An efficient no-reference method for stereoscopic images quality assessment

W. Hachicha, M. Kaaniche, A. Beghdadi, *Senior Member IEEE,*
and F. A. Cheikh, *Senior Member IEEE*

Abstract

The widespread use of 3D acquisition and display technologies has increased the interest of stereo image dataset in various application fields. As a result, it becomes necessary to have an efficient 3D quality assessment method to measure the human perception of stereoscopic images. While most of the state-of-the-art methods belong to the class of full-reference methods which require the original stereo images to be able to assess the quality, we propose in this paper a no-reference quality metric ~~which does not require any information of the original stereo images.~~ More precisely, after applying wavelet decompositions on the stereo images as well as its associated estimated disparity map and, resorting to appropriate statistical modeling, relevant features are extracted and used to predict the 3D quality of stereo images. Simulations, carried out on the standard Live 3D image quality database, show that our proposed design model achieves significant improvement compared to the state-of-the-art 3D quality assessment methods.

Index Terms

Stereo image, quality assessment (QA), no-reference 3D QA, wavelet transform, statistical features.

I. Introduction

The recent advances in acquisition and display technologies have increased the demand for 3D imaging through the use of stereoscopic images. Such images correspond to a pair of views, called left and right images, obtained by recording the same scene from two different viewpoints. These images allow 3D

W. Hachicha is with DOSIsoft, 45-47 avenue Carnot, 94230 Cachan, France. E-mail: walid.hachicha@dosisoft.com. M. Kaaniche and A. Beghdadi are with L2TI-Institut Galilée, Université Paris 13, Sorbonne Paris Cité, 99 avenue J. B. Clément, Villetaneuse, France, 93430. E-mail: mounir.kaaniche@univ-paris13.fr, azeddine.beghdadi@univ-paris13.fr. F. A. Cheikh is with the Norwegian Colour and Visual Computing Lab, Gjøvik University College, Norway. E-mail: faouzi.cheikh@hig.no.

reconstruction of the observed scene. This can be achieved by estimating the disparity map between the two views. For this reason, stereo images (SI) play a crucial role in various application fields such as 3DTV, computer games, and tele-presence [1]. Thus, the increasing of consumer 3D visual contents requires the design of efficient and reliable methods to assess the quality of these contents.

Since subjective quality assessment (QA) involves psychophysical experiments [2], [3], and so, is time-consuming and costly, it is more interesting to find an objective QA metric which is consistent with subjective human experience. Objective quality metrics are classified into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR). FR methods requires the original (i.e reference) image to be able to predict the quality of the distorted image [4], [5]. In the RR approaches, only some information or features of the reference image are used [6]. Finally, NR methods do not require any information of the original image to assess the quality of the distorted one [7], [8]. This class of methods is more interesting since the reference image is not available in many applications [9].

While the case of image (mono-view) quality assessment (IQA) has been extensively studied, the progress on the 3D stereoscopic image quality assessment (SIQA, also known as 3D QA) task is limited [10]. Indeed, in addition to the fist classification criterion based on the amount of information of the reference image that is available to assess the quality (FR, RR, and NR), SIQA methods are also divided into two categories. The first category of methods are based on 2D IQA models which are applied independently to the left and right views. For instance, in [11], [12], the authors perform FR 2D objective metrics on the two views to predict the quality of each image. Then, the final 3D quality score is obtained by combining, by various means (average, weighted sum, ...), the obtained 2D quality scores. In [13], the authors apply SIFT [14] and RANSAC [15] algorithms in order to extract some features (edges, corners) within each view. Then, these features are matched and used to compute the SI quality score. Moreover, a FR QA metric is also developed in [16], where the two views are firstly processed based on a HVS model by applying contourlet transform and contrast sensitivity function. Then, the difference between models of the left and right views is used to predict the quality.

While the first category of methods do not use the dispairy/depth information that can be generated from the stereo images, the second category of methods aims at exploiting this 3D information to deduce the 3D QA model. To this end, in [17], the authors propose a FR 3D QA method by combining the quality scores obtained with the stereo images as well as the disparity map. The quality score of the SI is computed by using the 2D QA metrics C4 [18] and SSIM [4]. Concerning the disparity information, a distortion measure is employed. In their paper, the authors conclude that the 2D SSIM mterics of the stereo pairs are enhanced by including the disparity information. In [19], You, *et al.* have investigated

eleven 2D image quality metrics applied on the stereo images as well as the disparity information. It has been shown that combining SSIM of the left and right views, with MAD (mean absolute difference) of the disparity map leads to a good prediction of the 3D QA. In [20], Zhu, *et al.* proposed a 3D QA method based on wavelet decomposition, contrast conversion and masking. Their main contribution consists in defining a depth sensitivity function to compute a perceptual quality metric. FR methods, based on the mechanisms of binocular fusion and suppression theories, have also been investigated [21], [22]. In order to avoid the estimation of the disparity/depth maps, Yang, *et al.* [23] proposed a FR 3D quality metric, inspired from the PSNR criterion, that is defined on the absolute difference between the two views. Instead of focusing on the quality of left and right views, other works have chosen to perform the 3D QA process on a cyclopean image. Indeed, in [24], this image is firstly computed as the average of the left image and the disparity compensated right image. Then, the quality metric is obtained by comparing the sensitivity coefficients of the original and distorted cyclopean images and studying the coherence of their associated disparity maps. In [25], the cyclopean image is generated by taking a weighted sum of the left image and the disparity-compensated right image. Then, FR 2D QA models are applied to the original and distorted cyclopean images to obtain the 3D quality score.

It is important to note here that these different approaches, and generally most of the state-of-the-art ones are FR methods. However, only few works have been devoted to the case of RR and NR 3D QA models, which are more interesting in practice. To this respect, Hewage, *et al.* [26] proposed a RR 3D QA method by computing the PSNR between the reference and distorted edges maps, extracted from the depth information. A RR 3D QA is also developed in [27] by using the entropy of the left and right views as well as their mutual information. In [28], Shao, *et al.* proposed a binocular guided quality lookup and visual codebook to develop a NR quality metric by simply pooling. Ryu and Sohn [29] developed also a NR quality metric by adopting a top-down approach modeling the binocular quality perception of the Human Visual System in the context of blockiness and blurriness. In [30], a NR metric is given by using the segmented local features of artifacts and disparity. However, their method is designed especially for JPEG compressed stereo images. Similarly, Akther, *et al.* [31] proposed a NR 3D QA method that extracts features from the SI as well as an estimated disparity map. Recently, Chen, *et al.* [32] proposed a NR QA method based on *statistical* features extraction from a cyclopean image [25] as well as the estimated disparity map and its associated uncertainty map. While the uncertainty map is modeled by using the log-normal distribution, the authors have applied a local mean subtraction and divisive variance normalization on the cyclopean image and the disparity map in order to model them with the Generalized Gaussian (GG) distribution.

