https://doi.org/10.2352/ISSN.2470-1173.2021.11.HVEI-110

Deep Quality evaluator guided by 3D Saliency for Stereoscopic Images

Oussama Messai +, Aladine Chetouani •, Fella Hachouf +, Zianou Ahmed Seghir *
+ Laboratoire ARC, Université des Frères Mentouri Constantine 1, Algérie.
• PRISME Laboratory, University of Orleans, France.
**** Computing Department, University of Abbes laghrour Khenchela, Algeria.***

Abstract

Due to the use of 3D contents in various applications, Stereo Image Quality Assessment (SIQA) has attracted more attention to ensure good viewing experience for the users. Several methods have been thus proposed in the literature with a clear improvement for deep learning-based methods. This paper introduces a new deep learning-based no-reference SIQA using cyclopean view hypothesis and human visual attention. First, the cyclopean image is built considering the presence of binocular rivalry that covers the asymmetric distortion case. Second, the saliency map is computed taking into account the depth information. The latter aims to extract patches on the most perceptual relevant regions. Finally, a modified version of the pre-trained vgg-19 is fine-tuned and used to predict the quality score through the selected patches. The performance of the proposed metric has been evaluated on 3D LIVE phase I and phase II databases. Compared with the state-of-the-art metrics, our method gives better outcomes.

Introduction

The use of 3D technology is becoming increasingly attractive for various academic and industrial applications (e.g., entertainments, 3D visualization, robotic navigation, medical surgery, etc.) [1]. However, treatments usually applied to transmit or display the content (i.e. compression, transmission, etc.) may impact the perceived quality. Hence, it seems important to have efficient methods that assess the quality or visual discomfort in order to guarantee a high quality of experience.

In general, Stereoscopic Image Quality Assessment (SIQA) metrics can be divided into two classes: subjective and objective methods. Subjective assessment is based on opinion scores given by human observers and is mostly expressed in terms of Mean Opinion Score (MOS) or Difference Mean Opinion Score (DMOS). This approach is effective and reliable for assessing perceptual quality, but it has some drawbacks, such as time consuming, high cost. Meanwhile, objective assessment offers an automatic evaluation and thus a lot of efforts has been dedicated to design accurate SIQA methods. Existing objective methods can be divided into three categories: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) metrics. FR methods utilize the reference stereo image, RR methods use only partial information of the reference image, while for NR metrics, the reference stereo image is completely unavailable. In order to design SIQA metrics, modeling Human Visual System (HVS) is important to simulate the visual judgment. However, the HVS is a complex visual process and still an open question for researchers. For SIQA development, many researchers have combined classical 2D metrics [2] or have used fusion hypothesizes of the perceived left and right eye signals called cyclopean view [3, 4]. Meanwhile, in most of the suggested SIQA approaches the visual attention is not explored.

In this study, we focus on the NR-SIQA, since the original stereoscopic images are not accessible in most practical situations. We take a step forward to develop an approach that can simulate the human judgment taking into account binocular rivalry and the observer selectively through the use of cyclopean image and visual attention, respectively.

The rest of this paper is divided as follows. Section presents related work. The proposed approach is described in Section . Section shows experimental results. Finally, section concludes the paper.

Related Work

Several metrics based on the use of 2D Image Quality Assessment (IQA) metrics have been proposed in the literature. Different strategies have been applied to derive a quality score through these 2D metrics (mean, weight, etc.) [5, 6]. Most of these methods do not consider the asymmetrical distortion case and thus fail to estimate the quality on the latter. The issue of asymmetric distortion in stereoscopic images is related to binocular rivalry/suppression which occurs when the eyes of a viewer see different scenes. This phenomenon often causes fatigue and visual discomfort to the observers [7]. However, metrics that exploit HVS characteristics such as cyclopean view have shown better results, especially for asymmetric distortions [8, 3].

