

Blind Quality Index of Depth Images Based on Structural Statistics for View Synthesis

Yipo Huang , Leida Li , *Member, IEEE*, Hancheng Zhu , and Bo Hu 

Abstract—The quality of depth images is crucial for virtual view synthesis. However, the quality assessment of depth images is still largely unexplored. This letter presents a blind quality metric of Depth image based on Structural Statistics (DSS). The design philosophy is inspired by the fact that structural distortion in the depth images usually leads to geometric distortion, which is the main cause for degraded quality of synthesized views. Specifically, the statistical features for shape and orientation are calculated based on discrete orthogonal moments and gradients, generating two groups of quality-aware features. Then, the quality model is built from the extracted statistical features using a regression module. The experimental results demonstrate the effectiveness of the proposed metric.

Index Terms—Quality assessment, view synthesis, depth image, structural distortion.

I. INTRODUCTION

WITH the rapid development of video technologies, virtual view synthesis has become increasingly popular [1], [2]. In view synthesis, the input depth images play a paramount role in the warping process, which have significant influence on the quality of synthesized views [3]. In practice, depth images usually undergo several stages before view synthesis, including acquisition, preprocessing, compression, transmission and post-processing. Each stage may introduce certain types of distortions, which will in turn degrade the quality of depth images. For example, during the acquisition, blur or white noise may be introduced due to the limited imaging sensor, and bit error is usually caused in the transmission stage. Moreover, the most existing image quality assessment (IQA) metrics are devised for natural images and thus are limited in evaluating the quality of depth images. Therefore, it is highly desirable to design accurate quality metrics specially for depth images.

Manuscript received January 12, 2020; revised March 17, 2020; accepted April 15, 2020. Date of publication April 20, 2020; date of current version May 19, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61771473, Grant 61991451, and Grant 61379143, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20181354, in part by the Science and Technology Plan of Xi'an (0191122015KYPT011JC013), and in part by the Six Talent Peaks High-level Talents in Jiangsu Province under Grant XYDXX-063. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Daniel P. K. Lun. (*Corresponding author: Leida Li.*)

Yipo Huang, Hancheng Zhu, and Bo Hu are with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: cumt_hyp@cumt.edu.cn; zhuhanheng@cumt.edu.cn; Hubo0523@cumt.edu.cn).

Leida Li is with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: ldli@xidian.edu.cn).

Digital Object Identifier 10.1109/LSP.2020.2988830

Up to now, only several metrics have been proposed to evaluate the quality of depth images. Farid *et al.* proposed a Blind Depth Quality Metric (BDQM) based on the sharpness of the compression sensitive regions [4]. In [5], Xiang *et al.* first measured the misalignment error between the edges of color and depth images. Then the Bad Point Rate (BPR) was calculated as the quality score of depth image. Le *et al.* [6] proposed a depth quality metric by evaluating the local depth distortion with the guidance of local color characteristics. In [7], Li *et al.* proposed a quality index for depth images by modeling the Statistics of Edge Profiles (SEP). The Canny operator was first employed for locating edge profiles in depth images. Then statistical features are extracted from gradient magnitude and Laplace-of-Gaussian domains based on the located edge profiles. Finally, the random forest was adopted to build the quality model for depth maps.

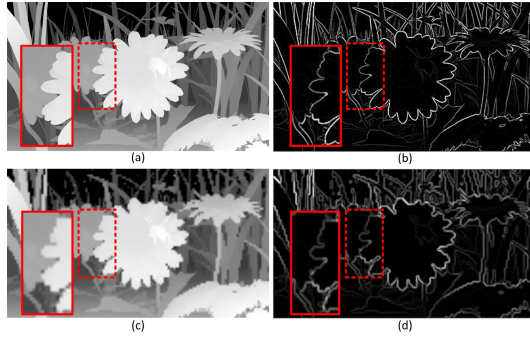
Inspired by the above efforts, this letter presents a blind quality metric of Depth images based on Structural Statistics (DSS). The design philosophy of the proposed metric is that the structures of depth images are related to the difference of depth distance between the foreground and background in the corresponding natural scene, which is crucial for view synthesis [8]. Specifically, a new depth structural statistics model is proposed based on multi-order Tchebichef moment. Furthermore, the orientation statistical features are calculated based on image gradients. Support vector regression (SVR) is adopted to model the underlying relation between statistical features of depth images and the ground truth quality. The superiority of the proposed metric is verified by experimental results on two public databases.

