ISSN 2007-9737

Experimental Assessment of Quality Metrics in Stereoscopic Imaging

Jesús Jaime Moreno Escobar[1], Oswaldo Morales Matamoros[1], Ricardo Tejeida Padilla[2], Hugo Quintana Espinosa[1]

[1] Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica,
Mexico

[2] Instituto Politécnico Nacional,
Escuela Superior de Turismo,
Mexico

jemoreno@esimez.mx

Abstract. The main objective of this work is to describe the state of the art of the Stereoscopic Image Quality Assessments (SIQA) from recent years, thus, it covers a compendium of models from 2007 to date. Furthermore, this paper summarizes 27 algorithms from 17 authors and their possible variations giving as a result of 280 stereoscopic metrics tested. This benchmarking is not only intended for researchers on the Image Quality Evaluation but also for researchers on the field on acquisition, processing and display of stereoscopic images. To this aim, we present not only a survey on image quality metrics but also psychophysical experiments on image databases available in this field. First, we sketch a general view of the importance of Stereoscopic Imaging. Thus, we propose different classifications in order to group the state of the art of SIQA. Then, we describe the performance of 280 metrics of SIQA using LIVE 3D Image Database. Results of algorithms are evaluated with the main purpose of being a reference for researchers in the Stereoscopic image quality field who want to perform further tests and proposing future models.

Keywords. Quality assessment databases, 3D image quality, stereoscopic image quality, JPEG2000 and stereoscopy.

1 Introduction

Stereoscopic coding and visualization systems are now an interesting field of research, but since the nineteenth century, several researchers presented some devices that displayed stereoscopic images. The Wendell's Stereoscope (Fig. 1(a), built in 1861) or View-Master (Fig. 1(b), commercially since 1935) are a good example of these devices. These devices captured two offset pictures (stereo-pair) separately showing left and right views to the appropriate eye of the observer. Then, when both views are combined into the brain, by means of stereopsis of binocular vision, they give the illusion of depth.

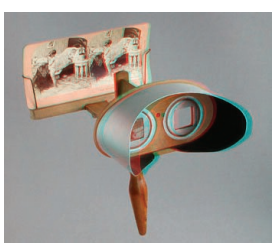
In general, these kind of stereoscopic past systems had 3 phases:

1. Capture (Coding),
2. Exposure (Decoding), and
3. Visualization (Displaying).

Nowadays, these stages have not changed, since they are just digitalized. Today, we capture stereoscopic images with digital cameras instead of plate cameras, we use polarized glasses instead of using anaglyph ones and the screen is no longer a piece of cardboard or reel, now it is a 3D television or 3D IMAX Screen.

Figure 2(c) shows the anaglyph image of *My heart is a jungle* (Fig. 2(a)), if this image is seen with anaglyph glasses, we could perceive it

ISSN 2007-9737



(a) Wendell's Stereoscope (1861)

(b) View-Master (1935)

Fig. 1. (a) Wendell's Stereoscope and (b) View-Master, images taken from [10] and [19], respectively

in different slices which depend on the apparent distance. This effect is similar like watching a version of its diorama, but digitalized(Figure 2(b)).

Maybe in future, stereoscopic scenarios with volumetric appearance and high quality will be part of everyday life, but now three-dimensional content is not simply a digital diorama, therefore the current challenge for researchers is to incorporate features based on the Human Visual System (HVS) for improving realism and immersiveness.

A general scheme of stereoscopic image quality assessing is depicted by Figure 3 (green block) and it is constituted by the following components:

1. Input: Stereoscopic Image, i.e. Left and Right views,
2. Process: Coding of Stereoscopic Algorithm ,
3. Output: Stereoscopic Representation of 3D Image, and
4. Feedback: Stereoscopic Image Quality Assessment (SIQA).

According to general systems theory, *Feedback* makes sure the efficiency of the *Process* [4]. So, the main objective of the SIQA is to quality assessment in the Left and Right views, then we can establish the degradation of the original stereo-pair. Coding of Stereoscopic Algorithm is to obtain the less possible degradation or fluctuations in the original source. Which is why, any kind stereoscopic image algorithm used in 3D projections, supports its evaluations employing a SIQA. So, we can recent evolution of stereoscopic algorithms is highly correlated with the evolution of the way to assess its visual quality. In this way, these methods of assessing stereoscopic quality have grown not only in number but also in importance.

The SIQA methods are based on the fact of evaluating two, or more, views of the same scene and the most of them employ either a 2D Image Quality Assessments (2DIQA) or slight adaptation of some HVS characteristics.

In this sense, it is reasonable to define SIQA as a variation of a 2DIQA, since in psychophysical experiments, the human observers subjectively assess the image quality [29, 13, 41] evaluating quality from a digital diorama, Figures 1(a) or 2(b), with lack of volumetric information of the scene.

SIQA algorithms can contribute to predict not only an objective response in general correlated with HVS but also the visual discomfort of an observer. In the early fifties, 3D Cinema was synonym of sickness and dizziness, sixty years later is related to blockbusters like *Avatar*[6].

Therefore, this paper is intended not only for SIQA researchers but also for researchers who study the visual discomfort or classical 2DIQA algorithms, since we classify and describe the most important features of SIQA algorithms and their combination with 2DIQA, resulting 280 metrics. Then, the main objective of this paper is to describe, in a general way, several SIQA algorithms in addition to compare them with the recent psychophysical experiments in the field. Which is why, we divided this work in four parts:

1. Current Stereoscopic Image Data Bases.
2. Classification of Stereoscopic Image Quality Assessments.
3. Description and Discussion of Stereoscopic Image Quality Assessments.
4. Exposition of Experimental Results.

(a) Original 2D Image. (b) Diorama seen sideways. (c) Anaglyph 3D Image.

Fig. 2. *My heart is a jungle* image, (a) and (b) images are taken from [2]

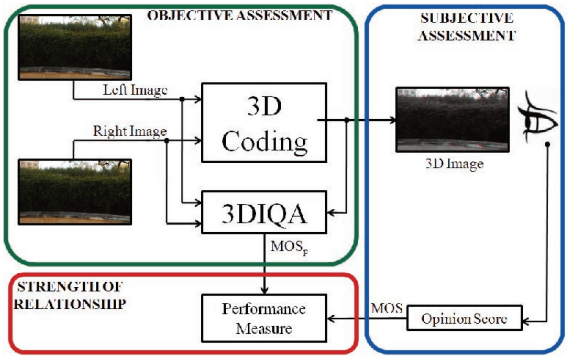


Fig. 3. Cybernetic diagram for describing a general system for stereoscopic image quality evaluation

2 Stereoscopic Image Data Bases

In SIQA field exists few image databases. Thus, in this section we highlight some features of the most used image databases in addition to mention some features of their psychophysical experiments, which contain standardized procedures from [20].

The stereoscopic images quality data obtained by these kind of experiments are based on the opinion score of an observer of individual quality judgments, which builds the database (Figure 3, blue block). Each attempt, the images are classified on a variable scale from excellent to bad. Then, making use of most common statistical metrics such as mean, standard deviation or variance, data are analyzed, giving as a result the Mean Opinion Scores (MOS). Different statistical procedures can be applied in the MOS of every stereoscopic image database psychophysical experiments, so we recommend to consult the citation for knowing them. Furthermore, MOS concentrates experimental results, which allow the description or comparison of any kind of stereoscopic assessment metric.

On the one hand we want to present three of the most used stereoscopic image databases:

— **LIVE 3D:** Laboratory for Image and Video Engineering of the University of Texas at Austin (USA) Stereoscopic image database the proposed by Moorthy et al. [29], available at *http://live.ece.utexas.edu*. Figure 4 shows the 20 Reference images used in this subjective study, shown only the left-views.

— **MMSPG:** Stereoscopic image database of the Multimedia Signal Processing Group of the École Polytechnique Fédérale de Lausanne (Switzerland) proposed by Goldmann et al. [13], Figure 5 shows four samples of reference images (only Left-views) used in this subjective study and it is available at:

http://mmspg.epfl.ch/.

— **FISE:** Stereoscopic image database of the Faculty of Information Science and Engineering of the Ningbo University (China) proposed by Wang et al.[41]. Figure 6 shows four samples of reference images (only

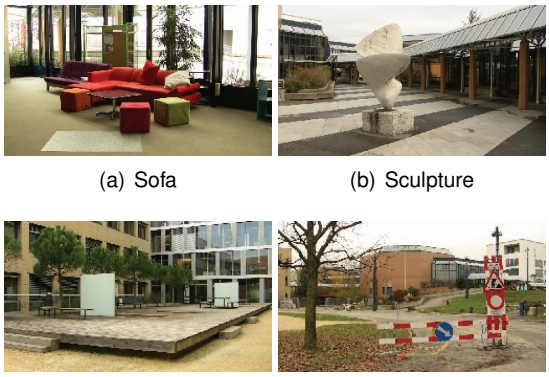
ISSN 2007-9737

Fig. 4. Left-images of the 20 source or reference views used in the subjective assessment (LIVE-3D) of Moorthy et al. [29]

Left-views) used in this subjective study. It is important to mention that [41] collected the imagery from the work made by [34], and they took just 2 of the 7 views that this image database originally had in order to have a stereo-pair [17] and it is available at: *http://vision.middlebury.edu/stereo/*

It is important to mention these three databases are not the only projects realized in the 3D/stereoscopic field. For example, [32] and [18] proposed another databases intended for improving the stereoscopic image quality. Both MMSPG and FISE image databases are not consistent with the size of reference and distorted images, since in some cases the size is different. So resizing images could change the precision of the subjective results. Thereby, this paper compares the psychophysical experiments of the LIVE 3D image database against a collected set of SIQA.

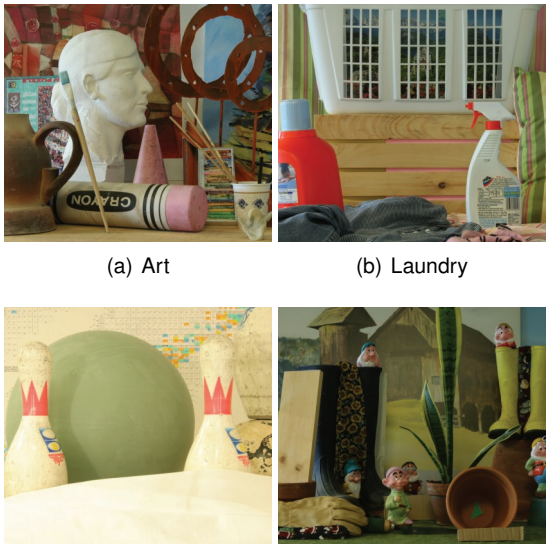
We want to mention their main features. Table 1 depicts the main characteristics of LIVE 3D, MMSPG and FISE image databases, where column *Features* refers to:



(a) Sofa (b) Sculpture

(c) Trees (d) Construction

Fig. 5. Left-views of sample images used in the subjective study (MMSPG) of Goldmann et al. [13]



(a) Art (b) Laundry

(c) Bowling1 (d) Dwarves

Fig. 6. Left-views of sample images used in the subjective study (FISE) of Wang et al. [41]

— Reference Images: Number of original or reference stereo-pairs.

— Distorted Images:Number of distorted stereo pairs.

— Format Images: Storage format of original and distorted stereo-pairs.

