Epipolar Shift Compensated Light Field Video Quality Metric

Nusrat Mehajabin, Dan Jin, Rui Yao, Thomas Dykstra, Mahsa Pourazad, Panos Nasiopoulos
Electrical and Computer Engineering
University of British Columbia
Vancouver, BC, Canada
e-mail: {nusratm, pourazad, panosn}@ece.ubc.ca,{dysjin, rruiyao, dykstr01}@student.ubc.ca

***Abstract*—Traditional video quality metrics are unsuitable for the Light Field (LF) video content as these metrics do not account for the structural and angular relationships among the various viewpoints found in LF content. While there is a growing amount of light field video content being produced for increasing application demand, there is currently no standardized objective method for measuring the quality of these videos. In this paper, we propose an objective quality metric for evaluating the spatial and angular quality of light field video content. We achieve this goal by leveraging the Epipolar Plane Images (EPI) along the horizontal, vertical, and diagonal views, on which we perform statistical analysis to determine the quality of the LF content. We also present our results and discuss our findings and future work on this topic.**

Keywords -Light Field; SSIM; PSNR; Objective Quality Metric; Epipolar Plane Image.

I. INTRODUCTION

Light field video is an interesting new technology that holds great promise. By recording video using multiple cameras [1] that are all pointing at the same scene, one can perform operations, such as changing perspective, "peeking" around objects in the foreground to see some of the background [2], and changing image focus in post-production [3], etc. Generating, transmitting, and rendering Light Field (LF) content is a growing field of research [4]. To meet different application demands, the industry must compress [5], synthesize, calibrate [6], and perform other operations on the original content. Therefore, there is a need for a quality metric to make sure that the processed content preserves the original spatial and angular relationships.

Prior to the growth of light field research and 3D imaging in general, a lot of research has been done on evaluating the quality of 2D images. Quality evaluations can be divided into two classes - subjective and objective methods. Subjective methods are valuable because perceived quality is ultimately intended to reflect human perception and the human visual system is often more sensitive to certain aspects of quality than others. There are different procedures defined for performing experiments to evaluate subjective quality, such as Single Stimulus Continuous Quality Evaluation (SSCQE) and Double Stimulus Continuous Quality-Scale (DSCQS) specified by the International Telecommunication Union – Radiocommunication (ITU-R) standard for images [7]. However, obtaining meaningful results from subjective experiments is expensive and time-consuming because the test environment and procedure must be consistent. Objective methods, on the other hand, can be automated. Examples of well-known methods include the Video Quality Metric (VQM) [8], which takes into consideration additional aspects of the human visual system and statistical methods such as Peak-Signal-to-Noise-Ratio (PSNR) and Structural Similarity Index (SSIM). Objective metrics can be categorized into Full Reference (FR), Reduced Reference (RR) and No Reference (NR) where FR metrics rely on complete information from a reference image and NR metrics derive quality from inherent attributes of the image [7].

There has been previous research to evaluate how the traditional 2D image quality metrics apply to 3D and light field content (which will be further described in Section 2). In terms of objective metrics, the typical approach has been to measure the PSNR or SSIM between each view of the reference light field image and the corresponding view in the processed image, followed by taking the average for the global metric [5]. However, light field content provides a lot of additional information, including structure across views, depth, and perspective. It is, therefore, worthwhile to create a quality metric that takes these inherent properties of light field content into consideration. In this paper, we propose an LF video quality metric that computes the horizontal, vertical, and diagonal EPIs of the reference and processed content to measure spatial and angular consistency. The major contribution of this paper is the inclusion of diagonal EPIs in the equation, which enables us to measure the angular consistency not only across horizontal and vertical views but also factors in the subtle angular changes introduced throughout the content.

The rest of the paper is organized as follows. Section II discusses the features and applicability of the state-of-the-art LF quality metrics. Section III presents the proposed method in detail. In Section IV, we discuss the experimental results. We conclude the paper in Section V.