In this paper, a novel NR 3D QA method is developed. The proposed method operates in the wavelet transform domain and adopts a statistical framework to predict the quality of stereoscopic images. While a natural scene statistics-based approach has also been recently developed by Chen *et al.* in [32], it is worth pointing out that the main contributions of the proposed work with respect to the other one [32] concern *both* the 2D texture feature extraction and the depth feature extraction techniques. More precisely, instead of fusing the two views in one image, referred to as a cyclopean image, and modeling it in the spatial domain as performed in [32], we first propose to extract the 2D features from both views in the wavelet transform domain. To this end, a *joint* wavelet decomposition, that exploits the redundancies between the two views, is developed to generate two compact multiresolution representations of the left and right images. Moreover, concerning the disparity map, we propose to apply a classical wavelet transform on the estimated map, and model its resulting wavelet subbands by using the Bernoulli Gaussian Generalized (BGG) distribution which is well adapted for modeling sparse signals. Indeed, unlike natural images, the disparity maps have different characteristics and exhibit much sparser structures. For this reason, and as it will be shown in Section III-B, the GG distribution may fail to model the disparity maps and the BGG one is found to be more appropriate for this modeling.

The remainder of this paper is organized as follows. In Section II, the joint wavelet decomposition, applied to the left and right images, is presented. Then, we describe in Section III the texture as well as the depth feature extraction techniques for the design of the NR 3D QA method. Finally, the performance of the proposed approach for various kinds of distortions are illustrated in Section IV and some conclusions are given in Section V.

II. Joint wavelet decomposition

Wavelet representation has ~~been~~ attracted a considerable attention in image analysis literature (in particular texture characterization and feature extraction) as it is more consistent with the characteristics of Human Visual System (HVS), that states that the ~~eye~~ decomposes images into different subbands and processes each independently [33]. For this reason, we first propose to generate two wavelet representations of the left image $I^{(l)}$ and the right one $I^{(r)}$. As these images correspond to the same 3D scene, and so contain redundant information, an efficient *joint* wavelet decomposition can be designed in order to exploit the cross-views redundancies. To this end, we will resort to the concept of Vector Lifting Scheme (VLS) [34]. We should note that the cross-views redundancies can be exploited by using the estimated disparity map. To this respect, we have used the disparity estimation (DE) method presented recently in [35]. Once the disparity map is available, we will now describe the principle of the joint wavelet

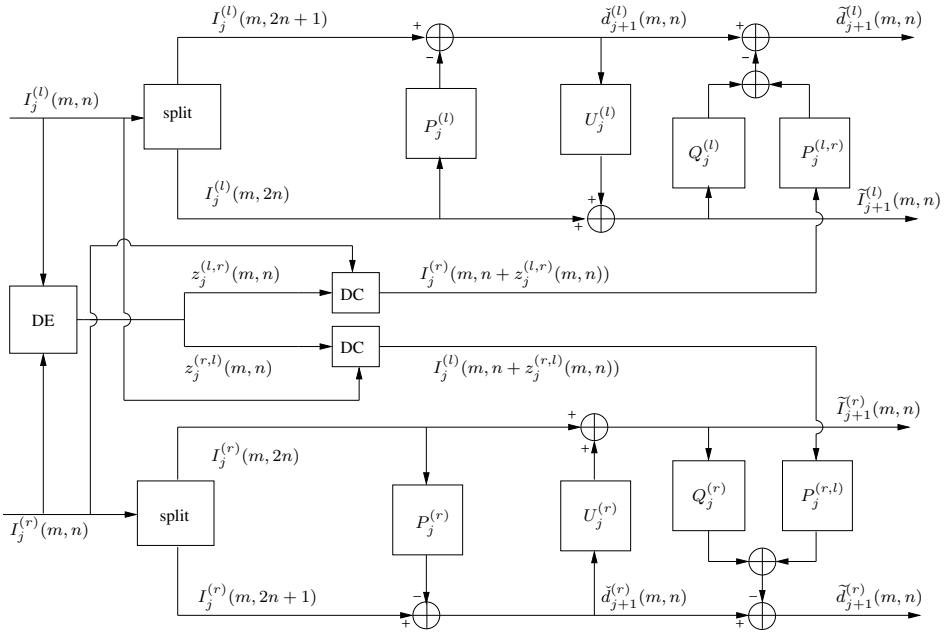


Fig. 1: The proposed joint wavelet decomposition.

decomposition.

The analysis structure based one the VLS is illustrated in Fig. 1. Indeed, for a given resolution level j, and for each image line m, two Predict-Update-Predict (PUP) lifting structures are applied on the input coefficients of the left and right images $I_j^{(l)}$ and $I_j^{(r)}$.

More precisely, concerning the right image, a first prediction step is performed to generate an intermediate detail signal $\check{d}_{j+1}^{(r)}$ (at resolution $j+1$) given by:

$$\check{d}_{j+1}^{(r)}(m,n) = I_j^{(r)}(m,2n+1) - \Big(\sum_{k \in \mathcal{P}_j^{(r)}} p_{j,k}^{(r)} I_j^{(r)}(m,2n-2k) \Big), \tag{1}$$

where the coefficients $p_{j,k}^{(r)}$ and the set $\mathcal{P}_j^{(r)}$ are respectively the weights and the support of the predictor of the odd samples of the right image $I_j^{(r)}(m,2n+1)$. After that, an update step is applied to compute the approximation coefficients $\widetilde{I}_{j+1}^{(r)}$ as follows:

$$\widetilde{I}_{j+1}^{(r)}(m,n) = I_j^{(r)}(m,2n) + \Big(\sum_{k \in \mathcal{U}_j^{(r)}} u_{j,k}^{(r)} \check{d}_{j+1}^{(r)}(m,n-k) \Big), \tag{2}$$

where the coefficients $u_{j,k}^{(r)}$ and the set $\mathcal{U}_j^{(r)}$ are respectively the weights and the support of the update operator. Finally, a second hybrid prediction step is performed to generate the final version of the detail

coefficients $\widetilde{d}_{j+1}^{(r)}$:

$$\widetilde{d}_{j+1}^{(r)}(m,n) = \check{d}_{j+1}^{(r)}(m,n) - \Big(\sum_{k \in \mathcal{Q}_j^{(r)}} q_{j,k}^{(r)} \widetilde{I}_{j+1}^{(r)}(m, n-k) + \sum_{k \in \mathcal{P}_j^{(r,l)}} p_{j,k}^{(r,l)} I_j^{(l)}(m, 2n+1+z_j^{(r,l)}(m,2n+1)-k) \Big), \quad (3)$$

where $\mathcal{Q}^{(r)}{}_j$ (resp. $\mathcal{P}_j^{(r,l)}$) and $q_{j,k}^{(r)}$ (resp. $p_{j,k}^{(r,l)}$) are the support and weights of the second intra (resp. inter)-image predictor for the right view, and $z_j^{(r,l)}$ is obtained by sampling and dividing by 2^j the initial estimated disparity map $z^{(r,l)}$ since the dimensions of the initial images are divided by 2^j at the j-th resolution level:

$$z_j^{(r,l)}(m,n) = \frac{1}{2^j} z^{(r,l)}(2^j m, 2^j n). \quad (4)$$

Note that the superscript (r,l) used with the disparity notation z_j is employed to distinguish the selected reference image during the disparity estimation process. Indeed, $z_j^{(r,l)}$ (resp. $z_j^{(l,r)}$) corresponds to the disparity values which allows to find for each pixel in the right (resp. left) image its homologous one in the left (resp. right) image. As it can be shown in Fig. 1 as well as Eq. (3), the disparity map $z_j^{(r,l)}$ is used in the disparity compensation (DC) process to compute $I_j^{(l)}(m, n + z_j^{(r,l)}(m,n))$ used in the second hybrid prediction stage.

