



HAL
open science

NO-REFERENCE QUALITY EVALUATION OF LIGHT FIELD CONTENT BASED ON STRUCTURAL REPRESENTATION OF THE EPIPOLAR PLANE IMAGE

Ali Ak, Suiyi Ling, Patrick Le Callet

► **To cite this version:**

Ali Ak, Suiyi Ling, Patrick Le Callet. NO-REFERENCE QUALITY EVALUATION OF LIGHT FIELD CONTENT BASED ON STRUCTURAL REPRESENTATION OF THE EPIPOLAR PLANE IMAGE. The 1st ICME Workshop on Hyper-Realistic Multimedia for Enhanced Quality of Experience, Jul 2020, London, United Kingdom. 10.1109/ICMEW46912.2020.9105975 . hal-02562117

HAL Id: hal-02562117

<https://hal.science/hal-02562117>

Submitted on 4 May 2020

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.

NO-REFERENCE QUALITY EVALUATION OF LIGHT FIELD CONTENT BASED ON STRUCTURAL REPRESENTATION OF THE EPIPOLAR PLANE IMAGE

Ali Ak, Suiyi Ling, Patrick Le Callet

Rue Christian Pauc, 44306, Nantes, France, firstname.lastname@univ-nantes.fr

ABSTRACT

As an emerging technology, Light Field (LF) has gained ever-increasing importance in the domain. In order to provide guidance for the development of perceptually accurate Light Field (LF) processing tools and supervise the entire streaming system, robust perceptual quality assessment metrics are required. Especially, No-Reference (NR) metrics are preferable to compare LF with different angular resolutions. Some metrics have been developed by extending commonly used 2D image quality metrics to the 4D LF domain with angular consistency terms. Nonetheless, although these models consistently show slightly improved performance, most of them are limited in evaluating the quality of LF using the sub-aperture views with additional terms on the angular domain. There is an evident lack of reliable quality metrics that are tailored to LF content. To remedy this lack, we propose a NR quality metric for LF contents based on representing EPI with structural descriptors, including the Histogram of Gradients and the Convolutional Sparse Coding based descriptors. The primary motivation resides in our observations that (1) LF related distortions on the angular domain are highly noticeable on the Epipolar Plane Image representations (EPI); (2) most of the distortions in EPI are structure-related. Extensive experiments on the MPI-LFA [1] LF image quality dataset demonstrate that our method provides competitive performance with the state-of-the-art NR image quality metrics.

Index Terms— Light field, Image quality assessment, No-reference, Histogram of oriented gradients, Convolutional sparse coding, Structural representation of EPI

1. INTRODUCTION

The search for ultra-realistic viewing experience has brought much focus to Light Field (LF) imaging in the last decades. As one of the very first studies, Michael Faraday *et al.* used 7D plenoptic function to define a complete light field [2]. It represents the radiance received on a point from each direction in 3D space at any given time for every wavelength. Lately, it has been simplified to a four-dimensional function. 4D function defines the emitted light from every point of the first plane to every point on the second plane in a two-plane

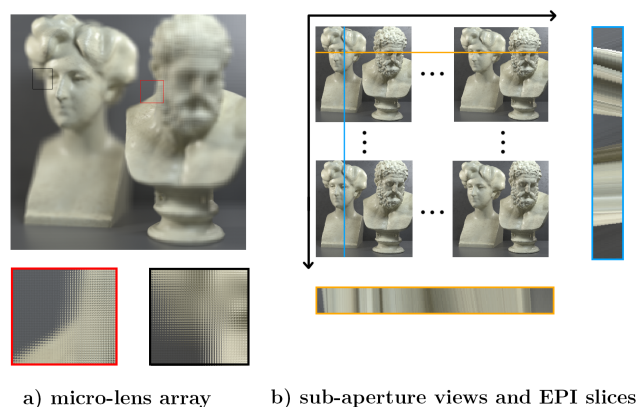


Fig. 1. Different representations of the same Light Field Images. LF image is from WIN5-LID dataset [3].

representation on the 3D space.

