Multidim Syst Sign Process (2013) 24:281–316
DOI 10.1007/s11045-012-0178-3

A perceptual metric for stereoscopic image quality assessment based on the binocular energy

Rafik Bensalma · Mohamed-Chaker Larabi

Received: 22 April 2011 / Revised: 7 January 2012 / Accepted: 28 January 2012 /
Published online: 21 February 2012

Abstract Stereoscopic imaging is becoming very popular and its deployment by means of photography, television, cinema. . .is rapidly increasing. Obviously, the access to this type of images imposes the use of compression and transmission that may generate artifacts of different natures. Consequently, it is important to have appropriate tools to measure the quality of stereoscopic content. Several studies tried to extend well-known metrics, such as the PSNR or SSIM, to 3D. However, the results are not as good as for 2D images and it becomes important to have metrics dealing with 3D perception. In this work, we propose a full reference metric for quality assessment of stereoscopic images based on the binocular fusion process characterizing the 3D human perception. The main idea consists of the development of a model allowing to reproduce the binocular signal generated by simple and complex cells, and to estimate the associated binocular energy. The difference of binocular energy has shown a high correlation with the human judgement for different impairments and is used to build the Binocular Energy Quality Metric (BEQM). Extensive experiments demonstrated the performance of the BEQM with regards to literature.

Keywords Stereoscopic quality assessment · Binocular energy · 3D perception · Simple and complex cells

1 Introduction

The Human Visual System (HVS) allows to analyze our complex environment represented in a spatiotemporal space of four dimensions (x, y, z, t). Its abilities related to psychological and physiological vision aspects highly contribute to the analysis and interpretation tasks. The binocular vision process is one important characteristic of the HVS and corresponds to

R. Bensalma · M.-C. Larabi (✉)
XLIM-SIC, UMR CNRS 7252, Université de Poitiers, Poitiers, France
e-mail: chaker.larabi@univ-poitiers.fr

R. Bensalma
e-mail: rafik.bensalma@univ-poitiers.fr

the combination of the left and right signals for the same area of the scene. This has always aroused interests of scientists coming from various domains thus leading to many experiments allowing to better understand this property and explain the involved factors (Hubel and Wiesel 1962, 1970; Barlow et al. 1967; Fleet et al. 1996; Ohzawa and Freeman 1986a,b). The visual perception in general and the binocular vision in particular are considered as very important research fields not only in physiology or psychology but nowadays in mathematics, computer vision and artificial intelligence.

From a market point of view, there are plenty of technologies including 3D capabilities as the 3D cinema, 3D television, 3D handheld devices, 3D medical imagery and so on. After the enthusiasm comes the questioning about mastering the quality of such a technology and identifying the 3D vision factors responsible for the visual fatigue. The main difficulty for such a medium lies in the unavailability of the perceived 3D image resulting from a binocular fusion of the left and right images made by the HVS. Even though a considerable progress has been made in quality assessment of 2D images, 3D is still an open field. Based on their experience in 2D quality assessment, several authors tried to adapt 2D metrics for stereoscopic images either by using the image-pair or by estimating the depth-maps (Campisi et al. 2007; Kaptein et al. 2008; Goldmann and Ebrahimi 2010; Tikanmäki et al. 2008; Hewage and Martini 2010; Xing et al. 2010). These approaches have proved to be quite ineffective in addressing the quality of 3D contents as the HVS does. This reveals that assessing 3D is a more complicated task than 2D because of the depth interpreted by the HVS. Therefore, an important effort has been devoted to the understanding of binocular vision mechanisms and development of metrics taking into account 3D perceptual properties.

Another problem, related to 3D quality assessment field, can be raised. It concerns the nature and availability of 3D stereoscopic databases. Indeed, in comparison to 2D, there is no consensual database to be used for the evaluation of metrics' performance. Moreover, several questions, linked to image registration, subjective paradigm, content type and acquisition protocols, remain open. Consequently, there is an important effort to put in this direction in order to allow a fair comparison of 3D metrics.

In this paper we propose a perceptual metric for the evaluation of registered stereoscopic images based on the binocular fusion of the HVS. The developed model tries to mimic the HVS by modeling the simple cells responsible for the local spatial frequency analysis and then the complex cells responsible for the generation of the binocular energy. This energy is used as an indicator of the quality of a stereoscopic image pair. This information is used for the construction of BEQM (Binocular Energy Quality Metric) allowing to evaluate the perceived quality difference between the original and the impaired stereo-pairs.

The remainder of this paper is organized as follows. Section 2 is a state-of-the-art of stereoscopic metrics reporting in a near exhaustive way the major contributions to 3D objective quality assessment. The proposed approach is described in Sect. 3 by giving first the principle of BEQM, then by detailing the adopted modeling for simple and complex cells. The formulation of the BEQM metric is then given in Sect. 4. A deep experimentation is conducted to demonstrate the appropriateness of the adopted modeling, to study the behavior of binocular energy for different impairments and to study the performance of the metric in comparison to the state-of-the-art in Sect. 5. This paper ends with some conclusions and future works.

2 State of the art of stereoscopic quality metric

The fascination generated by 3D image/video technology (3D cinema, 3DTV, 3D gaming) has created an increasing attraction of research activity regarding the quality evaluation field.

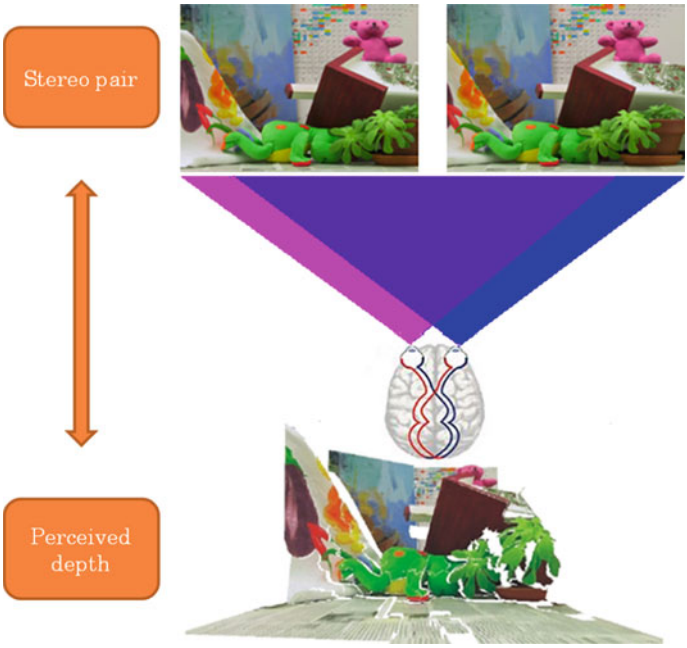


Fig. 1 Perceived quality from a stereoscopic image-pair

Even though 3D imaging is closely related to 2D, its quality assessment represents a more difficult problem to handle because of the third dimension that creates a sensation of depth. The main obstacle lies in the impossibility of an algorithm to access the 3D content as it is perceived by the human observer (cf. Fig. 1). Instead of this, an algorithm has access to a pair of images (left and right) and possibly the disparity or depth maps. The challenge in this field is to find a way to measure the perceptual quality of stereoscopic image pair as Human Visual System does and only by using the available data. To date, several algorithms (metrics) have been proposed in the quality assessment literature. This section is dedicated to their description and is addressing them under three categories. The first focuses on works using 2D metrics on stereoscopic images, the second reports on the exploitation of additional information like depth or disparity maps with 2D metrics and the third and last category deals with quality metrics development based only on 3D aspects.

2.1 Stereoscopic quality evaluation based on 2D metrics

The first approach that comes to mind when dealing with quality assessment of stereoscopic images is to use state-of-the-art 2D metrics. This approach has been followed by many researchers giving thus algorithms based on 2D metrics such as PSNR, SSIM (Wang et al. 2003), JND (Sarnoff Corporation 2003), DCTune (Watson 1993), PQS (Miyahara et al. 1998), NQM (Damera-Venkata et al. 2000), Fuzzy S7 (Weken et al. 2004), BSDM (Avcibas et al. 2002), IFC (Sheikh et al. 2005) and VIF (Sheikh and Bovik 2006) and many others. Most of these metrics have been used by You et al. (2010) where they applied and studied their performance on stereoscopic images. Four impairment types have been generated on stereoscopic images i.e. blur, noise, JPEG and JPEG 2000 compression. The pairs of impaired stereoscopic images were used as separate input for 2D metrics and the final score for a

stereo-pair is the average of right and left scores. The experimental results showed the highest correlation with regards to the human judgment when using the SSIM metric; although, the prediction performance of 3D impairments made by 2D metrics is less significant than that obtained on 2D images.

Another way has been opened by trying to exploit the disparity map for stereoscopic quality assessment. However, in the framework of 3DTV, authors of Campisi et al. (2007) and Kaptein et al. (2008) indicated that this disparity map does not improve the performance of metrics. Campisi et al. (2007) tried to check the validity of the previous remarks by applying a 2D metric on the left and right stereoscopic images and also on the disparity map. Their fidelity score is an average of images' scores combined with a score obtained on the disparity map by the same metric. Two important comments can be made about this approach: First, the 3D content of a disparity map cannot be interpreted by a 2D metric because it has not been developed in that sense and then, the usage of the disparity map on stereoscopic images generates a redundancy which will not improve the performance of all metrics and will depend on the the method of extracting disparity.