This decomposition will be applied by considering a prediction and update filters with spatial supports $\mathcal{P}_j^{(r)} = \{-1, 0\}$ and $\mathcal{U}_j^{(r)} = \{0, 1\}$. Their associated weights are given by: $p_{j,-1}^{(r)} = p_{j,0}^{(r)} = \frac{1}{2}$ and $u_{j,0}^{(r)} = u_{j,1}^{(r)} = \frac{1}{4}$, like the well-known 5/3 lifting scheme [36]. Concerning the second hybrid prediction stage, it is performed by taking $\mathcal{Q}_j^{(r)} = \{-1, 0\}$ and $\mathcal{P}_j^{(r,l)} = \{-1, 0, 1\}$. The coefficients $q_{j,k}^{(r)}$ and $p_{j,k}^{(r,l)}$ are optimized by minimizing the variance of the detail coefficients $\widetilde{d}_{j+1}^{(r)}$.

Once the decomposition strategy is handled for each line of the right image, a similar P-U-P based decomposition will also be applied on each line of the left image. For instance, by performing these decompositions along the lines and the columns of the right and left images, we generate one approximation subband as well as three detail subbands oriented horizontally, vertically, and diagonally for each image. Finally, two multiresolution representations of both images are obtained by repeating the same process on the right and left approximation subbands $I_{j+1}^{(r)}$ and $I_{j+1}^{(l)}$.

It should be noted that the key feature of such decomposition with respect to a classical lifting structure is that, for a given image, the prediction of the odd samples involves the neighboring samples of the same image *and also* neighbors of the matching samples taken from the *other* image thanks to the disparity compensation (DC) process. For this reason, the VLS tool has been found to be an efficient tool in

exploiting the inter-images dependencies, and, its efficiency in the context of stereo image coding has already been investigated in [34]. However, it is worth pointing out that this joint wavelet decomposition presents a main difference compared to that presented in [34]. Indeed, in the context of stereo image coding, it is mandatory (for a decoding purpose) that one image (for example the left one), referred to as a reference image, is decomposed by using a classical P-U lifting structure, and only, the other image (the right one) is decomposed by resorting to joint multiscale decomposition. However, this decoding constraint is not necessary in the context of stereo images quality assessment. For this reason, we have preferred to apply the joint decomposition on both images, to generate two compact multiresolution representations of the left and right views.

III. Proposed NR metric

Motivated by the benefits of exploiting the 3D information in the QA process, the proposed method, whose block diagram is illustrated in Fig. 2, belongs to the category of NR 3D QA methods described in Section I. Therefore, in addition to the left and right views, the disparity map will be employed to obtain a good prediction of SI quality. To this end, for each stereo image, the disparity maps are estimated by using the disparity estimation (DE) technique presented in [35].

As mentioned in Section II, the generated disparity fields are used to apply the joint wavelet decomposition on the stereo images, and so, produce two multiresolution representations of the left and right views. Then, some statistical features are extracted from the resulting wavelet subbands. This step is referred to as *texture feature extraction*. In addition to the visual contents of both views, further statistical features are also extracted from the disparity maps to exploit the 3D information. This step is referred to as *depth feature extraction*. Finally, both statistical features are combined and used in the quality prediction stage to assess the quality of each tested stereo images.

In what follows, the texture as well as the depth feature extraction strategies ~~will be~~ described.

A. Texture feature extraction

Wavelet representation has been found to be an effective way ~~that allows~~ to extract salient feature characterizing the texture of a given natural image. This can be achieved by adopting a statistical framework to model the different wavelet subbands. To this end, the Generalized Gaussian (GG) model has been extensively used to model distribution of the detail wavelet coefficients [37].

In our context of SIQA, and once the joint wavelet decomposition is performed, we will denote by $\left(I_j^{(l)}, I_j^{(l,o)}\right)_{\substack{o\in\{HL,LH,HH\}\\ j\in\{1,\ldots,J\}}}$ (resp. $\left(I_j^{(r)}, I_j^{(r,o)}\right)_{\substack{o\in\{HL,LH,HH\}\\ j\in\{1,\ldots,J\}}}$) the approximation as well as the three

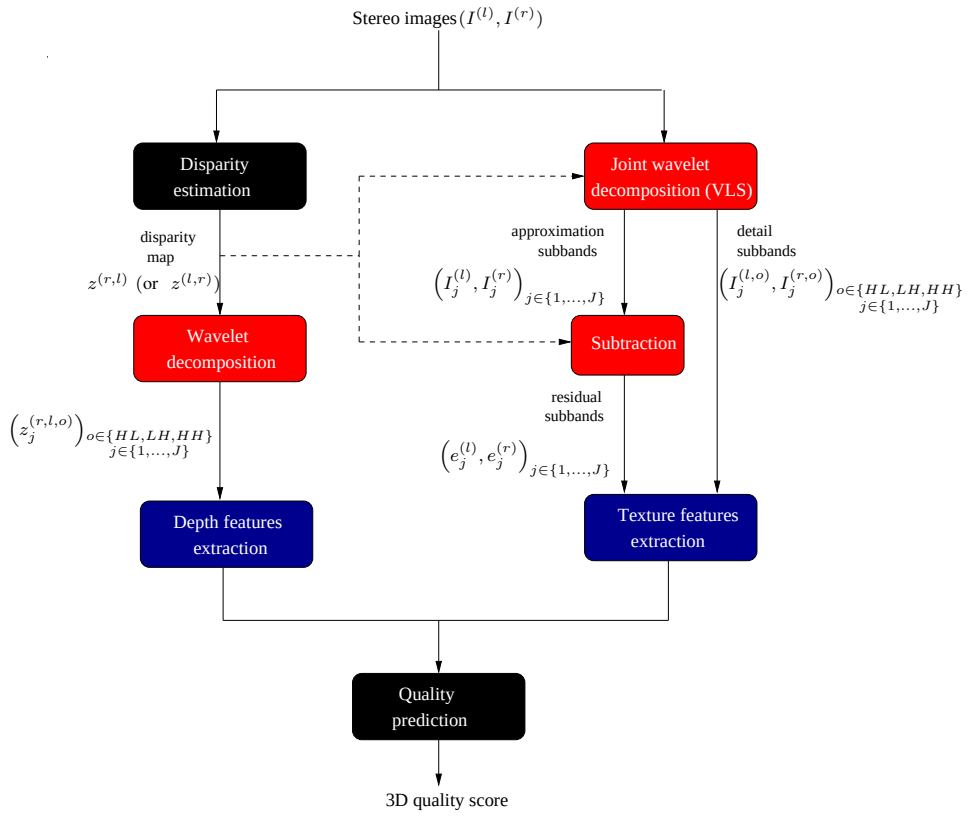


Fig. 2: Block diagram of the proposed NR 3D QA metric.

details subbands oriented horizontally (HL), vertically (LH) and diagonally (HH), at each resolution level j, of the left (resp. right) image wavelet representation. Note that J corresponds to the number of resolution levels used with the multiscale decomposition.