— Studio Images: Were the images captured in controlled conditions? Yes=Studio Images and No=Outdoor Images.

— Resolution: Size both for Reference and Distorted Images.

— Views: Views of the same scene.

— Distortions: Number of distortions, in the case of FISE are JPEG2000(JP2K), JPEG, Additive White Gaussian (WN) and Gaussian Blur (Blur). LIVE 3D use all these four noises in addition to Fast-Fading (FF).

— Observers: The subject pool considered in the study.

— Camera: Brand and model of the used camera.

— Capture Process: *Simultaneous* and *Not-Simultaneous* mean respectively that pictures were captured at the same time or not.

— Stereoscopic Display: Brand and model of the used stereoscopic display.

It is important to mention that this work is adequate to judge exclusively stereoscopic images containing so-called 2D artifacts, since all the artifacts of LIVE 3D, MMSPG and FISE image databases are 2D artifacts added separately and symmetrically to stereoscopic scene, i.e. the left and right images.

3 Classification of 3D/Stereoscopic Image Quality Assessments

Before describing a certain SIQA algorithm, we propose a classification. Note that there are many possible classifications and we just expose three of them.

Table 2 shows the 27 metrics from 17 authors described in this paper. It is important to highlight that we maintain the name of the image quality assessment that author gives it, in addition some of these author propose more than one metric. Independently of the author, a certain metric is

Table 1. Main features of LIVE 3D, MMSPG and FISE stereoscopic image databases

3D Image Database / Features	LIVE 3D	MMSPG	FISE
Reference Images	20	10	10
Distorted Images	365	100	20
Image Format	BMP	PNG	PNG
Studio Images	no	no	yes
Resolution	640×360	1920×1080	1390×1110
Views	2	10	2
Distortions	5	Not specified	4
Observers	32	20	20
Camera	$1\times$ Nikon D700	$2\times$ Canon HG20	$1\times$ Canon G1
Capture Process	Not-Simultaneous	Simultaneous	Not-Simultaneous
Stereoscopic Display	Viewsonic IZ3D	Hyundai S465D	Not specified

Table 2. Stereoscopic Image Quality Assessments

Algorithm	Metric
Akhter et al.[1]	$AkMOS_p$
Benoit et al.[3]	d_1
	d_2
	d_3
	Ddl_1
Bosc et al.[5]	Q_s
Campisi et al.[7]	A_v
	V_a
Chen et al.[9]	C_m
Gorley et al.[14]	$SBLC$
Gu et al.[15]	$ODDM_4$
Hewage et al.[16]	$PSNR_{edge}$
Jin et al.[21]	MSE_{ms}
	MSE_{dp}
Joveluro et al.[22]	PQM_{3D}
Mao et al.[24]	Q_{mao}
Shao et al.[35]	Q_{shao}
Shen et al.[38]	$HDPSNR$
Solh et al.[39]	$3VQM$
Yang et al.[48]	IQA
	SSA
You et al.[49]	$YouDMOS_p$
	OQ
	DQ_{map_1}
	DQ_{map_2}
	DQ_{map_3}
Zhu et al.[53]	e_i

referred by its name taking in to account in its corresponding row.

Thus, in this section all metrics in Table 2 are classified, then in section 4 they are described and in section 5 they are tested. Also, SIQA-$\mathcal{SET}$ will be call henceforth to the set of these 27 metrics.

Historically, several authors such as [45] describe the taxonomy of 2DIQA algorithms as follows:

— Full-Reference (FR): FR metrics gauge the quality of a presumably recovered or distorted image or view giving as a result a complete knowledge of the original or reference source.

— Reduced-Reference (RR): RR metrics predict the quality of a presumably recovered or distorted image or view giving as a result an incomplete knowledge of the original or reference source.

— No-Reference (NR): NR metrics assess the quality of a presumably recovered or distorted image or view without any knowledge of the original or reference source.

This classification cannot exactly be applied in the same from in the SIQA field, since it is impossible to obtain either original or distorted stereoscopic images, simply because they are perceived by a human observer. In this way, we are not able to obtain the cyclopean image that is

formed in the brain of the human beeing and we just can obtain information from left or right views of the so-called stereoscopic scene in order to predict other features, including the depth map.

Thus, our first classification is divide the SIQA-$\mathcal{SET}$ in global and local approaches. Global approaches take the information of whole image in order to compute the image quality, while local approaches measure the quality taking characteristics, features or information pixel by pixel or in certain regions of the image. So, once classified SIQA-$\mathcal{SET}$ would be as follows:

— Global Approaches: $AkMOS_p$, d_1, d_2, d_3, Q_s, A_v, V_a, C_m, $ODDM_4$, MSE_{dp}, Q_{shao}, $HDPSNR$, $3VQM$, IQA, SSA, $YouDMOS_p$, OQ, DQ_{map_1}, DQ_{map_2}, and DQ_{map_3}.

— Local Approaches: Ddl_1, $SBLC$, $PSNR_{edge}$, MSE_{ms}, PQM_{3D}, Q_{mao}, and e_i.

The second classification is based in the usage of the disparity map. Then, we can classify SIQA-$\mathcal{SET}$ in approaches which employ the disparity map and approaches which do not employ it. So, once classified SIQA-$\mathcal{SET}$ would be as follows:

— Approaches with disparity map: d_1, d_2, d_3, Ddl_1, Q_s, C_m, $PSNR_{edge}$, MSE_{ms}, MSE_{dp}, PQM_{3D}, $3VQM$, OQ, DQ_{map_1}, DQ_{map_2}, and DQ_{map_3} .

— Approaches without disparity map: $AkMOS_p$, A_v, V_a, $SBLC$, $ODDM_4$, Q_{mao}, Q_{shao}, $HDPSNR$, IQA, SSA, $YouDMOS_p$, and e_i.

Finally, the third classification is to divide the SIQA-$\mathcal{SET}$ in two groups, the first group combines features of 2D metrics, which be exchanged by any other, i.e. these kind of metrics could use interchangeably MSE or PSNR. In the second group are the metrics that do not use a 2D metric and can be consider as purely 3D image quality assessments. So, once classified SIQA-$\mathcal{SET}$ would be as follows:

— Approaches based on 2DIQA: d_1, d_2, d_3, A_v, V_a, $PSNR_{edge}$, MSE_{dp}, $YouDMOS_p$, and OQ.

— Stereoscopic Approaches: $AkMOS_p$, Ddl_1, Q_s, C_m, $SBLC$, $ODDM_4$, MSE_{ms}, PQM_{3D}, Q_{mao}, Q_{shao}, $HDPSNR$, $3VQM$, IQA, SSA, DQ_{map_1}, DQ_{map_2}, DQ_{map_3}, and e_i.

In Subsections 4.1 and 4.2 we describe all these metrics using this third classification. Table 3 shows an overview of the classification of all SIQA described in this paper.

Table 3. Overview of Classification of the SIQA-$\mathcal{SET}$

Metric	Information		Disparity Map		Base on	
	Global	Local	With	Without	2DIQA	SIQA
$AkMOS_p$	X			X		X
d_1	X		X		X	
d_2	X		X		X	
d_3	X		X		X	
Ddl_1		X	X			X
Q_s	X		X			X
A_v	X			X	X	
V_a	X			X	X	
C_m	X		X			X
$SBLC$		X		X		X
$ODDM_4$	X			X		X
$PSNR_{edge}$		X	X		X	
MSE_{ms}		X	X			X
MSE_{dp}	X		X		X	
PQM_{3D}		X	X			X
Q_{mao}		X		X		X
Q_{shao}	X			X		X
$HDPSNR$	X			X		X
$3VQM$	X		X			X
IQA	X			X		X
SSA	X			X		X
$YouDMOS_p$	X			X	X	
OQ	X		X		X	
DQ_{map_1}	X		X			X
DQ_{map_2}	X		X			X
DQ_{map_3}	X		X			X
e_i		X		X		X

4 Stereoscopic Image Quality Assessments

From Figure 3 green block, any metric of SIQA-$\mathcal{SET}$ assesses the stereoscopic image quality predicting the MOS or MOS_p. There are some algorithms, such as Q_s [5], which use a disparity map and left image for synthesizing the right image, so any SIQA needs two images with a certain disparity. In this section we describe two ways to divide the state of the art these SIQA algorithms.

So, it is important mentioning that for both *Metrics based on 2DIQA* and *Stereoscopic Metrics*

provided by the authors were coded ourselves in MatLab.

4.1 Metrics based on 2DIQA

The SIQA-$\mathcal{SET}$ was chosen based on their reported performance, in the same way we collected 29 2DIQA in order to provide a baseline of 2D metrics (2DIQA-$\mathcal{SET}$).

In the 2DIQA-$\mathcal{SET}$ we can find Statistical Image Quality Assessments (*St-IQA*), Full-Reference Image Quality Assessments (*FR-IQA*), and No-Reference Image Quality Assessments (*NR-IQA*). The fist twelve 2DIQA are part of the MetrixMux toolbox [40], while the rest of the metrics were collected from their respective authors.

On the one hand, for describing *St-IQA*, let $I(i,j)$ and $\hat{I}(i,j)$ be two images to be compared, being $I(i,j)$ the original reference or source image, which has to be considered with perfect and unquestionable quality and $\hat{I}(i,j)$ a distorted version of $I(i,j)$, whose quality in comparison to $I(i,j)$ is being evaluated.

In the other hand, both *FR-IQA* and *NR-IQA* algorithms just are listed in order to save space and we do not describe them here, so the reader is referred to the cited papers.

1. Mean-Squared Error (MSE, *St-IQA*), defined as:

$$MSE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[I(i,j) - \hat{I}(i,j)\right]^2. \quad (1)$$

2. Peak Signal-to-Noise Ratio (PSNR, *St-IQA*), defined as:

$$PSNR = 10\log_{10}\left(\frac{{\mathcal{I}_{max}}^2}{MSE}\right), \quad (2)$$

where ${\mathcal{I}_{max}}^2$ is the maximum possible error or the peak of the MSE value between $I(i,j)$ and $\hat{I}(i,j)$. That is, for an 8 bit per pixel (bpp) intensity image, ${\mathcal{I}_{max}}^2 = (2^8-1)^2 = 65025$ [43]. For 24bpp RBG images the PSNR is also defined by Eq. (2), where RGB MSE is the average of independent MSE values estimated in each 8bpp chromatic component, i.e. MSE_{Red}, MSE_{Green}, and MSE_{Blue}.