II. RELATED WORKS

LF images are subject to a wide variety of distortions during acquisition, processing, compression, storage, transmission, and rendering; any of these steps may result in

 ISBN: 978-1-61208-942-3

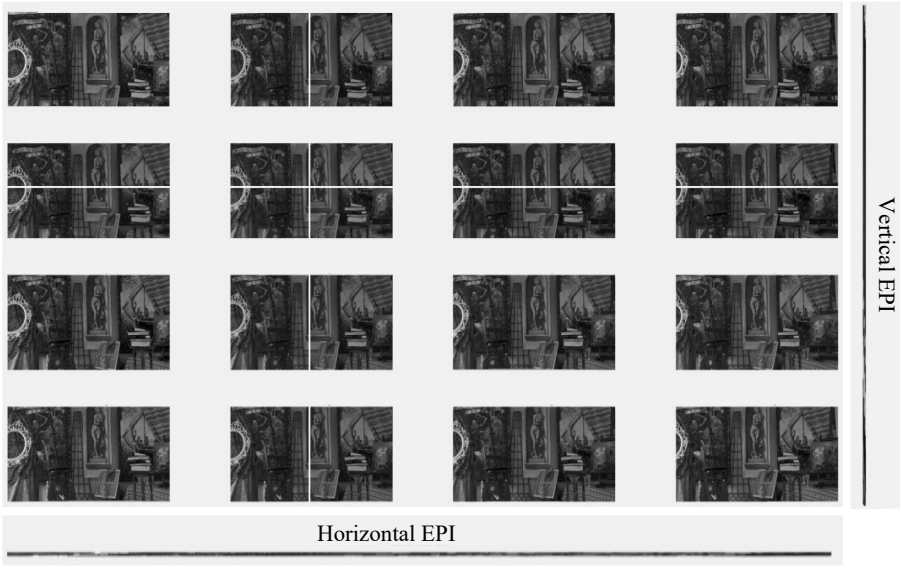

Figure 1. Horizontal and vertical EPIs of 'Painter' LF test sequence

visual quality degradation. The rapidly developing LF technology and consumer interest are pushing the need for objective quality evaluation of such contents.

One of the first works on LF quality evaluation by Adhikarla et. al [9] applies the traditional video metrics on individual light-field images and then averages the scores of overall images. They used Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure SSIM2D, which is widely used on 2D images, its extensions to angular domains – SSIM2D*1D, SSIM3D, and High Dynamic Range – Visible Difference Predictor (HDR-VDP-2), which stands out among perception-based quality metrics, the NTIA General Model for Video Quality Metric – VQM and the stereoscopic image quality metric – SIQM. To capture the full range of stereo quality metrics, they also included a stereoscopic video quality metric STSDLC. Another metric for the multiview video is MPPSNR, which computes the multi-resolution morphological pyramid decomposition on the reference and test images. Other metrics such as HDR-VDP-2, GMSD, STSDLC, and VQM perform well when comparing a distorted light field to a densely sampled reference LF. However, when a dense light field is not available, which is the case in camera array acquired LF, the usage of these metrics for quality assessment is not justified.

More recent works in LF quality metric domain, such as FR LFI-QA [10], measure the gradient magnitude similarity between the reference and processed content. Other methods [11] rely on depth maps. The accuracy of depth estimation, and consequently depth-map, depends on the method used and even with robust methods depth estimation is not always accurate. Hence, the proposed quality metric suffers inaccuracies too. [12] proposes an NR metric using horizontal and vertical EPIs to measure angular consistency. However, none of these methods fully exploit the structural and angular properties of LF. For example, for any given LF view [10] and [12] only consider the horizontal and vertical EPIs leaving the existing diagonal correlation underutilized. For dense LF content, this does not make a big difference as eventually the ray space is traversed twice. Though the relation is being indirectly factored into the quality metric using vertical and horizontal EPIs serially, for sparse LF content this indirect method cannot represent the quality accurately. Because of the wide baseline of the cameras and sparsely positioned cameras, we need additional scanning of the ray space to represent the quality accurately. Therefore, these methods are not suitable for sparse LF content.

