] and Raytrix[2]. Post-processing tasks like synthesizing a new view, refocusing, and depth estimation are examples of enhanced features available with this new technology [1]. However, through light field imaging huge amount of data are generated and efficient compression is required [2]. The evaluation of any compression technology requires reliable quality metrics. In particular, subjective evaluation of light field images is difficult and timely expensive because it is required to evaluate at least a representative number of views that can be generated for each light field image.

This work is funded by FCT through national funds and co-funded by FEDER PT2020 partnership agreement under the project PTDC/EEI-PRO/2849/ 2014 - POCI-01-0145-FEDER-016693, and under the project UID/EEA/50008/2019.

[1]https://www.lytro.com/

[2]http://www.raytrix.de/

In this paper, the reliability of six objective metrics is studied for the evaluation of a set of state-of-the-art compression methods, considering linear and perceptual spaces. Perceptual spaces are studied as the light field images used in this study have a 10-bit depth, and might have a higher dynamic range. In that case, according to [3] perceptual spaces can be more appropriate for the metrics computation. Moreover, a subjective study defined in [4] is used as ground truth for the metrics validation. In this study, five state-of-the-art codecs are subjectively evaluated. Finally, quality maps are also introduced to allow the observation of the quality variation between different image views. The paper is organized as described in the followings. Dataset, subjective tests and coding condition are introduced in section 2. Section 3 provides details about the objective evaluation. Section 4 benchmarks the objective metrics using the subjective results. Section 5 concludes the paper.

2. ENCODERS, DATASET, AND CODING CONDITION

2.1. Encoders

The first two methods employ the HEVC and VP9 encoders. In each encoder, the different light field image views (sub-aperture images of the lenslet image) are compressed in a pseudo-video sequence using a serpentine order for the different views [4]. Linear Approximation Prior (LAP) [5] exploits linearity between the sub-aperture images. For this purpose, sub-aperture images are divided into two non-overlap sets A and B. Sub-aperture images inside set A are arranged as a pseudo-video and are compressed using HEVC video encoder. Then, they are decoded and used to reconstruct dropped sub-aperture images that exist in set B using a global optimization strategy. In [6], sub-aperture images are considered as a multiview sequence and the multiview extension of HEVC (MV-HEVC) is used to exploit redundancies within sub-aperture images. In Sparse Predictive Coding (SPC) [7], lenslet images are decomposed into non-rectified sub-aperture images. Then, depth and geometry information of the scene are used to find displacements of center view's

segments to other views.

2.2. Dataset and coding condition

Five image contents from EPFL Light-Field Image Dataset [8] namely, $I01 = Bikes$, $I02 = Danger\_de\_Mort$, $I04 = Stone\_Pillars\_Outside$, $I09 = Fountain\_\&\_Vincent\_2$, and $I10 = Friends\_1$ have been selected for the evaluation.

Each raw lenslet image has 7728×5368 10-bit pixel resolution that after demosaicing and devignetting are decomposed into 15×15 sub-aperture images using the light field toolbox V0.4 [9]. Thereafter, sub-aperture images are converted to YCbCr color space using ITU-R Recommendation BT.709-6 [10] and, then, chroma subsampling from 4:4:4 to 4:2:2 is carried out. Compression is done targeting four compression ratios, $R1 = 0.75$, $R2 = 0.1$, $R3 = 0.02$ and $R4 = 0.005$ which are calculated by dividing the volume of the bitstream to the size of the uncompressed raw lenslet image considering 13×13 central sub-aperture images. Cropped images of central view of $Bikes$ compressed by HEVC in four bitrates are shown in Fig. 1.

2.3. Subjective quality assessment

A subjective test has been conducted for the above encoders by the Multimedia Signal Processing Group (MMSPG) in EPFL and the results have been published in [4]. In this study, the results of 10-bit images are considered. To conduct this test, the set-up was defined according the ITU-R Recommendation BT.500-13 [11]. The test conditions are summarized in Table. 1.