3. Structural Similarity Index (SSIM, *FR-IQA*), prosed by [47].
4. Multiscale SSIM Index (MSSIM, *FR-IQA*), prosed by [47].
5. Visual Signal-to-Noise Ratio (VSNR, *FR-IQA*), prosed by [8].
6. Visual Information Fidelity (VIF, *FR-IQA*), prosed by [44].
7. Pixel-Based VIF (VIFP, *FR-IQA*), prosed by [36].
8. Universal Quality Index (UQI, *FR-IQA*), prosed by [42].
9. Image Fidelity Criterion (IFC, *NR-IQA*), prosed by [37].
10. Noise Quality Measure (NQM, *FR-IQA*), prosed by [11].
11. Weighted Signal-to-Noise Ratio (WSNR, *FR-IQA*), prosed by [25].
12. Signal-to-Noise Ratio (SNR, *St-IQA*), defined as:

$$SNR = 10\log_{10}\frac{\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}[I(i,j)]^2}{MSE}. \quad (3)$$

13. Average Difference (AD, *St-IQA*), defined as:

$$AD = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} I(i,j) - \hat{I}(i,j). \quad (4)$$

14. Maximum Difference (MD, *St-IQA*), defined as:

$$MD = \max\left[I(i,j) - \hat{I}(i,j)\right]. \quad (5)$$

15. Normalized Absolute Error (NAE, *St-IQA*), defined as:

$$NAE = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left|I(i,j) - \hat{I}(i,j)\right|}{\sum_{i=1}^{N}\sum_{j=1}^{M} I(i,j)}. \quad (6)$$

16. Normalized Cross Correlation (NCC, *St-IQA*), defined as:

$$NCC = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(I(i,j)\times\hat{I}(i,j)\right)}{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(I(i,j)\right)^2}. \quad (7)$$

17. Structural Content (SC, *St-IQA*), defined as:

$$SC = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(I(i,j)\times\hat{I}(i,j)\right)}{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\hat{I}(i,j)\right)^2}. \quad (8)$$

18. Blind Image Quality Index (BIQI, *NR-IQA*), prosed by [28].
19. Blind/Referenceless Image Spatial Quality Evaluator Index (BRISQUE,), *NR-IQA*), prosed by [26].
20. Naturalness Image Quality Evaluator (NIQE, *NR-IQA*), prosed by [27].
21. No-Reference Peak Signal-to-Noise Ratio (NR-PSNR, *NR-IQA*), prosed by [31].
22. Perceptual Peak Signal-to-Noise Ratio (P^{2}SNR, *FR-IQA*), prosed by [30].
23. Feature-Similarity (FSIM, *FR-IQA*), prosed by [52].
24. Riesz-Transform Feature-Similarity (RFSIM, *FR-IQA*), prosed by [51].
25. Peak Signal-to-Noise Ratio with Contrast Sensitivity Function (PSNRHVSM, *FR-IQA*), prosed by [12].
26. JPEG Quality Score (JQS, *FR-IQA*), prosed by [46].
27. Practical Image Quality Metric (DCTEX, *FR-IQA*), prosed by [50].
28. Most Apparent Distortion (MAD, *FR-IQA*), prosed by [23].
29. Perceptual Quality Metric (PQM, *FR-IQA*), prosed by [22].

Once the 2DIQA-$\mathcal{SET}$ is defined, let us describe all the stereoscopic algorithms based on it.

— d_1, d_2, and d_3 are global disparity distortion measures described by Equation 9:

$$\begin{array}{l} d_1 = \mathcal{M}\cdot\sqrt{D_{dg}}, \\ d_2 = \mathcal{M}\cdot(D_{dg}), \\ d_3 = D_{dg}, \end{array} \quad (9)$$

where D_{dg} is computed using the correlation coefficient between the original disparity maps and the corresponding disparity maps processed after image degradation. $\mathcal{M}$ define the the averaged left and right image distortion measures.

— A_v and V_a are defined by Equation 10. For A_v algorithm, a 2DIQA is separately applied on the left and right views then for producing a single measure of stereoscopic assessment, the calculated quality results are averaged. While V_a uses some values in order to separately weight a predicted score on the left and right views , these weights are equal to 0.43 and 0.57 respectively, which are simmilar from the weights used in the central approach (0.5). The unequal weights in V_a can be helpful 3D image databases such as LIVE 3D, since stereoscopic images are taken sequentially with a 2D camera, thus there are often important differences between the left and the right views, due to movement between the left and right shots.

$$\begin{array}{l} A_v = \frac{1}{2}\left(2DIQA_{left} + 2DIQA_{right}\right), \\ V_a = 0.43\times 2DIQA_{left} + 0.57\times 2DIQA_{right}. \end{array} \quad (10)$$

— $PSNR_{edge}$ is an algorithm, which makes use of a specific kind of quality metric, where the reference or original image is barely employed, since for stereoscopic depth map transmission this algorithm only uses extracted information from edges. Different depth levels are represented by edges and contours of the depth map and quality evaluations can be used this information. This metric uses a *Sobel* filter both in original and distorted images in order to obtain four binary edge marks. Left and right edge marks are applied to the original and distorted stereo-pairs. Then, these

filtered source or original and presumably distorted stereo-pairs are tested using any image quality assessment of the 2DIQA-$\mathcal{SET}$. Finally, the individual results obtained both in original and distorted images are averaged.

$$PSNR_{edge} = f\left(2DIQA_{bem}\right). \tag{11}$$

Then, $PSNR_{edge}$ refers to full-reference 2DIQA rating for the depth map and $f\left(2DIQA_{bem}\right)$ refers to the 2DIQA quality rating for the side information (i.e. edge information/binary edge mask).

— MSE_{dp} measures the Mean Squared Error or any 2DIQA between the disparity difference between reference and distorted stereoscopic image. It is defined as:

$$MSE_{dp} = 2DIQA\left(DP_x, DP_y\right), \tag{12}$$

where DP_x, and DP_y are disparity map from reference and distorted image, respectively.

— $YouDMOS_p$, and OQ. The fist approach is obtained performing a nonlinear regression of a result of a certain metric (IQ) of the 2DIQA-$\mathcal{SET}$, using the following function:

$$YouDMOS_P = \frac{a_1}{1 + \exp\left[-a_2 \cdot (IQ - a_3)\right]}. \tag{13}$$

While the second approach, is a global combination, which computes two quality assessments of the distorted image. First, a result of a certain metric (IQ) and then, distorted disparity (DQ). This overall quality (OQ) is taken as the quality of the stereoscopic image using the following function:

$$OQ = a \cdot IQ^d + b \cdot DQ^e + c \cdot IQ^d + DQ^e, \tag{14}$$

where $a = 3.465$, $b = 0.002$, $c = -0.0002$, $d = -1.083$, and $e = 2.2$.

4.2 Stereoscopic Metrics

Similarly to the previous section, *Stereoscopic Metrics* are described by highlighting only the main features.

— $AkMOS_p$ is a no-reference perceptual quality assessment based on features local segmentation of artifacts and disparity, i.e. this metric extracts information from edge and non-edge areas in addition to evaluate blockiness based relative disparity estimation. $AkMOS_p$ is computed by the following equation:

$$AkMOS_p = \frac{4}{1 + \exp\left[-1.0217\left(\alpha \cdot D_Z + \beta \cdot B \cdot Z - 3\right)\right]} + 1, \tag{15}$$

where α and β are the model parameters, while D_Z, B and Z are the overall disparity feature, blockiness and zero crossing of each stereo-pair, respectively.

— Ddl_1 is obtained by evaluating Equation 17. The local SSIM measure map M_{map} is evaluated by measuring and fusing it with the distortion of the local disparity assessment using point-wise product. The disparity distortion is evaluated for each pixel p using the disparity map (DM) for both views (left and right) as follows (left view):

$$Ddl_{left} = M_{map\_left}\left(1 - \frac{\sqrt{DM_{or}\left(p\right)^2 - DM_{dg}\left(p\right)^2}}{255}\right). \tag{16}$$

Thus, Ddl1 is the average value of the N pixels of Ddl_{left} and Ddl_{right} maps and by averaging both results as follows:

$$Ddl1 = \frac{1}{2}\left(\frac{1}{N}\sum_N Ddl_{left} + \frac{1}{N}\sum_N Ddl_{right}\right). \tag{17}$$

— Q_s analyzes the location of the artifacts by means of masking images, which are obtained by evaluating the difference between original and synthesized views. Then, a threshold Th is applied in order to identify critical areas. Th is defined as follows:

$$Th = \frac{\max\left(I - I'\right)}{10}, \tag{18}$$

where I is the original image, and I' is the synthesized view. Then, this metric applies

SSIM measure only on these critical areas and the final score is the mean SSIM scores averaged by the number of pixels.

— C_m is a cyclopean image synthesized and disparity-compensated from stereo views and it is calculated by:

$$C_m(x,y) = W_L(x,y) \times I_L(x,y) + W_R[(x+d),y] \times I_R[(x+d),y], \quad (19)$$

where I_L and I_R are the left and right images respectively, and d is a disparity index that corresponds pixels from I_L to those in I_R. The weights W_L and W_R are calculated from the normalized Gabor filter magnitude responses.

— $SBLC$ or Stereo Band Limited Contrast employs RANSAC algorithm to extract regions with high spatial frequency both in the left and right images, then, matched points are found. Thus, Surrounding pixels of these points are calculated and pixels outside them are discarded. $SBLC$ is calculated as follows:

$$SBLC = \left(\frac{1}{p}\sum_{x=0}^{p}\frac{C_{Orig}(x)}{L_{Orig}}\right) - \left(\frac{1}{p}\sum_{x=0}^{p}\frac{C_{Comp}(x)}{L_{Comp}}\right), \quad (20)$$

where $C(x)$ is the the corresponding matched regions and then the average of matched regions founded both in the left and right images, L is the overall relative mean luminance, and p is the number of matched points in a certain region. $SBLC$ estimates $C(x)/L$ both in Original ($Orig$) and distorted ($Comp$) stereo-pairs.

— $ODDM_4$ is based on the ocular dominance theory and degree of parallax, the latter is calculated as follows:

$$d^{\theta} = \cos^{-1}\left(\frac{L(M)\cdot R(M)}{\|L(M)\|_2 \cdot \|R(M)\|_2}\right), \quad (21)$$

where L and R represent left and right views respectively, and M indicates the central region of the image.

Ocular dominance theory is introduced by predicting separately the values of left and right image quality, i.e. $Q_J(L)$ and $Q_J(R)$ are the JPEG Quality Score [46] of the left and right images. Hence, $ODDM_4$ is defined by:

$$ODDM_4 = Q_J(L) + Q_J(R) + d^{\theta}. \quad (22)$$

— MSE_{ms} is based on binocular human visual system, since it considers the cyclopean view and perceptibility of depth. This metric takes into account the masking effects of the Contrast Sensitive Function (CSF) and depth variability. Then, in order to differentiate the stereoscopic image structure, a downsampling into multi-scale images is applied to the left channel using a low pass filter. The size image is $H \times W$, while number of pixels in the i-th downsampled image is $(H \times W)/\left(2^{2i}\right)$. The final multi-scaled image of left channel is obtained as:

$$MSE_{ms} = \sum_{i}^{L} k_i MSE_{HVSM}\left(X_i^{left}, Y_i^{left}\right), \quad (23)$$

where X_i^{left} and Y_i^{left} are the i-th downsampled left image both in reference and distorted image, MSE_{HVSM} is MSE version of the metric PSNR-HVS-M proposed by [33]; k_i is a constant parameter.

— PQM_{3D} is based on Perceptual Quality Metric (PQM), which is developed to measure slight changes in noises inducted into an normal image. PQM is obtained subtracting the $PDM(f)$ from 1 and values less than 0 are equated to zero as the range of the metric is between 0 and 1, representing the worst and best qualities, respectively. The PQM is calculated as follows:

$$PQM = 1 - PDM(f), \quad (24)$$

and

$$PDM(f) = \frac{\sum_{t=1}^{T} W(t)\, PDM(t)}{\sum_{t=1}^{T} W(t)}, \quad (25)$$

where $PDM(t)$ is the distortion on all T block levels and finally weighted by the

weighting factor $W(t)$ to obtain the frame level Perceptual Distortion $PDM(f)$.

