DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

Using binocular and monocular properties for the construction of a quality assessment metric for stereoscopic images

Iana Iatsun, Mohamed-Chaker Larabi, Christine Fernandez-Maloigne
Departement XLIM-SIC UMR CNRS 7252, Bt. SP2MI, Teleport 2, Bvd Marie et Pierre Curie, BP 30179, 86962, Futuroscope, France

Abstract

More and more people are getting access to 3D; 3D-TV is a next expecting step in telecommunication. Before being presented to the public, any 3D content has to be coded, compressed and transmitted. All these treatments can impact the quality of the final product, thus it is essential to have a measurement tool for the estimation of the quality of stereoscopic content. Several studies have already modified the existing 2D quality metrics to be able evaluate 3D images, however the results are not satisfying. In this work, we propose a full-reference metric for a quality assessment of stereoscopic images employing the properties of binocular perception. The principle of the metric estimates the probability of fusion, that can be obtained perceptually. The quality of one view is assessed relatively to the other, and according to the result metric makes a decision, that is based on the binocular fusion properties. The comparison between views is performed only on the salient area. It is detected using the visual attention model based on the monocular depth cues and interest points. The metric has been tested on the publicly available dataset, and its results are coherent to the subjective scores.

Introduction

Image quality is an important factor in the design of any image processing system. The most common example is the application to compression algorithms in order to decrease required data size and keep high quality. Besides, quality evaluation metrics are used in acquisition, transmission and visualization processes. Numerous quality assessment metrics have been developed for 2D image and video [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Even though 3D quality assessment (QA) can seems close to 2D quality evaluation, it is far more complicated. The key issue is related to the limited access to the perceived 3D visual stimulus. The challenge is how to extract the features presenting the perceived 3D quality from the available information, like two views or depth/ texture.

First metrics for the evaluation the 3D quality were based on 2D quality metrics, like *PSNR*, *MSE SSIM* [15], *PQS* [16], *VSNR* [17], *IFC* [18], *UQI* [5], *VIF* [19], *NQM* [20]. The capabilities of these metrics to predict visual quality for stereoscopic images have been tested by *You et al.* [21]. Their significant disadvantage is the absolute neglecting of the binocular perception features [22]. However, nowadays it is clear that a simple extension of 2D QA cannot be enough to consider all the complexity of binocular vision. Moreover, 2D evaluation does not take into account the particularities of the fusion process [23]. The next wave of 3D QA metrics attempted to solve the drawbacks of algorithms purely based on 2D QM. Designed algorithms accounted for depth information and highlighted its importance for the perceived quality [24, 25, 26, 27, 28]. Nevertheless, it is doubting that only disparity/depth information is enough to model the binocular perception. Hence, it is getting clear that simple adjustment of 2D metrics cannot satisfy all the complexity of 3D perception.

So far, few metrics have been developed exploring the stereoscopic perception features. For example, the model using the binocular energy has been designed by *Bensalma et al.* [29]. Their algorithm estimates the difference of binocular energy between original and tested stereo couples. The binocular energy is obtained as step-by-step processing of the reference and tested images. The treatment includes the antagonist color transformation, and application of transforms linked to simple and complex cells. *Shao et al.* created a metric that considers the binocular fusion process [30]. Several image regions are defined by matching left and right channels pixel by pixel. These areas are associated to the binocular fusion region, binocular suppression and non-corresponding locations. Local phase and amplitude maps are calculated for the reference and distorted image. Using these local features, quality of each region is calculated independently. The final score of the metric is a weighting sum of these quality coefficients per area. However, the strong disadvantage of this metric is the requirement to train the coefficients every time the database is changed.

Binocular suppression theory takes its foundation from the rivalry concept. It states that one image can have the details of the scene, while the other has just minimum necessary information for disparity. For example, binocular fusion suppression has been used by *Wang et al.* in order to design a quality assessment metric [31]. The idea of *Zhao et al.* algorithm lies in the quality evaluation for each view of stereo-pair, and than in their combination according to binocular Just-Noticeable-Difference [32].

The idea of binocular dominance has been already considerably explored in the studies dedicated to compression [33, 34, 35], that motivates its application for a quality assessment [36]. The ocular dominance is performed by prediction of separate quality value for left and right views using JPEG Quality Scores [37]. The final score is given by the addition of view's scores and degree of parallax. Author of the work [38] also mimicked the binocular perception to combine the quality of left and right images. They applied a model of binocular perception proposed in [39] to luminance, contrast and structural similarity, and after combined all channels. Authors reported better results in comparison to conventional PSNR and SSIM metrics.

Since distortions are more visible at the edge boundaries, several authors suggested quality metrics following this idea. For instance, *Akhter et al.* created a no-reference perceptual metric

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

using the feature local segmentation and disparity [40]. The algorithm divides the stereo couple into blocks to characterize 3D impairments. In a similar direction, a stereoscopic image quality metric has been proposed by *Sazzad et al.* [41, 42]. Their metric is oriented towards quality evaluation of JPEG distortions on the images. Another model classifying differently the areas of images has been suggested in [43]. Authors applied Sobel filter on the left and right views in order to subdivide the reference and distorted images into edges, texture and smooth areas. Different weights are set for each region during applying SSIM metric. The final score of the metric is established as the combination of these three estimations.

A significant progress can be reported for the development of metrics with low-level aspects of the HVS, like luminance, texture masking and contrast sensitivity function, etc. However, the extension of objective metrics with the introduction of high-level HVS properties, as visual attention, goes much slower. Only few researchers have worked in this direction [44, 45, 46, 47, 48]. Two ideas exist about the relation of visual attention and quality assessment. The first hypothesis states that the artifacts inside the region of interest are more disturbing than in any other parts of the image. The second concept claims that people are more sensitive to the impacted area and tent to see regions of poor quality. So far, the overall quality assessment is heavily affected by such areas. Following the first idea, the designed algorithms pool the visual importance from eye-tracking results or attention models. Consequently, authors reported the improvements of the metrics performance after the introduction of saliency in this or another way [46, 47, 48, 49]. However, no trials to evaluate the impact of visual attention on the perceived image quality of 3D content have been done yet. In the proposed approach, we suggest to take benefit of the algorithm for saliency detection to design a quality assessment metric for stereoscopic images.