Chen *et al.* [3] have proposed a FR quality assessment model that utilizes the linear expression of cyclopean view influenced by binocular rivalry between left and right views. Authors of [9] have proposed a FR-SIQA model that combines two measures. They first measure the difference between the left and right reference images and the distorted ones. Then, they compute the difference between the pure stereo image disparity map and the deformed ones. You *et al.* [5] have developed a FR-SIQA model, where they incorporate 2D-IQA metrics with disparity information. Another FR-SIQA metric, named Binocular Energy Quality Metric (BEQM) has been proposed by Bensalma *et al* [4]. They have measured the stereoscopic image quality by calculating the binocular energy variation between the reference and distorted stereopairs. The authors of [10, 11] have proposed a PSNR-based stereo

Figure 1: **The flowchart of the proposed metric.**

IQA models. Hewage *et al*. [11] have extracted edge maps from the disparity maps of the reference and distorted stereo-pairs. The PSNR is then computed between the reference and test edge maps to assess the quality. Gorley *et al*. [10] have not used the disparity/depth information but they rather calculate quality scores on corresponded feature points of the left and right images provided by SIFT [12] and RANSAC [13].

Authors in [14] have deployed binocular perceptual information to perform a RR-SIQA. While Ma et al [15] have characterized the statistical properties of stereoscopic images in the reorganized Discrete Cosine Transform (RDCT) domain. In [16], a RR-SIQA method based on Natural Scene Statistics (NSS) and structural degradation has been also proposed.

Some NR-SIQAs have been developed by researchers. For instance, Akhter *et al.* [17] have designed reference-less SIQA algorithm. They extract features from the disparity map and the stereo-pairs. Zhou *et al.* [18] have simulated binocular phenomenon and use the well known k-Nearest Neighbors (KNN). Fang et al. [19] have proposed an unsupervised model for stereoscopic images. From the monocular and cyclopean view patches, they extract quality indicators in spatial and frequency domains. Then, Bhattacharyya distance is used to get a quality score.

Deep convolutional network predictors have been also proposed for NR-SIQA. Local patches are extracted and then combined to form global features [20]. Authors in [21] have modeled the human visual cortex using deep auto-encoder. In [22], the authors have considered the deep perception map and binocular weight model to predict the perceived stereo image quality. Xu et al [23] have simulated our human brain cognition process to propose NR-SIQA metric using deep encoder-decoder network. Authors in [24] have designed NR SIQA metric based on gradient quality-aware features derived from the cyclopean image and the disparity map. A Weak Learners (WL) regression model to map the features to quality scores has been adopted. Authors in [25] took a step forward to deploy visual saliency characteristics for FR SIQA metric. In their framework, they developed a 3D saliency map to assign weights to the stereo content and then used an existed 2D/Stereo IQA metrics to gauge the quality.

However, most of the above mentioned SIQA approaches do not explore the human visual attention to predict the quality. In this paper, we introduce a new SIQA method that uses the benefit of saliency map integrated with cyclopean image hypothesis and deep learning. The main scientific contribution of this paper is the combination of deep convolutional predictor with human 3D visual saliency for reference-less stereo IQA.

Proposed Method

The general framework of our method is summarized in Fig. 1. From a given stereo image, we first calculate the cyclopean image which allows to consider the binocular rivalry phenomenon that occurs on our brain. [7]. Then, we compute a 3D saliency map which aims to focus on regions that attract more our perception. After thresholding the 3D saliency map, small patches are extracted and fed into a CNN model in order to predict the overall quality of the stereo image. Each of these steps is described in this section.

Cyclopean image

The purpose of cyclopean image is to simulate the fusion of human brain of the perceived signal from the left and right eyes. To consider asymmetric distortions in stereo images, weight coefficients from left and right images are utilized for the fusion. As Gabor filter can be used to extract luminance and chromatic features and it is related to the function of primary visual cortex cells in primates [26], it can be deployed to compute the fusion weights and thus mimic the binocular suppression phenomenon.