To confirm the previous statements, Kaptein et al. (2008) have run subjective experiments using images with same objects at different depths. The used images have been impaired using a JPEG compression. By drawing the evolution of subjective scores versus object's depth and versus compression bitrate, authors concluded that there is no correlation between the depth and the 3D perceived quality of an object. At the same time, a high correlation has been noticed between the compression bitrate and 3D perceived quality; the latter result was foreseeable. These results may lead to consider that the depth information is unnecessary for quality evaluation task. However, it must be addressed carefully because (1) the experimental setup plays an important role in such experiments and (2) the definition of the depth itself has to be reconsidered. More recently, Goldmann and Ebrahimi (2010) demonstrated the ineffectiveness or more precisely the shortcoming of the usage of 2D metric for stereoscopic quality assessment. The results of their experiments conducted to the conclusion that 3D quality cannot be evaluated using 2D inspired quality metrics.

2.2 Stereoscopic quality evaluation using 2D metrics + 3D informations

To cope with the effectiveness of 2D metrics for the evaluation of stereoscopic images, several basic extensions have proposed in literature, in general, by integrating in various ways either the disparity map or the depth map in the metric scheme. You et al. (2010) focused on using the disparity map through three different approaches. In the first approach, 2D metrics are used only to compute the quality between the original and impaired disparity maps. The obtained results are more interesting that those obtained using the stereo pairs only. The second approach consists in combining the resulting quality maps obtained by the stereo-pairs (Q_o) with those of the disparity maps (Q_d). The pooling stage is ensured by a formula combining the contributions of Q_o and Q_d in addition to their mutual contribution ($Q_o.Q_d$). The third approach uses the average of the quality maps instead of pixel-wise combination while the fourth is a mixture of the second and the third approaches. The average of quality scores is used as a weighting factor for the pooling stage. The best results have been obtained using the SSIM metric.

Tikanmäki et al. (2008) have experimented in a similar way video quality metrics on color + depth sequences. Their algorithm consists in the application of the VSSIM (Wang et al. 2004) and PSNR metrics between the original and impaired image sequences where the second view is synthesized before and after compression. The validation of the approach is performed thanks to a subjective experiment. Finally, they concluded about the need of

metrics specific to stereoscopic content. In a quite different way, Hewage and Martini (2010) proposed an approach that computes the contours of both original and impaired depth maps. These contours are then binarized and compared by using a PSNR metric. The obtained results show similar performance with PSNR computed of image-pairs especially for higher bitrates.

Xing et al. (2010) have proposed an approach where the depth map, computed from the stereoscopic images, is weighted using an SSIM map measured between original and impaired left images. A threshold equal to 0.977 is defined for SSIM in order to take in to account or not the value of depth. The score of this metric is the average of the weighted depth map. The results have been compared to subjective scores and demonstrated a good correlation with the human judgement. Similarly, Boev et al. (2010) developed a metric as a product of the quality factor Q_s obtained by using SSIM metric between original and impaired disparity maps and the quality factor Q_m obtained by using SSIM metric between cyclopean images extracted from original and impaired images. The metric has been validated on different types of impairments using a subjective experiment. Results showed a coherence between the metric scores and the human judgement.

2.3 Stereoscopic quality evaluation based on a 3D analysis

Even though a great research effort has been made to extend 2D metrics to assess stereoscopic quality, it is still difficult to measure precisely the added-value of the simplistic usage of 3D information like disparity map, depth map or cyclopean image. It is important to highlight that the 3D perception has properties that 2D metrics cannot capture by only using such an information. Thus, there is a research focus on interpretation and modeling of real 3D criteria for quality assessment both for 3D graphic objets (Lavoué et al. 2006; Rittermann 2004; Cheng and Boulanger 2005) and stereoscopic images (Meesters et al. 2004).

3D graphic objects evaluation is not the focus of our research but it presents important similarities with the discussed topic. Rittermann (2004) proposed a metric for the quality evaluation of 3D graphic object by using some stereoscopic properties. The first step of the metric consists in capturing several view of the objet and generating the associated depth maps. After an image registration process, the metric extracts several features such as shape variation and statistical variation in the Fourier domain. The score of this metric is a linear combination of the extracted features. Similarly, Cheng and Boulanger (2005) proposed a metric based on geometrical properties of the object and using the JND (just noticeable difference) for redundancy reduction. The feature are extracted using a scale-space filtering and used to determine the perceptual impact during runtime. Refinement is applied only when the impact is higher than JND.

Akhter et al. (2010) proposed a no-reference metric for asymmetric JPEG compression based on a partitioning of the image-pair in fixed-size blocks to characterize 3D artifacts. The obtained blocks are classified into flat blocks (without contours) and active blocks (with contours). Two features are used to estimate the quality of a stereo-pair: First the blockiness considered as the average pixel-wise difference of a block and then the zero crossing that assigns a value of 1 when detected. These features are computed separately for flat and active blocks and are used in the matching process as a control parameter for the construction of the disparity map. The final score of the metric is a combination of the two features in addition to the difference of zero crossing obtained horizontally and vertically. In a quite different way, Gorley and Holliman (2008) proposed a metric based on the application of perceptual models i.e. the Michelson's contrast formula and a model of Stereo Band-Limited Contrast, on point matches between the right and left views of the stereo-pair.

Based on the disparity map, Kim et al. (Donghyun et al. 2009) proposed a quality metric for 3D stereoscopic videos. They first run a subjective experiment to determine the 3D quality criteria. Their metric is thus constructed as the mutual variation of a temporal average of chromatic difference and a temporal average of disparity difference. Olsson and Sjostrom (2007) proposed a metric based on the depth map. As a preliminary stage, they synthesized 2D stereo-pairs at different depth levels. The levels are chosen with regard to the camera distance and the focal information. For each of them, its pixels are identified using the algorithm described in Keita and Takeshi (2005). Finally, coding artifacts of each depth level are evaluated using the MSSIM metric.

2.4 Discussion

It is clear that the stereoscopic quality assessment field is relatively new and additional research effort is expected in order to achieve to a similar level of development as for 2D. To our knowledge, there is no perceptual 3D metric in the literature. Human 3D perception has to be explored deeply in order to understand and exploit phenomena like the binocular rivalry or binocular compensation for the quality prediction. This is what we propose in this paper.

Also, it is not clear how the different artifacts affect the human perception of depth. For example, JPEG 2000 tends to smooth/reduce the depth while JPEG creates false depth. This issue can be understood thanks to subjective experiments where the image database is acquired and displayed in a controlled way.

From a different point of view but very related to quality issues, stereoscopic compression opens the floor for other questions linked to the binocular compensation phenomenon. This latter is due to the quality difference between the right and left views as it is the case of many coders proposed in literature (Bensalma and Larabi 2010; Ellinas and Sangriotis 2004; Nath and Dubois 2006; Woo et al. 1999).

3 Proposed approach: from physiology to analytical modeling

3.1 BEQM metric principle

The perceptual quality of 3D images reconstructed by the HVS is highly dependent on the nature of the monocular signal coming from the retinal images. This signal is conveyed from the retina to the visual cortex where the binocular fusion occurs, giving thus the perceived image. The objective of our approach is to find an appropriate model for the binocular signal in order to evaluate its role on the perceptual quality. The proposed metric (BEQM: Binocular Energy Quality Metric) exploits the developed model by comparing the binocular signal quality of the original stereo-pair and an impaired stere-pair.

At the visual cortex level, the simple and complex cells are responsible for the binocular fusion as described in Fig. 2. Once the two retinal images are captured by the receptive fields of the ganglion cells, they are transfered to the left and right LGBs (Lateral Geniculate Body), then to the visual cortex. The simple cells are the first to receive this retinal information at the visual cortex (V1 area). These cells are characterized by elongated receptive fields having specific orientation and size. They work in pairs representing both eyes and each pair of receptive fields is connected to a complex cell in order to generate the binocular signal.

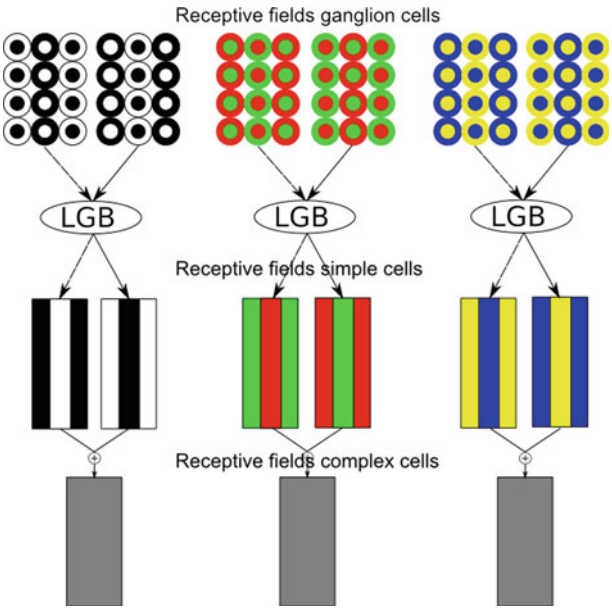


Fig. 2 Production of the binocular signal from a stereoscopic images pair. The receptive fields of the LGB to those of the complex cells

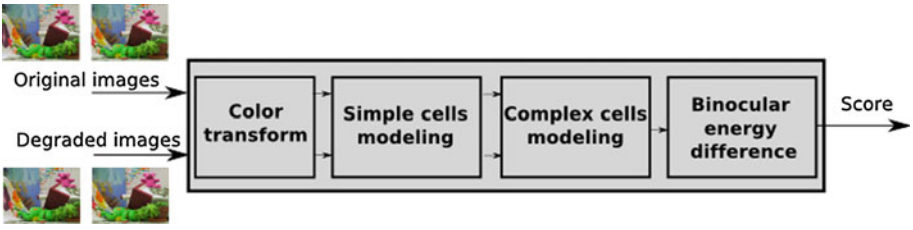
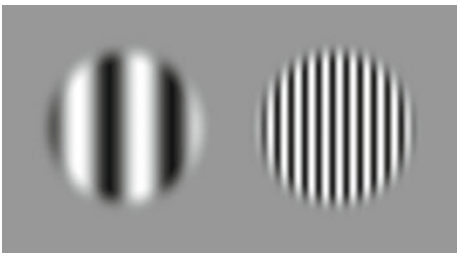


Fig. 3 Synoptic of the BEQM (Binocular Energy Quality Metric)

The quality of binocular signal determines the 3D perceived quality of a human. Since the binocular signal is generated from the monocular signals (left and right), its quality (i.e. perceived quality) is highly dependent on them. The focus of this work is to exploit the characteristics of this binocular signal in order to predict the quality of a stereo-pair having undergone some impairments such those due to compression, transmission and so on.