Since compression is closely linked to QA, the features of asymmetric compression have to be examined. Several studies have shown that overall perception quality for a stereo couple can be higher than the quality of the worst view of the pair [35]. Only few stereoscopic quality metrics have been designed in this direction. Moreover, their observations relate mostly to the artifacts induced by compression, although the combination of other types of distortions has to be also investigated. So far, in this work we propose a stereoscopic quality assessment metric for an asymmetric and symmetric images. Designed metric does not require either calculated, or provided disparity/depth information. Moreover, metric considers such high-level vision processes, as visual attention.

This paper is organized as follows: section 2 presents the visual attention detection approach, allowing to extract the salient features from a single view of stereo-pair. Section 3 describes the algorithm of the proposed stereoscopic metric, highlighting its binocular and monocular perception features. The metric results on publicly available database are described in section 4. This paper ends with some concluding remarks and give ideas about futures works.

Visual attention model for stereoscopic images

In the literature, most of the visual attention models for 3D are based on stereo depth or disparity. Generally, this information is accessible by having either two views or, texture and depth representation. However, depth information can be extracted using simple 2D images following monocular cues. For our goals, we select a visual attention model able to detect salient area for stereoscopic pair using only one view [50]. The framework of applied model is depicted on Figure 1.

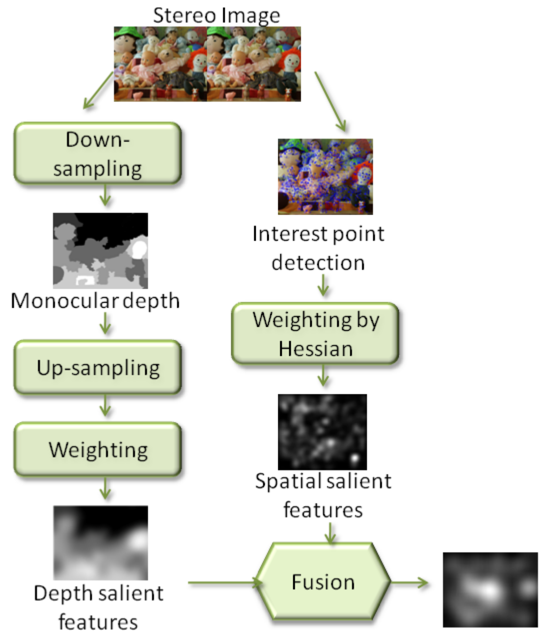

Figure 1: Flowchart of applied 3D saliency algorithm [50].

At the first stage of saliency detection, spatial visual features are extracted from a single view of stereo-pair (left one in our case[35]). The spatial saliency is detected using the similarities between interest and gaze points [51, 52, 53]. At next step, depth features are detected using monocular cues [54, 55, 56]. The latter does not require neither a priori information, nor learning process. Procedures of down-sampling and up-sampling are optional and allow to reduce complexity. Finally, a fusion of spatial and depth saliency features is performed. Details of saliency detection is briefly described below.

Spatial features

The similarity between interest points (Harris, SIFT and SURF) and gaze points, for a given set of parameters, have been demonstrated by *Nauge et al.* [53]. The spatial visual features are extracted by relying on interest points, considering the influence of central bias for still images.

SURF IP detectors have been selected due to their efficiency. The latter is based on calculating approximate Hessian response for image points in order to detect blob structures. For the scale-space analysis a pyramid of filters is used to approximate the LoG (Laplacian of Gaussian). In order to localize interest points in the image through different scales, a non-maximum suppression is applied. The descriptor extraction describes the distribution of the intensity content within the interest point neighborhood and its variation. The following parameters *hessianThreshold = 800, nOctaves=9, nOctaveLayers=1* have been set-up for *SURF* IP detector to achieve our goals. Extracted interest points are characterized by their position (x_i,y_i) and size (S_i). These information are

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

exploited for the construction of the spatial saliency by applying a kernel of Gaussian as described in equation 1.

$$SM_P = I_{Hessian} * G_{S_i} \quad (1)$$

where

$$G_S(x_i, y_i) = \frac{S_i}{2\pi\sigma^2} exp - \left(\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2} \right) \quad (2)$$

Several researchers have shown the existence of the *"central bias"* during watching 2D still images [57, 58]. They stated that gaze fixations are biased towards the center of the scene. Therefore, the visual attention model accounts for the central bias. Its effect is presented by adding one more feature map that is a central Gaussian kernel.

Monocular depth

The process of monocular depth evaluation [54] contains two main phases: building of hierarchical representation of the image and choosing segments to compose final depth order partition. The estimation of low level depth cues are integrated into the construction of binary partition tree (BPT) [59]. At the first step of segmentation, each pixel of image is considered to be a separate region. Hence, computational cost is directly proportional to the image resolution. This region is characterized by its color, shape, area and depth, that influence on a similarity measure. The depth ordering is based on BPT pruning, where leaves are represented by regions. They are iteratively merged based on similarity measure. Since depth information from T-junctions and convexity can be contradictory, a conflict resolution step is applied using a probabilistic model. Finally, a depth map is obtained pruning BPT, as described in equation 3.

$$D_{mono} = f_{BPT}(M) \quad (3)$$

The result of monocular depth estimation is a segmented image according to its depth position, so number of depth levels are limited. In order to obtain a depth saliency map, the weighting procedure is applied. The first-plan objects are getting the highest importance. Few levels after are considered as main plan and others as a background. Afterwards the depth saliency map is obtained by applying a Gaussian filter G to D_{mono} in order to smooth the segmentation results for a better integration in the fusion.

