



HAL
open science

Investigating Epipolar Plane Image Representations for Objective Quality Evaluation of Light Field Images

Ali Ak, Patrick Le Callet

► **To cite this version:**

Ali Ak, Patrick Le Callet. Investigating Epipolar Plane Image Representations for Objective Quality Evaluation of Light Field Images. European Workshop on Visual Information Processing, Oct 2019, Rome, Italy. 10.1109/EUVIP47703.2019.8946194 . hal-02335328

HAL Id: hal-02335328

<https://hal.science/hal-02335328v1>

Submitted on 28 Oct 2019

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.

INVESTIGATING EPIPOLAR PLANE IMAGE REPRESENTATIONS FOR OBJECTIVE QUALITY EVALUATION OF LIGHT FIELD IMAGES

Ali Ak, Patrick Le-Callet

IPI, LS2N, University of Nantes, France

ABSTRACT

With the ongoing advances in Light Field(LF) technology, research in LF acquisition, compression, processing has gained momentum. This increased the need for objective quality evaluation of LF content. Many processing algorithms are still optimized against peak signal to noise ratio(PSNR). Lately, several attempts have been made to improve objective quality evaluation such as extending 2D metrics to 4D LF domain. However, there is still a great room for improvement. In this paper, we experiment with existing 2D image quality metrics on the Epipolar Plane Image representations of LF content to reveal characteristics of LF related distortions. We discuss the challenges and suggest possible directions towards a LF image quality evaluation on EPI representations.

Index Terms— light field, epipolar plane image, objective quality evaluation

1. INTRODUCTION

The goal to produce ultra-realistic displays brought a lot of attention to Light Fields (LF) in the research community. Light Fields were introduced by Michael Faraday in 1846 on his lecture titled as "Thoughts on Ray Vibrations" [1]. In the last century, 7D plenoptic function is introduced. It describes the radiance received from every point to every point in 3D space at any given time at every wavelength. Later, it was simplified to 4D plenoptic function which allowed for easier computations. It is defined by 2 planes in space, parallel to each other. 4D function defines emitted light from any given point in the first plane to any point in the second plane. A comprehensive overview of the current literature for LF imaging has been done by Wu et al.[2].

Light Field content can be represented in many different ways. In Fig.1 sub-aperture views and a corresponding Epipolar Plane Image representation is presented. Sub-aperture view representation is the most common one where each view is an image from a slightly distant point of view depending on the baseline. The baseline is defined as the distance between each view. In this representation, we can observe the spatial information over each view. EPI representations, on the other hand, reveals information about the angular domain. It describes the shift of pixel information

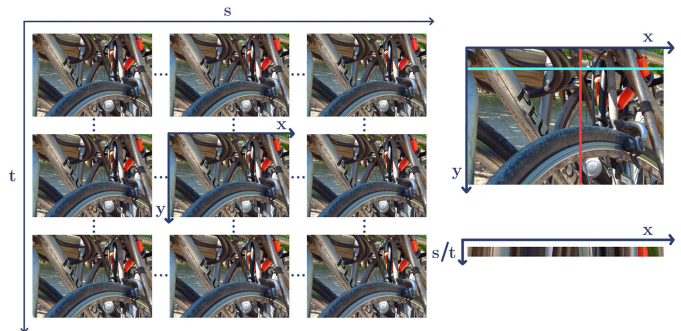


Fig. 1. 2 different Light Field image representation. On the left, each sub-aperture views. On the right, an EPI slice

over an angular axis. It is expected to have a continuous pixel value across several lines under the assumption of Lambertian surfaces on the image. For non-Lambertian surfaces, we may observe overlapping lines in EPIs because of the reflection and refraction. In case of a discontinuity in the angular domain, lines on the EPI becomes distorted.

From acquisition to the display of the LF content, there are different processing steps. Most of these processing tools add a different type of distortion to the processed LF image. Captured LF data needs to be coded, transmitted, and decoded at the display device. This process also results in unwanted distortions which lower perceived quality. Quality evaluation metrics are necessary to evaluate the performance of the developed tools for each stage.