Sub-aperture view, EPI, micro-lens array are different types of representations for LF images. In Fig. 1, examples of each representation over the same image are presented. While sub-aperture views are better at displaying spatial domain information, EPI representations reveal the angular domain information. It is expected to have continuous pixel values along the lines and gradients on each EPI.

From acquisition to display, LF content needs to go through several processing steps. Each processing step might introduce various types of distortion. The effect of distortions on the perceived quality of the produced LF image needs to be measured in order to develop perceptually satisfying processing tools. Although subjective experiments are the most accurate way of assessing the quality of an image, it is time-consuming and expensive. Thus, objective image quality metrics are required to predict the perceived quality of the images. Objective image quality metrics can be further categorized as full-reference and no-reference metrics. While full-reference metrics have access to non-distorted images to predict the quality of an image, no-reference metrics predict the quality of the image by only distorted images.

In LF imaging, artifacts generally appear on the angular domain. For this reason, many of the published LF image quality datasets such as MPI-LFA [1] or WIN5-LID [3] focus on distortions which result in angular domain inconsistencies.

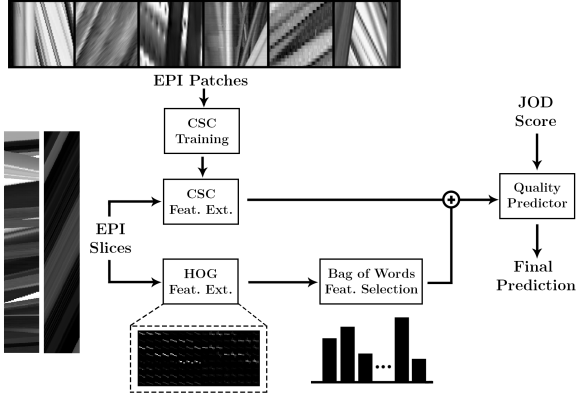


Fig. 2. Diagram of the proposed algorithm

Distortions in the angular domain are highly visible on EPI representations [4]. They are presented as broken lines and gradient structures, e.g., structure-related distortions, within EPI slices.

Based on the discussions above, in this work, we propose a no-reference LF image quality metric based on quantifying the structure-related distortions within EPI. On the one hand, Histogram of Oriented Gradient (HOG) descriptors [5] is extracted, and a HOG based bag-of-words codebook is trained to represent the overall structural statistics of EPI. On the other hand, to further measure the amount of non-natural structures in EPI, a Convolutional Sparse Coding (CSC) codebook is trained on a set of EPI patches that contain significant LF related distortions. After extracting the features using the learned codebook, the amount of non-natural structural distortions within EPI is then measured. With the higher-level structural representations, the quality scores are predicted utilizing Support Vector Regression (SVR). The proposed model has been evaluated on the MPI-LFA dataset [1] and provides competitive results with state of the art. The rest of the paper is organized as follows. In Sec. 2 related LF image quality works are presented. In Sec. 3 our proposed metric is explained in details. Experimental setup and results are given in Sec. 4. Finally, we discuss our findings and conclude in Sec. 5.

2. RELATED WORK

Image quality assessment has been studied extensively in the previous decades [6]. Thus, there are numerous 2D objective image quality metrics. Some of these metrics have been extended to LF domain [1]. Even though some of the extended metric predictions have a high correlation with the subjective scores, there is still room to improve.