Thus, PQM_{3D} is the average of individual PQM qualities of color and depth images, rendered into left and right views.

— Q_{mao} is summarized in the following steps:

1. Find the gradient magnitudes using *Sobel* operator both on left and right channels of the original and the distorted images, respectively.
2. Determine thresholds on the left and right channel.
3. Classify into edge, texture, and smooth regions each pixel of left and right images of the original and distorted images.
4. Use SSIM in order to evaluate six individual quality assessments of the images obtained in step 3).
5. The final score Q_{mao} of the stereo-pair is the combination of the results of the previous step.

— Q_{shao} extracts distortion-specific features using only the distorted stereoscopic image. In this way, Q_{shao} is a two-phase feature fusion procedure, namely *Training phase* and *Test phase*. First, this metric employs four distortion categories (Gaussian blur, White noise, JPEG compression and JPEG2000 compression) in order to predict which kind of noise distorted the original stereo-pair. Then, a Support Vector Regression (SVR) is used to predict the relationship between stereoscopic features and subjective scores.

— $HDPSNR$ includes a value of stereo vision into the classical PSNR formula and it is expressed by:

$$HDPSNR = \left(10 \log \frac{255^2}{S}\right) dB, \tag{26}$$

where S is a vector summation or Minkowski summation, defined by:

$$S = \left(\sum_{n=1}^{N} |e_n|^2\right)^{\frac{1}{2}}, \tag{27}$$

here e_n is the difference of left and right images after they are decomposed in N contourlet subbands and weighted by the following contrast sensitivity function, being f is the spatial frequency:

$$CSF(f) = (0.205 + 0.511) \exp^{-0.204f}. \tag{28}$$

— $3VQM$ is a combination of three distortion measures:

1. Temporal Inconsistencies (TI),
2. Spatial Outliers (SO), and
3. Temporal Outliers (TO).

Thus, $3VQM$ is defined as follows:

$$3VQM = K\left[1 - SO\left(SO \cap TO\right)\right]^a \left(1 - TI\right)^b \left(1 - TO\right)^c, \tag{29}$$

where SO, TO, and TI are normalized to range 0 to 1 and a,b, and c are determined by training. K is a constant for scaling $3VQM$ ranges. $(SO \cap TO)$ avoids to take into account a the outlier distortion more than once. Finally, Equation 29 is apply to left and right images obtaining two $3VQM$ matrices, which are averaged.

— IQA, and SSA assess stereo images from the perspective of average image quality and stereo sense, respectively.

IQA is defined as the arithmetic mean of the Left and Right image gauged by PSNR as follows:

$$IQA = \frac{PSNR_L + PSNR_R}{2}, \tag{30}$$

SSA contains the absolute disparity image, namely it is the different information from the stereo-pair and is defined as:

$$SSA = PSNR_M = 10 \log \frac{255^2}{MSE_M}, \tag{31}$$

here MSE_M is as follows:

$$MSE_M = \frac{\sum_M \left[|R_O - L_O| - |R_P - L_P|\right]^2}{M}, \tag{32}$$

where R_O and L_O refers to original stereo-pairs, while R_P and L_P are the processed or recovered ones. M is the number of nonzero-pixels in the original absolute disparity image ($R_O - L_O$).

— DQ_{map_1}, DQ_{map_2}, and DQ_{map_3}. All these three approaches are local combinations, which compute a quality map of the disparity image resulting an approximate distribution of the degradation on the distorted disparity image. Equation 33 computes the quality map on the disparity image in three different ways.

$$\begin{aligned} DQ_{map_1} &= \left(D - \overline{D}\right)^2, \\ DQ_{map_2} &= \left|D - \overline{D}\right|, \\ DQ_{map_3} &= 1 - \frac{\sqrt{D^2 - \overline{D^2}}}{255}, \end{aligned} \tag{33}$$

where D and $\overline{D}$ denote the original disparity image and the distorted disparity image, respectively.

e_i is an overall visual quality measure between original stereo-pair (org) and the distorted stereo-pair (dst) and is calculated by:

$$e_i = \sum_k A_k \sqrt[4]{\sum_{\theta,x,y} \left| r^{org}_{k,\theta}(x,y) - r^{dst}_{k,\theta}(x,y) \right|^4} + B_z \sqrt[4]{\sum_{x,y} |r^{org}_z(x,y) - r^{dst}_z(x,y)|^4}, \tag{34}$$

where $r_{k,\theta}$ is the output through Human Visual System (HSV) formulated as a subset $s_{k,\theta}$ at scale k and phase θ after wavelet decomposition and r_z the perceptual response to depth of HVS. A_k and B_z are weight coefficients that are determined experimentally.

4.3 Discussion

Once SIQA-$\mathcal{SET}$ is described, we found some similarities among SIQA that we want to highlight.

In Equation 9, [3] defined $\mathcal{M}$, which is employed in d_1 and d_2, in the same way that [7] did for A_v (Equation 10) in addition to [48] for IQA (Equation 30). Other authors such as [49] for $YouDMOS_p$ (Equation 13) and OQ (Equation 14) or [3] for V_a (Equation 9) just modify these averaging algorithms either weighting individual qualities of the stereo-pair or performing a nonlinear regression.

In this sense, [21] proposed MSE_{dp} (Equation 12) in the same way that [3] did it for d_3 (Equation 9).

Other SIQA algorithms such as $PSNR_{edge}$ [16], Q_s [5], $SBLC$ [14], or Q_{mao} [24] use tools of image processing particularly within edge detection algorithms, since they employ Sobel operators, RANSAC algorithm or location of artifacts.

SSA [48] and $HDPSNR$ [38] modify the well-known algorithm PSNR using respectively either an absolute squared difference of stereo-pair or a contrast sensitivity function after a contourlet transformation.

In the case of DQ_{map_3}, [49] eliminated the parameter M_{map_left} of the Equation 16 of Ddl_1 [3] in order to improve its performance in local distortions such as JPEG or JPEG2000.

Q_{shao} [35] is the only algorithms that employs a phase of *Training*, which is time-consuming task and had to be repeated every time the image database is changed or another kind of distortion is introduced.

While $AkMOS_p$ [1], e_i [53], C_m [9], and PQM_{3D} [22] weight right and left views using some features of HSV such as Normalize Gabor Filters or responses of HSV in the Wavelet Domain. In the particular case of $AkMOS_p$, this metric performs a particular nonlinear regression including local segmentation of artifacts and disparity.

Finally, $ODDM_4$ [15], MSE_{ms} [21], and $3VQM$ [39] employ a perceptual 2DIQA, but not as a simple average, these metrics add another characteristics or weights such as degree of parallax, depth variability or temporal inconsistencies.

5 Experimental Results

The evaluation results of every observer group (MOS) and image quality metric (MOS_p) are normalized to the scope from 0 to 1 according to following equation:

$$\widetilde{MOS_p} = \frac{MOS_p - MOS_p^{\min}}{MOS_p^{\max} - MOS_p^{\min}}, \tag{35}$$

where MOS_p denotes the calculated value of each metric, $\widetilde{MOS_p}$ denotes the normalized value. $MOS_p^{\min}$ and $MOS_p^{\max}$ are minimum and maximum values, which are founded after predicting the image quality across all LIVE 3D image database, respectively. We also employ Equation 35 for normalizing MOS results of the LIVE 3D image database.

From Figure 3 red block, *Strength of Relationship*(SR) indicates how related are two effects to trend or not to the same response.

So, we compare SR and a normalized MOS or MOS_p giving as a result a performance measure (PM), Thus, we used the following PM's:

— Pearson's Linear Correlation Coefficient (LCC),

— Kendall's Rank Ordered Correlation Coefficient (KROCC),

— Spearman's Rank Ordered Correlation Coefficient (SROCC), and

— Root-Mean-Squared Error (RMSE).

In this way, we use two kind of non-parametric correlation SROCC and KROCC, but the most common indicator is SROCC. Pearson's Correlation is a linear measure for estimating SR, when parametric or same nature data are used. In some cases, results of image quality assessments have no linear relationship because they have not the same nature, which is why, it is not quite convenient to use Linear Correlation Coefficient, for example MSE and PSNR are the same assessment but the latter in logarithmic scale and regardless their same nature, LCC estimates different correlation.

Perfect o good correlation coefficient value with human perception is close to 1 for any correlation coefficient. Furthermore, for obtaining better performance or lower RMSE, the closer to zero the better.

Besides, we employ three ways for expressing our results:

— Scatter plots depict the relationship between subjective results (normalized MOS) and objective results (normalized MOS_p) of a certain SIQA, listed in Figures 7 to 12.

— Overall performance tables show the results of the strength of relationship of a part of SIQA (the best results per PM) across not only all LIVE 3D image database but also every single distortion, listed in Tables 4 to 10.

— Correlation performance tables show the results of the strength of relationship of the top-ten SIQA of all 17 authors, listed in Tables 11 to 14.

5.1 Metrics based on 2DIQA

In this subsection, we expose the results not only of SIQA based on 2DIQA but also of certain authors that propose more than one SIQA. We use $\leftarrow$ to refer that we applied a certain 2DIQA algorithm into a SIQA.

— d_1, d_2, d_3 and Ddl_1. We obtain 88 metrics after combing d_1, d_2, and d_3 with 2DIQA-$\mathcal{SET}$ in addition to perform Ddl_1. If we correlate these 88 variations with the all 365 images of the LIVE 3D image database we obtain the results of the Table 4, which shows that the best linear correlation is obtained by $d_1 \leftarrow FSIM$ (91.69%), see also Figure 7(a). While $d_2 \leftarrow UQI$ is the best ranking metric, since both in SROCC and KROCC, it obtains the best correlation with the human observers. Also, based on the results of $d_1 \leftarrow BRISQUE$ is clear that for the set of distortions considered, this metric is the most accurate. Considering just distortions in the field of image compression, JPEG2000 and JPEG, we can highlight $d_2 \leftarrow UQI$ is the best metric in either linear or rank (not-linear) correlation.

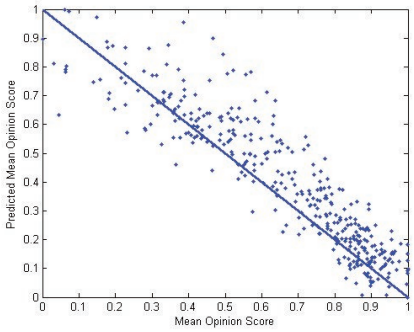
Table 4. Overall performance of the metrics proposed by [3] in predicting perceived stereoscopic image quality: Linear Correlation Coefficient (LCC), Spearman's Rank Ordered Correlation Coefficient (SROCC), Kendall's Rank Ordered Correlation Coefficient (KROCC) and Root Mean Squared Error (RMSE)

Distortion	SIQA	2DIQA	PM	Value
ALL	d_1	FSIM	LCC	0.9169
	d_2	UQI	SROCC	0.9335
	d_2	UQI	KROCC	0.7659
	d_1	BRISQUE	RMSE	0.1485
JP2K	d_2	UQI	LCC	0.9304
	d_2	UQI	SROCC	0.9104
	d_2	UQI	KROCC	0.7405
	d_3	MSE	RMSE	0.1583
JPEG	d_2	UQI	LCC	0.7620
	d_2	UQI	SROCC	0.7268
	d_2	UQI	KROCC	0.5212
	d_3	MSE	RMSE	0.1080
WN	Ddl_1	none	LCC	0.9330
	d_1	MSSIM	SROCC	0.9403
	d_1	MSE	KROCC	0.7861
	d_1	BRISQUE	RMSE	0.1001
Blur	d_2	UQI	LCC	0.9558
	d_2	UQI	SROCC	0.9306
	d_1	MSSIM	KROCC	0.7758
	d_1	BRISQUE	RMSE	0.1442
FF	d_2	UQI	LCC	0.8549
	d_2	UQI	SROCC	0.8162
	d_2	UQI	KROCC	0.6245
	d_2	NR-PSNR	RMSE	0.1116

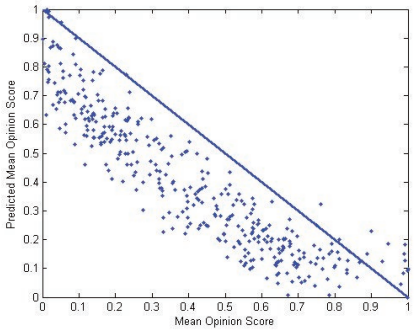
Figures 7(b) and 7(c) depict the scatter plots both for $d_2 \leftarrow UQI$ and $d_1 \leftarrow BRISQUE$, where we show the dispersion of the results obtained by these metrics.