First, the resulting wavelet details subbands of both images are modeled by using the GG distribution whose probability density function is given by:

$$\forall \xi \in \mathbb{R}, \qquad \forall v \in \{l, r\}, \qquad \tilde{g}_j^{(v,o)}(\xi) = \frac{\beta_j^{(v,o)} (\omega_j^{(v,o)})^{1/\beta_j^{(v,o)}}}{2\Gamma(1/\beta_j^{(v,o)})} e^{-\omega_j^{(v,o)} |\xi|^{\beta_j^{(v,o)}}} \tag{5}$$

where Γ is the gamma function defined by:

$$\forall x > 0, \qquad \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt, \tag{6}$$

$\omega_j^{(v,o)}$ and $\beta_j^{(v,o)}$ are respectively the scale and shape parameters for the left view (i.e "v" $= l$) and the right one (i.e "v" $= r$). These parameters can be estimated using the method of moments or maximum likelihood technique [37]. In addition, the variance $\sigma_j^{(v,o)}$ as well as the kurtosis $\kappa_j^{(v,o)}$ of each detail wavelet subband are also computed. In this way, these retained statistical parameters are grouped into

one feature vector.

In order to study the ability of this texture-based statistical features to distinguish different distortions, five common distortion types have been applied on a reference stereo images. These distortions correspond to a white noise (WN), JPEG compression, JPEG2000 compression, and fast fading (FF) model based on the Rayleigh fading channel. An example of stereo images affected by these distortions is illustrated in Fig. 3, and, the GG distributions of their associated horizontal detail subbands are shown in Fig. 4. While each kind of distortion has its own impact on the distribution of the wavelet subbands, the adopted statistical features may lead to a confusion between some distortions, like the blur and JPEG2000 ones as shown in Fig. 4.

To alleviate this shortcoming, the information given by the approximation subbands will be exploited in addition to the statistical parameters of the detail subbands. However, it is known that the GG distribution (5) is not well suited for modeling the coefficients of the approximation subbands. For this reason, during the joint wavelet decomposition process described in the previous section, we propose to generate *residual* subbands from the approximation ones of both left and right views. More precisely, similarly to a prediction error, the residual subbands of the left and right views, denoted respectively by $e_j^{(l)}$ and $e_j^{(r)}$, are computed by using disparity compensation (DC) followed by the subtraction operation:

$$\begin{cases} e_j^{(l)}(m,n) = I_j^{(l)}(m,n) - I_j^{(r)}(m, n + z_j^{(l,r)}(m,n)) \\ e_j^{(r)}(m,n) = I_j^{(r)}(m,n) - I_j^{(l)}(m, n + z_j^{(r,l)}(m,n)) \end{cases} \tag{7}$$

Thus, the GG distribution parameters of these residual subbands as well as their variance and kurtosis ones are further extracted. These parameters will be respectively denoted by $\beta^{e_j^{(v)}}$, $\omega^{e_j^{(v)}}$, $\sigma^{e_j^{(v)}}$ and $\kappa^{e_j^{(v)}}$, with $v \in \{l, r\}$.

To confirm the relevance of these statistical features, Fig. 5 illustrates the distributions of the resulting residual subbands for the different retained distortions applied on the same stereo images considered previously. Thus, unlike Fig. 4, this Figure shows that the previous confusion that may occur between the blur and JPEG2000 distortions is avoided, and also, the different distortions are better distinguished. Therefore, the proposed statistical-based texture feature vector is given by:

$$\mathbf{f} = (\mathbf{f}^{(l)}, \mathbf{f}^{(r)})$$

$$\text{where} \qquad \forall\ v \in \{l, r\}, \quad \mathbf{f}^{(v)} = \left(\beta_j^{(v,o)}, \omega_j^{(v,o)}, \sigma_j^{(v,o)}, \kappa_j^{(v,o)}, \beta^{e_j^{(v)}}, \omega^{e_j^{(v)}}, \sigma^{e_j^{(v)}}, \kappa^{e_j^{(v)}}\right)_{\substack{o \in \{HL,LH,HH\} \\ j \in \{1,\ldots,J\}}} \tag{8}$$

B. Depth features extraction

Compared to natural images, there are relatively few works that have studied the statistics of the disparity/depth maps. For instance, Hibbard [38] and Liu *et al.* [39] have observed that the distribution of the disparity follows a Laplacian shape. Note that this distribution corresponds to the particular case of the GG one when its shape parameter is equal to one ($\beta_j = 1$). While Hibbard have modeled the fixation distance, Liu *et al.* have used accurate range maps to compute the disparities. However, in the context of NR SIQA, only the distorted SI are available, and so, accurate range maps (or ground truths) can not be exploited to design a NR quality metric. Therefore, a disparity map should be estimated directly from the given distorted SI. To this end, Chen *et al.* [32] have proposed recently to estimate the disparity map by using a SSIM-based stereo matching algorithm. Then, a pre-processing step, which consists of a local mean substraction and divisive normalization, is applied to the estimated disparity maps in order to model them by a GG distribution and extract its associated statistical parameters. However, we have observed that the pre-processed disparity maps are not well fitted by a GG distribution.

To alleviate this drawback, we propose in this paper to resort a more general and appropriate model. More precisely, we first assume that the disparity maps are available by applying a given DE technique. As mentioned before, we used the method presented recently in [35] to generate the disparity maps. This method has the advantage of producing a dense and smooth disparity maps while preserving the depth discontinuities. Note that the efficiency of this method with respect to other popular methods such as graph cuts and variational approaches have already been shown and studied in [35]. It is worth noting, that among these different DE techniques, it is important to employ an efficient technique that guarantees the smoothness property of the estimated maps. Indeed, this property allows us to interpret such maps as a natural image which will be the object of a wavelet decomposition in order to be efficiently exploited in the feature extraction step. Once the disparity maps are available, we propose then to generate their wavelet representations by applying the well-known 5/3 lifting scheme [36] over J resolution levels. Finally, statistical features will be extracted from the different wavelet subbands. To model the distribution of these sparse representations, we propose to use the Bernoulli-Generalized Gaussian (BGG) model whose probability density function is given by [40]:

$$\forall \xi \in \mathbb{R}, \qquad g_j^{(z)}(\xi) = (1 - \epsilon_j^{(z)})\delta(\xi) + \epsilon_j^{(z)} \tilde{g}_j^{(z)}(\xi), \tag{9}$$

where $\tilde{g}_j^{(z)}(\xi)$ represents the probability density function of a GG distribution (characterized by its shape $\beta_j^{(z)}$ and scale $\omega_j^{(z)}$ parameters), $\epsilon_j^{(z)} \in [0, 1]$ denotes the mixture parameter and δ is the Dirac distribution. Note that the notation $z^{(r,l)}$ (or $z^{(l,r)}$) has been simply replaced in this part by z since it has been observed

that the disparities $z^{(r,l)}$ and $z^{(l,r)}$ presents similar statistics.