Moreover, there have been several attempts to develop specific LF objective quality metrics for LF images. Tamboli *et al.* [7] proposed a full-reference image quality evaluation algorithm where both spatial and angular quality were

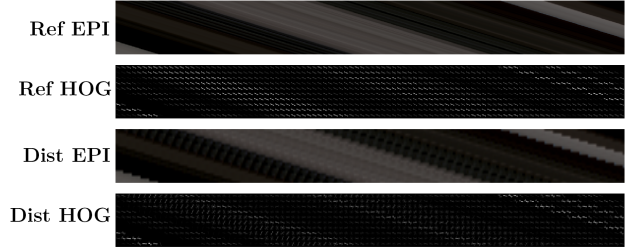


Fig. 3. Reference and Distorted EPI slices are presented with their corresponding HOG Feature Maps. High variance and low magnitude of HOG features on distorted EPI can be observed.

measured and combined to predict the final quality score. Authors have used Wavelet decomposition to predict the spatial quality component. Optical flow has been used to predict the angular quality component.

Shan *et al.* [8] proposed a no-reference LF image quality metric. The proposed model is composed of a color component, texture component, and a depth component. These three components are used to fit a support vector regression model to predict the final quality score.

Lately, BELIF [9] has been proposed based on quantifying distortions on Cyclopean images. In their framework, spatial features are extracted with the Tucket decomposition, while angular features are extracted from the structural similarity between the cyclopean image pairs. Finally, these features are used to fit a regression model to predict the final quality score. One disadvantage of this approach is that the model only considers horizontally adjacent pairs of views to evaluate the angular consistency.

3. PROPOSED MODEL

The proposed model utilizes CSC and histogram of oriented gradients to represent the structures in the EPI slices of LF contents. Extracted features are fed to a regression model where the final score is predicted. The general diagram of the proposed model is shown in Fig. 2. 360 EPI slice patches with obvious LF distortions are extracted from the MPI-LFA dataset [1] and fast convolutional sparse coding implementation [10] has been used to learn a dictionary that is composed of non-natural structure elements. Then, the learned dictionary is used to extract representative features from distorted EPI slices. More information is given in Sec. 3.2. Additionally, HOG features have been extracted from EPI slices. Afterward, the structural statistics of EPI are represented utilizing a Bag of Words (BoG) model constructed with the HOG features. Detailed explanations regarding the procedure of HOG feature extraction, descriptions are given in Sec. 3.1. In the final stage, extracted HOG and CSC features are concatenated and inputted to a regression model, which are explained in detail in Sec. 3.3. The regression model is trained with a 1000-fold cross-validation. Details are given in Sec. 4.2.

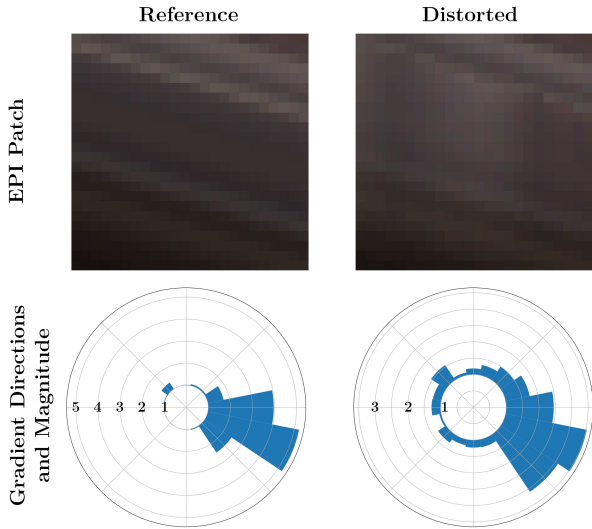


Fig. 4. Patches from Reference and Distorted LF EPI slices are presented in the first row. Corresponding HOG directions and magnitudes are plotted on the circular histogram.

3.1. Histogram of Oriented Gradients based Bag Of Words Model

HOG has been used as a feature descriptor for numerous computer vision tasks such as human detection [5]. It reveals the underlying structural information on the image and could be employed to quantify non-natural structure-related distortions in the task of image quality assessment [11]. An example is depicted in Fig. 3, where one distorted EPI slice and its corresponding reference are presented. By comparing visualized HOG maps, it can be observed that the structure-related distortions lead to higher variance in HOG feature blocks. To have a closer comparison, we further show the circular histogram of directional gradients of a pair of EPI $24 \times 24px$ sized patches in Fig. 4. As observed, the gradients of the distorted patch are spread out in diverse directions with lower magnitudes, while the ones of the reference patch are gathered in fewer directions ending out in higher magnitudes of each bin (gradient direction).