— A_v and V_a. We could evaluate 58 variations of these two algorithms and we obtain the results of the Table 5, which shows that the best linear correlation is obtained by $A_v \leftarrow UQI$ (93.94%), see also Figure 8(a). In Figure 8(b), $A_v \leftarrow MAD$ is the best ranking metric, since both in SROCC and KROCC it obtains the best correlation with the human observers, in addition, this metric is the most accurate since it obtained the less RMSE.

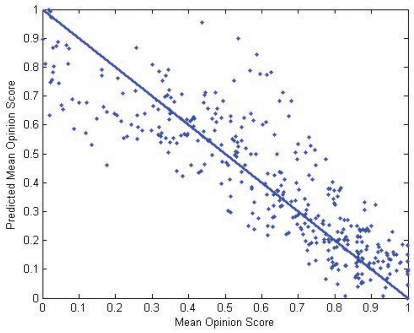
For JPEG2000 Distortion, $A_v \leftarrow MAD$ is the best ranking metric, i.e. to average right and left MAD qualities, if we change the weighted parameters of the stereo-pair, namely $V_a \leftarrow MAD$, we get the best ranking metric for JPEG distortion.



(a) $d_1 \leftarrow FSIM$



(b) $d_2 \leftarrow UQI$



(c) $d_1 \leftarrow BRISQUE$

Fig. 7. $MOS\ v_s\ MOS_p$ (both normalized). MOS_p is predicted by (a) d_1 using FSIM, (b) d_2 using UQI, and (c) d_1 using BRISQUE

— $PSNR_{edge}$. Despite our modification of this algorithm, which included using not only PSNR but also any metric of the 2DIQA-$\mathcal{SET}$ (Table 6), we obtain that $PSNR_{edge} \leftarrow NCC$ is slightly correlated with HVS, Figure 9,

ISSN 2007-9737

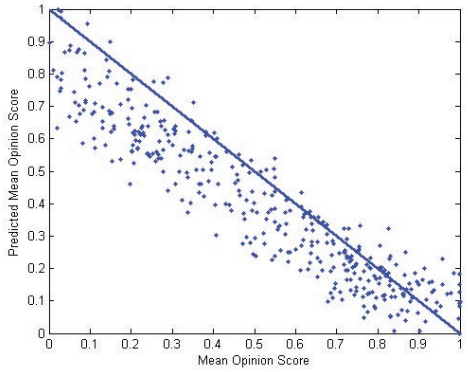
Table 5. Overall performance of the metrics proposed by [7] in predicting perceived stereoscopic image quality: Linear Correlation Coefficient (LCC), Spearman's Rank Ordered Correlation Coefficient (SROCC), Kendall's Rank Ordered Correlation Coefficient (KROCC) and Root Mean Squared Error (RMSE)

Distortion	SIQA	2DIQA	PM	Value
ALL	A_v	UQI	LCC	0.9371
	A_v	MAD	SROCC	0.9394
	A_v	MAD	KROCC	0.7772
	A_v	MAD	RMSE	0.0732
JP2K	A_v	UQI	LCC	0.9441
	A_v	MAD	SROCC	0.9247
	A_v	MAD	KROCC	0.7663
	A_v	MAD	RMSE	0.0630
JPEG	A_v	MAD	LCC	0.7686
	V_a	MAD	SROCC	0.7388
	V_a	MAD	KROCC	0.5408
	A_v	MAD	RMSE	0.0529
WN	A_v	MAD	LCC	0.9523
	A_v	MAD	SROCC	0.9497
	A_v	MAD	KROCC	0.8044
	A_v	MAD	RMSE	0.0805
Blur	A_v	MAD	LCC	0.9660
	A_v	MAD	SROCC	0.9537
	A_v	MAD	KROCC	0.8362
	V_a	BIQI	RMSE	0.0759
FF	A_v	UQI	LCC	0.8787
	A_v	UQI	SROCC	0.8328
	A_v	UQI	KROCC	0.6447
	A_v	MAD	RMSE	0.0861

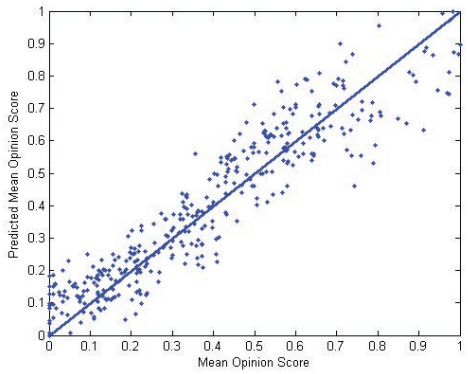
which incidentally is the best metric in terms of linear correlation. Furthermore, $PSNR_{edge} \leftarrow VIFP$ is the best overall ranking metric with about 15 percent less correlated with HVS than other metrics described previously.

— MSE_{ms}, and MSE_{dp}. From Table 7, when we average individual qualities of right and left depth maps, i.e. MSE_{dp}, it obtained the best results for LCC (Fig. 10(a)) and RMSE in all distortions. In this way, $MSE_{dp} \leftarrow UQI$ is highly linear correlated with the opinion of observers for the included image compression distortions. The ranking obtained by MSE_{ms} is better correlated with HVS than the one gotten by MSE_{dp}, across all kind of distortions of LIVE 3D.

Figure 10(b) shows that not always an excellent ranking means accuracy of results, even the difference between estimations is



(a) $A_v \leftarrow UQI$



(b) $A_v \leftarrow MAD$

Fig. 8. MOS v_s MOS_p (both normalized). MOS_p is predicted by (a) A_v using UQI, and (b) A_v using MAD

important, MSE_{ms} ranks the 90% of the results (SROCC) in the same order that an observer could do it.

— $YouDMOS_p$, OQ, DQ_{map_1}, DQ_{map_2}, and DQ_{map_3}. UQI measures the degree of linear correlation between original and distorted signals [42], when it is combined with a nonlinear regression of average of the quality of single views, using Equation 13, we found the best results in all distortions, except White Noise.

The scatter plot of Figure 11(a) depicts that $YouDMOS_p$ results tend to be more

Table 6. Overall performance of the metric proposed by [16] in predicting perceived stereoscopic image quality: Linear Correlation Coefficient (LCC), Spearman's Rank Ordered Correlation Coefficient (SROCC), Kendall's Rank Ordered Correlation Coefficient (KROCC) and Root Mean Squared Error (RMSE)

Distortion	SIQA	2DIQA	PM	Value
ALL	$PSNR_{edge}$	NCC	LCC	0.5772
	$PSNR_{edge}$	VIFP	SROCC	0.7976
	$PSNR_{edge}$	VIFP	KROCC	0.5958
	$PSNR_{edge}$	AD	RMSE	0.2084
JP2K	$PSNR_{edge}$	NQM	LCC	0.7749
	$PSNR_{edge}$	NQM	SROCC	0.8045
	$PSNR_{edge}$	NQM	KROCC	0.6082
	$PSNR_{edge}$	SC	RMSE	0.1641
JPEG	$PSNR_{edge}$	NQM	LCC	0.4916
	$PSNR_{edge}$	NQM	SROCC	0.4860
	$PSNR_{edge}$	NQM	KROCC	0.3376
	$PSNR_{edge}$	VSNR	RMSE	0.0913
WN	$PSNR_{edge}$	NQM	LCC	0.7999
	$PSNR_{edge}$	VIFP	SROCC	0.8616
	$PSNR_{edge}$	VIFP	KROCC	0.6652
	$PSNR_{edge}$	NAE	RMSE	0.2084
Blur	$PSNR_{edge}$	SSIM	LCC	0.8114
	$PSNR_{edge}$	NQM	SROCC	0.8385
	$PSNR_{edge}$	NQM	KROCC	0.6646
	$PSNR_{edge}$	NAE	RMSE	0.1156
FF	$PSNR_{edge}$	NCC	LCC	0.7738
	$PSNR_{edge}$	NCC	SROCC	0.7022
	$PSNR_{edge}$	NCC	KROCC	0.5188
	$PSNR_{edge}$	AD	RMSE	0.1449

Table 7. Overall performance of the metrics proposed by [21] in predicting perceived stereoscopic image quality: Linear Correlation Coefficient (LCC), Spearman's Rank Ordered Correlation Coefficient (SROCC), Kendall's Rank Ordered Correlation Coefficient (KROCC) and Root Mean Squared Error (RMSE)

Distortion	SIQA	2DIQA	PM	Value
ALL	MSE_{dp}	UQI	LCC	0.7962
	MSE_{ms}	none	SROCC	0.8952
	MSE_{ms}	none	KROCC	0.7022
	MSE_{dp}	BIQI	RMSE	0.1754
JP2K	MSE_{dp}	UQI	LCC	0.8512
	MSE_{ms}	none	SROCC	0.8608
	MSE_{ms}	none	KROCC	0.6620
	MSE_{dp}	MSE	RMSE	0.1583
JPEG	MSE_{dp}	UQI	LCC	0.5769
	MSE_{dp}	UQI	SROCC	0.5779
	MSE_{dp}	UQI	KROCC	0.4085
	MSE_{dp}	VSNR	RMSE	0.1031
WN	MSE_{dp}	UQI	LCC	0.8832
	MSE_{ms}	none	SROCC	0.9310
	MSE_{ms}	none	KROCC	0.7665
	MSE_{dp}	BIQI	RMSE	0.1130
Blur	MSE_{dp}	FSIM	LCC	0.8531
	MSE_{ms}	none	SROCC	0.9318
	MSE_{ms}	none	KROCC	0.7717
	MSE_{dp}	BPSNR	RMSE	0.2192
FF	MSE_{dp}	PSNRHVSM	LCC	0.7244
	MSE_{ms}	none	SROCC	0.6859
	MSE_{ms}	none	KROCC	0.4998
	MSE_{dp}	JQS	RMSE	0.1766

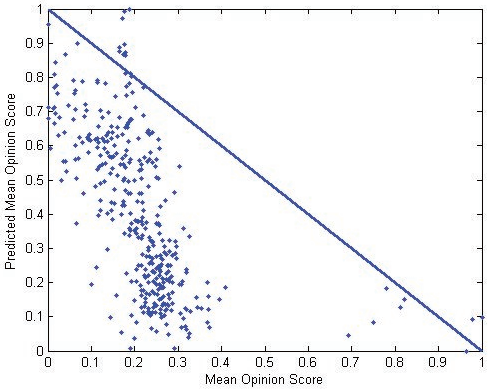


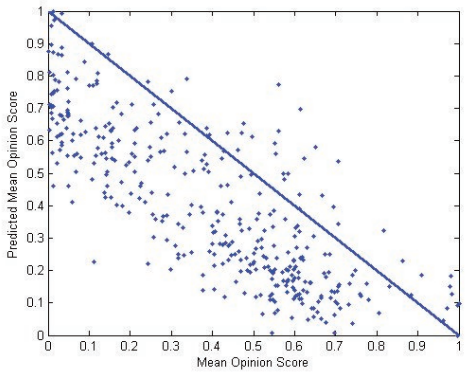
Fig. 9. $MOS\ v_s\ MOS_p$ (both normalized). MOS_p is predicted by $PSNR_{edge}$ using NCC

concentrated along perfect correlation. In brief, Table 8 shows that $YouDMOS_p \leftarrow UQI$ performs better results in terms of any kind of correlation coefficient but it is not as accurate as DQ_{map_2} and viceversa, which is why in Figure 11(b) the results of DQ_{map_2} are more dispersed and its results are closer to the perfect result than $YouDMOS_p$ results.