$$SM_D = D_{mono} * G \quad (4)$$

Fusion of visual features

For our goal we selected a combination approach given the good fusion result of visual features [50] - *Global Non-Linear Normalization followed by Normalization* (*GNLNS*) [60]. This combination process takes advantage from maps with few saliency peaks and neglect the uniform distributed maps. It is described by equation 5, where M_i stands for the maximum value of i-map, and m_i is the mean of other maximum values of this i-map.

$$SM_{out} = \sum_i \left[(N(SM_i)) \cdot (M_i - m_i)^2 \right] \quad (5)$$

Proposed quality assessment metric for S3D images

The perceived quality of stereoscopic image depends strongly on the quality of each view of the stereo-pair and their possibility to be fused. The proposed approach seeks to present and evaluate these features of stereo image. Moreover, the idea behind the designed metric is to consider the characteristics of the human visual system (HVS), specifically binocular perception.

The perceived image quality of stereoscopic content depends considerably on the features of binocular perception. Such particularities as binocular rivalry and masking effect have been demonstrated to impact the image quality evaluation [39, 61]. The left and right views of stereo-pair can be coded (or compressed) symmetrically. In this case, the perceived quality depends on the type of distortion and its mutual effect. Depending on the degree of distortion, the conventional fusion is normally possible, since images remain pretty the same. In case of asymmetrical processing, the quality of one view is lower than the other, due to compression or down-sampling [35]. According to the binocular suppression theory [62, 63], in such case the perceived quality is closer to the higher value. Nevertheless, the difference in quality between left and right cannot be unlimited, since after a certain threshold fusion starts to be impossible or degradation becomes prevailing. This threshold of tolerant asymmetry between view's PSNR has been shown to be dependent on the screen type, and it is equal to 21 dB for parallax barrier display and 33 dB for polarized projection screen [64, 65, 66]. Hence, the account for the coding character of stereo pair is an important feature of the proposed 3D quality metric.

The proposed quality metric for 3D images considers the HVS features and characteristics of stereo content. The flowchart of the proposed metric is depicted on Figure 2. It is a full-reference metric, hence the couple of reference and test images are required.

The algorithm of our quality metric starts from the detection of the area of interest. For this goal, the saliency detection model, described in previous section is used. Its significant advantage is the possibility to detect the area of interest from a single view, using the data obtained with a monocular depth cues and SURF interest points (IP). It has been proved in the literature, that the eye dominance does not affect the fusion process, so either left, or right view of reference stereo-pair can be used for the saliency detection. After saliency has been evaluated for one view of the reference (left in our case), we propose to project the salient area on the second original image. This projection can be performed thanks to SURF detectors used for saliency evaluation algorithm. Each interest point in the salient area of the left view corresponds to an interest point on the right view. Since the salient neighborhood of this IP is known, the right view is reconstructed, as given on Figure 3. Such technique allows to perform saliency detection only once, and takes the advantage of the detected interest point for its projection on the second view.

At the next step, from the obtained saliency maps the binary masks are created. By this, we mean that only pixels with saliency values higher than a threshold are kept. This procedure is performed equally for the left and right views of reference and test images. The example of the obtained maps are presented on Figure 4. The threshold value was fixed at the level of 0.3.

In order to obtain the quality score of stereoscopic pair (Q_s),

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

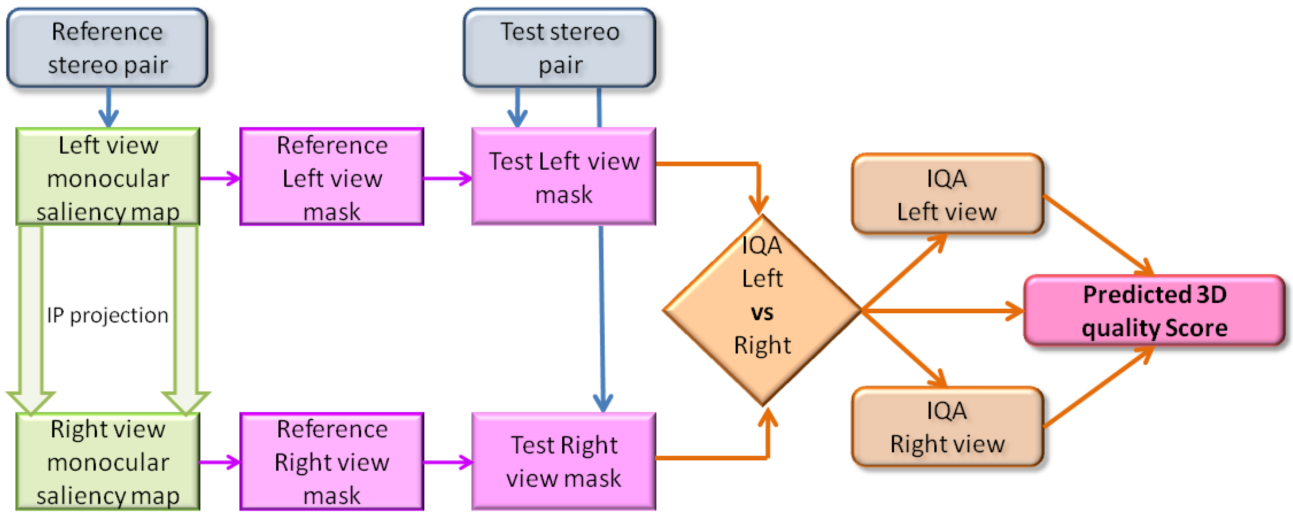

Figure 2: Flowchart of the proposed quality metric for 3D images.