Objective image quality assessment for 2D images has attracted considerable attention over the last several decades[3]. Quality evaluation of videos and sparse multi-view content also attracted attention[4]. There have been attempts to create a metric for LF image quality assessment[5]. Authors combined estimated spatial quality with angular quality to provide a score.

In this paper, we experiment with some of the existing 2D quality evaluation metrics on EPI representations. Rest of the paper is organized as follows: In Section 3, we investigate existing objective image quality metrics and select a set of metrics to experiment on EPI representations. In Section 2,

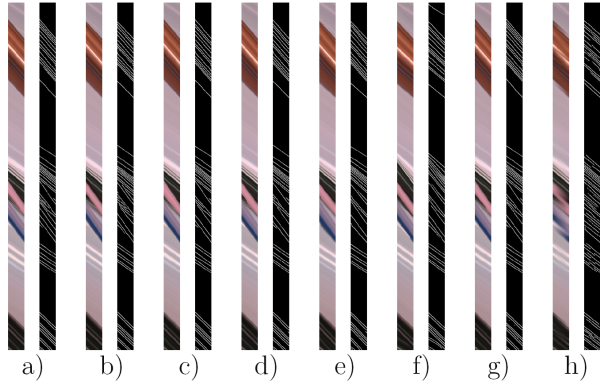


Fig. 2. 7 different distortions on a single EPI slice from the FVV dataset[6] and their corresponding edge maps. a)Reference, b)3D-HEVC, c)Multi-View Video, d)HEVC Test Model, e)JPEG200, f)Lossless edge depth coding, g) Color channel correlation based coding, h)Z-LAR-RP

we investigate existing LF image quality datasets and LF related distortions. In Section 4, we share our experiments and discuss the acquired results. Finally in Section 5 we conclude with our findings.

2. LIGHT FIELD RELATED DISTORTIONS

Different stages in LF visual chain may introduce a different type of distortion[2]. Even though there is numerous type of distortions in common with 2D images or videos, there exist some distortions specific to LF content as well. Recently, very few LF image quality datasets have been published. Some of them consider common distortions[5] such as gaussian blur, additive noise, etc. Some publicly available datasets consider LF specific distortions[6][7][8] such as compression, view synthesis related distortions. There is still a need for datasets who offers acquisition and display related distortions with the corresponding subjective scores.

In Free Viewpoint Video dataset(FVV)[6], seven different compression algorithms have been used to code depth maps. Coded depth maps are used to synthesize intermediate views to generate horizontal parallax videos with 50 frames. Generated videos are mirrored to 100 frames and then evaluated subjectively. Even though FVV dataset consists only horizontal parallax, it has high angular resolution and LF related distortions. In Fig. 2, an EPI slice from the "balloons" image is presented. 7 mid-level distortions are also presented with different annotations. Extracted edge maps from each distorted image are also displayed next to them. Each coding algorithm introduces a different type of distortion to the EPI image. Some algorithms produce more subtle distortions than other algorithms which result in higher perceived quality in the subjective experiment.

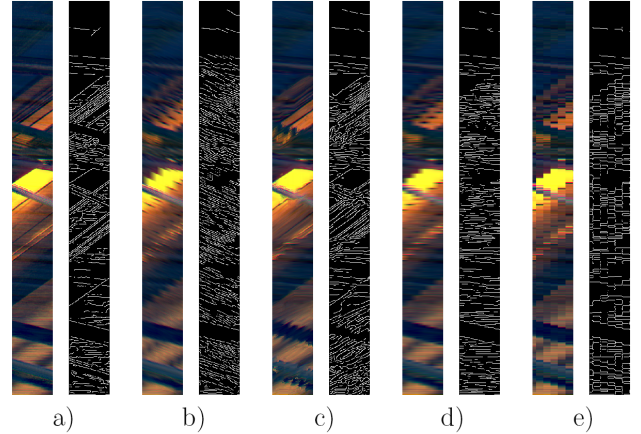


Fig. 3. 4 of 7 different distortions on a single EPI slice from the dataset by Adhikarla et al.[9] and their corresponding edge maps a)Reference, b)Depth map interpolation, c)Optical flow interpolation, d)Linear interpolation, e)Nearest neighborhood interpolation