To conduct the subjective test, only 97 sub-aperture images out of 169 are selected. Each participant was asked to score each stimulus regarding to an alongside uncompressed reference with a 7-point scale based on recommendation of ITU-R Recommendation BT.500-13 [11]. In this way, scores range from -3 (much worse) to +3 (much better). Training steps and more information can be found in [4].

Table 1: Subjective test condition.

Feature	Value
Bit depth	10
Display	Eizo ColorEdge CG318-4K
Size of display	31.1in
Resolution of display	4096×2160
Setup	ITU-R Recommendation BT.500-13
Light	adjustable neon lamps of 6500 K color temperature
Color of the background walls	mid grey
Illumination level measured on the screens	15 lux
Distance of the subjects from the monitor	7 times the height of the displayed content
Monitor calibration1	$sRGB$ Gamut
Monitor calibration2	$D65$ white point
Monitor calibration3	20 cd/m^2 brightness
Monitor calibration4	minimum black level of 0.2 cd/m^2
Methodology	passive

3. OBJECTIVE QUALITY ASSESSMENT

Six full reference quality metrics were computed to assess the performance of the state-of-the-art encoders including PSNR_Y, PSNR_YUV, SSIM_Y [12], MS-SSIM_Y [13], FSIM_Y [14], and HDRVDP-2_Y [15]. These metrics were computed in three spaces: Linear, perceptually uniform (PU), and perceptual quantized (PQ). The perceptual spaces PU and PQ are used because the images are at 10-bit depth and according to [3] it might be more appropriate if they are represented at a high dynamic range.

3.0.1. Linear

In the linear domain, original uncompressed and compressed images are used to calculate the objective metrics.

3.0.2. PU

To obtain metrics in the PU domain, luminance (L) of reference and distorted images are normalized and scaled in the PU domain using the following equation:

$$V(L) = \frac{pu(L) - pu(L_{min})}{pu(L_{max}) - pu(L_{min})}$$
$$L_{min} = 0,\ L_{max} = 1023 \quad (1)$$

(a) $R4$ (b) $R3$ (c) $R2$ (d) $R1$

Fig. 1: Cropped images of central view of $Bikes$ compressed by HEVC in four bitrates.

To obtain $pu(x)$, $pu2\_encode$ [3] function has been used.

3.0.3. PQ

To obtain metrics in PQ domain, luminance (L) of reference and distorted images are normalized and scaled in the PQ domain using the following equation [16]:

$$V(L) = \left(\frac{0.8359+18.8516L_p}{1+18.6875L_p}\right)^{78.8438}$$
$$L_p = \left(\frac{L}{L_{max}}\right)^{0.1593} \quad (2)$$
$$L_{min} = 0,\ L_{max} = 1023$$

3.1. Performance evaluation

Objective metrics were computed for all 13×13 sub-aperture images. To evaluate the overall performance of the compression methods, the average of the objective metrics is usually computed. However, the variation of the objective metrics is an important factor that must be considered for a better assessment of an encoder performance. To address this, in addition to the mean values in each compression ratio, standard deviation, and the maximum and minimum of the objective metrics values obtained for the different sub-aperture images are also represented in the rate-distortion curves shown in Fig. 2.

For a better analysis of the diversity of the objective metrics for the different sub-aperture images, distortion maps per image view can be used. These maps represent the variation of the metric between different views. As examples, $PSNR_Y$ and $FSIM_Y$ maps for the content $I01$ at the compression ratio of $R2$ are shown in Fig. 3. As can be seen, the metrics reveal some fluctuation of quality between different sub-aperture images, which should not happen in a reliable encoder.