II. PROPOSED DEPTH QUALITY METRIC

In view synthesis, the warping operation generally introduces geometric distortions due to the structural distortions in the depth images. The underlying reason is that the shape of object boundary is destroyed by the distorted depth images, which in turn leads to edge misalignment between the input depth and color images. Further, the real orientation of depth difference is easily distorted by structural distortions in the depth images. These facts naturally lead us to consider the shape and orientation features for evaluating the quality of depth images.

A. Structure Representation

The gradient has been demonstrated to be effective for structure representation in the IQA community [9]–[11]. The gradient



map of a depth image (**D**) is first calculated as:

$$\mathbf{G}_D = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}, \quad (1)$$

$$\mathbf{G}_x = [-1 \ 0 \ 1] * \mathbf{D}, \quad \mathbf{G}_y = [-1 \ 0 \ 1]^T * \mathbf{D}, \quad (2)$$

where \mathbf{G}_x and \mathbf{G}_y denote the horizontal and vertical gradients, T and $*$ represent the transpose and convolution operators, respectively.

Fig. 1 shows a reference depth image and its distorted version, together with their gradient maps. It can be seen from Fig. 1(c) that the structure is obviously destroyed by downsampling blur compared to image (a). For example, in the edge regions of the flowers, the shape is quite unnatural. Moreover, the structural distortion can be better viewed in the gradient map, as shown in Fig. 1(d). By calculating the gradients, we easily obtain the structural information of the depth images, and the structural distortions can be better represented using the gradient maps.

B. Shape Statistics

Different distortions usually cause different shapes of structures [12]. Discrete Tchebichef moments have been demonstrated to be effective for shape description [13]. In this letter, a novel statistical model is proposed using the multi-order Tchebichef moments, where the objective is to describe the different shapes of structures destroyed by diversified distortions. Specifically, the gradient map is first partitioned into equal-size blocks with $n \times n$ pixels. Then, the Tchebichef moments up to the $[(n - 1) + (n - 1)]th$ order of the m^{th} block are computed as:

$$\mathbf{T}_m = \begin{pmatrix} C_{00} & C_{01} & \cdots & C_{0(n-1)} \\ C_{10} & C_{11} & \cdots & C_{1(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ C_{(n-1)0} & C_{(n-1)1} & \cdots & C_{(n-1)(n-1)} \end{pmatrix}, \quad (3)$$

where C_{00} represents the direct current (DC) coefficient of the moments, and the others are alternating current (AC) coefficients. In this work, n is set to 8.

The different orders in Tchebichef moments have varying characteristics in describing the shape [12]. To capture the shape of structure, the Tchebichef moment coefficients of each block

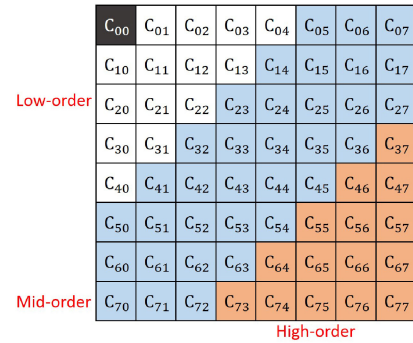


Fig. 2. Tchebichef moment coefficients collected along three orders.