In literature, several physiological models have been described (Foster et al. 1983; Liu et al. 1992; Pollen and Ronner 1981. They try to model some properties of the simple and complex cells in order to describe the binocular energy. However, these models are not really used in computer vision applications. So, we try, in this work, to propose an analytical model aiming to compute the binocular signal that we use afterward to estimate the perceptual quality of an impaired stereo-pair. A simplified synoptic of our work is given in Fig. 3. The first step consists in an antagonist color transform that tries to mimic the HVS (cf. Fig. 2). The second models some properties of the simple cells used in the complex cells modeling (step 3). The quality score provided by our metric is the difference of binocular energy between the original and impaired stereo-pairs.

Fig. 4 Luminance gratings



3.2 Modeling the simple cells

3.2.1 Simple cells behavior

The functional role of the simple cells has been addressed in many previous works from different research fields either to understand their behavior or to model their properties. The pioneering definition of their receptive fields structure has been made by Hubel and Wiesel (1962) in the 1950s. The first assumption about their role was based on their appearance i.e. similarity with borders, contours and bars in a given environment. The second major explanation, that can be applicable to signal processing, was based on a local spatial frequency analysis.

According to Hubel and Wiesel (1970) and Campbell et al. (1969), receptive fields of simple cells are linear spatial filters characterized by their elongated shape composed of two antagonistic regions ON (excited) and OFF (inhibited) inherited from ganglion cells (cf. Fig. 2). Each ON or OFF region of a simple cell is connected to the LGB Kuffler (1953) where it gets the retinal information. Barlow et al. (1967) have demonstrated in their early work that cells in the primary cortex respond preferentially to shifted bars (cf. Fig. 4); which reveals the existence of a disparity detector.

For instance, if we take the gratings of Fig. 4, the simple cell is optimally activated when the size of white (resp. black) bar is the same than its ON (resp. OFF) area. Several physiological experiments demonstrated that these cells can be modeled using linear filters from their impulse response measured on the visual cortex. DeAngelis et al. (1991) have approximated the impulse response using a *Gabor* wavelet as described by Eq. 1.

$$\psi^k(x_1, x_2) = g(x_1, x_2)\exp[-i(x_1 cos\alpha_k + x_2 sin\alpha_k)] \tag{1}$$

This type of spatial arrangement is described mathematically by a two-dimensional *Gabor* function where the ON and OFF regions correspond respectively to peaks and hollows of the Gabor function as illustrated in Fig. 5.

Motivated by physiological studies of the visual perception, several sampling functions have been proposed, particularly directional wavelets. The latter can be non adaptive as Curvelets (Candès et al. 2006), Contourlets (Do and Vetterli 2001) and Complex wavelets (Kingsbury 1997) or adaptive by taking into account the signal geometry as Wedgelets (Donghyun et al. 2009), Bandelets (Le Pennec and Mallat 2005) and Grouplets (Mallat and Peyré 2006). The visual characteristics being modeled by the previously cited functions can be summarized as follows:

1. **Multi-resolution :** allows an image decomposition into perceptual channels similarly to the HVS from the low frequencies to high frequencies.
2. **Spatial location :** The image element should be localizable in the spatial and frequency spaces.

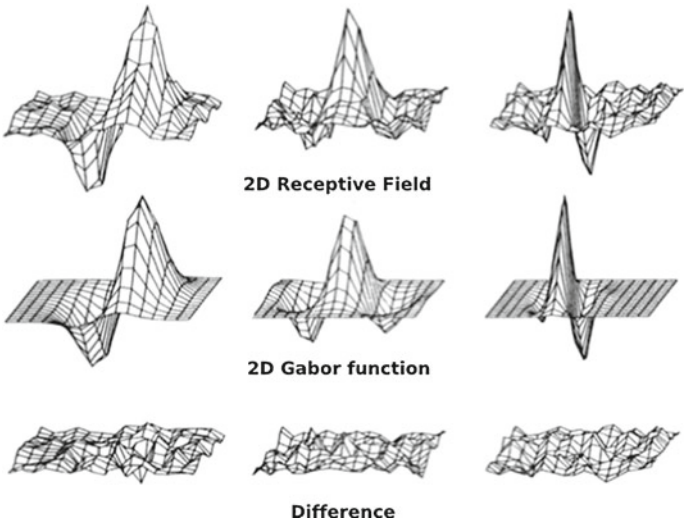


Fig. 5 Comparison between measured responses of simple cells and their Gabor modeling (Jones and Palmer 1987)

3. **Sampling type :** For some applications (compression for example), the representation should avoid redundancy.

3.2.2 Proposed model for simple cells

Our modeling tries to follow the description given in Fig. 2. The first step consists in a color space conversion from RGB (acquisition color space) to CIE$L^*a^*b^*$ (Schanda 2007). The latter is perceptually correct and represents, in our case, the color antagonism as in the HVS. So after this operation, each image of the stereo-pair is represented with a single channel of luminance L^* and two perpendicular (important criterion for the rest of the work) channels of chrominance a^* and b^*.

As the visual field is characterized by two types of information i.e. binocular and monocular, two types of simple cells exist in the visual cortex. First, the monocular simple cells having monocular receptive fields as illustrated by Fig. 6a and serving as recipient for monocular information from left and right retinas called occluded information. Then, binocular simple cells (organized in pairs) having binocular receptive fields (cf. Fig. 6b) and characterized by their size, phase and orientation (Palmer and Davis 1981). Each pair has a phase difference of 90° (Foster et al. 1983; Liu et al. 1992; Pollen and Ronner 1981). A pair of simple cells is connected to a complex cell of the same type (monocular in Fig. 6c or binocular in Fig. 6d).

In literature, several analytical models can be found for simple cells (Fleet et al. 1996; Ohzawa and Freeman 1986a,b). The response of a pair of simple cells is often represented as a complex cell $C(x) = \rho(x)e^{\phi(x)}$ (cf. Fig. 7). In order to model this spatial-frequency response, characterized by a size, an amplitude, a phase and an orientation, we studied different directional wavelets mentioned in Sect. 3.2.1. We then retained three transforms: Discrete Wavelet Transform (DWT) (Mallat 1989), Complex Wavelet Transform (CWT) (Kingsbury 1997) and Bandelet transform (Le Pennec and Mallat 2005).

In order to ease the understanding of the proposed model, we start by giving a brief description of transforms mentioned above. The Complex Wavelet Transform (CWT) has

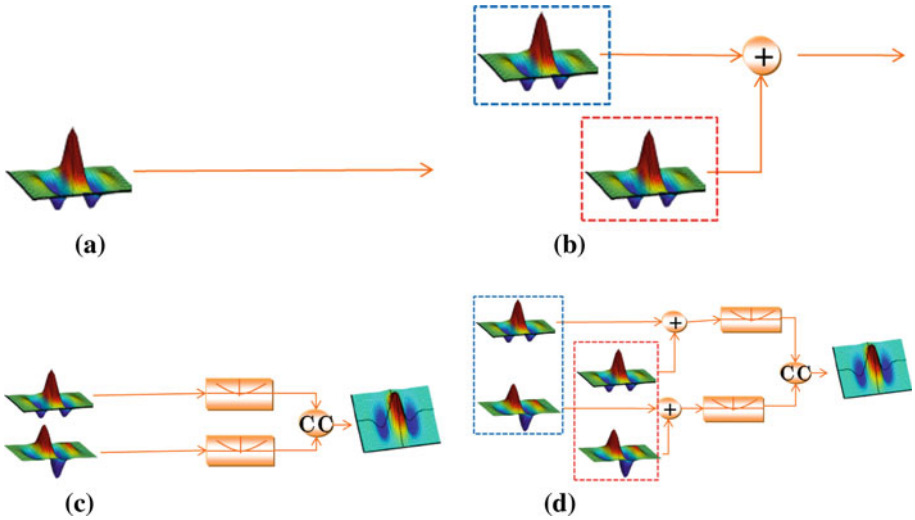


Fig. 6 Descriptive scheme of simple and complex cells. **a** Monocular simple cell. **b** Binocular simple cell. **c** Monocular complex cell. **d** Binocular complex cell

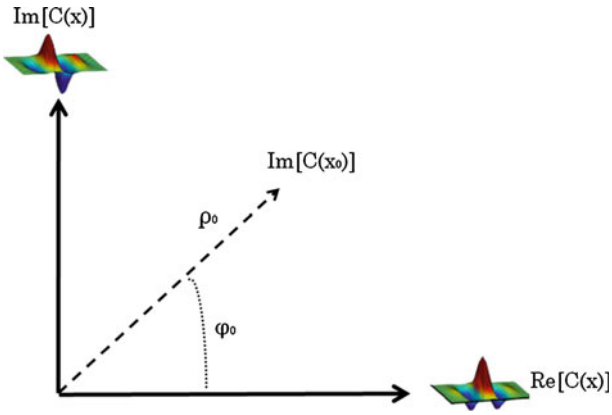


Fig. 7 Illustration of the response of a quadrature pair of simple cells

been introduced in order to overcome the drawbacks of the Discrete Wavelet Transform (DWT) (Mallat 1989). The latter may generate only few coefficients around a strong singularity. Moreover, the DWT is not invariant to translation which may have some problems for stereoscopic images where the disparity causes a shift between the same objects on both views.