Inspired by the model used in [3], we construct a cyclopean image over three channels Red, Green, and Blue (RGB) rather than one gray channel to maintain the distortion effects on the three channels. Where these channels together represent the perceived spatial information by the observer. The utilized formula is as follows:

$$C(x,y)_n = w_l(x,y)_n \times I_l(x,y)_n + w_r(x+d,y)_n \times I_r(x+d,y)_n \quad (1)$$

where C refers to the cyclopean image, I_l and I_r are the left and right RGB views, respectively and n indicates the color channel. w_l and w_r are the weighting coefficients computed from Gabor filter responses for the left and right eye, respectively. The weights are computed from the Gabor filter bank responses over eight orientations. d is the disparity index that matches pixels from left image I_l with those in right image I_r.

Fig. 2 shows RGB cyclopean image formed from the left image that is not distorted and the right image that is distorted. It is worth noting that this asymmetric distortion is stated clearly into the cyclopean image (the red boxes).

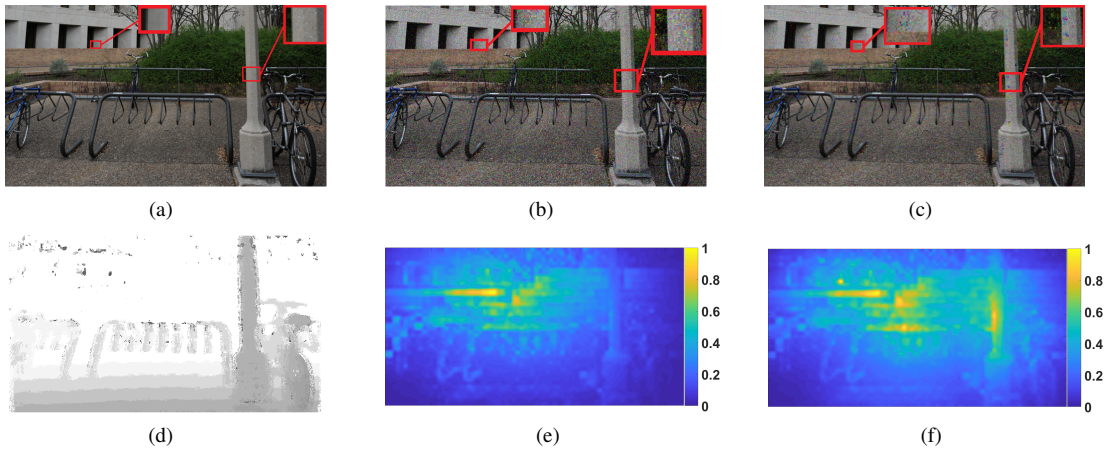


Figure 2: **(a) Left view without distortion, (b) Right view with White Noise (WN) distortion.(c) RGB cyclopean image example, (d) estimated depth using disparity map, (e) 2D saliency map and (f) the used 3D saliency map.**

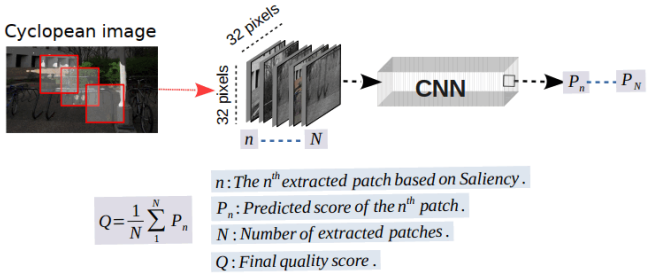


Figure 3: **Patch extraction and the measure followed in the proposed metric.**

3D Saliency map

Visual attention/saliency is an important characteristic of our HVS, since it represents the regions of the image in which the observer focus the most. Hence, salient regions impact more the subjective scores given by the observers and thus the quality of a given image is highly related to these regions. However, the saliency was already used for 2D images [27] or 3D meshes [28, 29]. Although, the usage of Saliency map is a step further to HVS simulation, it remains unconsidered in the most recent NR-SIQA metrics.