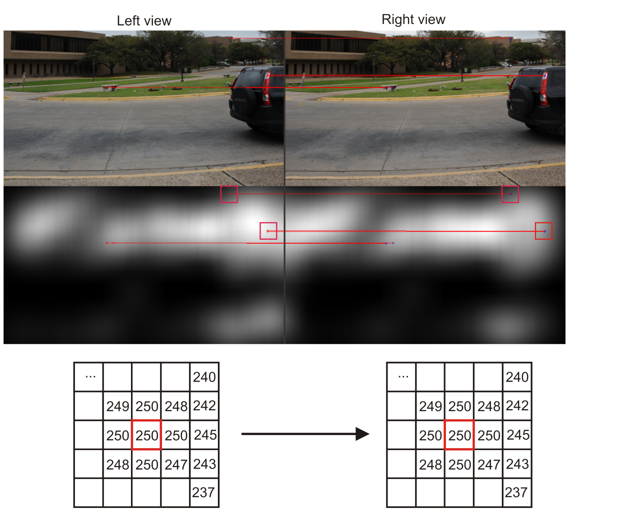

Figure 3: Explanation of the saliency map projection. Only few interest points are detected for visibility.

we evaluate the quality difference between left and right masked test views ($IQA_{left-right}$). For this goal, in our model stereoscopic metric MSSIM [9] has been chosen. Depending on what is value of $IQA_{left-right}$, the stereoscopic metric follows different algorithms. So far, if the obtained $IQA_{left-right}$ is higher or equal to 0.9, it means that the left and right images are characterized by identical or similar quality, hence the evaluation of one view is enough to predict the image quality of the stereo pair. If the difference between left and right view quality is significant, the $IQA_{left-right}$ is lower than 0.9, thus the stereo pair is asymmetrically coded. Consequently, we propose to estimate this asymmetry, and check whenever it is possible to fuse this stereo pair or no. To do so, we set up the second threshold value for the quality difference between left and right views: $IQA_{left-right}$ should be higher than 0.6. We claim that stereo images having asymmetry level within this threshold can be fused according to the binocular suppression theory, and overall quality is closer to higher value. To obtain the score of the perceived image quality, we apply a mathematical model of binocular perception suggested in [39] and described with equation 6. The parameters for this case are $w = 0.4$ and $n = 2$.

$$Q_s = [w(IQ_{low})^n + (1-w)(IQ_{high})^n]^{1/n} \tag{6}$$

Afterwards, if the quality difference between left and right images is considerable, and $IQA_{left-right}$ is less than 0.6, the low quality starts to be dominant decreasing thus the overall quality score. To model this process, the parameters in equation 6 are fixed at $w = 0.8$ and $n = 2$.

$$Q_s = \begin{cases} q_{high} & \text{,if } IQA_{left-right} > 0.9 \\ \sqrt{0.4 \cdot (Q_{low})^2 + 0.6 \cdot (Q_{high})^2} & \text{,if } 0.6 < IQA_{left-right} \leq 0.9 \\ \sqrt{0.8 \cdot (Q_{low})^2 + 0.2 \cdot (Q_{high})^2} & \text{,if } IQA_{left-right} \leq 0.6 \end{cases} \tag{7}$$

Results and Discussion

The proposed algorithm has been tested on the *"Live 3D" II phase* database, that consists of 8 reference and 360 impacted images with registered human score. The examples of images are given on Figure 5. The dataset includes 3 symmetrically and 6 asymmetrically distorted stereo couples. The stereo pairs are impacted by five different distortions, like White Noise, JPEG, JPEG2000, Gaussian Blur and Fast Fading.

The results computed on Live 3D database are presented on Table 1. The outcome of the proposed metric is compared to the average of 2D quality evaluation on the left and right views (marked as QM_{avr}). Such averaging is also performed for the stereo-pair after masking, that are marked $QM_{avr+mask}$. We applied Pearson linear correlation (*PCC*), Spearman correlation coefficient (*SCC*), root-mean-squared error (*RMSE*) and Kendall

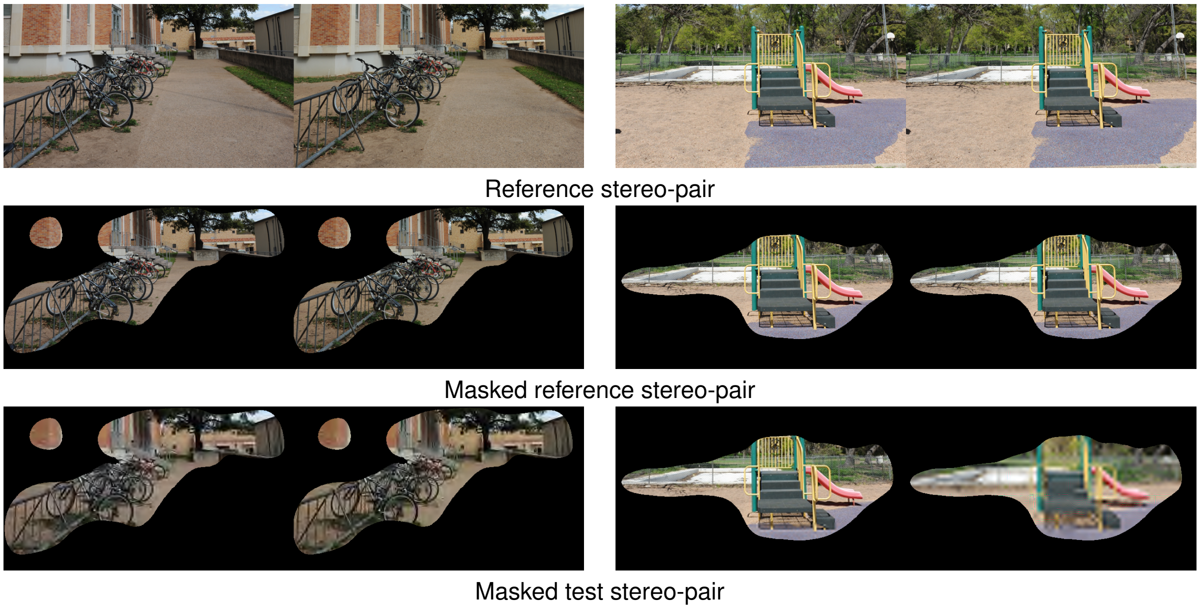

Figure 4: Example of created mask for stereo-pair

Table 1: Performance of 3D IQA algorithms overall on the database *3D LIVE IIphase* . Perfect correlation case $PCC \to \pm 1$, $SCC \to \pm 1$, $KCC \to 1$, $RMSE \to 0$.