Based on the discussion above, HOG is exploited in this work to represent and quantify distorted EPI patches. Given an EPI slice, it is first divided into a set of 10×10 local blocks. Then HOG descriptors are calculated block-wise. More specifically, directional gradients of each pixel are calculated with a simple kernel $[-1, 0, 1]$, and a signed histogram with 16 orientations that covers $(0, 360)$ degrees is further employed to represent each local block.

After getting the local HOG features, a Bag Of World (BOW) dictionary is trained using the extracted HOG features to obtain a global HOG based representation for the entire EPI slice. Formally, a set of extracted HOG feature vectors

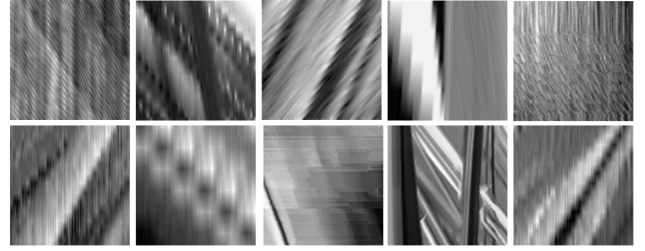


Fig. 5. Examples of patches selected for the CSC training. In total 360 patches have been used to represent 6 type of LF related distortions from the MPI-LFA dataset [1].

f_{HOG} is clustered into K clusters $\{c_1, \dots, c_K\}$ using the K-means algorithms. With the obtained BOW, each EPI slice image could be then encoded as a higher-level structural representation $f_{HOG-BOW} = \{\theta_1, \dots, \theta_k, \dots, \theta_K\}$, where θ_k is defined as

$$\theta_k = \frac{\sum_i^{n_p} \mathbf{1}(p_i) \in c_k}{n_p}, \quad (1)$$

$\mathbf{1}(b)$ is an indicator function that equals to 1 if the specified binary clause b is true, n_p is the number of patches within the EPI slice, and p_i is the i_{th} patch of the set.

3.2. Convolutional Sparse Coding

CSC optimizes a sparse representation for an image using a set of convolutional filters, *i.e.*, the convolutional sparse dictionary. It is proven in [12] that the image quality assessment procedure within the human visual system also adheres to the strategy of ‘sparse coding’ [13]. Based on this fact, CSC was exploited to quantify structure-related artifacts in several immersive applications [14, 15, 16]. We collected a set of distorted patches from EPI slices and trained a CSC codebook, where each element within the codebook represent a certain type of non-natural structure that could be observed in EPI slices. Details are provided below.

To learn a CSC dictionary that is representative for the non-natural structure-related distortions within the EPI slices, we manually collected more than 1k EPI patches ($100 \times 100 px$) from a validation set of the MPI-LFA dataset. These patches cover six types of LF-related distortions and five levels of distortions for each type (details of the dataset are provided in Sec. 3.3). Among these, we further selected 360 patches that contain visible structural distortions. More explicitly, only patches that were agreed by two experts have remained. Instances of patches contain different types of artifacts are presented in Fig. 5.

With the collected distorted dataset, the CSC dictionary was learned via the following objective function:

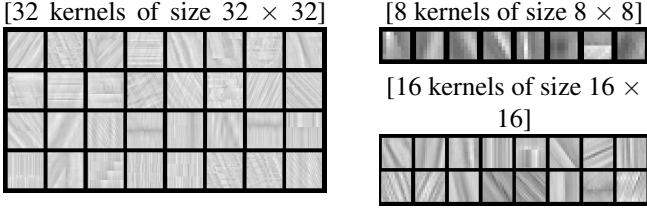


Fig. 6. Kernels learned using CSC on three different scales.