According to this observation, 3D saliency has been used in this study to extract perceptual relevant patches instead of all patches. In this study, the 3D saliency method proposed in [30] has been employed. This method is based on the integration of the depth information and 2D saliency maps. The saliency map of the luminance, color and texture from one view are first computed [31]. Then, the depth map is calculated through the left and right views. After applied a normalization, the 3D saliency map is finally given by averaging these maps. To show the difference between the 2D saliency map and the 3D one, we compute non-depth saliency map (i.e. 2D saliency) and depth saliency map (i.e. 3D saliency). From Fig. 2, we can see that the 3D saliency map gives more importance to near objects than the 2D one.

The 3D saliency map is thresholded to extract patches of size 32x32x3 from the cyclopean image only on the most salient regions. After several tests, the threshold has been fixed to 0.5. The extracted patches are then fed to a CNN model to predict their quality.

Network architecture

Different models have been proposed in the literature. Some of them are commonly used because they provide high performance in several tasks (AlexNet [32], VGG16 [33], ResNet [34], etc.). In this paper, the proposed approach is based on a modified version of the pre-trained vgg-19 [33].

This model consists in total of 19 main layers, where the input size is changed to be 32 $\times$ 32 pixels RGB patches (i.e. size of the extracted patches). The Fully Connected (FC) layers are replaced by three other FC layers of size 128, 10 and 1, respectively. The first two FC layers are followed by a ReLU layer as activation function. The last FC layer is a regression layer of size 1, since our target is a quality index.

During the learning, the CNN model is trained for 50 epochs with a learning rate of 0.01. Stochastic Gradient Descent (SGD) with a momentum equals to 0.9 are used as optimization function. The human scores are normalized in the form of DMOS to min-max normalization [0,1]. The closer to 0 the better quality of the stereo image is.

After-all, the quality index of a given stereo image is computed by averaging the predicted scores over the extracted saliency patches.

Experimental Results

Datasets and Training Protocol

Two databases have been used to evaluate the performance of our metric: 3D LIVE phase I (LIVE-P1) and 3D LIVE phase II (LIVE-P2). LIVE-P1 [35] consists of 365 distorted stereo images of size 360 x 640 pixels. Five degradation types are considered (White Noise: WN, JPEG2000: JP2K, JPEG, Fast Fading: FF and Blur). All the distortions are carried out symmetrically. LIVE-P2 [3] contains 360 distorted stereo images with the same size as phase I. This database includes symmetric as well as asymmetric distortion stereo-pairs.

During the saliency-guided extraction of patches, we allow maximum of 80% overlapping between patches as shown in Fig. 3 for data augmentation. 5-fold cross validation technique is used to train and test our model. To this end, we split randomly the dataset

Table 1: **Saliency ablation test of the proposed approach.**

	LIVE-P2		
Method	SROCC	PLCC	RMSE
Saliency guided	0.977	0.979	2.290
Without saliency	0.961	0.963	3.014

Table 2: **Overall performance comparison on LIVE-P1 and LIVE-P2.**

		LIVE-P1			LIVE-P2		
Type	Metric	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
FR	Benoit [9]	0.899	0.902	7.061	0.728	0.748	7.490
	You [5]	0.878	0.881	7.746	0.786	0.800	6.772
	Gorley [10]	0.142	0.451	14.635	0.146	0.515	9.675
	Chen [3]	**0.916**	**0.917**	**6.533**	**0.889**	**0.900**	**4.987**
	Hewage [11]	0.501	0.558	9.364	0.501	0.558	9.364
	Bensalma [4]	0.874	0.887	7.558	7.558	0.769	7.203
RR	RR-BPI [14]	-	-	-	0.867	0.915	4.409
	RR-RDCT [15]	0.905	0.906	6.954	0.809	0.843	6.069
	Ma [16]	**0.929**	**0.930**	**6.024**	**0.918**	**0.921**	**4.390**
NR	Akhter [17]	0.383	0.626	14.827	0.543	0.568	9.294
	Zhou [18]	0.901	0.929	6.010	0.819	0.856	6.041
	Fang [19]	0.877	0.880	7.191	0.838	0.860	5.767
	DNR-S3DIQE [20]	0.935	0.943	-	0.871	0.863	-
	3D-AdaBoost [24]	0.930	0.939	5.605	0.913	0.922	4.352
	DBN [22]	0.944	0.956	4.917	0.921	0.934	4.005
	DECOSINE [21]	0.953	0.962	-	0.941	0.950	-
	PAD-Net [23]	0.973	0.975	3.514	0.967	0.975	2.446
	Proposed	**0.980**	**0.982**	**3.054**	**0.977**	**0.979**	**2.290**