Algorithm	***PCC***	***SCC***	***KCC***	***RMSE***
$PSNR_{avr}$	0.687	0.679	0.503	13.671
$VSNR_{avr}$	0.652	0.625	0.453	8.874
$SSIM_{avr}$	0.748	0.738	0.550	7.488
$MSSIM_{avr}$	0.618	0.731	0.558	11.462
$PSNR_{avr+mask}$	0.571	0.683	0.506	18.541
$VSNR_{avr+mask}$	0.660	0.749	0.558	8.479
$SSIM_{avr+mask}$	0.768	0.758	0.562	7.226
$MSSIM_{avr+mask}$	0.795	0.776	0.599	6.836
Proposed metric	**0.823**	**0.815**	**0.620**	**6.401**

correlation (KCC) to evaluate the performance of the proposed metric. Linear correlation coefficient and root-mean-square-error have been calculated after a non-linear logistic regression.

It can be noticed from the results in Table 1, that the proposed metric presents the best results. We would like to highlight that no separation for symmetric and asymmetric distortion before applying the metrics has been done. It can be observed accordingly that applying masking on stereo pairs brings considerable improvement for 2D stereo metrics.

Since several distortions are applied to the images in the tested database, the investigation of the metric is performed on different artifacts. Hence, Table 2 reflects the metric's efficiency on the distortion of JPEG type. The results of KCC and SCC are noticeably lower than for overall performance. JPEG compression creates the local artifacts, linked to DCT transformation with segmentation block of 8×8 pixels. These localized impairments of block type can impact depth perception. Since 2D metrics do not account for any depth information, the drop in performance is important. Although, the proposed model shows the best performance due to its stage of left versus right evaluation. This procedure helps to incorporate the possibility of image fusion.

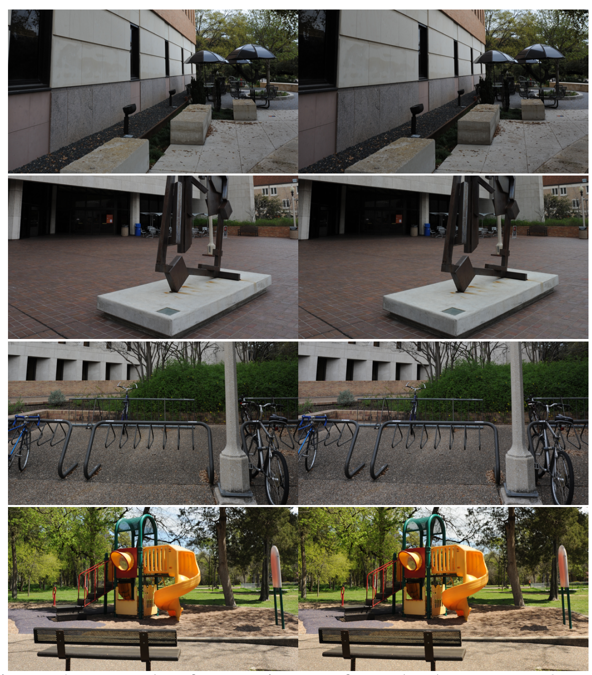

Figure 5: Example of stereo-images from database *"Live 3D" II phase*

The JPEG2000 degradation presents the blur caused by the quantification of the wavelet coefficient. JPEG2000 artifact can smooth edges, that impacts depth perception. It can be observed from Table 3 that 2D algorithms perform better on images with distortion JPEG2000 than JPEG.

From the all tested distortions, Blur and White Noise are global distortions, thus their effect on depth perception is less important. The results on the metric's performance are presented for white noise distortion in Table 4 and in Table 5 for blur respectively. The best performance for all experimented metrics can be noticed from the Tables.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

Table 2: Performance of 3D IQA algorithms on on images with the JPEG distortion. Perfect correlation case $PCC \to \pm 1$, $SCC \to \pm 1$, $KCC \to 1$, $RMSE \to 0$.

Algorithm	PCC	SCC	KCC	RMSE
$PSNR_{avr}$	0.231	0.492	0.343	10.548
$VSNR_{avr}$	0.338	0.371	0.245	5.965
$SSIM_{avr}$	0.557	0.551	0.391	6.085
$MSSIM_{avr}$	0.672	0.678	0.471	5.426
$PSNR_{avr+mask}$	0.459	0.504	0.348	6.965
$VSNR_{avr+mask}$	0.415	0.622	0.452	6.667
$SSIM_{avr+mask}$	0.672	0.672	0.474	5.422
$MSSIM_{avr+mask}$	0.693	0.680	0.485	5.282
Proposed metric	**0.792**	**0.827**	**0.621**	**4.472**

Table 3: Performance of 3D IQA algorithms on images with the JPEG2000 distortion. Perfect correlation case $PCC \to \pm 1$, $SCC \to \pm 1$, $KCC \to 1$, $RMSE \to 0$.