Two ways have been introduced for computing the CWT. The first approach is based on the Hilbert transform () where the image is analyzed using a floating-point wavelet as described by Mallat (1989). The obtained coefficients are considered as the real part of the CWT. The imaginary part is obtained by applying the DWT on the Hilbert transform result. The second approach is called the dual-tree method (Selesnick et al. 2005). It consists in analyzing the image by two different DWTs. In order to obtain the real and imaginary parts of the CWT, two couples of filters, each composed by a low-pass and a high pass filters, are applied. The first couple computes the real part of the CWT and the second part computes the imaginary part of the CWT. We retained this approach for the development of our model.

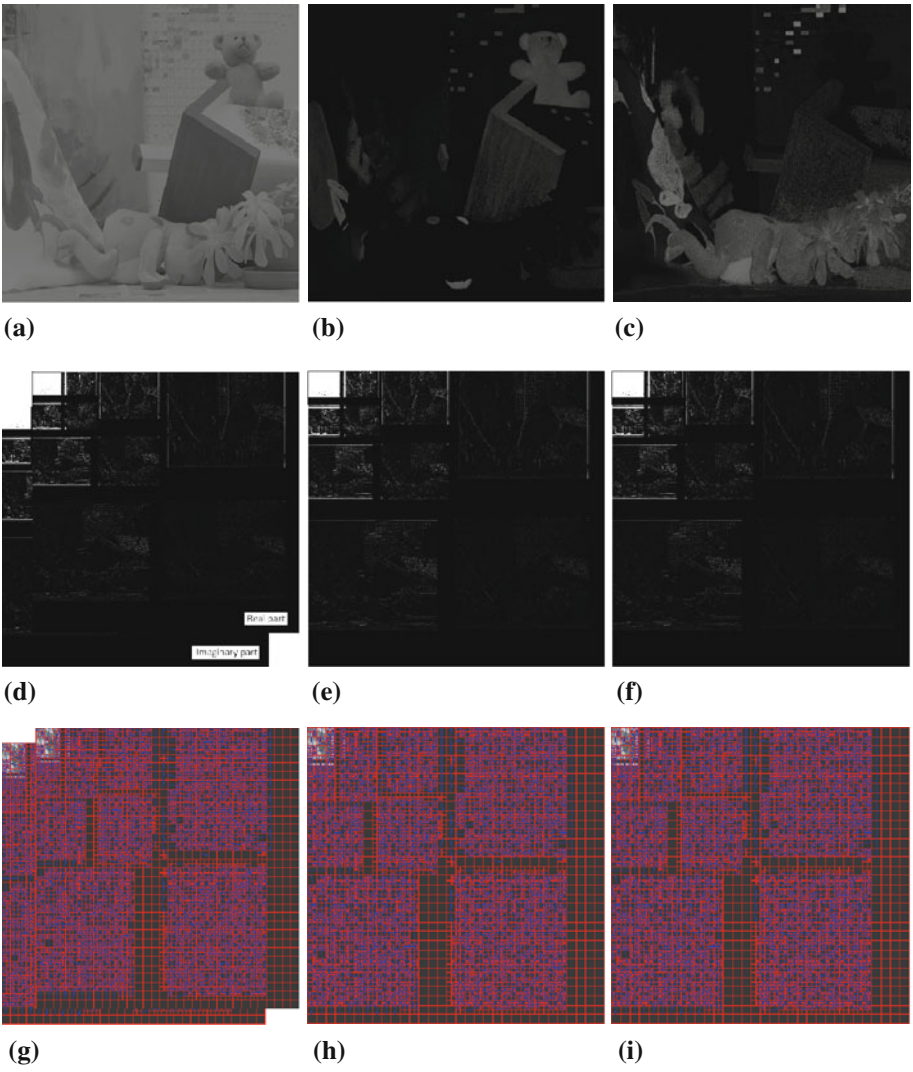


Fig. 8 Application of the CWT on L^* and DWT on a^* and b^*, and bandeletization of the obtained results. **a** L* component. **b** a* component. **c** b* component. **d** CWT(L*). **e** DWT(a*). **f** DWT(b*). **g** Bandelet transform of L*. **h** Bandelet transform of a*. **i** Bandelet transform of b*

The aim is to represent the two pairs of stereoscopic images (original and impaired) using a set of complex functions (right: $C_r(x) = \rho_r(x)e^{\phi_r(x)}$ and left: $C_l(x) = \rho_l(x)e^{\phi_l(x)}$). In the first step, the real and imaginary parts of the response are separated using the CWT on the luminance component (see Fig. 8d) and the DWT on the chromatic components (see Fig. 8e, f). Therefore, the luminance is composed of the real and imaginary components obtained by CWT and the chrominance is organized in the same way with DWT(a^*) as the real part and DWT(b^*) as the imaginary part.

The next step consists in describing the characteristics of the simple cells starting from the coefficients obtained previously. We opted for the bandelet transform which is very close to the behavior of a simple cell (Mallat and Peyré 2006). It splits up the obtained subbands in

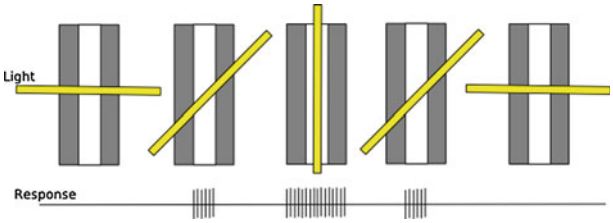


Fig. 9 Response associated to different excitations of the receptive field of a simple cell

a quadtree of variable size following the image geometry and an orientation is computed and assigned to each block depending on the coefficients. The block is called a dyadic square and is characterized by its size, amplitude and orientation as a simple cell. This relation has been identified previously by (Peyre 2005). Figure 8g–i give the results of the bandelet transform respectively on the CWT coefficients for luminance and DWT for chrominance.

As mentioned previously in Sect. 3.2.1, a simple cell is characterized by a receptive field of type ON/OFF. This important characteristic determines the intensity of the simple cell response. In addition to this, the orientation and size play an important role as shown in Fig. 9. If the stimulus undergoes an impairment that changes its characteristics (position, size, orientation), the response of the simple cell will decrease depending on the introduced gap.

If we consider the pair of stereoscopic images as the perceived stimulus, the model should have a similar behavior in front of impairments since it is built with transforms (wavelets) that are sensitive to spatial impairments. It means that the amplitude $\rho(x)$ and phase $\phi(x)$ of the complex function will be different (see Fig. 7) because the coefficients of the real and imaginary parts have changed.

The proposed metric lies on the availability of the reference image-pair. Therefore, the complex functions (dyadic square pairs) of the original stereo-pair must be the same than those of the impaired stereo-pair to be able to compare them. The next step consists in the computation of the dyadic squares orientation depending on the impaired wavelet coefficients.

Finally, the output of the proposed model for simple cells (simple cells response) is sent to complex cells as described in Fig. 6c, d. The binocular energy generated by the complex cell is directly related to the responses of simple cells received as input. In other words, if the stimulus is not impaired, the simple cells response is optimal leading to an optimal response of the complex cell and *vice versa*. To complete the description of the proposed model, the next section is devoted to complex cells modeling with the aim to estimate the binocular energy.

3.3 Modeling the complex cells

3.3.1 Complex cells behavior

Complex cells do not have the same characteristics than simple cells except the size. Complex cells are not sensitive to the orientation and position of a stimulus unlike simple cells. Moreover, they are sensitive to motion direction which must be perpendicular to the stimulus. However, this last property is not used in our model because we address still images only. Figure 10 illustrates the response of a complex cell to different stimuli. Receptive fields of complex cells do not have an antagonist representation (ON/OFF) which imply a lack of sensitivity to monocular phase. It is thus insensitive to stimulus position.