into training (80%) and test (20%) sets 5 times. This procedure allows to avoid bias and show the ability of the method over the whole database.

Evaluation Criteria

The performance has been measured across three metrics: The *RMSE*, the Spearman's rank-order correlation coefficient (*SROCC*), and the Pearson linear correlation coefficient (*LCC*) between the predicted quality scores (objective scores) and the subjective ones (*DMOS*). *LCC* and *RMSE* measures the assessment accuracy, while *SROCC* evaluates the prediction notability. Higher values for *LCC* and *SROCC* (closer to 1) and lower values for *RMSE* (closer to 0) indicate superior linear rank-order correlation and better precision with respect to human quality judgments, respectively.

Comparison with the State-of-the-Art

The results obtained have been compared with several FR and NR SIQA, among them, a very recent blind metric based on the use of an auto-encoder, called PAD-Net [22]. To illustrate the individual contribution of 3D saliency map, we first conducted a no saliency based test, where we fed all possible patches of the cyclopean image to the deep predictor. The patches were sequentially extracted by sliding over the whole scene from left to right with stride of 32 pixel. Table 1 demonstrates the superiority of the saliency guided method over all the three indexes. This test supports the usage of the proposed saliency map for the SIQA. Moreover besides accuracy improvement, the saliency guidance approach may also decrease the cost and run-time, since the approach uses recommended patches rather than using all patches of the scene.

Table 2 shows the results of these methods on both datasets. Best metric of each category (FR, RR and NR) is represented on bold and the best one whatever the category is with a gray background. As can be seen, our metric outperforms all the compared FR and NR metrics on both databases. The best FR metric is the one proposed by Chen et al [3], while the method proposed by Ma et al [16] achieved the best performance among the compared RR methods. The improvements in term of PLCC relative to the best metrics of each category (Chen for FR and Ma for RR) are higher than 7% for FR and 5.6% for RR on LIVE-P1, while they are higher than 8.8% for FR and 6.3% for RR on LIVE-P2.

Table 3: **SROCC performance for symmetric and asymmetric distorted images on LIVE-P2. Best result of each category is highlighted in bold.**

Metric	Type	Symmetric	Asymmetric
Benoit [9]		0.860	0.671
You [5]		0.914	0.701
Gorley [10]		0.383	0.056
Chen [3]	FR	**0.923**	**0.842**
Hewage [11]		0.656	0.496
Bensalma [4]		0.841	0.721
Akhter [17]		0.420	0.517
3D-AdaBoost [24]	NR	0.898	0.917
PAD-Net [23]		**0.982**	0.954
Proposed		0.969	**0.979**

Table 4: **Performance of the proposed metric under different train-test partitions on LIVE-P1 and LIVE-P2.**

	LIVE-P1			LIVE-P2		
Partition	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
80%-20%	0.980	0.982	3.054	0.977	0.979	2.290
70%-30%	0.976	0.979	3.285	0.973	0.976	2.435
50%-50%	0.970	0.975	3.637	0.967	0.970	2.733

Globally, the performance are higher for LIVE-P1, since this database contains only symmetric distorted stereo images. However, we can notice that our method achieves close results for both datasets. Indeed, the PLCC and SROCC values obtained for LIVE-P1 are respectively 0.982 and 0.980, while those obtained for LIVE-P2 are 0.977 and 0.979, respectively.