Algorithm	PCC	SCC	KCC	RMSE
$PSNR_{avr}$	0.491	0.633	0.458	12.545
$VSNR_{avr}$	0.611	0.612	0.458	5.909
$SSIM_{avr}$	0.664	0.696	0.503	7.334
$MSSIM_{avr}$	0.810	0.807	0.608	5.745
$PSNR_{avr+mask}$	0.551	0.628	0.446	9.132
$VSNR_{avr+mask}$	0.668	0.756	0.563	7.301
$SSIM_{avr+mask}$	0.725	0.710	0.529	6.755
$MSSIM_{avr+mask}$	0.825	0.815	0.629	5.537
Proposed metric	**0.840**	**0.817**	**0.636**	**5.313**

Table 4: Performance of 3D IQA algorithms on images with the white noise distortion. Perfect correlation case $PCC \to \pm 1$, $SCC \to \pm 1$, $KCC \to 1$, $RMSE \to 0$.

Algorithm	PCC	SCC	KCC	RMSE
$PSNR_{avr}$	0.470	0.627	0.492	11.154
$VSNR_{avr}$	0.554	0.658	0.506	8.915
$SSIM_{avr}$	0.881	0.875	0.692	5.059
$MSSIM_{avr}$	0.818	0.805	0.621	6.157
$PSNR_{avr+mask}$	0.789	0.615	0.478	8.564
$VSNR_{avr+mask}$	0.768	0.770	0.564	6.856
$SSIM_{avr+mask}$	0.862	0.856	0.669	5.414
$MSSIM_{avr+mask}$	0.918	0.915	0.744	4.242
Proposed metric	**0.938**	**0.934**	**0.780**	**3.690**

Table 5: Performance of 3D IQA algorithms on images with the blur distortion. Perfect correlation case $PCC \to \pm 1$, $SCC \to \pm 1$, $KCC \to 1$, $RMSE \to 0$.

Algorithm	PCC	SCC	KCC	RMSE
$PSNR_{avr}$	0.551	0.815	0.658	9.984
$VSNR_{avr}$	0.785	0.820	0.607	7.124
$SSIM_{avr}$	0.827	0.796	0.615	7.825
$MSSIM_{avr}$	0.774	0.778	0.593	8.810
$PSNR_{avr+mask}$	0.799	0.846	0.650	11.564
$VSNR_{avr+mask}$	0.870	0.885	0.699	6.846
$SSIM_{avr+mask}$	0.846	0.832	0.633	7.420
$MSSIM_{avr+mask}$	0.803	0.807	0.625	8.296
Proposed metric	**0.933**	**0.902**	**0.723**	**4.990**

Fast-fading (FF) artifact appears during transmission of JPEG2000 compressed image over a Rayleigh fading channel. Hence it is a mixture of two impairment sources. As depicted on Table 6, the proposed metric performs relatively good on FF type of distortions also.

Table 6: Performance of 3D IQA algorithms on images with the fast fading distortion. Perfect correlation case $PCC \to \pm 1$, $SCC \to \pm 1$, $KCC \to 1$, $RMSE \to 0$.

Algorithm	PCC	SCC	KCC	RMSE
$PSNR_{avr}$	0.567	0.711	0.537	14.547
$VSNR_{avr}$	0.774	0.786	0.608	7.684
$SSIM_{avr}$	0.837	0.824	0.649	6.292
$MSSIM_{avr}$	0.848	0.809	0.637	6.081
$PSNR_{avr+mask}$	0.735	0.721	0.545	9.125
$VSNR_{avr+mask}$	0.769	0.834	0.668	7.350
$SSIM_{avr+mask}$	0.867	0.840	0.661	5.732
$MSSIM_{avr+mask}$	0.874	0.832	0.659	5.589
Proposed metric	**0.880**	**0.896**	**0.714**	**5.223**

Conclusion

In this work, we have presented a full-reference quality assessment metric for stereoscopic images. The proposed metric benefits from visual attention information and simulates the features of binocular perception. The saliency is extracted from a single of stereo pair. The innovative procedure of quality comparison between two views allows to evaluate the possibility of fusion, so depth perception. Depending on the similarity between left and right views, the algorithm applies different evaluation procedures. This assessment approach imitates the visual perception quality. A significant contribution of this work is the evidence that visual attention can improve the performance of quality assessment. The quality evaluation is performed only on the area of interest. This approach demonstrates the gain in efficiency in comparison to quality assessment of the full images.

The proposed quality assessment algorithm has been tested on the publicly available dataset with several different types of distortion. Obtained results demonstrated high correlation with subjective scores by using the state-of-the-art techniques. The best efficiency on stereo pairs distorted by white noise and blur, while the less performance have been obtained for stereo couples with JPEG compression artifact. Overall, the suggested model shows the best performance in comparison to the tested quality assessment.

As a perspective, the investigation of the distortion impact on 3D image quality is lacking. This information can improve significantly the performance of 3D IQA. Although, there is no accepted theory about the impact of artifacts on the perception of depth, thus overall image quality. Consequently, the modeling of 3D image quality binocular perception is a great challenge on the way to comprehensive evaluation of quality of experience.

References

[1] Pedersen, M. and Hardeberg, J. Y., "Full-reference image quality metrics: Classification and evaluation," *Foundations and Trends® in Computer Graphics and Vision* **7**(1), 1–80 (2012).

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

[2] Girod, B., "What's wrong with mean-squared error?," in [*Digital images and human vision*], Watson, A. B., ed., 207–220, MIT press, London, England (1993).

[3] Wang, Z., Bovik, A. C., and Lu, L., "Why is image quality assessment so difficult?," in [*Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*], **4**, IV–3313, IEEE (2002).

[4] Wang, Z. and Bovik, A. C., "Mean squared error: love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine* **26**(1), 98–117 (2009).

[5] Wang, Z. and Bovik, A. C., "A universal image quality index," *IEEE Signal Processing Letters* **9**(3), 81–84 (2002).

[6] Chen, G.-H., Yang, C.-L., Po, L.-M., and Xie, S.-L., "Edge-based structural similarity for image quality assessment," in [*Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], **2**, II–II, IEEE (2006).

[7] Chen, G.-H., Yang, C.-L., and Xie, S.-L., "Gradient-based structural similarity for image quality assessment," in [*Proceedings of IEEE International Conference on Image Processing (ICIP)*], 2929–2932, IEEE (2006).