[3] https://sourceforge.net/projects/hdrvdp/files/simple_metrics/1.0/

4. OBJECTIVE QUALITY METRICS BENCHMARKING

In this section, correlation between objective metrics and subjective scores are reported. The MOS values against six objective metrics are plotted in Fig. 4. To evaluate the representation provided by the objective metrics of the subjective evaluation, MOS_p values are predicted from objective metrics by using a logistic function, defined as follows:

$$MOS_p(i) = b(1) + \frac{b(2)}{1 + exp(-b(3) \times (MR(i) - b(4)))} \quad (3)$$

where $MOS_P(i)$ is a representation for the predicted MOS for the ith image. $b(j)$ are the regression parameters and MR represents objective metric result. The initial values for $b(1)$ to $b(4)$ are MOS_{min}, MOS_{max}, MR_{max} and MR_{min}, respectively. Then, predicted MOS (MOS_p) are compared to the MOS values that are considered as ground truth. To assess performance of the objective metrics, five measures including Pearson correlation coefficient (PCC), Spearman Rank-Order Correlation Coefficient ($SROCC$), Kendall Rank-Order Correlation Coefficient ($KROCC$), Root-Mean-Squared Error ($RMSE$), and Outlier Ratio (OR) [3] were computed. The performance of the objective metrics in all spaces is summarized in table 2.

5. CONCLUSIONS

In this work, the performance of some objective metrics, notably, PSNR-Y, PSNR-YUV, SSIM-Y, MS-SSIM-Y, FSIM-Y, and HDRVDP-2-Y were studied. The FSIM-Y computed in

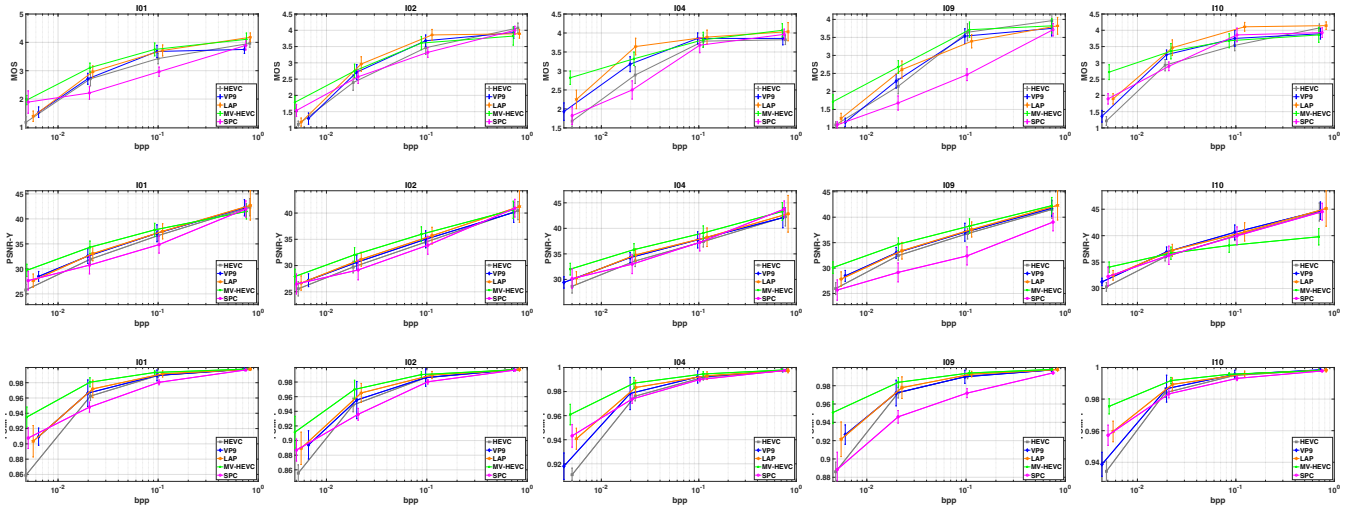

Fig. 2: MOS and Rate-distortion curves for the PSNR-Y and FSIM-Y metrics.