$$\min_{D, Z} \frac{1}{2} \|y - \sum_{e=1}^E D_e Z_e\|_2^2 + \alpha \sum_{e=1}^E \|Z_e\|_1 \quad (2)$$

$$\text{s.t. } \|D_e\|_2^2 \leq 1 \quad \forall e \in \{1, \dots, E\},$$

where y is an input image, D_e is the e_{th} element of the CSC dictionary, Z_e is the feature map with respect to the kernel D_e , α is a parameter that balances the reconstruction loss and the sparsity, E is the number of elements in the dictionary and \cdot indicates the convolution operation.

To speed up the optimization procedure, one of the out-of-the-shelf algorithms [10] was applied with an adaptive batch-like learning strategy equipped with a parallel computation setting. We divided the collected EPI patches into 30 batches, where each batch contain one distortion level for one distortion type. After the learning procedure, we obtained 293 kernels in total. By eliminating the noisy kernels based on the energy [10], we kept only 56 kernels from the candidate set, including 32 kernels with size of 32×32 , 16 kernels with size of 16×16 , and 8 kernels with size of 8×8 . The remained kernels are displayed in Fig. 6. It could be observed that the non-natural structural distortions could be well captured by most of the kernels (see Fig. 5 for comparison).

Using the learned dictionary, for a given $m \times n$ EPI slice, it could then be represented by a $m \times n \times E$ tensor of feature maps $Z_{EPI} = [Z_1; \dots; Z_e; \dots; Z_E]$, where each map Z_e is the response of using kernel D_e . With the feature maps, a CSC-based feature descriptor f_{csc} could be then computed using as done in [14, 15]:

$$f_{csc} = (f_{act}(Z_1), \dots, f_{act}(Z_E)), \quad (3)$$

where f_{act} is defined as

$$f_{act}(Z_e) = \frac{\sum_{i=1}^m \sum_{j=1}^n \mathbf{1}(e(i, j) > \epsilon)}{m \times n}, \quad (4)$$

ϵ is a threshold for selecting activated pixels. Function $f_{act}(\cdot)$ aggregates the number of pixels which are above the threshold ϵ in each sparse feature map Z_e corresponding to each kernel D_e . Intuitively, this function counts the number of pixels that are activated by the corresponding kernel. In other words, since the kernels are trained to capture stitching-related artifacts, this process can be interpreted as the computation of certain types of artifacts in the entire image and thus can be used to indicate perceived quality.

3.3. Quality Prediction

After feature normalization, the extracted HOG, CSC-based representations were concatenated and fed into a Support Vector Regression (SVR) model to predict the final quality scores utilizing a basic radial basis function kernel as done in [9].

4. EXPERIMENTS

4.1. Experimental Setup and Evaluation Criteria

We conduct our evaluations on MPI-LFA dataset [1]. The MPI-LFA dataset contains 350 LF images evaluated by 40 subjects by pair-wise comparison. Out of 350 LF images in the dataset, there are 14 reference LF images. 336 distorted LF image has been generated by applying HEVC-3D encoder [17], linear interpolation, nearest-neighbor interpolation, optical flow estimation, quantized depth maps, and angular domain gaussian blur. Six different severity levels have been used to generate distorted LF images.

LF images in the MPI-LFA dataset has only horizontal parallax with 101 sub-aperture views. The spatial resolution of each sub-aperture view is 960×720 px. Thus, each LF image consists of 720 EPI slice with 101×960 px resolution. In order to reduce the computational complexity, we considered only 36 EPI slices to extract features from each LF image.