[8] Brooks, A. C., Zhao, X., and Pappas, T. N., "Structural similarity quality metrics in a coding context: exploring the space of realistic distortions," *IEEE Transactions on Image Processing* **17**(8), 1261–1273 (2008).

[9] Wang, Z., Simoncelli, E. P., and Bovik, A. C., "Multiscale structural similarity for image quality assessment," in [*Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*], **2**, 1398–1402, Ieee (2003).

[10] Osberger, W., *Perceptual Vision Models for Picture Quality Assessmnet and Compression Applications*, PhD thesis, Queensland University of Technology, Brisbane (1999).

[11] Chen, S., Beghdadi, A., and Chetouani, A., "Color image assessment using spatial extension to CIE DE2000," in [*Proceedings of International Conference on Consumer Electronics (ICCE)*], 1–2, IEEE (2008). Digest of Technical Papers.

[12] Maalouf, A. and Larabi, M.-C., "A no-reference objective color image sharpness metric," in [*Proceedings of European Signal Processing Conference (EUSIPCO)*], 1019–1022 (2010).

[13] Nauge, M., Larabi, M.-C., and Fernandez, C., "A reduced-reference metric based on the interest points in color images," in [*Picture Coding Symposium (PCS), 2010*], 610–613, IEEE (2010).

[14] Nauge, M., Larabi, C., and Fernandez, C., "Quality estimation based on interest points through hierarchical saliency maps," in [*3rd European Workshop on Visual Information Processing*], 186–191, IEEE (2011).

[15] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

[16] Miyahara, M., Kotani, K., and Algazi, V., "Objective picture quality scale (PQS) for image coding," *IEEE Transactions on Communications* **46**(9), 1215–1226 (1998).

[17] Chandler, D. M. and Hemami, S. S., "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing* **16**(9), 2284–2298 (2007).

[18] Sheikh, H. R., Bovik, A. C., and De Veciana, G., "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing* **14**(12), 2117–2128 (2005).

[19] Sheikh, H. R. and Bovik, A. C., "A visual information fidelity approach to video quality assessment," in [*1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*], 23–25 (2005).

[20] Damera-Venkata, N., Kite, T. D., Geisler, W. S., Evans, B. L., and Bovik, A. C., "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing* **9**(4), 636–650 (2000).

[21] You, J., Xing, L., Perkis, A., and Wang, X., "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in [*Proceedings of International Video Processing and Quality Metrics for Consumer Electronics*], (2010).

[22] Campisi, P., Le Callet, P., and Marini, E., "Stereoscopic images quality assessment," in [*Proceedings of 15th European Signal Processing Conference (EUSIPCO'07)*], (2007).

[23] Goldmann, L. and Ebrahimi, T., "3D quality is more than just the sum of 2D and depth," in [*IEEE International Workshop on hot topics in 3D*], IEEE (2010).

[24] Olsson, R. and Sjostrom, M., "A depth dependent quality metric for evaluation of coded integral imaging based 3D-images," in [*Proceedings of 3DTV Conference*], 1–4, IEEE (2007).

[25] Alexandre, B., Patrick, L. C., Patrizio, C., and Romain, C., "Quality assessment of stereoscopic images," *EURASIP journal on image and video processing* **2008** (2009).

[26] Kaptein, R. G., Kuijsters, A., Lambooij, M. T., IJsselsteijn, W. A., and Heynderickx, I., "Performance evaluation of 3D-TV systems," in [*IS&T/SPIE Electronic Imaging Proceedings*], International Society for Optics and Photonics (2008). 680819.

[27] Xing, L., You, J., Ebrahimi, T., and Perkis, A., "A perceptual quality metric for stereoscopic crosstalk perception," in [*Proceedings of 17th IEEE International Conference on Image Processing (ICIP)*], 4033–4036, IEEE (2010).

[28] Hewage, C. T. and Martini, M. G., "Reduced-reference quality metric for 3D depth map transmission," in [*Proceedings of 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*], 1–4, IEEE (2010).

[29] Bensalma, R. and Larabi, M.-C., "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing* **24**(2), 281–316 (2013).

[30] Shao, F., Lin, W., Gu, S., Jiang, G., and Srikanthan, T., "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics.," *IEEE Transactions on Image Processing: a publication of the IEEE Signal Processing Society* **22**(5), 1940–1953 (2013).

[31] Wang, X., Kwong, S., and Zhang, Y., "Considering binocular spatial sensitivity in stereoscopic image quality assessment," in [*Visual Communications and Image Processing Conference (VCIP)*], 1–4, IEEE (2011).

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

[32] Zhao, Y., Chen, Z., Zhu, C., Tan, Y.-P., and Yu, L., "Binocular just-noticeable-difference model for stereoscopic images," *IEEE Signal Processing Letters* **18**(1), 19–22 (2011).

[33] Seuntiens, P., Meesters, L., and Ijsselsteijn, W., "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Transactions on Applied Perception (TAP)* **3**(2), 95–109 (2006).

[34] Aflaki, P., Hannuksela, M. M., Hakkinen, J., Lindroos, P., and Gabbouj, M., "Subjective study on compressed asymmetric stereoscopic video," in [*Proceedings of 17th IEEE International Conference on Image Processing (ICIP)*], 4021–4024, IEEE (2010).

[35] Beni, P. A., *Compression and Subjective Quality assessment of 3D video*, PhD thesis, Tampere University of Technology, Tampere (2013).

[36] Gu, K., Zhai, G., Yang, X., and Zhang, W., "A new no-reference stereoscopic image quality assessment based on ocular dominance theory and degree of parallax," in [*Proceedings of 21st International Conference on Pattern Recognition (ICPR)*], 206–209, IEEE (2012).