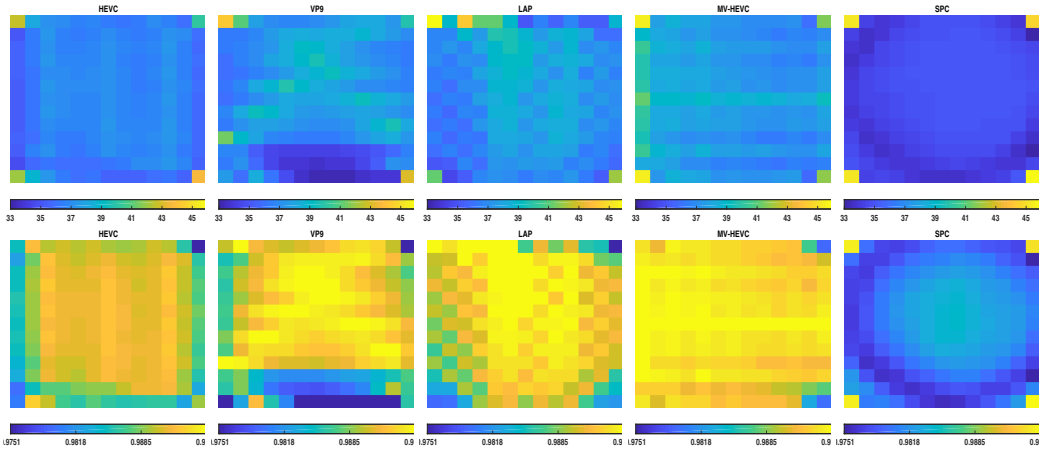

Fig. 3: Distortion maps for PSNR-Y (first row) and FSIM-Y (second row) for $R2$ bpp.

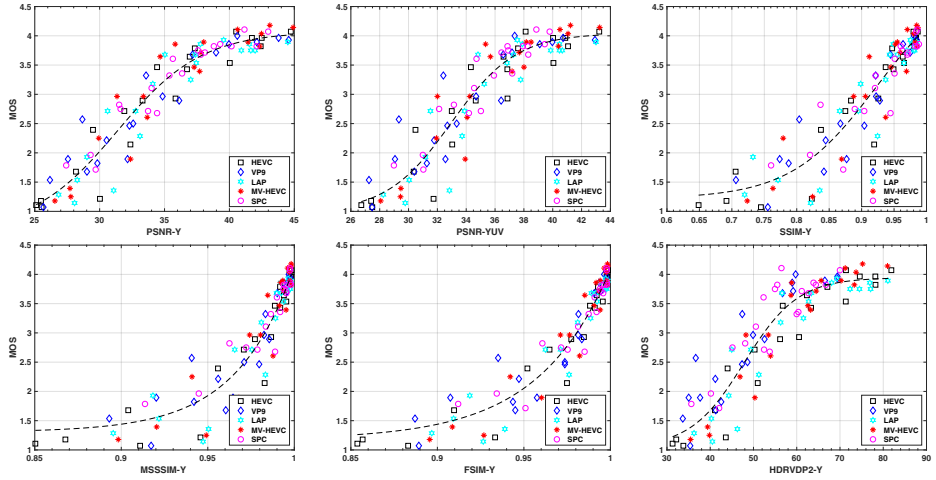

Fig. 4: Objective versus subjective values in the linear space.

Table 2: Measures between MOS_p and MOS values.