For the MPI-LFA dataset, the authors used the just objectionable differences(JOD) as the scale. Zero(0) JOD score means having no quality difference, while negative values indicate an observable quality difference. There are robust evaluation methods for image quality metrics, such as the model proposed by Krasula et al[18]. Global correlation measures such as PCC, SCC do not consider the uncertainty in the subjective scores, and objective metrics need to be mapped to the subjective quality experiment range. The proposed approach by Krasula et al. resolves these problems. However, in order to use the proposed evaluation methodology, we need to have access to statistical information about the subjective test. if MOS scores, and not individual scores, are the only reported results of subjective experiments, it is not enough to run such full comprehensive evaluation. Unfortunately, MPI-LFA dataset does not provide enough statistical information about the results rather than the JOD scores, we utilized the cross-validation methodology for the evaluation of the proposed model

Cross-validation is another alternative to reliably evaluate the generalization performance of a machine-learning-based quality assessment system [19]. Therefore, the proposed model was evaluated through 1000-fold cross-validation as done in [14, 20, 21]. At each fold, the whole dataset was randomly divided into 80% for training and 20% for testing, without overlap between them for 1000 times. The median Pearson Linear Correlation Coefficient (PCC_m), median Spearman Rank-Order Correlation Coefficient (SCC_m), median Root Mean Squared Error ($RMSE_m$) between the sub-

jective and objective scores are reported across the 1000 runs for performance evaluation.

4.2. Experimental Results

The overall performances are shown in Table 1. The results are compared to several 2D full-reference metrics [22] [23] [24] [25] [26], 2D no-reference metrics [27] [28], multi-view and LF quality metrics [29] [9]. As observed, the proposed NR model outperforms all the traditional 2D and 3D metrics, and achieves competitive performance compared to the state-of-the-art LF metric, *i.e.*, BELIF[9].

| Metrics | SCC _m | PCC _m | RMSE _m |
|-----------------|------------------|------------------|-------------------|
| PSNR | 0.8078 | 0.7830 | 1.2697 |
| SSIM [22] | 0.7027 | 0.7123 | 1.4327 |
| VIF [23] | 0.7843 | 0.7861 | 1.2618 |
| FSIM [24] | 0.7776 | 0.7679 | 1.3075 |
| MSSIM [25] | 0.7675 | 0.7518 | 1.3461 |
| IWSSIM [26] | 0.8124 | 0.7966 | 1.2340 |
| BRISQUE [27] | 0.6724 | 0.7597 | 1.1317 |
| NFERM [28] | 0.6454 | 0.7451 | 1.1036 |
| MW-PSNR R. [29] | 0.7217 | 0.6757 | 1.5048 |
| MW-PSNR F. [29] | 0.7232 | 0.6770 | 1.5023 |
| BELIF [9] | 0.8854 | 0.9096 | 0.7877 |
| Proposed | 0.8942 | 0.9005 | 0.8916 |

Table 1. Performance results on the MPI-LFA dataset.

4.3. Ablation Study

In order to better understand the contributions of each structure representation to the proposed framework, an ablation study has been conducted. Results are summarized in Table 2. When CSC features are removed, HOG features alone has a lower correlation compared to removing CSC features being removed. These observations demonstrate that both feature descriptors contribute to quantifying the structural distortions better.

| | SCC _m | PCC _m | RMSE _m |
|-----|------------------|------------------|-------------------|
| HOG | 0.7782 | 0.7845 | 1.2690 |
| CSC | 0.8088 | 0.8143 | 1.1740 |

Table 2. Performances of ablative models using only the CSC or the HOG descriptor.

5. CONCLUSION AND DISCUSSION

HOG features are utilized in other computer vision tasks to detect shapes and structures on the 2D images [5]. As suggested in [4], structural feature descriptors are useful to reveal the distortions on EPI representations. Additionally, we adopted CSC to learn a representative dictionary for distorted EPIs. Contrary to the traditional optimization procedure, we

have developed a novel batch-like approach to split the training set into different parts by the severity-level of each individual distortion type. This allows the model to converge to a representative kernel dictionary for the complete set of distortions, and the learning procedure could be sped up through parallel computation.