Several models have been proposed to deal with the behavior of complex cells: Fleet et al. (1996) and Ohzawa and Freeman (1986b), to name a few. Similarly to simple cells, two types

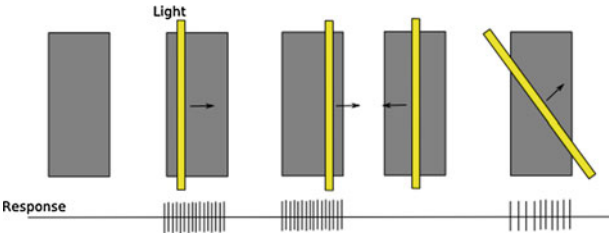


Fig. 10 Response associated to different excitations of the receptive field of a complex cell

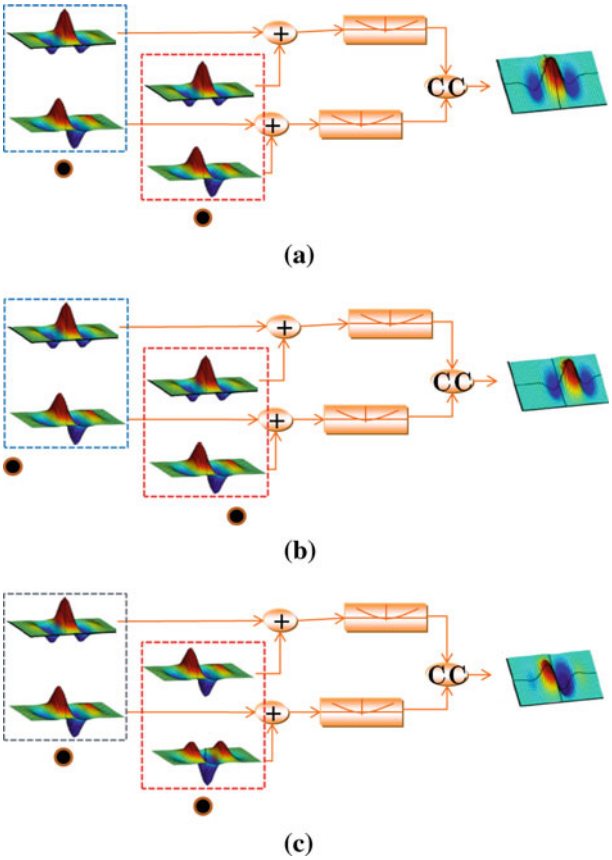


Fig. 11 Inheritance of position and phase properties from simple cells. **a** Zero disparity. **b** Position shift. **c** Phase shift

of complex cells exist: monocular complex cells and binocular complex cells. Receptive fields of monocular complex cells receive as input the signal of two monocular simple cells of the same retina (left or right), in order to compute the monocular energy that corresponds to occluded regions. Whereas, receptive fields of binocular complex cells receive as input two binocular simple cells in order to generate the binocular energy. The couple of simple cells are grouped in a quadrature phase.

In the models described by Jones and Palmer (1987), Foster et al. (1983) and Liu et al. (1992), simple cells used as input of complex cells have the same spatial relation between their monocular receptive fields (amplitude, size, orientation and phase shift). The complex cell inherits the simple cells' properties apart their sensitivity to the orientation and spatial arrangement. The binocular energy generated by a complex cell depends on the position disparity and phase shift between simple cells as illustrated by Fig. 11.

3.3.2 Proposed model for complex cells

Once the simple cells modeled, the next step is dedicated to the calculation of the generated binocular energy. Three parameters are used for this calculation, namely position, phase and orientation shifts. Our model is based essentially on those developed by Fleet et al. (1996) and Ohzawa and Freeman (1986b). The binocular energy, in our case, is calculated by using two pairs of dyadic squares (representing simple cells) from the left image ($C_l(x) = \rho_l(x)e^{\phi_l(x)}$) and two pairs from the right ($C_r(x) = \rho_r(x)e^{\phi_r(x)}$). $\rho_i(x)$ is the monocular amplitude where i refers to left (l) and right (r). It is obtained from the dyadic squares pair belonging respectively to the real and imaginary parts of the luminance and the chrominance and is computed as follows:

$$\rho_l(x) = |C_l(x)| = \left[\mathrm{Re}(C_l(x))^2 + \mathrm{Im}(C_l(x))^2\right]^{1/2} \tag{2}$$

$$\rho_r(x) = |C_r(x)| = \left[\mathrm{Re}(C_r(x))^2 + \mathrm{Im}(C_r(x))^2\right]^{1/2} \tag{3}$$

$\phi_i(x)$ is the monocular phase obtained from a pair of dyadic squares where i refers to left (l) and right (r). Monocular Phase is one of the important characteristics of dyadic squares with the size, the position and the orientation.

$$\phi_l(x) = arg(C_l(x)) = arctan\left(\frac{\mathrm{Im}(C_l(x))}{\mathrm{Re}(C_l(x))}\right) \tag{4}$$

$$\phi_r(x) = arg(C_r(x)) = arctan\left(\frac{\mathrm{Im}(C_r(x))}{\mathrm{Re}(C_r(x))}\right) \tag{5}$$

In order to match the pairs of dyadic squares of the left image with those of the right image, we used the formula given in Eq. 6.

$$\begin{aligned} E(x) &= |C_l(x) + C_r(x)|^2 \\ &= [\mathrm{Re}(C_l(x)) + \mathrm{Re}(C_r(x))]^2 + [\mathrm{Im}(C_l(x)) + \mathrm{Im}(C_r(x))]^2 \end{aligned} \tag{6}$$

$E(x)$ models the response of a binocular complex cell. When the complex cell is monocular, Eq. 6 will use a pair of dyadic squares coming from one of the stereoscopic images. Knowing that the two pairs of dyadic squares, belonging to the left and right images respectively, have similar characteristics, a polar representation is adopted giving the following equation:

$$E(x) = \rho_l^2(x) + \rho_r^2(x) + \rho_l(x)\rho_r(x)cos(\Delta\phi(x)) \tag{7}$$

$E(x)$ is the binocular energy generated by a complex cell and $\Delta\phi(x)$ is the inter-ocular phase shift as expressed in Eq. 8. This modeling is correct when the two pairs of dyadic squares from the left and right images have the same position and orientation; which is not always the case.

$$\Delta\phi(x) = \phi_l(x) - \phi_r(x) \tag{8}$$

If the disparity is varied by slightly shifting the position of a stimulus, the associated amplitude and phase will vary as well. However, in various models, the amplitude variation is considered as negligible in comparison to the phase variation.Therefore, the binocular energy is expressed as a function of the disparity considered as the inter-ocular phase shift. Figure 12 gives an example of a pair of dyadic squares represented by its real and imaginary parts. On Fig. 12d, the monocular phase signal increases linearly with the spatial position x. For small changes in disparity, inter-ocular phase shift depends on the gradient curvature. If the inter-ocular phase signal is rapidly increasing with a small disparity, a difference will be generated (see Eq. 8). The derivative of the monocular phase signal $\phi_d'(x)$ (see Fig. 12e) is hence important and is often called the instantaneous spatial frequency (Adelson and Bergen 1985).

When two pairs of dyadic squares, belonging to the left and right images, have not the same position, the response $C_r(x)$ becomes a shifted version of the left response $C_l(x)$ i.e. $C_l(x) = C_r(x - d)$ where d is the disparity. This disparity is expressed by a phase shift $\phi_l(x) = \phi_r(x - d)$. Knowing that we consider in this work only registered stereoscopic images, the disparity is then only horizontal as described in Fig. 13.

In the model developed by Fleet et al. (1996), the phase shift is expressed as a Taylor series of the inter-ocular phase shift (Eq. 9) where $O[d^2]$ is the second order term.

$$\begin{aligned}\Delta\phi(x,d) &= \phi_l(x) - \phi_r(x)\\ &= \phi_l(x) - \phi_l(x-d)\\ &= d \times \phi_l'(x) + O[d^2]\end{aligned} \tag{9}$$

One can notice that the inter-ocular phase shift is expressed as a product of the disparity d and the instantaneous spatial frequency. The combination of Eq. 7 and 9 gives the formula of Eq. 10 that represents a function taking into account the retinal disparity. When there is no disparity, the inter-ocular phase shift is equal to zero leading to the maximization of the binocular energy ($cos(0) = 1$). Inversely, an increase of the inter-ocular shift leads to a decrease of the binocular energy.

$$E(x) = \rho_l^2(x) + \rho_r^2(x) + \rho_l(x)\rho_r(x)cos(d \times \phi_l'(x)) \tag{10}$$

It has been demonstrated that the orientation is an important parameter in the calculation of the binocular energy (Ohzawa and Freeman 1986a,b). Also, the disparity has a high correlation with the orientation of the simple cells (Fleet et al. 1996). An increasing of the disparity is translated by an important orientation difference between two pairs of simple cells (dyadic squares in our case) as expressed in Eq. 11 and illustrated by Fig. 14. This leads to the expression of the inter-ocular phase shift using the disparity d and the orientation difference $\Delta\omega$ as given in Eq. 12.

$$\begin{aligned}C_r(x) &= e^{i\Delta\omega}C_l(x-d)\\ &= \rho_l(x-d)e^{\phi_l(x-d)+\Delta\omega}\end{aligned} \tag{11}$$

$$\begin{aligned}\Delta\phi(x) &= \phi_l(x) - \phi_r(x-d) - \Delta\omega\\ &\approx d \times \phi_l' - \Delta\omega + O[d^2]\end{aligned} \tag{12}$$