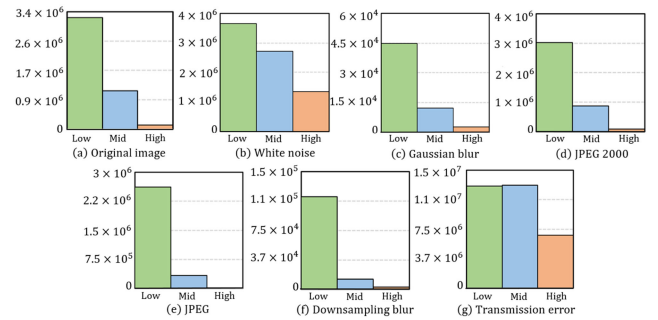


Fig. 3. Statistic histograms of the energy distribution within the three orders in the original depth image and its six distorted versions.

are divided into three regions with different orders, including low-, mid- and high-order, as shown in Fig. 2. Then, the energy of each order is computed based on the corresponding AC coefficients as:

$$E_m^p = \sum_{n=1}^N C_n^2, \quad (4)$$

where C_n denotes the AC coefficient of the moments, N is the total number of AC coefficients in the regions of corresponding order, p denotes different orders, including low-, mid- and high-order.

Finally, the sum of the energy of all the blocks belonging to the p order is calculated as follows:

$$E^p = \sum_{m=1}^M E_m^p, \quad (5)$$

where M is the total number of blocks.

Fig. 3 shows the histograms of the energy distribution within the three orders in the original depth image Fig. 1(a) and its six distorted versions. It can be seen from Fig. 3 that the depth images with different distortions have different statistical characteristics, which indicates the effectiveness of the proposed shape statistics for portraying the distortions in the depth images.

C. Orientation Statistics

The orientation of structures is a vital factor affecting the quality of color images [14]. The gradient magnitudes of depth

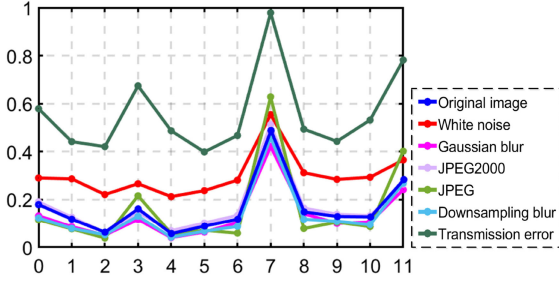


Fig. 4. Orientation distributions in the original depth image and its six distorted versions.

images have very similar statistical characteristics with the corresponding color images, which has been demonstrated in [15]. Inspired by this, the orientation statistics are further calculated as:

$$\mathbf{O}_D = \arctan \left(\frac{\mathbf{G}_y}{\mathbf{G}_x + \tau} \right) \frac{180^\circ}{\pi}, \quad (6)$$

where τ donates a small constant for numerical stability.

The range of structure orientation is $[-180^\circ, 180^\circ]$. We divided it into equally-spaced segments, and the range of each segment is represented as s_0, s_1, \dots, s_k . Then the statistical cumulative gradient amplitudes within each orientation segment are computed as:

$$\mathbf{S}_k = \sum_{X=1}^H \sum_{Y=1}^W \mathbf{G}_D \cdot f_0(\mathbf{O}_D(X, Y), s_k), \quad (7)$$

where the size of the depth image is $H \times W$, and the function f_0 is defined as:

$$f_0(\theta, \psi) = \begin{cases} 1, & \theta \in \psi \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where θ and ψ represent the $\mathbf{O}_D(X, Y)$ and s_k , respectively. In this letter, k is empirically set to 11.

Fig. 4 illustrates the orientation distributions in the original depth image Fig. 1(a) and its six distorted versions. It is easily observed that the shapes of the curves depend on the distortion types. Therefore, the orientation statistical features are sensitive to distortions, which also confirms that different types of distortions in depth images can be effectively represented by the proposed features.

Furthermore, considering the multi-scale characteristic of the human visual system (HVS) [16], [17], the Gaussian low-pass function [18] is adopted to generate 5 scales of the depth image and the aforementioned quality features are extracted accordingly. In this letter, a total of 75 quality-aware features are extracted from each depth image, including 15 multi-scale shape statistical features and 60 multi-scale orientation statistical features.

D. Quality Prediction

In this work, the SVR model [19] is employed for learning the quality model based on statistical features of training images. Finally, the quality scores are predicted using the trained model.

In implementation, the Radial Basis Function (RBF) is adopted as the SVR kernel.