III. PROPOSED QUALITY METRIC

In order to design an LF quality metric for camera array-based (sparse) content, we leverage the horizontal, vertical, and diagonal EPIs. This way, we cover the ray space multiple times, and the quality metric can detect even the subtlest inconsistencies in the processed LF.

A. LF and EPIs

LF is described using a standard two-plane parameterization. Rays are defined using two parallel planes Π and Ω. The first plane Ω denotes image coordinates (x, y) ∈ Ω. The second plane Π contains the focal points (s, t) ∈ Π of all cameras. An entire 4D light field can thus be described by a function

$$R\ (s,t,x,y) \rightarrow L(s,t,u,v) \quad (1)$$

where $L(s,t,u,v)$ defines the intensity of the corresponding ray defined by the intersection (u,v) with the image plane and (s,t) with the focal plane, respectively. Furthermore, (u,v) can be treated as the spatial, and (s,t) can be treated as the angular resolution of the LF. Hence, there will be $s \times t$ sub-aperture views each having the resolution of $u \times v$. These sub-aperture views are slightly shifted from each other depending on the distance between the cameras. These disparities among the views can be estimated on the 2D slices

Figure 2. Diagonal EPI for views I_{1,1} ,I_{2,2} ,I_{3,3} , and I_{4,4} . The three figures on the left show an overlay of each pair of views along the diagonal and the matching Scale Invariant Feature Transform (SIFT) features that were detected.

$\sum_{t^*v^*} L$ from the 4D light field structure. This is achieved by setting t to a fixed value $t^*$and v to a fixed value v^* hence generating the EPIs.

$$S_{t^*v^*}: \sum_{t^*v^*} L \rightarrow R \quad (2)$$

$$(u,s) \rightarrow S_{t^*v^*}(u,s) := L(s,t^*,v^*,u) \quad (3)$$

Other slices with different fixed coordinates such as s and u, are defined analogously. Traditionally, EPIs have been used to estimate the depth and synthesize novel views. However, the EPIs can also be used to measure the spatial quality by comparing the original EPI and the processed EPI. The slopes created due to disparity among views can be measured to determine the angular quality of LF. To make the metric robust, we include the complete ray space multiple times in the form of horizontal, vertical, and diagonal EPIs.

B. Generating Horizontal, Vertical & Diagonal EPIs

The horizontal and vertical EPIs are generated as described in Shi et. al. [12]. Each view in the light field image is denoted as $I_{\{u,v\}}(s,t)$ where (s,t) is the spatial coordinate and (u,v) is the angular coordinate.

Each horizontal EPI is denoted as $E_{\{v^*,t^*\}}(s,u)$ where v^* and t^* are the fixed angular and spatial coordinates. Each row of the horizontal EPI contains the row of pixels from view $I_{\{u^*,v^*\}}(s,t^*)$. Similarly, the vertical EPI is denoted as $E_{\{u^*,s^*\}}(t,v)$. Each row of the vertical EPI contains the column of pixels from view $I_{\{u^*,v^*\}}(s^*,t)$.

Figure 1 shows an example of a horizontal and a vertical EPI. A total of 1088×4 horizontal EPIs and 2048×4 vertical EPIs are generated for each frame (all the views) of the LF.