	PCC	SROCC	KROCC	RMSE	OR
PSNR-Y	0.9514	0.9466	0.8037	0.2967	0.20
PSNR-YUV	0.9310	0.9230	0.7618	0.3523	0.23
SSIM-Y	0.9350	0.9305	0.7662	0.3419	0.22
MSSSIM-Y	0.9465	0.9384	0.7817	0.3120	0.23
FSIM-Y	0.9648	0.9520	0.8126	0.2541	0.14
HDRVDP2-Y	0.9285	0.9014	0.7329	0.3574	0.26
pu-PSNR-Y	0.9339	0.9262	0.7785	0.3494	0.20
pu-PSNR-YUV	0.9135	0.9021	0.7374	0.3983	0.24
pu-SSIM-Y	0.9325	0.9274	0.7675	0.3492	0.22
pu-MSSSIM-Y	0.9478	0.9368	0.7854	0.3086	0.20
pu-FSIM-Y	0.9529	0.9397	0.7951	0.2934	0.18
pq-PSNR-Y	0.9333	0.9258	0.7776	0.3511	0.22
pq-PSNR-YUV	0.9131	0.9022	0.7369	0.3994	0.23
pq-SSIM-Y	0.9297	0.9250	0.7679	0.3571	0.21
pq-MSSSIM-Y	0.9467	0.9369	0.7854	0.3116	0.21
pq-FSIM-Y	0.9511	0.9366	0.7886	0.2988	0.21

the linear space shows a better correlation with subjective test results. However, the popular PSNR-Y has a very similar performance and can be considered as a very reliable alternative to FSIM-Y. In this study, none of the perceptual spaces led to metrics with better performance than the linear space. However, this might be due to the reduced dynamic range of the testing content.

In the assessment of the light field encoding methods, it is important to study the variations of the objective metrics for the different image views, in addition to the metrics' mean values. Hence, the standard variation was represented in the rate distortion plots and the maximum and minimum values of the objective metrics' are considered as well. Moreover, distortion maps are proposed to assess the variation of the quality through the different light field image views. In the subjective test of [4], as images are displayed as a sequence some artifacts might be concealed and affect the subjective scores. Moreover, the effect of the frame rate and refocusing might also influence the subjective test outcome. New methodologies for subjective and objective assessment that take into consideration the quality variation between different views of the light field image will be studied as future work.

6. REFERENCES

[1] Hadi Amirpour, Antonio Pinheiro, Manuela Pereira, and Mohammad Ghanbari, "Analysis of motion vectors and parallel computing in pseudo-sequence based light field image compression methods," in *Proc. SPIE 10752, Applications of Digital Image Processing XLI*, September 2018, vol. 10752, pp. 10752 – 10752 – 12.

[2] H. Amirpour, M. Pereira, and A. Pinheiro, "High efficient snake order pseudo-sequence based light field image compression," in *2018 Data Compression Conference*, March 2018, pp. 397–397.

[3] Philippe Hanhart, Marco V. Bernardo, Manuela Pereira, António M. G. Pinheiro, and Touradj Ebrahimi, "Benchmarking of objective quality metrics for hdr image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 39, Dec 2015.

[4] Irene Viola and Touradj Ebrahimi, "Valid: Visual quality assessment for light field images dataset," p. 3, 2018.

[5] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 4562–4566.

[6] G. Tech, Y. Chen, K. Mller, J. R. Ohm, A. Vetro, and Y. K. Wang, "Overview of the multiview and 3d extensions of high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 35–49, Jan 2016.

[7] I. Tabus, P. Helin, and P. Astola, "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and JPEG 2000," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 4562–4566.

[8] M. Rerabek, L. Yuan, L. A. Authier, and T. Ebrahimi, "EPFL light-field image dataset," 2015.

[9] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1027–1034.

[10] ITU-R BT.709-6, "Methodology for the subjective assessment of the quality of television pictures," Jun 2012.

[11] ITU-R BT.500-13, "Parameter values for the HDTV standards for production and international programme exchange," Jun 2015.

[12] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[13] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, Nov 2003, vol. 2, pp. 1398–1402 Vol.2.

[14] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.

[15] Rafat Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," in *ACM SIGGRAPH 2011 Papers*, New York, NY, USA, 2011, SIGGRAPH '11, pp. 40:1–40:14, ACM.

[16] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," in *The 2012 Annual Technical Conference Exhibition*, Oct 2012, pp. 1–9.