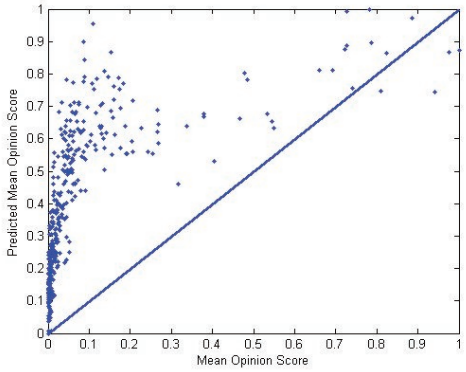
5.2 Stereoscopic Metrics

In this subsection, we sketch only the results of metrics that do are not based on a normal metric or they just are based in one feature of a certain normal image quality assessment.

Table 9 just contains the results of all stereoscopic metrics exposed in subsection 4.2. Where Ddl_1 linear correlates in 86.29%, Figure 12(a), being the assessment that estimates the best Linear Correlation. Furthermore, MSE_{ms} is the best ranked metric obtaining the best results both in SROCC and KROCC with 89.52% and 70.22%, respectively.



(a) $MSE_{dp} \leftarrow UQI$



(b) MSE_{ms}

Fig. 10. $MOS\ v_s\ MOS_p$ (both normalized). MOS_p is predicted by (a) MSE_{dp} using UQI, and (b) MSE_{ms}

Also, DQ_{map_2} is the most precise stereoscopic assessment for all whole of considered noises and image compression noises such as JPEG2000 and JPEG distortions.

Taking in to account only noises produced by a image compression coder, DQ_{map_3} is best metric in LCC and SROCC. Figure 12(b) depicts dispersion of the 365 results of DQ_{map_3}.

5.3 ALL SIQA-$\mathcal{SET}$

Regarding the overall experimental results, Table 10 depicts the performance of all SIQA of the SIQA-$\mathcal{SET}$ exposed in section 4.

Table 8. Overall performance of the metrics proposed by [49] in predicting perceived 3D image quality: Linear Correlation Coefficient (LCC), Spearman's Rank Ordered Correlation Coefficient (SROCC), Kendall's Rank Ordered Correlation Coefficient (KROCC) and Root Mean Squared Error (RMSE)

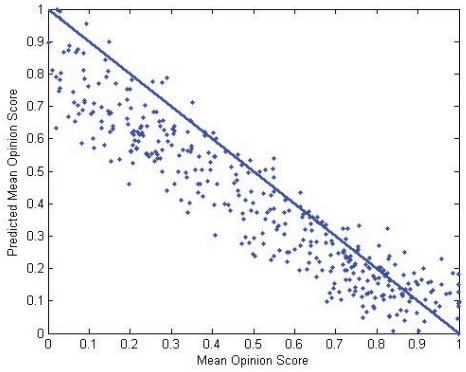
Distortion	SIQA	2DIQA	PM	Value
ALL	$YouDMOS_p$	UQI	LCC	0.9371
	$YouDMOS_p$	UQI	SROCC	0.9372
	$YouDMOS_p$	UQI	KROCC	0.7722
	DQ_{map_2}	none	RMSE	0.1289
JP2K	$YouDMOS_p$	UQI	LCC	0.9441
	$YouDMOS_p$	UQI	SROCC	0.9095
	$YouDMOS_p$	UQI	KROCC	0.7405
	DQ_{map_2}	none	RMSE	0.0961
JPEG	$YouDMOS_p$	UQI	LCC	0.7678
	$YouDMOS_p$	UQI	SROCC	0.7383
	$YouDMOS_p$	UQI	KROCC	0.5358
	DQ_{map_2}	none	RMSE	0.0742
WN	$YouDMOS_p$	SSIM	LCC	0.9326
	OQ	MSSIM	SROCC	0.9425
	OQ	MSSIM	KROCC	0.7911
	$YouDMOS_p$	NAE	RMSE	0.1237
Blur	$YouDMOS_p$	UQI	LCC	0.9517
	$YouDMOS_p$	MSSIM	SROCC	0.9282
	$YouDMOS_p$	AD	KROCC	0.7818
	OQ	VSNR	RMSE	0.2097
FF	$YouDMOS_p$	UQI	LCC	0.8787
	$YouDMOS_p$	UQI	SROCC	0.8328
	$YouDMOS_p$	UQI	KROCC	0.6447
	DQ_{map_2}	none	RMSE	0.1498

Thus, Figure 11(a) shows the metric $YouDMOS_p \leftarrow UQI$, this assessment linear correlates in 93.71% and obtains the best Linear Correlation.

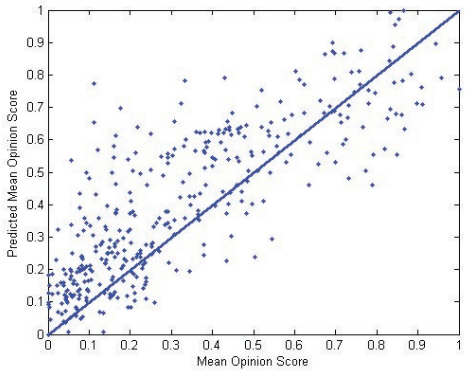
Furthermore, the best ranked metric is $A_v \leftarrow MAD$, because it is best correlated both in SROCC and KROCC with 93.94% and 77.72%, respectively.

Also, we can say that $A_v \leftarrow MAD$ is the most precise algorithm for all set of distortions considered and JPEG2000 and JPEG noises. Regarding only these image compression distortions, $A_v \leftarrow MAD$ is the best ranking metric, i.e. to average right and left MAD qualities, if we change the weighted parameters of the stereo-pair, namely $V_a \leftarrow MAD$, we get the best ranking metric for JPEG distortion.

In this paper we have presented several metrics (280) for gauging the quality of a stereo-pair intended for researchers interested



(a) $YouDMOS_p \leftarrow UQI$



(b) DQ_{map_2}

Fig. 11. $MOS\ vs\ MOS_p$ (both normalized). MOS_p is predicted by (a) $YouDMOS_p$ using UQI, and (b) DQ_{map_2}

in stereoscopic coding, visual discomfort or stereoscopic displaying. These researchers could want to found the best metric taking into account a certain respond in a certain distortion, for example.

The selected metric could be the best overall measure, which does not mean it would obtain the best results in all individual distortions. For those researchers who are interested in knowing the behavior of the top-ten overall SIQA-$\mathcal{SET}$, we propose Tables 11 to 14, which were made considering the following aspects:

Table 9. Overall performance of the *Stereoscopic Metrics* in predicting perceived stereoscopic image quality: Linear Correlation Coefficient (LCC), Spearman's Rank Ordered Correlation Coefficient (SROCC), Kendall's Rank Ordered Correlation Coefficient (KROCC) and Root Mean Squared Error (RMSE)

Distortion	SIQA	PM	Value
ALL	Ddl_1	LCC	0.8629
	MSE_{ms}	SROCC	0.8952
	MSE_{ms}	KROCC	0.7022
	DQ_{map_2}	RMSE	0.1289
JP2K	DQ_{map_3}	LCC	0.8604
	DQ_{map_3}	SROCC	0.8638
	MSE_{ms}	KROCC	0.6620
	DQ_{map_2}	RMSE	0.0961
JPEG	DQ_{map_3}	LCC	0.5750
	DQ_{map_3}	SROCC	0.5439
	DQ_{map_3}	KROCC	0.3806
	DQ_{map_2}	RMSE	0.0742
WN	Ddl_1	LCC	0.9330
	Ddl_1	SROCC	0.9380
	Ddl_1	KROCC	0.7829
	$ODDM_4$	RMSE	0.1079
Blur	Ddl_1	LCC	0.9056
	MSE_{ms}	SROCC	0.9318
	MSE_{ms}	KROCC	0.7717
	$ODDM_4$	RMSE	0.1304
FF	$ODDM_4$	LCC	0.7540
	$ODDM_4$	SROCC	0.7734
	$ODDM_4$	KROCC	0.5783
	$ODDM_4$	RMSE	0.1490

— Each Table represents only one Performance Measure, either LCC, SROCC, KROCC, or RMSE.

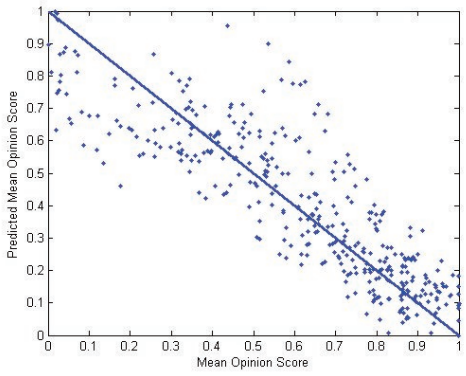
— In order to rank the 17 authors, we just chose the best metric (across all images in LIVE 3D) for those authors who proposed more than one SIQA.

— Regards overall performance, we eliminated the seven metrics that obtained the worst effects.

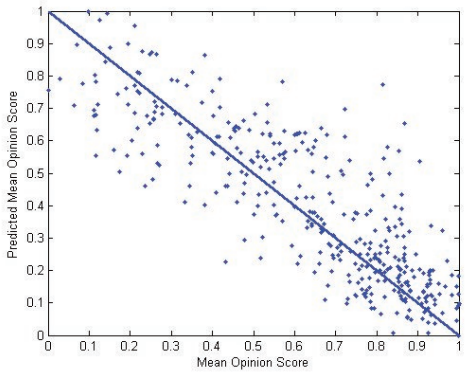
— Once we obtained the top-ten we indicated with **Bold** text the best metric, while with *italic* text the second best metric.

— Finally, we sorted alphabetically the top-ten by author name.