III. EXPERIMENTAL RESULTS

A. Experimental Setup

Different from the general depth image quality methods, the proposed quality index is to study the impact of distorted depth images on the quality of synthesized views. Therefore, the performance of the proposed DSS metric is evaluated on two public databases for view synthesis, including the MCL-3D database [20] and the SIAT synthesized video quality database [21]. For the MCL-3D database, we test 648 distorted depth images in total, which are generated based on nine different image-plus-depth sources. In this database, six different types of distortions are added to the depth images, including downsampling blur, white noise, Gaussian blurring, transmission error, JPEG and JPEG2000 compression. For the SIAT database, we test 80 video sequences with four different compression degrees. In implementation, 10 frames are uniformly sampled from each sequence, and the predicted scores of all extracted frames are averaged as the final quality score. For both databases, since we focus on depth distortion in this work, we only test the case that undistorted color images and distorted depth images are adopted for view synthesis. Moreover, the ground truth quality scores of depth images are based on the quality of the corresponding synthesized views, which are represented as the average values of human rating scores collected from a subjective experiment.

In this letter, four popular criteria are employed for performance evaluation. Specifically, Pearson linear correlation coefficient (PLCC) and root mean squared error (RMSE) are adopted for measuring prediction accuracy. Spearman rank order correlation coefficient (SRCC) and Kendalls rank correlation coefficient (KRCC) are used to measure monotonicity.

B. Performance Evaluation

To demonstrate the advantages of the proposed DSS metric, three recent depth image quality metrics are compared, including BDQM [4], BPR [5] and SEP [7]. Furthermore, five general-purpose no-reference (NR) IQA metrics for natural images are also employed for comparison, including NIQE [22], BRISQUE [23], BLIINDS-II [24], DESIQUE [25], and BIQA [26]. It is worth to emphasize that the aforementioned learning-based metrics are all retrained on the corresponding database for fair comparison. In implementation, 80% of depth images are randomly selected to train the quality model, while the remaining 20% depth images are used to evaluate performance. To avoid bias, the training-test process is conducted 1000 times, and the median performance values are reported. Table I lists the experimental results on the MCL-3D and SIAT databases, where the best performance values are marked in boldface.

From Table I, it can be observed that the proposed DSS achieves the best performances on the two databases, regardless of prediction accuracy and monotonicity. (1) On the MCL-3D database, among NR-IQA metrics, the performance of the BIQA [26] metric ranks the first. Among the three existing quality

TABLE I
COMPARISON WITH EXISTING QUALITY METRICS ON THE MCL-3D
AND SIAT DATABASES

Database	Metric	Criterion			
		PLCC	SRCC	KRCC	RMSE
MCL-3D	NIQE [22]	0.3918	0.3451	0.2332	1.6741
	BRISQUE [23]	0.6594	0.3516	0.2390	1.2046
	BLIINDS-II [24]	0.3880	0.3421	0.2304	1.6891
	DESIQUE [25]	0.5728	0.3699	0.2524	1.3134
	BIQA [26]	0.8094	0.6176	0.4392	0.9307
	BDQM [4]	0.3591	0.3623	0.2464	1.7031
	BPR [5]	0.5938	0.5539	0.4024	1.2637
	SEP [7]	0.8744	0.7983	0.6181	0.7722
	Proposed DSS	0.9263	0.8350	0.6667	0.5881
SIAT	NIQE [22]	0.3284	0.3591	0.2846	0.1817
	BRISQUE [23]	0.5045	0.4929	0.4103	0.1331
	BLIINDS-II [24]	0.4035	0.3993	0.3052	0.1551
	DESIQUE [25]	0.4338	0.3550	0.2760	0.1420
	BIQA [26]	0.5491	0.5874	0.4519	0.1324
	BDQM [4]	0.6152	0.5616	0.4443	0.1174
	BPR [5]	0.6569	0.6127	0.4975	0.1092
	SEP [7]	0.6888	0.6190	0.5000	0.1039
	Proposed DSS	0.8017	0.7381	0.5714	0.0880