[37] Wang, Z., Sheikh, H. R., and Bovik, A. C., "No-reference perceptual quality assessment of JPEG compressed images," in [*Proceedings of International Conference on Image Processing (ICIP)*], **1**, I–477, IEEE (2002).

[38] Ryu, S., Kim, D. H., and Sohn, K., "Stereoscopic image quality metric based on binocular perception model," in [*Proceedings of 19th IEEE International Conference on Image Processing (ICIP)*], 609–612, IEEE (2012).

[39] Meegan, D. V., Stelmach, L. B., and Tam, W. J., "Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery.," *Journal of Experimental Psychology: Applied* **7**(2), 143 (2001).

[40] Akhter, R., Sazzad, Z. P., Horita, Y., and Baltes, J., "No-reference stereoscopic image quality assessment," in [*IS&T/SPIE Electronic Imaging Proceedings*], International Society for Optics and Photonics (2010). 75240T.

[41] Sazzad, Z. P., Yamanaka, S., Kawayokeita, Y., and Horita, Y., "Stereoscopic image quality prediction," in [*International Workshop on Quality of Multimedia Experience (QoMEx)*], 180–185, IEEE (2009).

[42] Sazzad, Z., Akhter, R., Baltes, J., and Horita, Y., "Objective no-reference stereoscopic image quality prediction based on 2D image features and relative disparity," *Advances in Multimedia* **2012**, 8 (2012).

[43] Mao, X., Yu, M., Wang, X., Jiang, G., Peng, Z., and Zhou, J., "Stereoscopic image quality assessment model with three-component weighted structure similarity," in [*Proceedings of International Conference on Audio Language and Image Processing (ICALIP)*], 1175–1179, IEEE (2010).

[44] Ninassi, A., Le Meur, O., Le Callet, P., and Barba, D., "Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric," in [*IEEE International Conference on Image Processing (ICIP)*], **2**, II–169, IEEE (2007).

[45] You, J., Perkis, A., Hannuksela, M. M., and Gabbouj, M., "Perceptual quality assessment based on visual attention analysis," in [*Proceedings of the 17th ACM International Conference on Multimedia*], 561–564, ACM (2009).

[46] Liu, H. and Heynderickx, I., "Studying the added value of visual attention in objective image quality metrics based on eye movement data," in [*Proceedings of 16th IEEE International Conference on Image Processing (ICIP)*], 3097–3100, IEEE (2009).

[47] Moorthy, A. K. and Bovik, A. C., "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing* **3**(2), 193–201 (2009).

[48] Liu, H. and Heynderickx, I., "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology* **21**(7), 971–982 (2011).

[49] Nauge, M., *The exploration of statistical structures of the image for the prediction of visual saliency and perceived quality*, PhD thesis, University of Poitiers, Poitiers (December 2012).

[50] Iatsun, I., Larabi, C.-M., and Fernandez-Mologne, C., "Spatio temporal modeling of visual attention for stereoscopic 3D video," in [*Proceedings of IEEE International Conference on Image Processing (ICIP)*], IEEE (2014).

[51] Lowe, D. G., "Object recognition from local scale-invariant features," in [*Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV)*], **2**, 1150–1157, IEEE (1999).

[52] Privitera, C. M. and Stark, L. W., "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(9), 970–982 (2000).

[53] Nauge, M., Larabi, M.-C., and Fernandez-Maloigne, C., "A statistical study of the correlation between interest points and gaze points," in [*IS&T/SPIE Electronic Imaging Proceedings*], International Society for Optics and Photonics (2012). 829111.

[54] Palou, G. and Salembier, P., "Monocular depth ordering using T-junctions and convexity occlusion cues," *IEEE Transactions on Image Processing* **22**, 1926–1939 (2013).

[55] Dimiccoli, M. and Salembier, P., "Exploiting T-junctions for depth segregation in single images," in [*Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*], 1229–1232, IEEE (2009).

[56] Dimiccoli, M., *Monocular Depth Estimation for Image Segmentation and Filtering*, PhD thesis, Universitat Politecnica de Catalunya, Barcelona, Spain (2009).

[57] Tseng, P.-H., Carmi, R., Cameron, I. G., Munoz, D. P., and Itti, L., "Quantifying center bias of observers in free viewing of dynamic natural scenes," *Journal of vision* **9**(7), 4 (2009).

[58] Tatler, B. W., "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions," *Journal of Vision* **7**(14), 4 (2007).

[59] Salembier, P. and Garrido, L., "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *Image Processing, IEEE Transactions on* **9**(4), 561–576 (2000).

[60] Itti, L. and Koch, C., "Feature combination strategies for saliency-based visual attention systems," *Journal of Electronic Imaging* **10**(1), 161–169 (2001).

[61] Chen, M.-J., Su, C.-C., Kwon, D.-K., Cormack, L. K., and

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-211

Bovik, A. C., "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication* **28**(9), 1143–1155 (2013).

[62] Blake, R., "Threshold conditions for binocular rivalry.," *Journal of Experimental Psychology: Human Perception and Performance* **3**(2), 251 (1977).

[63] Anstis, S. and Rogers, B., "Binocular fusion of luminance, color, motion and flicker–two eyes are worse than one," *Vision Research* **53**(1), 47–53 (2012).

[64] Stelmach, L. B., Tam, W. J., Meegan, D. V., Vincent, A., and Corriveau, P., "Human perception of mismatched stereoscopic 3D inputs," in [*Proceedings of International Conference on Image Processing (ICIP)*], **1**, 5–8, IEEE (2000).

[65] Brust, H., Smolic, A., Mueller, K., Tech, G., and Wiegand, T., "Mixed resolution coding of stereoscopic video for mobile devices," in [*Proceedings of 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*], 1–4, IEEE (2009).

[66] Saygili, G., Gurler, C. G., and Tekalp, A. M., "Quality assessment of asymmetric stereo video coding," in [*Proceedings of 17th IEEE International Conference on Image Processing (ICIP)*], 4009–4012, IEEE (2010).