In the dataset by Adhikarla et al.[9], LF resolution is chosen as $960 \times 720 \times 101$. Several types of distortions are considered, such as transmission, reconstruction, and display related distortions. As in FVV dataset, there is only horizontal parallax but with a high angular resolution. In Fig. 3, an EPI slice with 4 different reconstruction distortions along its non-distorted version is presented. All of the presented distortions are highly visible on EPIs

3. OBJECTIVE IMAGE QUALITY EVALUATION

Numerous metrics have been proposed to assess the quality of images and videos over the last several decades[3]. On the other hand, PSNR is still used by averaging PSNR on sub-aperture images to evaluate the performance of many LF processing tools.

This paper focuses on full-reference image quality assessment (FR-IQA), where the reference image exists along with the evaluated distorted image. There are mainly two strategies for FR-IQA model design. One is using a bottom-up approach where different stages in the Human Visual System(HVS) is modeled separately to evaluate the image quality. These stages include visual masking, contrast sensitivities, and just noticeable differences. This results in computationally complex models. The second approach is called top-down, which tries to estimate the whole HVS behaviour with broad mathematical assumptions. This provides a computationally simple evaluation procedure while sacrificing accuracy. In our work, we focused on existing top-down approaches, and those utilize the structural information on the images.

Non-shift edge-based ratio(NSER)[10] evaluates the im-

age quality by the edge differences between the reference and distorted images. Natural image counter evaluation(NICE)[11] operates similarly, where the comparison is made between morphologically dilated edge maps. Gradient magnitude similarity deviation (GMSD) works with standard deviation of the image gradients to evaluate the image quality. Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR)[12] calculates the image quality based on morphological wavelet decomposition. PSNR is calculated between each band of decomposed reference and distorted image pairs.

GMSD works on directional image gradients to produce the final quality score. Prewitt filters along horizontal and vertical directions are used to convolve the image. They are defined as:

$$h_x = \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix}, h_y = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \end{bmatrix}$$

Both reference(r) and distorted(d) images are convolved with filters above. Gradient magnitudes at each pixel location for both images are calculated as follows:

$$m_r(i) = \sqrt{(r \otimes h_x)^2(i) + (r \otimes h_y)^2(i)} \quad (1)$$

$$m_d(i) = \sqrt{(d \otimes h_x)^2(i) + (d \otimes h_y)^2(i)} \quad (2)$$

These gradient magnitudes for each map then used to calculate gradient magnitude similarity(GMS) map as follows:

$$GMS(i) = \frac{2m_r(i)m_d(i)}{(m_r)^2(i) + (m_d)^2(i)} \quad (3)$$

Standard deviation pooling is used on GMS to calculate the final quality score. Convolution of the input images with Prewitt filters reveals the structural information on the image. This explains our motivation to choose GMSD to experiment.

NICE is slightly simpler, and it relies on edge detection algorithms. Both reference and distorted images are inputted to an edge-detection algorithm. Extracted edge maps are then dilated with a plus-sign kernel. Non-zero elements in the XOR maps between the dilated image pair is then used to calculate the final quality score.

4. EXPERIMENTS

4.1. Non-Uniform Distortion Distribution on EPI Slices

In Fig. 4, quality evaluation of 3 different distortions on the same LF image is plotted. NICE[11] metric is used as a quality score. A lower number represents better quality. We have

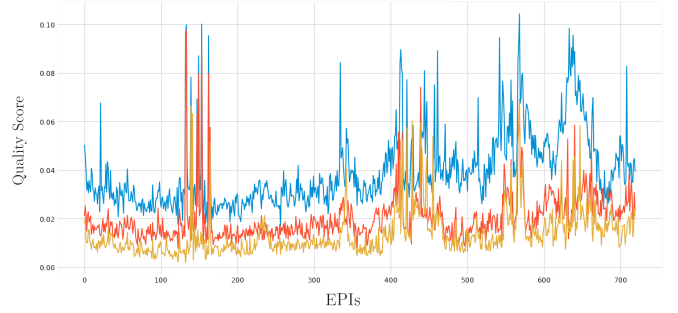


Fig. 4. Each line in the plot represents a different type of distortions on the image(“balloons” from FVV dataset). Every point in the X axis represents a single EPI slice. Quality scores are calculated based on NICE.