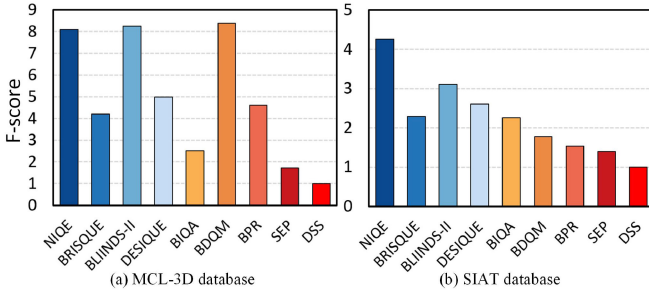


Fig. 5. F statistics of the compared metrics against the proposed DSS metric.

metrics of depth images, the SEP [7] metric performs better. (2) On the SIAT database, among the existing quality metrics, SEP [7] also achieves the best performance. By comparison, the proposed DSS significantly outperforms all the existing popular quality metrics on both databases.

C. Statistical Significance Analysis

In order to investigate the statistical significance of each compared metric relative to DSS, we conduct the F-test [27] and the experimental results on the two databases are shown in Fig. 5. It is easily observed that the F-scores of the existing metrics are all higher than 1, which demonstrates that the proposed DSS achieves better prediction accuracy than all the compared metrics.

Further, a summary of the statistical performance is listed in Table II. This is obtained by comparing the F-scores with a threshold ($F_{critical}$), which is related to the number of images and a confidence level [28]. In this letter, the confidence level is set to 95%, and the $F_{critical}$ values on the MCL-3D and SIAT databases are 1.138 and 1.693, respectively. It can be seen from Table II that the proposed DSS shows obvious superiority over all the existing metrics on the MCL-3D database. On the SIAT database, the proposed DSS performs statistically competitive with BPR [5] and SEP [7], and it outperforms the other metrics.

TABLE II
SUMMARY OF STATISTICAL PERFORMANCE BETWEEN THE PROPOSED DSS AND
THE COMPARED METRICS, WHERE 1(0) INDICATES THAT OUR METRIC
PERFORMS STATISTICALLY BETTER (COMPETITIVE)

Metric	MCL-3D	SIAT
NIQE [22]	1	1
BRISQUE [23]	1	1
BLIINDS-II [24]	1	1
DESIQUE [25]	1	1
BIQA [26]	1	1
BDQM [4]	1	1
BPR [5]	1	0
SEP [7]	1	0

TABLE III
COMPARISON WITH EXISTING DEPTH IMAGE QUALITY METRICS ON
DIFFERENT TYPES OF DISTORTIONS

PLCC	BDQM [4]	BPR [5]	SEP [7]	DSS
White noise	0.4504	0.3812	0.9108	0.9491
Gaussian blur	0.3029	0.7345	0.8651	0.9294
JPEG2000	0.5284	0.6582	0.8421	0.8054
JPEG	0.6626	0.6259	0.8970	0.8193
Downsampling blur	0.2524	0.5972	0.8691	0.9523
Transmission error	0.5680	0.5083	0.8439	0.9408
SRCC	BDQM [4]	BPR [5]	SEP [7]	DSS
White noise	0.4184	0.3510	0.8358	0.8928
Gaussian blur	0.3370	0.7281	0.7899	0.8214
JPEG2000	0.3538	0.5594	0.7703	0.7155
JPEG	0.6621	0.6017	0.8334	0.7478
Downsampling blur	0.2809	0.5276	0.7809	0.8822
Transmission error	0.4813	0.4173	0.7651	0.8929

It can be concluded from the table that our metric has the best statistical performance.

D. Performances on Different Types of Distortions

The proposed DSS metric is designed for evaluating the quality of depth images with diversified distortions. In order to further evaluate the performance of DSS for different distortions, we conduct the experiment on six different types of distortions of the MCL-3D database. For comparison, the performances of the three existing quality metrics for depth images are also included. The experimental results are reported in Table III.

It can be seen from Table III that the proposed DSS also delivers very promising performances on different types of distortions, especially for downsampling blur, white noise, transmission error and Gaussian blur. Therefore, the proposed DSS achieves the best performance in evaluating diversified distortions of depth images.