Evaluation of the proposed approach has been done via cross-validation on the MPI-LFA dataset. Even though it is more robust than a direct evaluation, other evaluation methods such as [18] could have been adopted with sufficient statistical information about the subjective experiment. Thus, we have utilized 1000 fold cross validation as suggested in the literature. Evaluation showed that the proposed metric is capable of predicting LF quality in correlation with the subjective opinions and provided a competitive result in comparison to existing metrics.

Acknowledgment

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

6. REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Ali Ak and Patrick Le Callet, “Investigating Epipolar Plane Image Representations for Objective Quality Evaluation of Light Field Images,” in *European Workshop on Visual Information Processing*, Rome, Italy, Oct. 2019.
- [5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, June 2005, vol. 1, pp. 886–893 vol. 1.
- [6] 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.

- [7] 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.
- [8] L. Shan, P. An, C. Meng, X. Huang, C. Yang, and L. Shen, "A no-reference image quality assessment metric by multiple characteristics of light field images," *IEEE Access*, vol. 7, pp. 127217–127229, 2019.
- [9] L. Shi, S. Zhao, and Z. Chen, "Belif: Blind quality evaluator of light field image with tensor structure variation index," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 3781–3785.
- [10] Michal Šorel and Filip Šroubek, "Fast convolutional sparse coding using matrix inversion lemma," *Digit. Signal Process.*, vol. 55, no. C, pp. 44–51, Aug. 2016.
- [11] Suiyi Ling, Patrick Le Callet, and Zitong Yu, "The role of structure and textural information in image utility and quality assessment tasks," *Electronic Imaging*, vol. 2018, no. 14, pp. 1–13, 2018.
- [12] Ayyoub Ahar, Adriaan Barri, and Peter Schelkens, "From sparse coding significance to perceptual quality: A new approach for image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 879–893, 2017.
- [13] Bruno A Olshausen and David J Field, "Sparse coding of sensory inputs," *Current opinion in neurobiology*, vol. 14, no. 4, pp. 481–487, 2004.
- [14] Suiyi Ling, Gene Cheung, and Patrick Le Callet, "No-reference quality assessment for stitched panoramic images using convolutional sparse coding and compound feature selection," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018, pp. 1–6.
- [15] Suiyi Ling and Patrick Le Callet, "How to learn the effect of non-uniform distortion on perceived visual quality? case study using convolutional sparse coding for quality assessment of synthesized views," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 286–290.
- [16] Suiyi Ling, Jing Li, Patrick Le Callet, and Junle Wang, "Perceptual representations of structural information in images: application to quality assessment of synthesized view in ftv scenario," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1735–1739.
- [17] G. Tech, Y. Chen, K. Müller, J. Ohm, A. Vetro, and Y. 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.
- [18] L. Krasula, K. Fliegel, P. Le Callet, and M. Klíma, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–6.
- [19] Paolo Gastaldo, Rodolfo Zunino, and Judith Redi, "Supporting visual quality assessment with machine learning," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 54, 2013.
- [20] Soo-Chang Pei and Li-Heng Chen, "Image quality assessment using human visual dog model fused with random forest," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3282–3292, 2015.
- [21] Dingquan Li, Tingting Jiang, Weisi Lin, and Ming Jiang, "Which has better visual quality: The clear blue sky or a blurry animal?," *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1221–1234, 2018.
- [22] 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.
- [23] Hamid Sheikh and Alan Bovik, "A visual information fidelity approach to video quality assessment," 01 2005.
- [24] 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.
- [25] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale 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.
- [26] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, May 2011.
- [27] Z. Chen, W. Zhou, and W. Li, "Blind stereoscopic video quality assessment: From depth perception to overall experience," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 721–734, Feb 2018.
- [28] Deepti Ghadiyaram and Alan C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *Journal of Vision*, vol. 17, no. 1, pp. 32–32, 01 2017.
- [29] D. Sandić-Stanković, D. Kukulj, and P. Le Callet, "Dibr synthesized image quality assessment based on morphological wavelets," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015, pp. 1–6.