From Table 11 $A_v \leftarrow UQI$ (Fig. 8(a)) and



(a) Ddl_1



(b) DQ_{map_3}

Fig. 12. $MOS\ v_s\ MOS_p$ (both normalized). MOS_p is predicted by (a) Ddl_1, and (b) DQ_{map_3}

$YouDMOS_p \leftarrow UQI$ (Fig. 11(a)) obtained the best results in linear correlation coefficient not only in overall performance (93.71%) but also in all individual distortions, except in White Noise. For White Noise $d_1 \leftarrow FSIM$ (Fig. 7(a)) is the best metric with 93.07%.

Table 12 shows the performance across SIQA-$\mathcal{SET}$ in estimating subjective 3D/stereoscopic image quality using SROCC. Where $A_v \leftarrow MAD$(Fig. 8(b)) obtained the best results in overall performance (93.94%), JPEG2000 (92.47%), White Noise (94.97%), and Gaussian Blur (95.37%). While $YouDMOS_p \leftarrow UQI$ (Fig. 11(a)) got the best results in JPEG (73.83%), and Fast Fading distortion (83.28%).

Table 10. Overall performance across SIQA-$\mathcal{SET}$ in predicting perceived 3D image quality: Linear Correlation Coefficient (LCC), Spearman's Rank Ordered Correlation Coefficient (SROCC), Kendall's Rank Ordered Correlation Coefficient (KROCC) and Root Mean Squared Error (RMSE)

Distortion	SIQA	2DIQA	PM	Value
ALL	$YouDMOS_p$	UQI	LCC	0.9371
	A_v	MAD	SROCC	0.9394
	A_v	MAD	KROCC	0.7772
	A_v	MAD	RMSE	0.0732
JP2K	$YouDMOS_p$	UQI	LCC	0.9441
	A_v	MAD	SROCC	0.9247
	A_v	MAD	KROCC	0.7663
	A_v	MAD	RMSE	0.0630
JPEG	A_v	MAD	LCC	0.7686
	V_a	MAD	SROCC	0.7388
	V_a	MAD	KROCC	0.5408
	A_v	MAD	RMSE	0.0529
WN	A_v	MAD	LCC	0.9523
	A_v	MAD	SROCC	0.9497
	A_v	MAD	KROCC	0.8044
	A_v	MAD	RMSE	0.0805
Blur	A_v	MAD	LCC	0.9660
	A_v	MAD	SROCC	0.9537
	A_v	MAD	KROCC	0.8362
	V_a	BIQI	RMSE	0.0759
FF	A_v	UQI	LCC	0.8787
	A_v	UQI	SROCC	0.8328
	A_v	UQI	KROCC	0.6447
	A_v	MAD	RMSE	0.0861

From Table 13 $A_v \leftarrow MAD$ (Fig. 8(b)) obtained the best results in linear correlation coefficient not only in overall performance (77.72%) but also in all individual distortions, except in Fast Fading distortion. For Fast Fading distortion $YouDMOS_p \leftarrow UQI$ (Fig. 11(a)) is the best metric with 64.47%.

Table 14 shows the performance across SIQA-$\mathcal{SET}$ in prediction of the perceived stereoscopic image quality using Root Mean Squared Error. Where $A_v \leftarrow MAD$ (Fig. 8(b)) obtained the best results not only in overall performance (0.0732), but also in JPEG2000 (0.0630), JPEG (0.0529), White Noise (0.0805), Gaussian Blur (0.0919), and Fast Fading distortion (0.0861).

Table 11. Performance across SIQA-$\mathcal{SET}$ in predicting perceived stereoscopic image quality: Linear Correlation Coefficient (LCC). **Bold** indicates the best metric, while *italics* the second best

Author	SIQA	2DIQA	JP2K	JPEG	WN	Blur	FF	ALL
Benoit et al.[3]	d_1	FSIM	*0.9119*	*0.6259*	**0.9307**	*0.9358*	*0.7834*	*0.9169*
Bosc et al.[5]	Q_s	*none*	0.0259	0.1563	0.8866	0.1853	0.0882	0.4115
Campisi et al.[7]	A_v	UQI	**0.9441**	**0.7678**	0.9199	**0.9517**	**0.8787**	**0.9371**
Gu et al.[15]	$ODDM_4$	*none*	0.7728	0.4461	0.9223	0.7024	0.7540	0.7460
Hewage et al.[16]	$PSNR_{edge}$	NCC	0.6737	0.3293	0.7997	0.8027	0.7738	0.5772
Jin et al.[21]	MSE_{dp}	UQI	0.8512	0.5769	0.8832	0.8523	0.6327	0.7962
Joveluro et al.[22]	PQM_{3D}	*none*	0.1393	0.2415	0.8477	0.0444	0.1765	0.4790
Mao et al.[24]	Q_{mao}	*none*	0.7189	0.1290	0.7701	0.7527	0.4413	0.7082
Yang et al.[48]	IQA	*none*	0.7665	0.1187	*0.9244*	0.7690	0.6993	0.7002
You et al.[49]	$YouDMOS_p$	UQI	**0.9441**	**0.7678**	0.9199	**0.9517**	**0.8787**	**0.9371**

Table 12. Performance across SIQA-$\mathcal{SET}$ in predicting perceived stereoscopic image quality: Spearman's Rank Ordered Correlation Coefficient (SROCC). **Bold** indicates the best metric, while *italics* the second best

Author	SIQA	2DIQA	JP2K	JPEG	WN	Blur	FF	ALL
Benoit et al.[3]	d_2	UQI	*0.9104*	0.7268	0.9248	0.9306	*0.8162*	0.9335
Campisi et al.[7]	A_v	MAD	**0.9247**	*0.7364*	**0.9497**	**0.9537**	0.7720	**0.9394**
Gorley et al.[14]	$SBLC$	*none*	0.6744	0.4431	0.6219	0.6229	0.2133	0.5963
Gu et al.[15]	$ODDM_4$	*none*	0.8131	0.4202	0.9206	0.6577	0.7734	0.7223
Hewage et al.[16]	$PSNR_{edge}$	VIFP	0.7802	0.2360	0.8616	0.7958	0.5027	0.7976
Jin et al.[21]	MSE_{ms}	AD	0.8608	0.4484	0.9310	*0.9318*	0.6859	0.8952
Joveluro et al.[22]	PQM_{3D}	*none*	0.0239	0.1329	0.9167	0.1398	0.3360	0.2667
Mao et al.[24]	Q_{mao}	*none*	0.7460	0.1629	0.7790	0.6279	0.3599	0.7253
Yang et al.[48]	IQA	*none*	0.7993	0.1212	*0.9316*	0.9020	0.5875	0.8340
You et al.[49]	$YouDMOS_p$	UQI	0.9095	**0.7383**	0.9255	0.9252	**0.8328**	*0.9372*

Table 13. Performance across SIQA-$\mathcal{SET}$ in predicting perceived stereoscopic image quality: Kendall's Rank Ordered Correlation Coefficient (KROCC). **Bold** indicates the best metric, while *italics* the second best

Author	SIQA	2DIQA	JP2K	JPEG	WN	Blur	FF	ALL
Benoit et al.[3]	d_2	UQI	*0.7405*	0.5212	0.7570	0.7697	*0.6245*	0.7659
Campisi et al.[7]	A_v	MAD	**0.7663**	**0.5377**	**0.8044**	**0.8362**	0.5909	**0.7772**
Gorley et al.[14]	$SBLC$	*none*	0.4608	0.3065	0.4468	0.4141	0.1510	0.4201
Gu et al.[15]	$ODDM_4$	*none*	0.6089	0.2666	0.7456	0.5071	0.5783	0.5284
Hewage et al.[16]	$PSNR_{edge}$	VIFP	0.5899	0.1596	0.6652	0.6222	0.3694	0.5958
Jin et al.[21]	MSE_{ms}	AD	0.6620	0.2869	*0.7665*	*0.7717*	0.4998	0.7022
Joveluro et al.[22]	PQM_{3D}	*none*	0.0152	0.0893	0.7473	0.0929	0.2263	0.1869
Mao et al.[24]	Q_{mao}	*none*	0.5418	0.1216	0.5791	0.4525	0.2535	0.5294
Yang et al.[48]	IQA	*none*	0.5918	0.0735	0.7665	0.7333	0.4168	0.6296
You et al.[49]	$YouDMOS_p$	UQI	0.7405	*0.5358*	0.7570	0.7636	**0.6447**	*0.7722*

Table 14. Performance across SIQA-$\mathcal{SET}$ in predicting perceived stereoscopic image quality: Root Mean Squared Error (RMSE). **Bold** indicates the best metric, while *italics* the second best

Author	SIQA	2DIQA	JP2K	JPEG	WN	Blur	FF	ALL
Benoit et al.[3]	d_1	BRISQUE	0.1935	0.1750	*0.1001*	0.1442	*0.1278*	0.1485
Campisi et al.[7]	A_v	MAD	**0.0630**	**0.0529**	**0.0805**	**0.0919**	**0.0861**	**0.0732**
Gorley et al.[14]	$SBLC$	*none*	0.2980	0.1923	0.4733	0.3872	0.5295	0.3750
Gu et al.[15]	$ODDM_4$	*none*	0.1041	0.1654	0.1079	*0.1304*	0.1490	0.1315
Hewage et al.[16]	$PSNR_{edge}$	AD	0.1995	0.1875	0.2949	0.2203	0.1449	0.2084
Jin et al.[21]	MSE_{dp}	BIQI	0.1840	0.1726	0.1130	0.2714	0.1778	0.1754
Mao et al.[24]	Q_{mao}	*none*	0.3899	0.4327	0.3921	0.3730	0.3014	0.3783
Shen et al.[38]	$HDPSNR$	*none*	0.2035	0.2237	0.2981	0.2162	0.2549	0.2415
Yang et al.[48]	IQA	*none*	0.2081	0.0931	0.4136	0.3473	0.4275	0.2932
You et al.[49]	DQ_{map_2}	VSNR	*0.0961*	*0.0742*	0.1273	0.2506	0.1498	*0.1289*

6 Conclusions and Future Work

This paper describes 27 algorithms SIQA exposed by 17 authors, summarizing the research made in 3D/stereoscopic image quality field in the recent years. Nine metrics of this SIQA-$\mathcal{SET}$ can be combined with any 2DIQA, then they were separated from rest. These metrics were defined as *Metrics based on 2DIQA* and they were tested with 29 2DIQA, having a total of 262. Thus, we considered that the remaining 18 metrics were grouped as *Stereoscopic Metrics*.

For *Metrics based on 2DIQA* $YouDMOS_p$ using UQI got the best linear correlation, 94%, with the opinion of an observer, same percentage obtained by A_v using MAD but employing a no-linear correlation. For *Stereoscopic Metrics*, Ddl_1 got the best linear correlation (86%) with the opinion of an observer, whereas in the 89% MSE_{dp} similarly ranked as a human observer, if no-linear correlation is employed.

The difference between $YouDMOS_p$ and A_v is that the first use a nonlinear regression function of the average of certain 2DIQA metric while the latter is just the mean of one of 29 2DIQA applied to stereo-pair, in this way $YouDMOS_p$ with UQI is linearly better than A_v with UQI for just 0.000611%.

So, our results of these 27 algorithms in the field of SIQA could lead to conclude that *Metrics based on 2DIQA* can assess the perceptual quality of third dimensional or stereoscopic images. The implication of results of the presented research should be considered with caution, since the first matter to observe is that the majority of the *Stereoscopic Metrics* are only adaptations of 2DIQA, which add some features such as depth variances from the disparity map, for instance. Any perceptual feature is included in the manner that this disparity information is taken, namely any algorithm incorporates disparity masking.