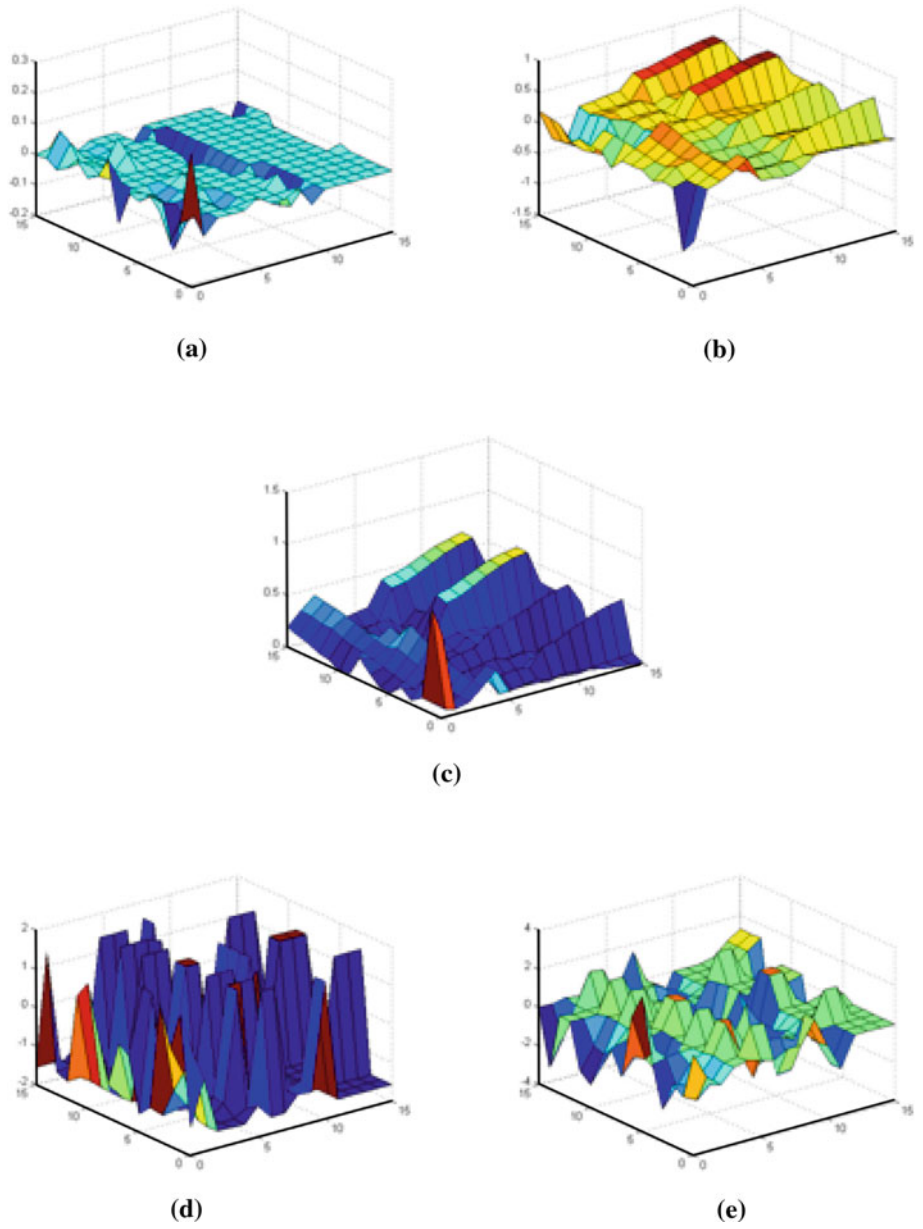


Fig. 12 Example of a pair of dyadic squares belonging to one image of the stereo-pair. **a** Real part of a pair of dyadic squares. **b** Imaginary part of a pair of dyadic squares. **c** Amplitude of the complex function. **d** Inter-ocular phase of a pair of dyadic squares. **e** Instantaneous phase.

By integrating the position shift with the orientation difference, the binocular energy becomes :

$$E(x) = \rho_l^2(x) + \rho_r^2(x) + \rho_l(x)\rho_r(x)cos(d \times \phi_l'(x) - \Delta\omega) \tag{13}$$

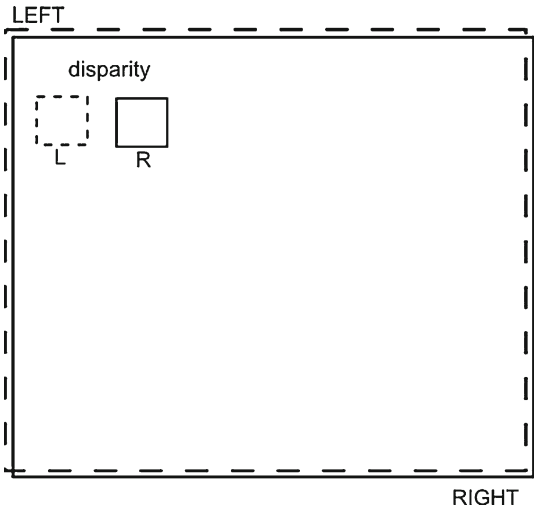


Fig. 13 Illustration of the binocular disparity using a superimposition of the *left* and *right* views. L is the object on the *left* view and R is the object on the *right* view

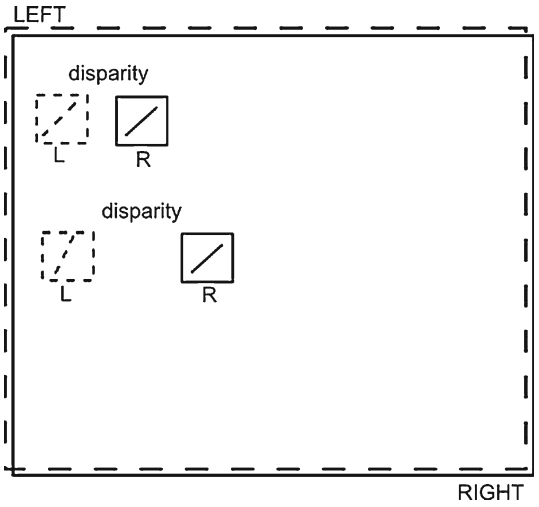


Fig. 14 Illustration of the orientation variation in relation to binocular disparity using a superimposition of the left and right views. L is the object on the *left* view and R is the object on the *right* view

The binocular energy is maximized when the phase shift is equivalent to the product of disparity and instantaneous spatial frequency ($d \times \phi_l^{'}(x) = \Delta\omega$). Notice that even if the dyadic squares have important values of disparity and phase shift, the binocular energy remains maximized while the cosine argument is close to zero. Several physiology works (Field and Tolhurst 1986) demonstrated that the simple cells properties (local disparity, amplitude, orientation) depend on oculomotor elements like the vergence and accommodation.

Back to the focus of this paper i.e. evaluate the quality of stereoscopic images, the idea is to find for each simple cell (dyadic square) from the left image $C_l(x) = \rho_l(x)e^{\phi_l(x)}$ a simple

cell from the right image $C_r(x) = \rho_r(x)e^{\phi_r(x)}$ that maximizes the binocular energy. The metric is formulated in the next section in order to ease its comprehension.

4 BEQM formulation

In this section, we summarize the proposed modeling in order to give a comprehensive formulation of the metric BEQM. We proposed a stereoscopic matching model based on the simple and complex cells properties responsible for the retinal image fusion. The binocular energy $E_j(C_r, C_l)$ measured between two cells (represented by dyadic squares) and belonging respectively to the left and right images, depends on their characteristics i.e. amplitude, phase shift and the inter-ocular phase difference. The impairments on stereoscopic image pairs will result in a variation of the described characteristics leading thus to a variation of the generated binocular energy in comparison to the original image-pair.

The first step consists in calculating the binocular energy generated by the original image-pair. This energy is represented by a set $Z = \{E_j(C_r, C_l)\}$ of couples of complex functions belonging respectively to the left and right original images. After the application of the CWT and DWT on the impaired stereoscopic images, the geometry (quadtree) of the original image-pair is applied to decompose the coefficient into dyadic squares. For each dyadic square, the amplitude, phase and orientation are computed using the actual coefficients of the impaired image-pair allowing to compute the binocular energy of the impaired image-pair represented by $Z' = \{E'_j(C'_r, C'_l)\}$.

As illustrated in Fig. 15, the BEQM metric between two stereoscopic image pairs (I_r, I_l) and (I'_r, I'_l) is obtained thanks to weighted sums of binocular energy difference of each pair of simple cells (α_j) of each subband (β_i) and for luminance and chrominance (γ_k). BEQM is formulated as follows:

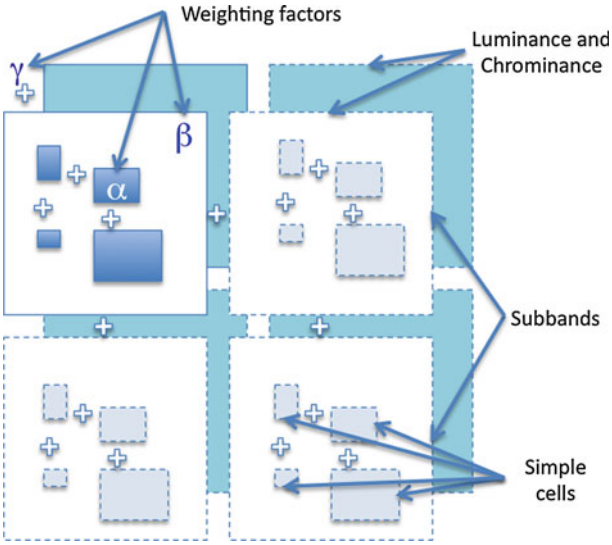


Fig. 15 Description of parameters used in the formulation of the BEQM metric

$$Score(I_r, I_l, I'_r, I'_l) = \sum_{k=1}^{C} \gamma_k \sum_{i=1}^{S} \beta_i \sum_{j=1}^{N_i} \alpha_j \frac{E_{ijk}(C_r, C_l) - E'_{ijk}(C'_r, C'_l)}{E_{ijk}(C_r, C_l) + E'_{ijk}(C'_r, C'_l)} \quad (14)$$

where S is the number of subbands equal to $S = (3 \times L) + 1$ with L the number decomposition levels , N_i is the number of simple cells for the i^{th} subband and C is the number of components equal to 2 in our case. The weighting factors α_j, β_i and γ_k have been set to the same value in order to give the same contribution in the computation of the binocular energy. The sum of weighting factors is equal to one.