Therefore, the final depth feature vector $\mathbf{f}^{(z)}$ will be composed of the mixture, scale and shape parameters of the BGG model as well as the variance and kurtosis ones of the different detail subbands of the estimated disparity maps:

$$\mathbf{f}^{(z)} = \left(\epsilon_j^{(z,o)}, \beta_j^{(z,o)}, \omega_j^{(z,o)}, \sigma_j^{(z,o)}, \kappa_j^{(z,o)}\right)_{\substack{o\in\{HL,LH,HH\}\\ j\in\{1,\ldots,J\}}} \tag{10}$$

~~It is important to emphasize here that the motivation of using the BGG model is the following. Indeed, by conducting many experiments on the disparity maps corresponding to the stereo images of the live 3D database~~ [10], we have observed that the distributions of disparities are generally centered at zero and ~~have high peaks.~~ This observation allows us to conclude that the disparity map can be interpreted as a sparse signal. ~~For such sparse representation, the GG model may fail to model its distribution and the BGG one becomes more appropriate for that purpose.~~ To confirm this, we first illustrate in Fig. 6 the disparity maps associated to the different distorted stereo images shown in Fig. 3. Note that these disparity maps are estimated by selecting the right image as a reference image. Then, by taking the disparity map given by Fig. 6(d) as an example, we focus on the modeling of the diagonal detail coefficients by displaying their histogram (in blue) as well as the probability density function of the theoretical GG and BGG models (in red). It can be seen that the BGG model fits well the detail coefficients distribution, and so, it is more accurate than the GG one.

C. Stereo image quality evaluation

Once the texture and depth features are extracted, two procedures can be generally used to obtain a measure of an image quality [7], [32], [41]. In the first one, the features are directly mapped to a quality score thanks to a regression ~~module.~~ However, in the second one, a two-stage framework is performed. First, the probability of most likely distortion type affecting the image is computed. A quality score relative to each distortion is obtained by using ~~regression.~~ After that, the final predicted quality score is deduced as the dot product of the distortion probability vector and the vector of distortion quality scores. It should be noted here that such two-stage procedure has been found to be efficient for NR IQA [7], and, has also been adopted recently for NR SIQA [32]. For this reason, and for fair comparison, we have employed the two-stage approach to predict the quality of stereo images.

IV. Experimental results

Our simulations are performed on the popular LIVE 3D image quality database presented in [10]. This database is composed of 20 reference images and 365 distorted stereo images with their associated

Difference Mean Opinion Scores (DMOS). To this end, five common distortion types are considered: compression using JPEG and JPEG2000 coding techniques, Gaussian blur, additive white Gaussian noise (WN) and fast fading (FF) based on the Rayleigh fading channel model. More precisely, each of the JPEG, JPEG2000, WN and FF distortions are applied on 80 pairs while the remaining 45 stereo images are distorted by the blur one. Moreover, symmetric distortions (i.e. same amount of distortions) are used with the left and right views. Furthermore, for each distortion type, different degradation levels of visual contents are achieved by varying a given control parameter within a pre-defined range [10].

To evaluate the performance of the proposed NR 3D QA method, we ~~will~~ consider the following SIQA developed by: Gorley *et al.* [13], Shen *et al.* [16], Benoit *et al.* [17], You *et al.* [19], Zhu *et al.* [20], Yang *et al.* [23], Hewage *et al.* [26], Akhter *et al.* [31], and Chen *et al.* [32]. Let us recall that all these approaches have been summarized in Section I. Moreover, as aforementioned, most of the reported 3D QA methods, in particular [13], [16], [17], [19], [20] and [23], belong to the class of FR methods. However, the approach developed in [26] is a RR method, and those presented in [31], [32] are NR methods.

Since a two-stage-based prediction ~~module~~ has been adopted in this work as mentioned in Section III-C, and also for fair comparison with [32], SVM and SVR are considered as classifier and regression models to distinguish the different distortion types and predict the stereo image quality, respectively. To this end, the LIBSVM package [42] has been employed. More precisely, the classification is performed by using 1000 iterations of the train-test step. At each iteration, 80% of the dataset is randomly chosen as a training data while the remaining 20% of the dataset is used for testing. In this manner, we have obtained a mean classification accuracy of 94.9%, which proves that the proposed 3D features are discriminative and allow to distinguish the different distortion types.

While the interest of the proposed texture and depth feature extraction techniques has already been shown and discussed in Section III (Figures 4-7), we will now compare our proposed method to the state-of-the-art 3D QA ones. To this respect, the following standard performance measures are evaluated: Spearman's Rank Ordered Correlation Coefficient (SROCC), Pearson's linear correlation coefficient (LCC) and root-mean-squared error (RMSE). Note that RMSE shows the prediction consistency of a given QA method, and, LCC and SROCC reflect its good correlation with human judgments. Thus, lower values of RMSE and higher LCC and SROCC values (closer to 1) show good performance.

The performance of the different SIQA methods for each distortion type in terms of LCC, SROCC and RMSE are given in Table I, II and III, respectively. Moreover, Table IV shows the performance of these methods for all the database images. From these tables, it can be seen that the NR 3D QA

methods developed by Akther *et al.* [31] and Chen *et al.* [32] outperform most of the FR and RR 3D QA approaches except the FR ones proposed by Benoit *et al.* [17] and Yang *et al.* [23]. In addition, as it has already been outlined in previous studies, it should be noted that the performance of these methods (FR, RR and NR ones) is significantly low for the JPEG distortion type. However, our proposed method yields better performance compared to these RR and FR state-of-the-art QA methods. Most importantly, our proposed NR 3D QA approach achieves a significant improvement compared to the homologous ones (i.e. NR ones). For instance, when considering, the JPEG distortion, our approach results in a SROCC (resp. LCC) value of about 0.87 (resp. 0.88) while those obtained with the NR methods developed by Akhter *et al.* and Chen *et al.* are equal to respectively equal to 0.67 and 0.62 (resp. 0.72 and 0.69). Therefore, the obtained results show that the proposed 3D quality assessment method yields a good metric for predicting the quality of stereo images.

V. Conclusion

In this paper, we have focused on the quality assessment issue of stereoscopic images. More precisely, a no-reference 3D quality metric has been developed. The idea consists of using statistical features of the left and right views as well as the disparity map to assess the quality of a distorted stereo images. To this end, we have resorted to efficient wavelet decomposition and appropriate distribution models to extract discriminative features. Experimental results show the effectiveness of the proposed 3D quality assessment technique compared to the recent state-of-the-art methods. In future work, it would be interesting to investigate other features by resorting to joint statistical modeling to exploit the inter-views dependencies. Moreover, a more general dataset should be constructed to take into account other types of 3D distortions by including for example stereo pairs with different resolutions, asymmetric distortions that may result from stereo image coding applications [43], and, vertical disparity component that will be obtained in the case where the stereo images are not rectified.