It is important to realize that observers employed not only in LIVE 3D but also in MMSPG or FISE image databases judge the stereoscopic image quality watching some slices, apparently separated, of a 2D scenario, which is a disadvantage for the *Stereoscopic Metrics*. Also, another disadvantage for *Stereoscopic Metrics* is that the distortions in LIVE 3D image database are not designed or applied stereoscopically, since they separately distorted the left and right images.

Some distortions that LIVE 3D image database considers, such as Gaussian Blur and Additive White Gaussian Noise, are global distortions and therefore, they would not affect too much the perception of depth. Not only *Metrics based on 2DIQA* correlates extremely well with these distortions but also some *Stereoscopic Metrics* do it well, such as d_1 or MSE_{dp}. However for those distortions with localized artifacts, the performance both of *Metrics based on 2DIQA* and *Stereoscopic Metrics* is lower, especially for the local blocking artifacts caused by a JPEG compression. Furthermore, some irregularities in terms of the depth map appear when localized distortions are evaluating, which is why the presented state-of-the-art SIQA-$\mathcal{SET}$does not correlate well. For JPEG compression distortion, the performance of $YouDMOS_p$ and A_v is unexpectedly good in spite of being dependant functions on a monoscopic image quality.

If we take in to account that *Stereoscopic Metrics* are simple designs somehow based on a certain 2DIQA, we also can realize that the gap between *Stereoscopic Metrics* and *Metrics based on 2DIQA* can be filled proposing assessments with some features of the best correlated metrics.

Acknowledgment

This article is supported by National Polytechnic Institute (Instituto Poliécnico Nacional) of Mexico by means of Project No. 20190046 granted by Secretariat of Research and Postgraduate(Secretería de Investigación y Posgrado), National Council of Science and Technology of Mexico (CONACyT) and LABEX Σ-LIM France, Coimbra Group Scholarship Programme granted by University of Poitiers and Region Poitou-Charentes (France). The research described in this work was carried out at the Superior School of Mechanical and Electrical Engeniering (Escuela Superior de Ingeniería Mecánica y Eléctrica), Campus Zacatenco.

References

1. **Akhter, R., Parvez Sazzad, Z. M., Horita, Y., & Baltes, J. (2010).** No-reference stereoscopic image quality assessment. *Stereoscopic Displays and Applications XXI*, pp. 75240T–75240T–12.
2. **Behance (2013).** Crafts, fine arts.
3. **Benoit, A., Le Callet, P., Campisi, P., & Cousseau, R. (2008).** Quality assessment of stereoscopic images. *EURASIP Journal on Image and Video Processing*, Vol. 2008, No. 1, pp. 659024.
4. **Bertalanffy, L. V. (1989).** *Teoría General de los Sistemas*.
5. **Bosc, E., Pepion, R., Le Callet, P., Koppel, M., Ndjiki-Nya, P., Pressigout, M., & Morin, L. (2011).** Towards a new quality metric for 3-D synthesized view assessment. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 5, No. 7, pp. 1332–1343.
6. **Cameron, J. (2009).** *Avatar. Lightstorm Entertainment Dune Entertainment and Ingenious Film Partners*, Vol. Distributed by 20th Century Fox.
7. **Campisi, P., Callet, P. L., & Marini, E. (2007).** Stereoscopic images quality assessment. *Proceedings of the15th European Signal Processing Conference (EUSIPCO)*.
8. **Chandler, D. & Hemami, S. (2007).** Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, Vol. 16, No. 9, pp. 2284–2298.
9. **Chen, M.-J., Su, C.-C., Kwon, D.-K., Cormack, L. K., & Bovik, A. C. (2012).** Full-reference quality assessment of stereoscopic images by modeling binocular rivalry. *Forty-Sixth Annual Asilomar Conference on Signals, Systems, and Computers*.
10. **Connia, S. (2013).** 150 years of Steroscopy.
11. **Damera-Venkata, N., Kite, T., Geisler, W., Evans, B., & Bovik, A. (2000).** Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, Vol. 9, pp. 636–650.
12. **Egiazarian, K., Astola, J., Ponomarenko, N., Lukin, V., F.Battisti, & Carli, M. (2006).** Two new full-reference quality metrics based on HVS. *Proceedings of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pp. 4.
13. **Goldmann, L., De Simone, F., & Ebrahimi, T. (2010).** Impact of acquisition distortion on the quality of stereoscopic images. *Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics*.
14. **Gorley, P. & Holliman, N. (2008).** Stereoscopic image quality metrics and compression. *Stereoscopic Displays and Applications XIX*, pp. 680305–680305–12.
15. **Gu, K., Zhai, G., Yang, X., & Zhang, W. (2012).** A new no-reference stereoscopic image quality assessment based on ocular dominance theory and degree of parallax. *21st International Conference on Pattern Recognition (ICPR)*, pp. 206–209.
16. **Hewage, C. & Martini, M. (2010).** Reduced-reference quality metric for 3D depth map transmission. *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010*, pp. 1–4.
17. **Hirschmuller, H. & Scharstein, D. (2007).** Evaluation of cost functions for stereo matching. *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR2007)*, pp. 1–8.
18. **Huan, M., Minazuki, A., & Hayashi, H. (2012).** Study on 3D image assessment using motion capture system. *International Conference on Advanced Applied Informatics (IIAIAAI)*, pp. 182–186.
19. **Image3D (2013).** Everybody looks.
20. **ITU (2002).** *Bt-500-11: Methodology for the Subjective Assessment of the Quality of Television Pictures*.
21. **Jin, L., Boev, A., Gotchev, A., & Egiazarian, K. (2012).** 3D-DCT based multi-scale full-reference quality metric for stereoscopic video. *Proceedings of the Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*.
22. **Joveluro, P., Malekmohamadi, H., Fernando, W. A. C., & Kondoz, A. (2010).** Perceptual video quality metric for 3d video quality assessment. *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4.
23. **Larson, E. C. & Chandler, D. M. (2010).** Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, Vol. 19, No. 1, 011006.
24. **Mao, X., Yu, M., Wang, X., Jiang, G., Peng, Z., & Zhou, J. (2010).** Stereoscopic image quality

assessment model with three-component weighted structure similarity. *International Conference on Audio Language and Image Processing (ICALIP)*, pp. 1175–1179.

25. **Mitsa, T. & Varkur, K. (1993).** Evaluation of contrast sensitivity functions for formulation of quality measures incorporated in halftoning algorithms. *IEEE International Conference on Acustics, Speech and Signal Processing*, Vol. 5, pp. 301–304.
26. **Mittal, A., Moorthy, A., & Bovik, A. (2012).** No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, Vol. 21, No. 12, pp. 4695–4708.
27. **Mittal, A., Soundararajan, R., & Bovik, A. (2013).** Making a "Completely Blind " image quality analyzer. *IEEE Signal Processing Letters*, Vol. 20, No. 3, pp. 209–212.
28. **Moorthy, A. & Bovik, A. (2010).** A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, Vol. 17, No. 5, pp. 513–516.
29. **Moorthy, A. K., Su, C.-C., Mittal, A., & Bovik, A. C. (2012).** Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*.
30. **Moreno, J. (2012).** P^2SNR: Perceptual full-reference image quality assessment for JPEG2000. *Data Compression Conference (DCC)*, pp. 406–406.
31. **Moreno, J. & Fernandez, C. (2013).** NRP-SNR: No-reference peak signal-to-noise ratio for JPEG2000. *Data Compression Conference (DCC)*, pp. 511–511.
32. **Park, M., Luo, J., & Gallagher, A. (2012).** Toward assessing and improving the quality of stereo images. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 6, No. 5, pp. 460–470.
33. **Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., & Carli, M. (2011).** Modified image visual quality metrics for contrast change and mean shift accounting. *11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, pp. 305–311.
34. **Scharstein, D. & Szeliski, R. (2002).** A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, Vol. 47, pp. 7–42.
35. **Shao, F., Lin, W., Gu, S., Jiang, G., & Srikanthan, T. (2013).** Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. *IEEE Transactions on Image Processing*, Vol. 22, No. 5, pp. 1940–1953.
36. **Sheikh, H. & Bovik, A. (2006).** Image information and visual quality. *IEEE Transactions on Image Processing*, Vol. 15, No. 2, pp. 430–444.
37. **Sheikh, H., Bovik, A., & Cormack, L. (2005).** No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing*, Vol. 14, No. 11, pp. 1918–1927.
38. **Shen, L., Yang, J., & Zhang, Z. (2009).** Stereo picture quality estimation based on a multiple channel hvs model. *2nd International Congress on Image and Signal Processing (CISP)*, pp. 1–4.
39. **Solh, M., Bauza, J. M., & AlRegib, G. (2011).** 3vqm: A 3d video quality measure. *IEEE Transactions on Image Processing*.
40. **Visual Communications Laboratory of Cornell University (2010).** MeTriX MuX visual quality assessment package.
41. **Wang, X., Yu, M., Yang, Y., & Jiang, G. (2009).** Research on subjective stereoscopic image quality assessment. *Multimedia Content Access: Algorithms and Systems III*, pp. 725509–725509–10.
42. **Wang, Z. & Bovik, A. (2002).** A universal image quality index. *IEEE Signal Processing Letters*, Vol. 9, No. 3, pp. 81–84.
43. **Wang, Z. & Bovik, A. (2009).** Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, Vol. 26, No. 1, pp. 98–117.
44. **Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004).** Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612.
45. **Wang, Z. & Bovik, A. C. (2006).** *Modern Image Quality Assessment*. Morgan & Claypool Publishers: Synthesis Lectures on Image, Video, & Multimedia Processing, 1 edition.
46. **Wang, Z., Sheikh, H. R., & Bovik, A. (2002).** No-reference perceptual quality assessment of jpeg compressed images. *International Conference on Image Processing*, volume 1, pp. I–477–I–480 vol.1.
47. **Wang, Z., Simoncelli, E., & Bovik, A. (2003).** Multiscale structural similarity for image quality assessment. *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 2, pp. 1398–1402.

48. Yang, J., Hou, C., Zhou, Y., Zhang, Z., & Guo, J. (2009). Objective quality assessment method of stereo images. *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp. 1–4.

49. You, J., Xing, L., Perkis, A., & Wang, X. (2010). Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis. *Proceedings of the International Workshop on Video Processing and Quality Metrics*.

50. Zhang, F., Li, S., Ma, L., & Ngan, K. N. (2010). Limitation and challenges of image quality measurement, pp. 774402–774402–8.

51. Zhang, L., Zhang, D., & Mou, X. (2010). Rfsim: A feature based image quality assessment metric using riesz transforms. *17th IEEE International Conference on Image Processing (ICIP)*, pp. 321–324.

52. Zhang, L., Zhang, D., Mou, X., & Zhang, D. (2011). Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, Vol. 20, No. 8, pp. 2378–2386.

53. Zhu, Z. & Wang, Y. (2009). Perceptual distortion metric for stereo video quality evaluation. *WSEAS Trans. Sig. Proc.*, Vol. 5, No. 7, pp. 241–250.

Article received on 27/02/2019; accepted on 25/03/2019.
Corresponding author is Jesús Jaime Moreno Escobar.