IV. CONCLUSION

In this letter, we have presented a blind quality metric for depth images based on structural statistics. Two groups of statistical features are constructed to portray the structural shape and orientation characteristics of depth images. A SVR model is trained for predicting the quality of depth images for view synthesis. Experimental results on two public view synthesis databases demonstrate the advantages of the proposed metric over the relevant state-of-the-arts.

REFERENCES

- [1] A. Q. de Oliveira, M. Walter, and C. R. Jung, "An artifact-type aware DIBR method for view synthesis," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1705–1709, Nov. 2018.
- [2] Y. Zhang, Y. Zhao, and L. Yu, "Block-based in-loop view synthesis for 3-D video coding," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 441–444, Apr. 2014.
- [3] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of DIBR-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 914–926, Apr. 2018.
- [4] M. S. Farid, M. Lucenteforte, and M. Grangetto, "Blind depth quality assessment using histogram shape analysis," in *Proc. 3DTV Conf., True Vision Capture, Transmiss. Display 3D Video*, 2015, pp. 1–5.
- [5] S. Xiang, L. Yu, and C. W. Chen, "No-reference depth assessment based on edge misalignment errors for T + D images," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1479–1494, Mar. 2016.
- [6] T. H. Le, S. W. Jung, and C. S. Won, "A new depth image quality metric using a pair of color and depth images," *Multimedia Tools Appl.*, vol. 76, no. 9, pp. 11 285–11 303, 2017.
- [7] L. Li, X. Chen, J. Wu, S. Wang, and G. Shi, "No-reference quality index of depth images based on statistics of edge profiles for view synthesis," *Inform. Sci.*, vol. 516, pp. 205–219, Dec. 2019.
- [8] Y. Zhou, L. Li, S. Wang, J. Wu, Y. Fang, and X. Gao, "No-reference quality assessment for view synthesis using DoG-based edge statistics and texture naturalness," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4566–4579, Sep. 2019.
- [9] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.
- [10] G. Ding, Y. Guo, K. Chen, C. Chu, J. Han, and Q. Dai, "DECODE: Deep confidence network for robust image classification," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3752–3765, Aug. 2019.
- [11] Y. Guo, G. Ding, and J. Han, "Robust quantization for general similarity search," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 949–963, Feb. 2018.
- [12] L. Li, W. Lin, and H. Zhu, "Learning structural regularity for evaluating blocking artifacts in JPEG images," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 918–922, Aug. 2014.
- [13] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.
- [14] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.
- [15] C. Su, L. K. Cormack, and A. C. Bovik, "Color and depth priors in natural images," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2259–2274, Jun. 2013.
- [16] Y. Yang, W. Cao, S. Wu, and Z. Li, "Multi-scale fusion of two large-exposure-ratio images," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1885–1889, Dec. 2018.
- [17] G. Ding, W. Chen, S. Zhao, J. Han, and Q. Liu, "Real-time scalable visual tracking via quadrangle kernelized correlation filters," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 140–150, Jan. 2018.
- [18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [19] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Jan. 2011.
- [20] R. Song, H. Ko, and C. C. Kuo, "MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source," *J. Inf. Sci. Eng.*, vol. 31, no. 5, pp. 1593–1611, 2015.
- [21] X. Liu, Y. Zhang, S. Hu, S. Kwong, C. J. Kuo, and Q. Peng, "Subjective and objective video quality assessment of 3D synthesized views with texture/depth compression distortion," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4847–4861, Dec. 2015.
- [22] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [23] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [24] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
- [25] Y. Zhang and D. M. Chandler, "An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes," in *Proc. SPIE, Image Quality Syst. Perform.*, vol. 8653, p. 86530J, Feb. 2013.
- [26] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
- [27] L. Li, D. Wu, J. Wu, H. Li, W. Lin, and A. C. Kot, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.
- [28] L. Li, H. Zhu, G. Yang, and J. Qian, "Referenceless measure of blocking artifacts by Tchebichef kernel analysis," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 122–125, Jan. 2014.