References

[1] I. Feldmann, W. Waizenegger, N. Atzpadin, and O. Schreer, "Real-time depth estimation for immersive 3D videoconferencing," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Tampere, June 2010, pp. 1–4.

[2] J.-S. Lee, L. Goldmann, and T. Ebrahimi, "Paired comparison-based subjective quality assessmentof stereoscopic images," *Multimedia Tools and Applications*, vol. 67, no. 1, pp. 31–48, 2012.

[3] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "Assessment of stereoscopic crosstalk perception," *IEEE Transactions on Multimedia*, vol. 14, no. 2, pp. 326–337, 2012.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions onImage Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[5] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[6] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 239–256, 2008.

[7] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[8] S. Varadarajan and L. J. Karam, "An improved perception-based no-reference objective image sharpness metric using iterative edge refinement," in *International Conference on Image Processing*. San Diego, CA, USA: IEEE, 2008, pp. 401–404.

[9] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: what does the future hold?" *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 675–696, 2011.

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

[11] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *European Signal Processing Conference*, Poznan, Poland, 2007.

[12] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, and A. M. Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-D video," *Electronics Letters*, vol. 44, no. 16, pp. 963–965, 2008.

[13] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2008, pp. 680 305–680 305.

[14] D. G. Lowe, "Object recognition from local scale-invariant features," in *International conference on Computer vision*, vol. 2. Kerkyra, Greece: Ieee, 1999, pp. 1150–1157.

[15] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[16] L. Shen, J. Yang, and Z. Zhang, "Stereo picture quality estimation based on a multiple channel HVS model," in *International Congress on Image and Signal Processing*, Tianjin, China, 2009, pp. 1–4.

[17] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP journal on image and video processing*, vol. 2008, 2009.

[18] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *International Conference on Image Processing*, vol. 3. IEEE, 2003, pp. III–185.

[19] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, AZ, USA, 2010.

[20] Z. Zhu and Y. Wang, "Perceptual distortion metric for stereo video quality evaluation," *WSEAS Transactions on Signal Processing*, vol. 5, no. 7, pp. 241–250, 2009.

[21] W. Hachicha, A. Beghdadi, and F. A. Cheikh, "Stereo image quality assessment using a binocular just noticeable difference model," in *IEEE International Conference on Image Processing*, Melbourne, Australia, 2013.

[22] G. Jiang, J. Zhou, M. Yu, Y. Zhang, F. Shao, and Z. Peng, "Binocular vision based objective quality assessment method for stereoscopic images," *Multimedia Tools and Applications*, vol. 74, no. 18, pp. 8197–8218, 2015.

[23] J. Yang, C. Hou, Y. Zhou, Z. Zhang, and J. Guo, "Objective quality assessment method of stereo images," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*. Potsdam, Germany: IEEE, 2009, pp. 1–4.

[24] A. Maalouf and M.-C. Larabi, "Cyclop: A stereo color image quality assessment metric," in *International Conference on Acoustics, Speech and Signal Processing*. Prague, Czech Republic: IEEE, 2011, pp. 1161–1164.

[25] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

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

[27] F. Qi, D. Zhao, and W. Gao, "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2338–2344, 2015.

[28] F. Shao, W. Lin, S. Wang, G. Jiang, and M. Yu, "Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook," *IEEE Transactions on Broadcasting*, vol. 61, no. 2, pp. 154–165, 2015.

[29] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, vol. 24, no. 4, pp. 591–602, 2014.

[30] Z. M. P. Sazzad, S. Yamanaka, Y. Kawayokeita, and Y. Horita, "Stereoscopic image quality prediction," in *International Workshop on Quality of Multimedia Experience*. San Diego, CA, USA: IEEE, 2009, pp. 180–185.

[31] R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," in *IS&T/SPIE Electronic Imaging*, vol. 7524, 2010.

[32] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, 2013.

[33] Z. Gao and Y. F. Zheng, "DWT-based image quality metric," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 910–922, 2008.

[34] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet, "Vector lifting schemes for stereo image coding," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2463–2475, 2009.

[35] W. Miled, J.-C. Pesquet, and M. Parent, "A convex optimization approach for depth estimation under illumination variation," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 813–830, 2009.

[36] W. Sweldens, "The lifting scheme: a new philosophy in biorthogonal wavelet construction," in *Wavelet Applications in Signal and Image Processing III, SPIE*, San-Diego, CA, USA, September 1995, pp. 68–79.

[37] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, February 2002.

[38] P. B. Hibbard, "A statistical model of binocular disparity," *Visual Cognition*, vol. 15, no. 2, pp. 149–165, 2007.

[39] Y. Liu, A. C. Bovik, and L. K. Cormack, "Disparity statistics in natural scenes," *Journal of Vision*, vol. 8, no. 11, p. 19, 2008.

[40] M. Kaaniche, A. Fraysse, B. Pesquet-Popescu, and J.-C. Pesquet, "A bit allocation method for sparse source coding," *IEEE Transactions on Image processing*, vol. 23, no. 1, pp. 137–152, January 2014.

[41] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[42] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, p. 27, 2011.

[43] W. Hachicha, M. Kaaniche, A. Beghdadi, and F. A. Cheikh, "Efficient inter-view bit allocation methods for stereo image coding," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 765–777, 2015.

TABLE I: Performance evaluation of the proposed NR metric: Linear Correlation Coefficient (LCC).

Algorithm	Category	JP2K	JPEG	WN	Blur	FF
Gorley [13]	FR	0.485	0.312	0.796	0.852	0.364
Shen [16]	FR	0.503	0.389	0.898	0.684	0.483
Benoit [17]	FR	0.939	0.640	0.925	0.948	0.747
You [19]	FR	0.877	0.487	0.941	0.919	0.730
Zhu [20]	FR	0.807	0.379	0.517	0.777	0.503
Yang [23]	FR	0.920	0.640	0.930	0.930	0.740
Hewage [26]	RR	0.904	0.530	0.895	0.798	0.669
Akhter [31]	NR	0.905	0.729	0.904	0.617	0.660
Chen [32]	NR	0.907	0.695	0.917	0.917	0.735
Proposed	NR	**0.965**	**0.882**	**0.961**	**0.969**	**0.842**

TABLE II: Performance evaluation of the proposed NR metric: Spearman's Rank Ordered Correlation Coefficient (SROCC).

Algorithm	Category	JP2K	JPEG	WN	Blur	FF
Gorley [13]	FR	0.420	0.015	0.740	0.749	0.366
Shen [16]	FR	0.213	0.244	0.891	0.658	0.266
Benoit [17]	FR	0.910	0.603	0.929	0.931	0.698
You [19]	FR	0.859	0.438	0.939	0.882	0.588
Zhu [20]	FR	0.770	0.292	0.465	0.793	0.475
Yang [23]	FR	0.902	0.601	0.937	0.928	0.695
Hewage [26]	RR	0.855	0.500	0.896	0.690	0.544
Akhter [31]	NR	0.865	0.675	0.913	0.554	0.639
Chen [32]	NR	0.863	0.617	0.919	0.878	0.652
Proposed	NR	**0.924**	**0.865**	**0.949**	**0.933**	**0.768**

TABLE III: Performance evaluation of the proposed NR metric: Root-mean-squared-error (RMSE).