The challenge with generating diagonal EPIs is to find the corresponding pixels in the diagonally aligned view. To determine which row of pixels in each view belongs to a particular diagonal EPI, the vertical offset is required. Therefore, we need to compensate for the shift experienced by the diagonal translation. For every pair of views along the diagonal, we detect all matching Scale Invariant Feature Transform (SIFT) features [13]. The matching features are used to calculate the average vertical offset and are used to determine the row of pixels in the next diagonal view to include in the EPI. To the best of our knowledge, this is the first work that uses SIFT to compensate for the shift in EPIs and create diagonal EPIs to measure angular consistency of LF content. Figure 2 shows an example of the matching SIFT features corresponding to each pair of views along the diagonal that contains $I_{\{1,1\}}, I_{\{2,2\}}, I_{\{3,3\}}$, and $I_{\{4,4\}}$. For the frame (all views) depicted in Figure 2, we have generated 28×2048 diagonal EPIs.

After generating all the EPIs, we have traversed the ray space three times and covered all the neighboring views of each view at least once. In this way, the quality metric can register even the subtle inconsistencies in the angular domain.

C. EPI Similarity

In order to quantify spatial quality, we compare the average PSNR and SSIM between the horizontal, vertical, and diagonal EPIs of the original and processed LF. This differs from the traditional quality metrics for LF content where the average PSNR or SSIM is calculated between each view separately and averaged to report quality. This method also provides insight into how the spatial relationship is maintained across the LF and applications requiring camera positions from content.

D. EPI Gradient & Average Kurtosis

We measure the angular distortion or deterioration using the pixel-wise gradient from horizontal, vertical, and diagonal EPIs. For each horizontal, vertical, and diagonal EPI, the gradient at each pixel is calculated using the Sobel

 ISBN: 978-1-61208-942-3

TABLE I. LF SPATIAL QUALITY FROM EPI SIMILARITY

PSNR					
	QP45	**QP40**	**QP35**	**QP30**	**QP25**
Horizontal EPI	30.64	33.14	35.64	37.89	39.95
Vertical EPI	30.91	33.43	35.90	38.07	40.04
Diagonal EPI	29.21	31.37	32.95	33.80	35.39
Average PSNR	**30.26**	**32.65**	**34.83**	**36.59**	**38.46**
SSIM					
	QP45	**QP40**	**QP35**	**QP30**	**QP25**
Horizontal EPI	0.89	0.94	0.96	0.97	0.98
Vertical EPI	0.88	0.93	0.95	0.97	0.98
Diagonal EPI	0.85	0.90	0.92	0.94	0.96
Average SSIM	**0.88**	**0.92**	**0.95**	**0.96**	**0.97**

gradient operator. A histogram of the gradients is generated, and we use the average kurtosis as a metric to describe the amount of distortion introduced from compression.

IV. EXPERIMENTS AND EVALUATION

LF content can experience degradation from different operations, such as compression, calibration, super-resolution, and other operations. In this paper, we validate the proposed quality metrics by using LF videos compressed at different Quantization Parameter (QP) levels of 25, 30, 35, 40, and 45 using the MV-HEVC, Multi-view extension of High Efficiency Video Coding [14], to compress light field content. We use camera array-based LF test sequences [6]. We report the results for the 'Painter' test sequence in this paper. Other test sequences show consistent performance.

TABLE II. KURTOSIS OF GRADIENT DIRECTION HISTOGRAM

	Reference	**QP45**	**QP40**	**QP35**	**QP30**	**QP25**
Horizontal EPI	1.36	1.23	1.26	1.27	1.29	1.30
Vertical EPI	1.40	1.24	1.26	1.28	1.29	1.31
Diagonal EPI	1.37	1.24	1.26	1.28	1.29	1.31
Average Kurtosis	**1.38**	**1.24**	**1.26**	**1.28**	**1.29**	**1.30**

Table 1 contains the average PSNR and SSIM values between all horizontal, vertical, and diagonal EPIs from the compressed LF with respect to the original video. The results indicate that the EPI similarity correlates with different amounts of compression.