We also show in Table 3 the performance of our metric on symmetric and asymmetric distorted stimuli. As can be seen, some metrics totally fail to predict the quality for asymmetric distorted images, while they obtain high correlations for symmetric distorted images (Benoit, You and Bensalma). PAD-Net gets the best performance for symmetric distorted images, while our metric obtains the second and the first best correlations for symmetric and asymmetric distorted images, respectively. However, according to the Table 2 our metric achieves the best global results. Moreover, high accuracy on asymmetric distortions is more challenging, since most of the existing method fail.

Then, we evaluate the performance of our method according to the size of the training set. Table 4 shows the correlations achieved for a training set of size 50%, 70% and 80%. Performances remain stable with a slight decrease. The diminution is similar for both datasets. Overall, the proposed method performs better than other models on natural scene stereoscopic images from 3D LIVE database. However, this study did not focus on particular class of images since the model can be re-trained and deployed on any desired type of class. Thus, predictions on specific class or individual scenes is not examined, where it can be addressed in future work for deeper analysis.

To exhibit the prediction responses against human score (objective scores predicted by our method vs. subjective scores), we show in Fig. 4 and 5 the scatter plots obtained on LIVE-P1 and LIVE-P2, respectively. For both datasets, the distribution of the predicted scores is in accordance with the DMOS for all the con-

sidered degradation types.

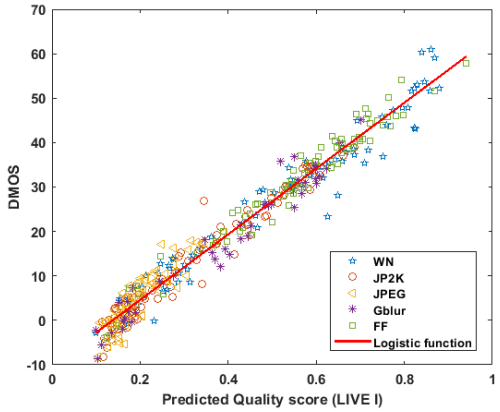


Figure 4: **Scatter plot of subjective scores DMOS against objective scores from the proposed metric on LIVE-P1.**

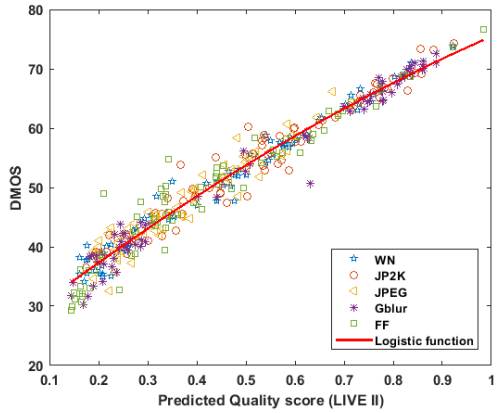


Figure 5: **Scatter plot of subjective scores DMOS against objective scores from the proposed metric on LIVE-P2.**

Conclusion

In this paper, a blind stereoscopic IQA based on the use of cyclopean image and saliency map has been proposed. Cyclopean image has been introduced to consider asymmetrical distortion, while the saliency aims to focus on the most perceptual relevant regions. Patches have been then selected on the saliency regions of the cyclopean image and fed as input to a modified version of the pre-trained vgg-19 model to estimate their quality. Experimental results demonstrate the efficiency of our metric since it outperforms all the compared FR and NR SIQA of the state-of-the-art on LIVE P1 and LIVE-P2. The proposed approach takes a step further toward the best perceptual modeling for quality judgment than the methods which do not consider visual attention information.

In future, we tend to examine different pre-trained CNN models for further improvements and more study the impact of the saliency method.

Acknowledgments

The computation in this work was funded by TIC-ART (Traitement des Images au serviCe de l'ART) project, Centre-Val de Loire department, France.

References

[1] Lauren A Christopher, Albert William, and Aaron A Cohen-Gadol, "Future directions in 3-dimensional imaging and neurosurgery: stereoscopy and autostereoscopy," *Neurosurgery*, vol. 72, no. suppl_1, pp. A131–A138, 2013.