Algorithm	Category	JP2K	JPEG	WN	Blur	FF
Gorley [13]	FR	11.323	6.212	10.197	7.562	11.569
Shen [16]	FR	12.275	6.022	7.294	10.554	10.882
Benoit [17]	FR	4.426	5.022	6.307	4.571	8.257
You [19]	FR	6.206	5.709	5.621	5.679	8.492
Zhu [20]	FR	7.681	6.068	14.720	9.127	10.736
Yang [23]	FR	4.421	5.019	6.298	4.570	8.252
Hewage [26]	RR	5.530	5.543	7.405	8.748	9.226
Akhter [31]	NR	5.4836	4.4736	7.0929	11.387	9.3321
Chen [32]	NR	5.402	4.523	6.433	5.898	8.322
Proposed	NR	**3.247**	**2.950**	**4.377**	**3.301**	**6.266**

TABLE IV: Performance evaluation of the proposed NR metric for all the database images.

Algorithm	Category	LCC	SROCC	RMSE
Gorley [13]	FR	0.451	0.142	14.635
Shen [16]	FR	0.574	0.068	13.547
Benoit [17]	FR	0.902	0.899	7.063
You [19]	FR	0.881	0.879	7.747
Zhu [20]	FR	0.626	0.639	12.782
Yang [23]	FR	0.870	0.900	7.060
Hewage [26]	RR	0.830	0.814	9.139
Akhter [31]	NR	0.427	0.383	14.827
Chen [32]	NR	0.895	0.891	7.247
Proposed	NR	**0.932**	**0.924**	**5.840**

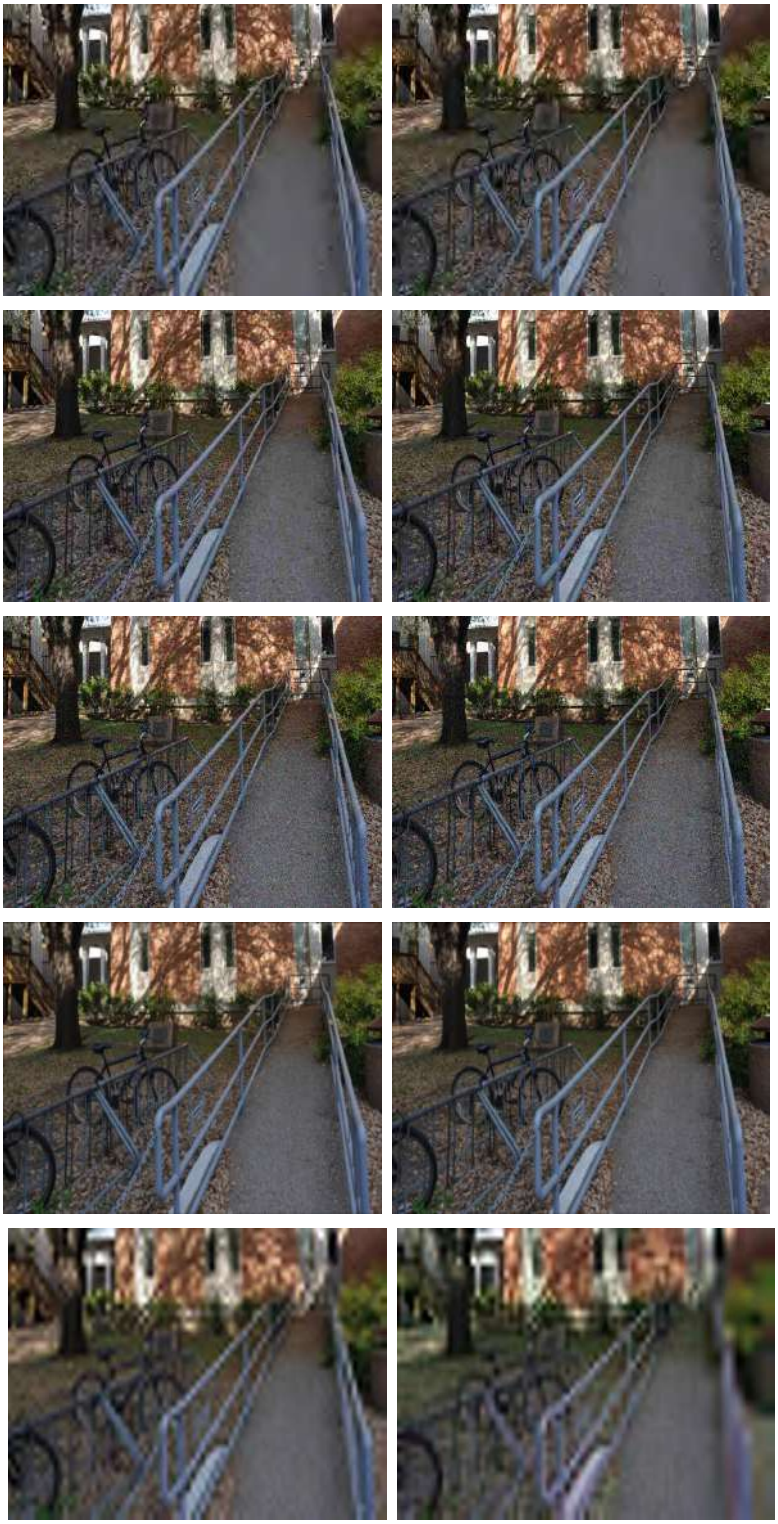


Fig. 3: The different distortion types obtained with a given stereo image using (from top to bottom): JPEG 2000, JPEG, WN, Blur and Fast fading.

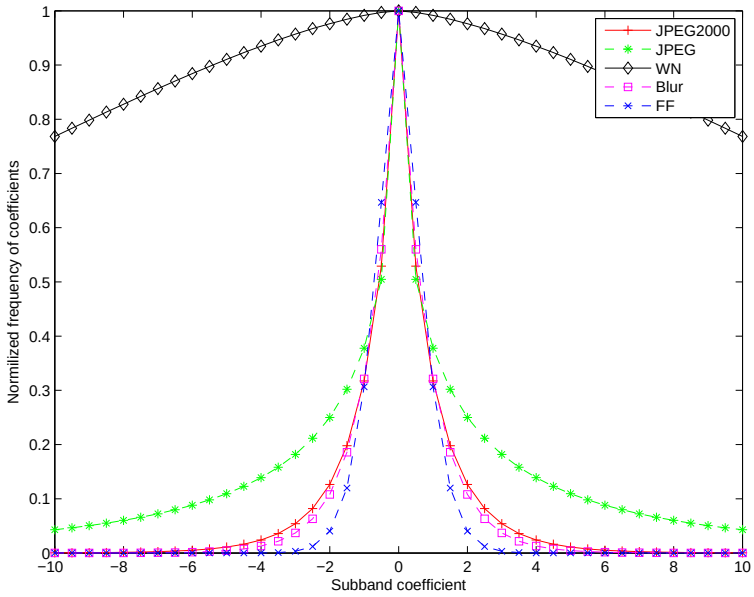


Fig. 4: GG distributions of the first scale horizontal detail subband of the stereo image illustrated in Figure 3 with different distortion types.