observed that the same EPI slices contribute to lower quality despite the fact that different source of distortions is introduced. EPI slices which belong to occluded regions contribute more to lower the perceived quality. This suggests that averaging quality scores across all EPI slices may not be the optimal solution for pooling. Additionally, lower quality EPI slices might be used as a selective process to identify distorted regions on the spatial domain.

4.2. Edge Detection on EPIs

	PCC	SROCC
Canny	0.7145	0.6552
Sobel	0.6457	0.5951
Prewitt	0.6450	0.6038

Table 1. NICE[11] metric performance on the dataset by Adhikarla et al[9]. Metric is incorporated with different edge detector algorithms[13][14].

Few edge detectors are commonly used in literature such as Canny[13], Sobel[14]. Even though most of the methods have been reviewed extensively in the literature[15], their performance on EPI for quality evaluation has not been investigated yet. We have experimented with three different edge detector algorithms and evaluated their performance on revealing distortions on EPIs. We incorporated different edge detectors on NICE[11] metric. In table 1, the effect of different edge detectors on the performance of NICE metric is reported. For evaluation, 2 different source images have been chosen from the dataset by Adhikarla et al[9], which resulted in 800 comparisons. Even though it does not reflect the full performance of the metric on the dataset, it is enough comparisons to evaluate the edge-detector performances for NICE algorithm.

4.3. Sub-aperture view vs EPI

	EPI	View
MW-PSNR	0.7698	0.7921
GMSD	0.7410	0.6715
NICE	0.5122	0.4310

Table 2. Pearson Correlations of MW-PSNR, GMSD, NICE metrics on EPIs and Sub-Aperture views. FVV dataset has been used for evaluation.

In table 2, Pearson correlation coefficients with subjective scores are presented for MW-PSNR[12], GMSD[16] and NICE[11] metrics on FVV dataset[6]. All metrics have been tested on the same dataset with both LF representations. We observed that GMSD and NICE provide higher correlation when it is used with EPI images as input. On the other hand, MW-PSNR provides a lower correlation with subjective opinions when used with EPI inputs. Low spatial resolution of EPIs provides only 2 scales for morphological wavelet decomposition, which is utilized by MW-PSNR metric. This results in low accuracy in terms of prediction for the MW-PSNR metric. Gradient provides structural information regarding the image, and in EPIs this information reveals the distortions in the angular domain. This results in higher accuracy for GMSD and NICE, which relies on filtering the image with Prewitt and Canny kernels.

4.4. Individual Distortions

	MW-PSNR		GMSD	
	EPI	View	EPI	View
3D-HEVC	0.4665	0.7137	0.3644	0.5110
Multi-View Video	0.9549	0.9484	0.8649	0.8746
HEVC Test Model	0.9584	0.8662	0.7974	0.7059
JPEG2000	0.8268	0.7687	0.7109	0.6643
Lossless edge depth	0.8016	0.8635	0.8288	0.8186
Color corr. based	0.7940	0.9099	0.9058	0.7315
Z-LAR-RP	0.7195	0.8147	0.7811	0.7317

Table 3. Pearson Correlations of MW-PSNR and GMSD metrics on EPIs and Sub-Aperture views for individual distortions available in FVV dataset.

Another experimentation has been made on individual distortions available on the FVV dataset. Some of the available distortions are more subtle on the EPIs. Both metrics, MW-PSNR and GMSD, has been used to evaluate the quality of individual distortions with both EPI slices and GMSD as an input. Pearson correlation coefficients of the evaluations are reported in table 3. Low visibility of 3D-HEVC coding distortions on the EPI slices results in lower accuracy of the quality evaluation for both metrics when used with EPI slices. Multi-View video coding distortions is very easy for both of

the metrics to predict, and there are no statistically significant differences between different input types. We observed that HEVC test model coding and JPEG2000 coding results in more visible distortions on many of the EPI slices. This provides a better correlation with subjective scores for both metrics on EPIs.