[2] Sid Ahmed Fezza, Aladine Chetouani, and Mohamed-Chaker Larabi, "Using distortion and asymmetry determination for blind stereoscopic image quality assessment strategy," *J. Visual Communication and Image Representation*, vol. 49, pp. 115–128, 2017.

[3] M. Chen, Che-Chun Su, Do-Kyoung Kwon, Lawrence K Cormack, and Alan C Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[4] R. Bensalma and Mohamed-Chaker Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing*, vol. 24, no. 2, pp. 281–316, 2013.

[5] J. You, Liyuan Xing, Andrew Perkis, and Xu Wang, "Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis," in *Proc. of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA*, 2010.

[6] Anush Krishna Moorthy and Alan Conrad Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[7] R. Blake, David H Westendorf, and Randall Overton, "What is suppressed during binocular rivalry?," *Perception*, vol. 9, no. 2, pp. 223–231, 1980.

[8] Aladine Chetouani, "Full reference image quality metric for stereo images based on cyclopean image computation and neural fusion," in *2014 IEEE Visual Communications and Image Processing Conference*. IEEE, 2014, pp. 109–112.

[9] A. Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau, "Quality assessment of stereoscopic images," *EURASIP journal on image and video processing*, vol. 2008, no. 1, pp. 1–13, 2009.

[10] P. Gorley and Nick Holliman, "Stereoscopic image quality metrics and compression," in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 680305–680305.

[11] CTER Hewage, Stewart T Worrall, Safak Dogan, and AM Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-d video," *Electronics letters*, vol. 44, no. 16, pp. 963–965, 2008.

[12] David G Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Ieee, 1999, vol. 2, pp. 1150–1157.

[13] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in computer vision*, pp. 726–740. Elsevier, 1987.

[14] Feng Qi, Debin Zhao, and Wen Gao, "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Transactions on multimedia*,

vol. 17, no. 12, pp. 2338–2344, 2015.

[15] Lin Ma, Xu Wang, Qiong Liu, and King Ngi Ngan, "Reorganized dct-based image representation for reduced reference stereoscopic image quality assessment," *Neurocomputing*, vol. 215, pp. 21–31, 2016.

[16] Jian Ma, Ping An, Liquan Shen, and Kai Li, "Reduced-reference stereoscopic image quality assessment using natural scene statistics and structural degradation," *IEEE Access*, vol. 6, pp. 2768–2780, 2017.

[17] R. Akhter, ZM Parvez Sazzad, Yuukou Horita, and Jacky Baltes, "No-reference stereoscopic image quality assessment," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 75240T–75240T.

[18] Wujie Zhou, Weiwei Qiu, and Ming-Wei Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of 3-d images," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 404–415, 2017.

[19] Meixin Fang and Wujie Zhou, "Toward an unsupervised blind stereoscopic 3d image quality assessment using joint spatial and frequency representations," *AEU-International Journal of Electronics and Communications*, vol. 94, pp. 303–310, 2018.

[20] Heeseok Oh, Sewoong Ahn, Jongyoo Kim, and Sanghoon Lee, "Blind deep s3d image quality evaluation via local to global feature aggregation," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4923–4936, 2017.

[21] Jiachen Yang, Kyohoon Sim, Xinbo Gao, Wen Lu, Qinggang Meng, and Baihua Li, "A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1314–1328, 2018.

[22] Jiachen Yang, Yang Zhao, Yinghao Zhu, Huifang Xu, Wen Lu, and Qinggang Meng, "Blind assessment for stereo images considering binocular characteristics and deep perception map based on deep belief network," *Information Sciences*, vol. 474, pp. 1–17, 2019.

[23] Jiahua Xu, Wei Zhou, Zhibo Chen, Suiyi Ling, and Patrick Le Callet, "Predictive auto-encoding network for blind stereoscopic image quality assessment," *arXiv preprint arXiv:1909.01738*, 2019.

[24] Oussama Messai, Fella Hachouf, and Zianou Ahmed Seghir, "Adaboost neural network and cyclopean view for no-reference stereoscopic image quality assessment," *Signal Processing: Image Communication*, p. 115772, 2020.