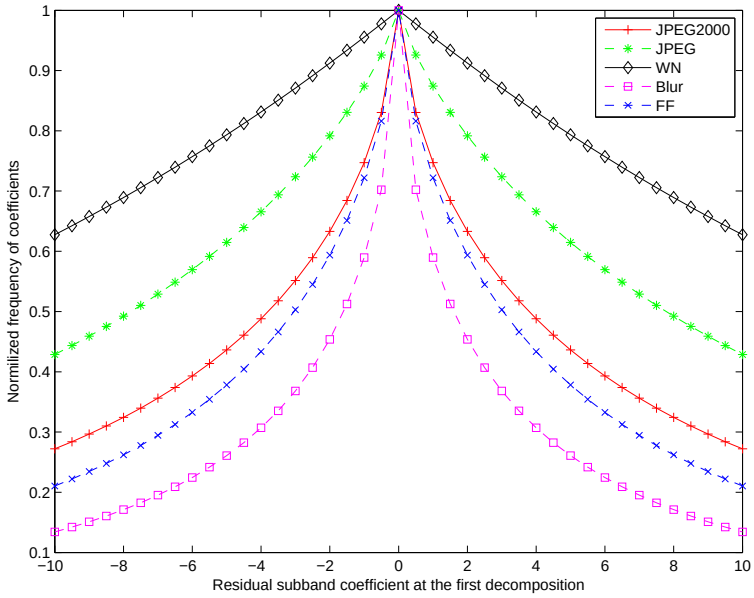
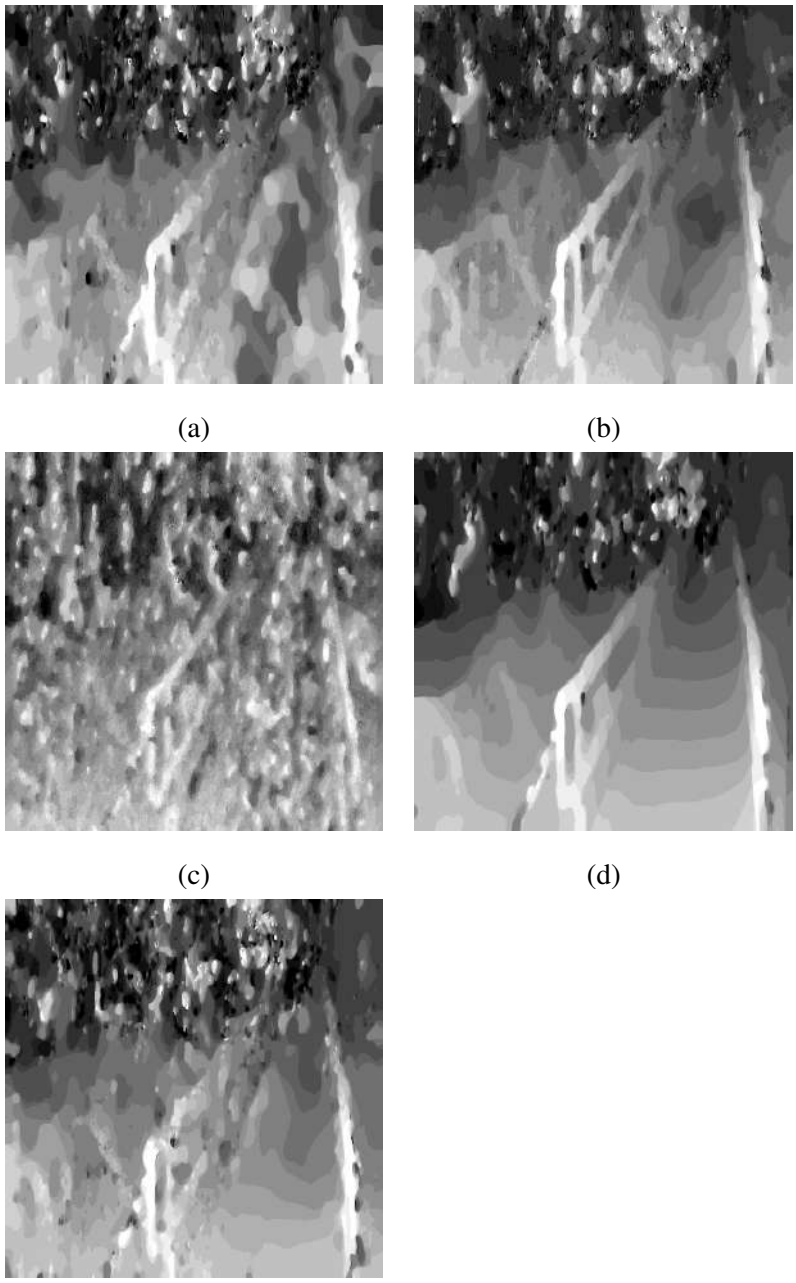
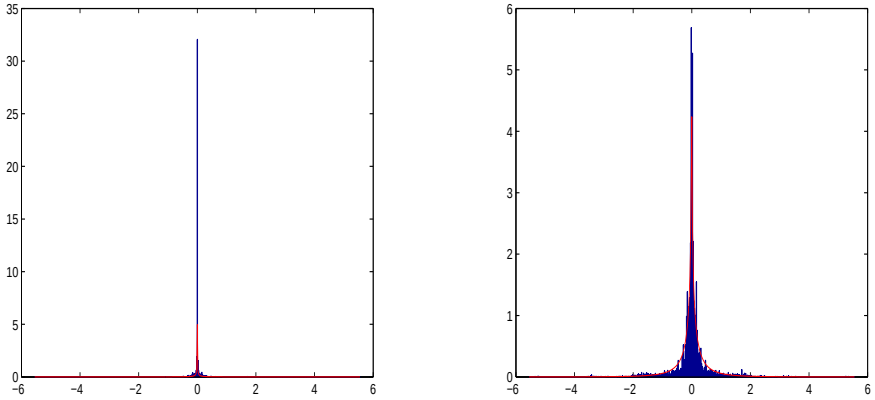


Fig. 5: GG distribution of the first scale residual subbands of the stereo image illustrated in Figure 3 with different distortion types.



(a) (b) (c) (d) (e)

Fig. 6: Disparity maps estimated from the distorted SI shown in Fig. 3 using (a) JPEG 2000, (b) JPEG, (c) WN, (d) Blur and (e) Fast fading.



(a) $\beta_1^{(z,HH)} = 0.01$ (b) $\epsilon_1^{(z,HH)} = 0.28$ and $\beta_1^{(z,HH)} = 0.41$

Fig. 7: Modeling the distribution of the diagonal detail wavelet coefficient of the disparity map shown in Fig. 6 using the (a) GG model (b) BGG model.