The pixel-wise gradient calculation is presented in Figure 3 in 10° bins (36 bins in total). We can adjust the number of bins depending on the level of accuracy desired. We can quantify the results of gradient direction using kurtosis. Table 2 contains the average kurtosis values for all horizontal, vertical, and diagonal EPIs. The results show that with increasing QP levels or more compression, the kurtosis of the gradient histogram decreases. Figure 3 illustrates this relationship. The gradient histograms show the count of

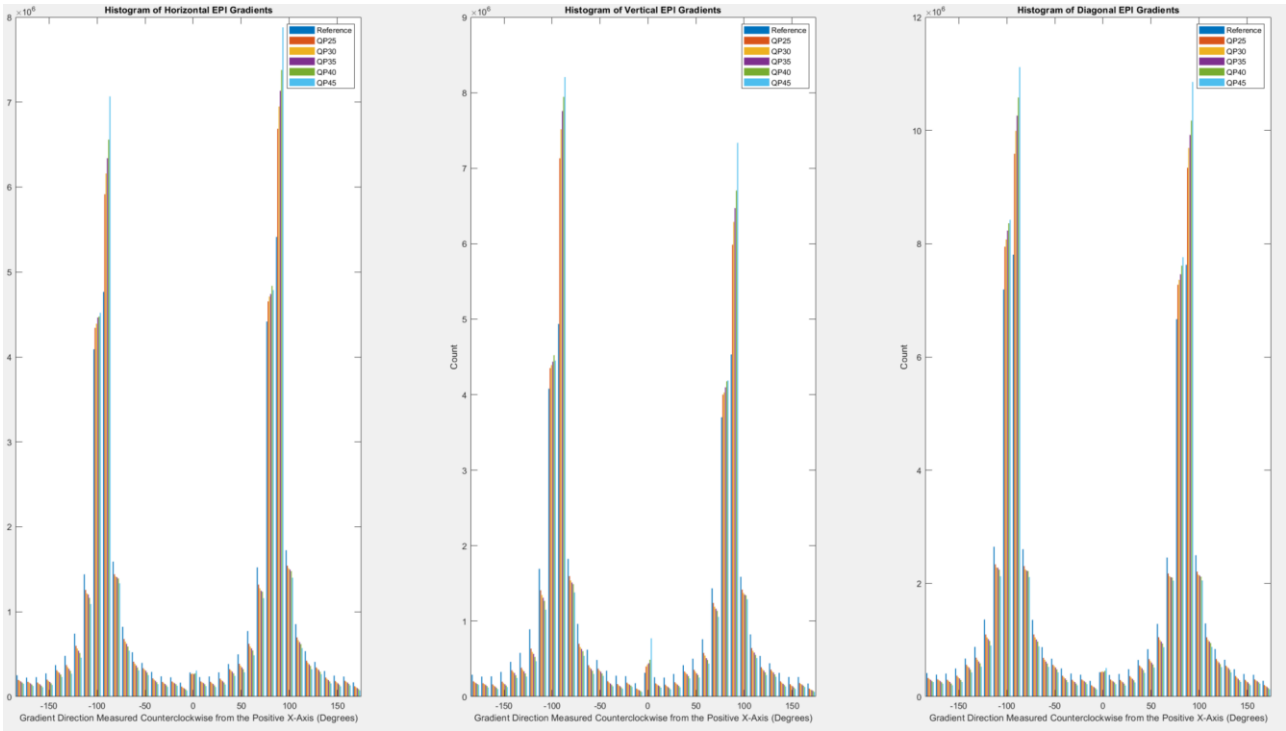

Figure 3. Gradient histograms of the horizontal, vertical, and diagonal EPIs. The histograms show that the gradient direction is more concentrated at ±90° for higher QP values.