[25] Yun Liu, Jiachen Yang, Qinggang Meng, Zhihan Lv, Zhanjie Song, and Zhiqun Gao, "Stereoscopic image quality assessment method based on binocular combination saliency model," *Signal Processing*, vol. 125, pp. 237–248, 2016.

[26] John G Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision research*, vol. 20, no. 10, pp. 847–856, 1980.

[27] A. Chetouani, "Convolutional neural network and saliency selection for blind image quality assessment," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2835–2839.

[28] Ilyass Abouelaziz, Aladine Chetouani, Mohammed El Hassouni, Longin Jan Latecki, and Hocine Cherifi, "No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling," *Pattern Recognition*, vol. 100, pp. 107174, 2020.

[29] Mohamed Hamidi, Aladine Chetouani, Mohamed El Haziti, Mohammed El Hassouni, and Hocine Cherifi, "Blind robust 3d mesh watermarking based on mesh saliency and wavelet transform for copyright protection," *Information*, vol. 10, no. 2, 2019.

[30] Junle Wang, Matthieu Perreira Da Silva, Patrick Le Callet, and Vincent Ricordel, "Computational model of stereoscopic 3d visual saliency," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2151–2165, 2013.

[31] Yuming Fang, Zhenzhong Chen, Weisi Lin, and Chia-Wen Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3888–3901, 2012.

[32] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," *CoRR*, vol. abs/1404.5997, 2014.

[33] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[35] A.K. Moorthy, Che-Chun Su, Anish Mittal, and Alan Conrad Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013.

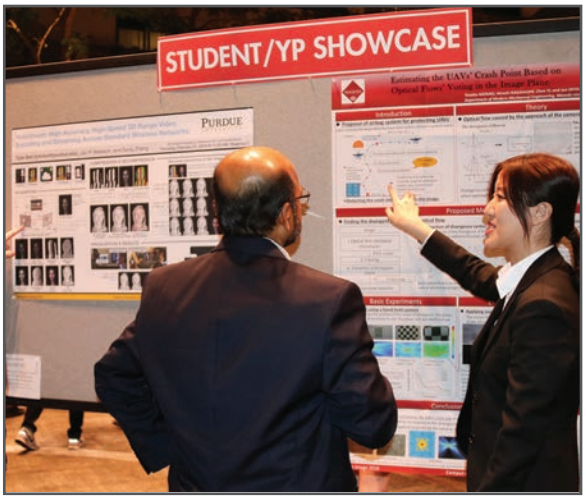
JOIN US AT THE NEXT EI!

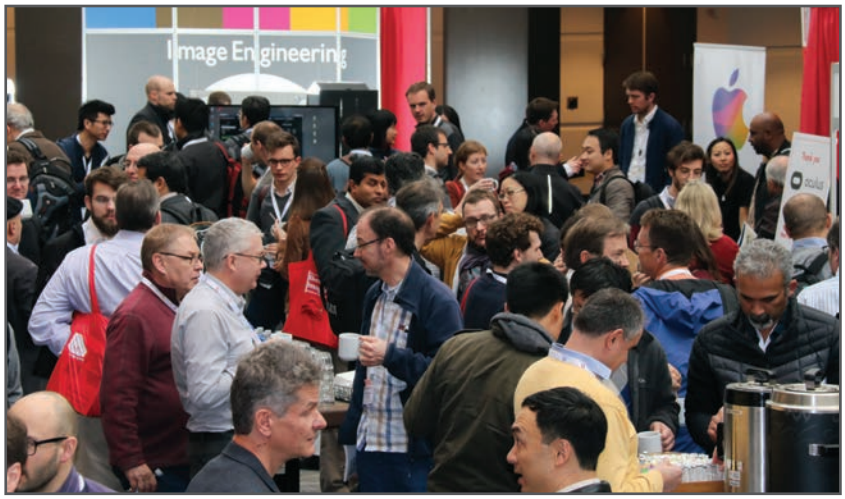
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org