5 Experimental results and discussion

This section is devoted to the presentation of experimental results related to the BEQM metric in addition to their discussion. We used for the experiments a stereoscopic database containing 13 image-pairs from the university of Toyama (Akhter et al. 2010; Media Information and Communication Technology (MICT) Lab-oratory 2011). The main interest with this database is the availability of subjective scores allowing to study the correlation of our results with the human judgement. The subjective assessment has been run by the MICT laboratory (Toyama). Seven different levels of JPEG compression were applied for each of the images. There are 70 symmetric and 420 asymmetric coded image pairs of size 640 × 480. In the symmetric stereoscopic coding, an equal level of compression is applied to the left and right images whereas in the asymmetric coding, the compression level of the two images is different. According to the recommendation of the International Telecommunication Union (ITU) for still-image coding, an Absolute Category Rating (ACR : method has been used with a quality scale of five categories (Bad = 1, Poor = 2, Fair = 3, Good = 4 and Excellent = 5). Twenty-four observers have participated to the subjective test.

Also, two images, Teddy and Cones, have been used from the middlebury database (Scharstein and Szeliski 2002a,b) to confirm the results on another acquisition modality even though there is no subjective score. Figures 16 and 17 give an overview of the content of the fifteen image-pairs.

The experimental results are organized in three distinct parts: (1) analysis and discussion of the binocular energy maps obtained with the proposed model, (2) Study of the behavior of the binocular energy with regards to different types of impairments like noise and compression, and (3) a statistical study of the performance of BEQM using a VQEG procedure (VQEG 2008) and a comparison with other metrics.

5.1 Analysis of the binocular energy maps

Figures 18 and 19 show respectively the results of the proposed model for *Cactus* using one decomposition level and *Teddy* using one and 3 decomposition levels. They provide for each pair and each view a localization of the regions where the binocular energy is the most important (Figs. 18c, d, 19c, d), the binocular energy maps (Figs. 18e, f, 19e, f) and a 3D representation of the binocular energy maps (Figs. 18g, h, 19g, h).

It is known that human observers perceive more depth when the disparity increases between left and right views (Blake and Wilson 1991). Our model is developed to return high values of binocular energy in regions where objects are on the front and low values for regions with objects on the bottom (with regards to depth). Also, an important value of binocular energy translates an important excitation of the complex cell provoked by an important excitation of associated simple cells. On Fig. 18, we can notice that the matched

Fig. 16 Overview of the content of the Toyama database

Fig. 17 Teddy and Cone image-pairs from Middlebury database

regions correspond in majority to the cactus in the front of the image-pair leading thus to important values of binocular energy in this area. The same observation can be made for Fig. 19 where, at different decomposition levels, it can be noticed that the important values of binocular energy correspond to objects in the front.

This behavior of simple and complex cells explains the depth perception at the borders of objects. Closer is the object and higher is the contrast provoking thus an optimal projection

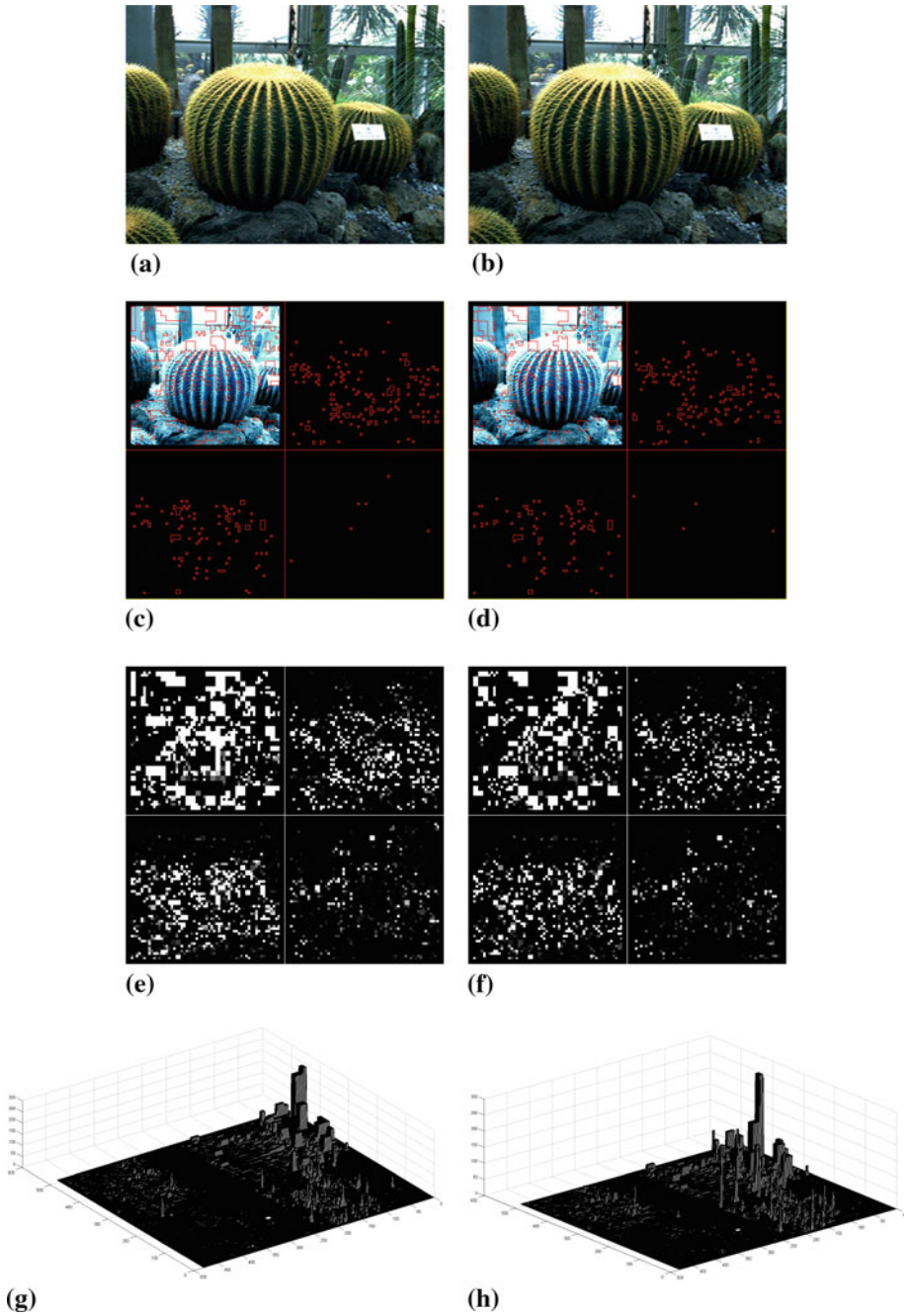


Fig. 18 Matching of the stereo-pair of *Cactus* and estimation of the associated binocular energy. **a** *Left* image. **b** *Right* image. **c** Regions of high binocular energy (*left*). **d** Regions of high binocular energy (*right*). **e** Binocular energy map (*left*). **f** Binocular energy map (*right*). **g** 3D representation of the binocular energy (*left*). **h** 3D representation of the binocular energy (*right*)