gradients from all EPIs and are presented in bins of 10°. The angle is measured counterclockwise from the positive $x-axis$. For each 10° bin, the histograms show the gradient counts at each QP level. The histograms show that with increasing QP levels, the gradients become more concentrated at $\pm 90°$, which is likely a result of the stepwise artifacts introduced in the EPI with more compression. Since the histogram is centered at 0°, it makes sense for the kurtosis to decrease with more compression since the tails of the histogram at $\pm 90°$ are larger relative to the center.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed an LF quality metric using horizontal, vertical, and diagonal EPIs. Our major contribution was the SIFT-assisted diagonal EPI translation for scanning the diagonal ray space of the LF. Such a quality metric is useful in determining the spatial and angular consistency of processed LFs with the original. We applied our metrics to compressed LFs and found the metric can accurately describe the quality of a sparse LF. An important next step would be further validation of our metrics by evaluating LFs that have undergone other distortions such as super-resolution, calibration, etc. We also intend to correlate our results with subjective testing scores.

REFERENCES

[1] M. Levoy and P. Hanrahan, "Light Field Rendering," In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, 1996, Aug 1, pp. 31-42.

[2] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-held Plenoptic Camera," 2005 (Doctoral dissertation, Stanford University) .

[3] C. Zhang, G. Hou, Z. Zhang, Z. Sun, and T. Tan, "Efficient auto-refocusing for light field camera," Pattern Recognit., vol. 81, pp. 176–189, Sep. 2018, doi: 10.1016/j.patcog.2018.03.020.

[4] C. Conti, L. D. Soares, and P. Nunes, "Light field coding with field-of-view scalability and exemplar-based interlayer prediction," IEEE Trans. Multimed., vol. 20, no. 11, pp. 2905–2920, 2018, doi: 10.1109/TMM.2018.2825882.

[5] N. Mehajabin, M. T. Pourazad, and P. Nasiopoulos, "An Efficient Pseudo-Sequence-Based Light Field Video Coding Utilizing View Similarities for Prediction Structure," IEEE Trans. Circuits Syst. Video Technol., 2021, pp. 1-16, doi: 10.1109/TCSVT.2021.3092282.

[6] N. Sabater et al., "Dataset and Pipeline for Multi-view Light-Field Video," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2017-July, pp. 1743–1753, 2017, doi: 10.1109/CVPRW.2017.221.

[7] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality analysis for 3D video using 2D video quality models," IEEE Trans. Consum. Electron., vol. 54, no. 4, pp. 1969–1976, 2008, doi: 10.1109/TCE.2008.4711260.

[8] P. Joveluro, H. Malekmohamadi, W. A. C. Fernando, and A. M. Kondoz, "Perceptual video quality metric for 3D video quality assessment," 3DTV-CON 2010 True Vis. - Capture, Transm. Disp. 3D Video, pp. 1–4, 2010, doi: 10.1109/3DTV.2010.5506331.

[9] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, and P. Didyk, "Towards a quality metric for dense light fields," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2, 2017, pp. 58–67.

[10] Y. Fang, K. Wei, J. Hou, W. Wen, and N. Imamoglu, "Light Filed Image Quality Assessment by Local and Global Features of Epipolar Plane Image," 2018 IEEE 4th Int. Conf. Multimed. Big Data, BigMM 2018, pp. 1-6, doi: 10.1109/BigMM.2018.8499086.

[11] P. Paudyal, F. Battisti, S. Member, M. Carli, and S. Member, "Reduced Reference Quality Assessment of Light Field Images," IEEE Trans. Broadcast., vol. 65, no. 1, pp. 152–165, 2019.

[12] L. Shi, W. Zhou, Z. Chen, and J. Zhang, "No-Reference Light Field Image Quality Assessment Based on Spatial-Angular Measurement," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 11, pp. 4114–4128, 2020, doi: 10.1109/TCSVT.2019.2955011.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," Proc. IEEE Int. Conf. Comput. Vis., vol. 2, pp. 1150–1157, 1999, doi: 10.1109/iccv.1999.790410.

[14] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro, and Y. K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 1, pp. 35–49, 2016, doi: 10.1109/TCSVT.2015.2477935.

 ISBN: 978-1-61208-942-3