5. CONCLUSION

We have run several experiments to evaluate existing metric performances on EPI representations. In our experiments, we have adopted two different metrics (NICE, GMSD) which rely on structural information such as gradient and extracted edge information. Although both metrics are not designed for EPI domain, we have observed that their performances increased compared to sub-aperture view performances.

On the other hand, metrics which utilize multi-scale decomposition cannot fully benefit from the low spatial resolution of EPIs. Multi-scale decomposition algorithms are designed to separate the image into several components based on spatial frequency and orientation ranges. This separation does not provide any information in the angular domain because an EPI slice is a line across an angular direction and does not contain any information regarding its neighborhood pixels in the opposite direction. In order to benefit from EPI representations, structural feature extraction, which is not based on natural image statistics is required. Currently, simple edge extraction algorithms do not provide a satisfactory correlation with subjective evaluations. Experimental results suggest that with a carefully designed feature extraction scheme, EPI representations might be useful for identifying numerous type of angular domain distortions.

Additionally, when both horizontal and vertical parallax is taken into consideration, the same pixel values need to be evaluated twice. This doubles the computational complexity. A selective mechanism on the spatial domain to locate the occluded areas can be beneficial. This may allow identifying EPI slices which contribute most to distortions on the angular domain.

Acknowledgment

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 765911 (RealVision)

6. REFERENCES

- [1] Michael Faraday Esq. D.C.L. F.R.S., “Liv. thoughts on ray-vibrations,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 28, no. 188, pp. 345–350, 1846.

- [2] Gaochang Wu, Belen Masia, Adrin Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, pp. 1–1, 08 2017.
- [3] A. de Angelis, A. Moschitta, F. Russo, and P. Carbone, "Image quality assessment: an overview and some metrological considerations," in *2007 IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement*, July 2007, pp. 47–52.
- [4] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2 PART 1, pp. 165–182, 2011.
- [5] Roopak R. Tamboli, Balasubramanyam Appina, Sumohana Channappayya, and Soumya Jana, "Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display," *Signal Processing: Image Communication*, 2016.
- [6] Dragana Sandi-Stankovi, Federica Battisti, Dragan Kukolj, Patrick Le Callet, and Marco Carli, "Free view-point video quality assessment based on morphological multiscale metrics," 06 2016.
- [7] Irene Viola and Touradj Ebrahimi, "Valid: Visual quality assessment for light field images dataset," *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–3, 2018.
- [8] L. Shi, S. Zhao, W. Zhou, and Z. Chen, "Perceptual evaluation of light field image," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 41–45.
- [9] Vamsi Kiran Adhikarla, Marek Vinkler, Denis Sumin, Rafal Mantiuk, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk, "Towards a quality metric for dense light fields," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3720–3729, 2017.
- [10] M. Zhang, X. Mou, and L. Zhang, "Non-shift edge based ratio (nser): An image quality assessment metric based on early vision features," *IEEE Signal Processing Letters*, vol. 18, no. 5, pp. 315–318, May 2011.
- [11] D. M. Rouse and S. S. Hemami, "Natural image utility assessment using image contours," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 2217–2220.
- [12] Dragana Sandi-Stankovi, Dragan Kukolj, and Patrick Le Callet, "Dibr-synthesized image quality assessment based on morphological multi-scale approach," *EURASIP Journal on Image and Video Processing*, vol. 2017, 07 2016.
- [13] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov 1986.
- [14] Nick Kanopoulos, Nagesh Vasanthavada, and Robert L Baker, "Design of an image edge detection filter using the sobel operator," *IEEE Journal of solid-state circuits*, vol. 23, no. 2, pp. 358–367, 1988.
- [15] G.T. Shrivakshan and Chandramouli Chandrasekar, "A comparison of various edge detection techniques used in image processing," *International Journal of Computer Science Issues*, vol. 9, pp. 269–276, 09 2012.
- [16] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, Feb 2014.