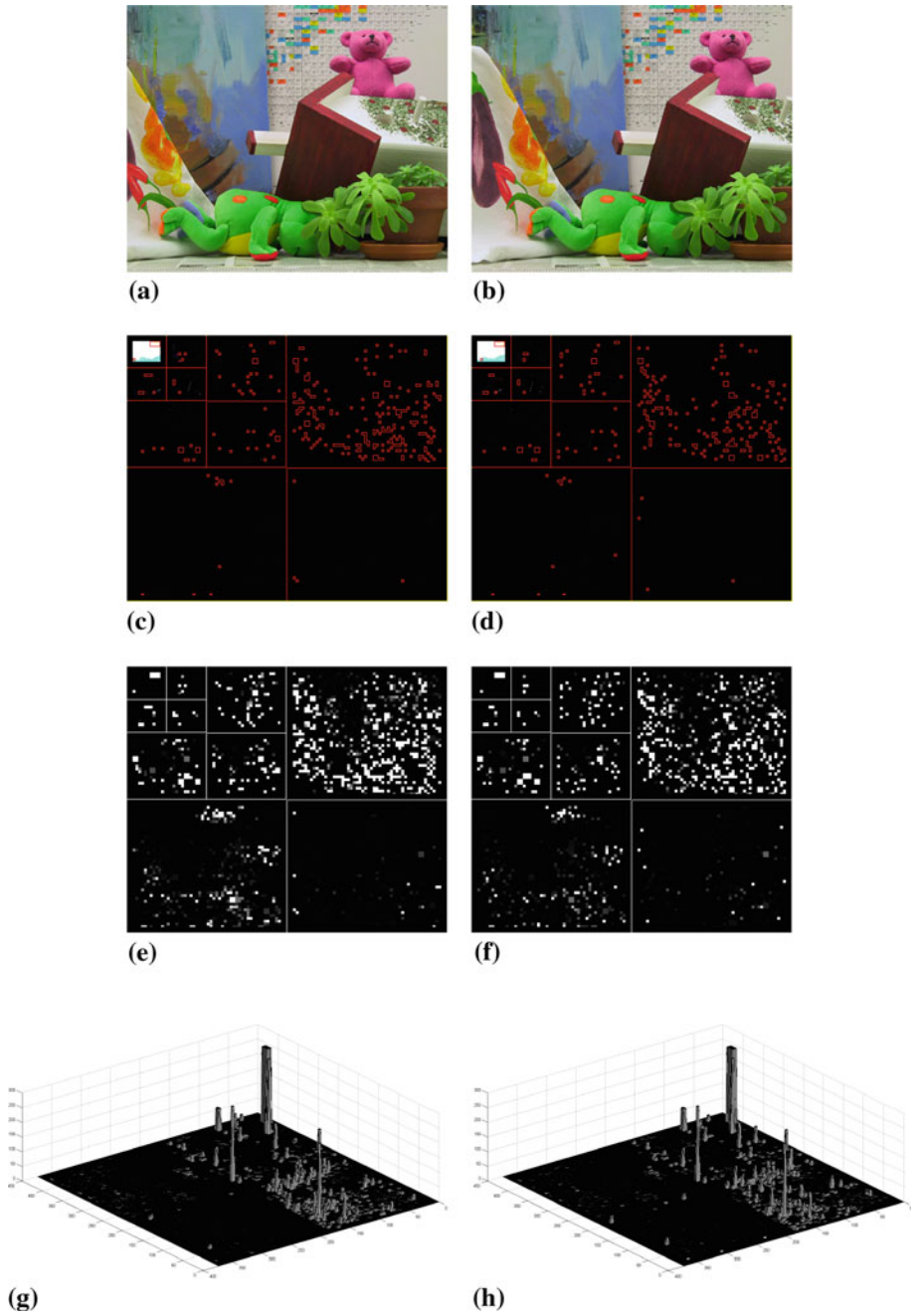


Fig. 19 Matching of the stereo-pair of *Teddy* and estimation of the associated binocular energy. **a** *Left* image. **b** *Right* image. **c** Regions of high binocular energy (*left*). **d** Regions of high binocular energy (*right*). **e** Binocular energy map (*left*). **f** Binocular energy map (*right*). **g** 3D representation of the binocular energy (*left*). **h** 3D representation of the binocular energy (*right*)

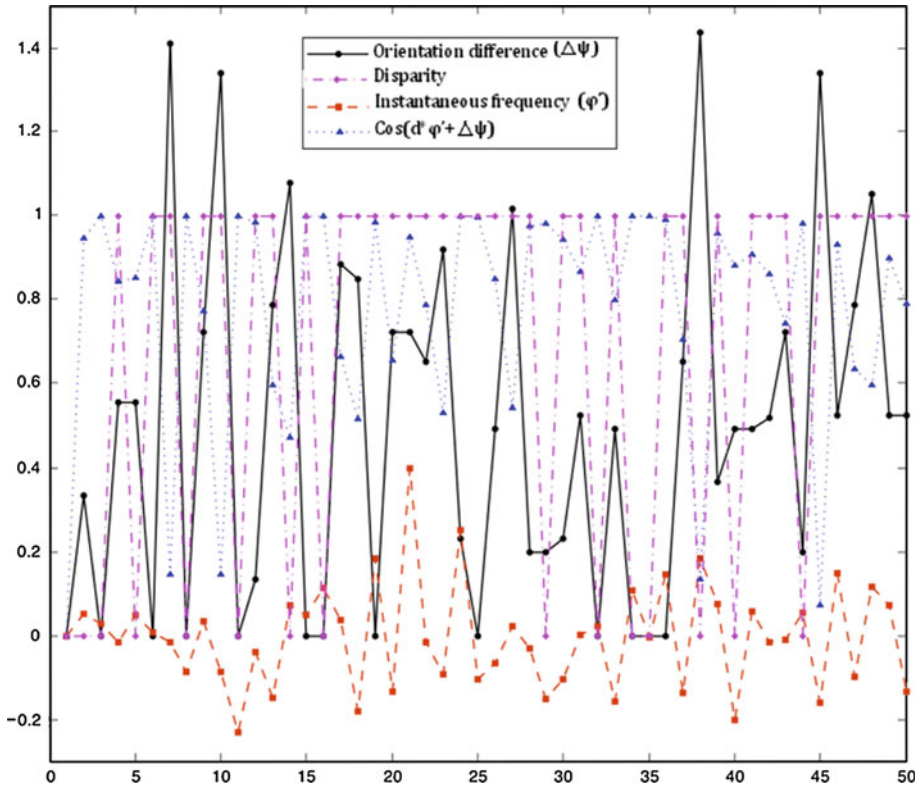


Fig. 20 Variation of the cosine used to compute the binocular energy function of the orientation difference, the disparity and the instantaneous frequency for all the pairs of dyadic squares (x-axis)

of the object's contours on the simple cell. This can be observed easily on the contour of the bear present in *Teddy* image-pair. From the point of view of the transforms that we used in our model, when an object is on the front, the generated wavelet coefficients (CWT and DWT) are high at the contours and low for uniform areas. These coefficients compose the real and imaginary parts of a simple cell modeling (dyadic squares pair) $C(x) = \rho(x)e^{\phi(x)}$ and generate a high amplitude for the calculation of the binocular energy. A similar reasoning can be made for objects in the back of the scene.

The orientation and inter-ocular phase are, as mentioned previously, very important for a simple cell. Remind that for a correct matching of a pair of simple cells, the inter-ocular phase shift is proportional to the product of disparity (d) and instantaneous frequency ($\phi^{'}$). Figure 20 gives the evolution of disparity, orientation difference and instantaneous frequency together with the evolution of the $cos(d \times \phi^{'} - \Delta\omega)$ influencing the binocular energy. This behavior can be observed on high frequency subbands of Figs. 18e, f, 19e, f. In the case of a low contrast inside the simple cell, the bandelet transform, used in our model, does not affect any orientation to it. In our model, we cope with this aspect by assigning an orientation difference of $\Delta\omega = 180°$ for these simple cells. This leads the cosine (consequently the binocular energy) to converge to 0 when the disparity is close to 0 for flat regions or those in the back of the scene.

5.2 Analysis of the model's behavior for different impairments

We have tested our model for different types of impairments in order to analyze its behavior and confirm or not its high correlation with the perceived quality. We selected two compression standards: (1) JPEG known for the blocking effect generated at low bitrates and (2) JPEG 2000 known for the blurring and ringing effects at low bitrates. In addition to compression, we used a noise to study the effect of impulse artifact on the binocular energy and consequently of the perceived quality. The three artifacts have been generated symmetrically (same strength on the left and right images) and asymmetrically (different strengths on the left and right images). Seven quality factors ranging between 10 and 100 have been retained for JPEG (cf. Fig. 21), 6 bitrates between 0.03 and 1 bpp for JPEG 2000 (cf. Fig. 23) and 9 noise density between 0.005 and 0.045 for salt & pepper noise (cf. Fig. 24). In order to increase the readability of figures, only six images are shown.

One important characteristic that has to be taken in the interpretation of the plots is the dominant eye. If we consider, a stereoscopic image pairs P_o encoded twice asymmetrically (P_1 and P_2) where the bitrate of P_{1R} equal to P_{2L} and the bitrate of P_{1L} equal to P_{2R} (R and L are respectively the right and left images). The difference of binocular energy between P_1 and the original pair P_o, and between P_2 and the original pair P_o are different. This can be confirmed on the plots and is explained by the notion of dominant eye.

In the binocular matching used in our model, one of the views plays the role of master while the other view is the slave. Therefore, the master regions are used as a reference for finding the best match in the slave image. We have chosen the right image as a master in our case. With this configuration, the score obtained with the pair having the greatest bitrate for the master is higher than the other pair. This can be confirmed easily on Figs. 21, 23 and 24. In order to confirm these observations, we modified our model to take the left image as the master. The results obtained with the modified model confirmed what is stated before.

5.2.1 JPEG

The JPEG compression standard is well-known for the blocking effect generated at low bitrates. Of course, this effect is due to the adopted scheme based on 8×8 DCT. In addition to 8×8 blocks visible on the left and right views, it smoothes the contour inside the blocks. From the point of view of our model, these two artifacts are interpreted differently. In the case of false contours (created by the blocks) crossing a pair of dyadic squares (represented by a complex function $C(x)$), there will be an important difference of amplitude and orientation with the original pair. Knowing that the used wavelet transforms (CWT and DWT) are sensitive to impairments, the false contours generate high coefficients. The latter are used by the bandelet transform, which is adaptive to the analyzed signal, to determine the orientation of the dyadic square, causing thus a variation of the orientation with regard to the original pair. The same effect can be measured on the amplitude of a dyadic square. So, two complex functions representing two regions in the right and left images have different amplitude and orientation leading to a variation/decrease of the binocular energy. From Fig. 21, we can notice that there is a coherence between the JPEG impairment and the variation of binocular energy. The curves show in majority an important gap for low bitrates of the right images. In addition to the dominant eye aspect mentioned previously, it is related to the non-linear perception of the HVS that responds to stimuli depending on their range. A similar observation can be made on the results of subjective assessment of JPEG asymmetric compression given in Fig. 22. This gap is mainly due to the compensation phenomenon in which the image of higher quality, in the stereoscopic pair, compensates the image of lower quality. Thus,

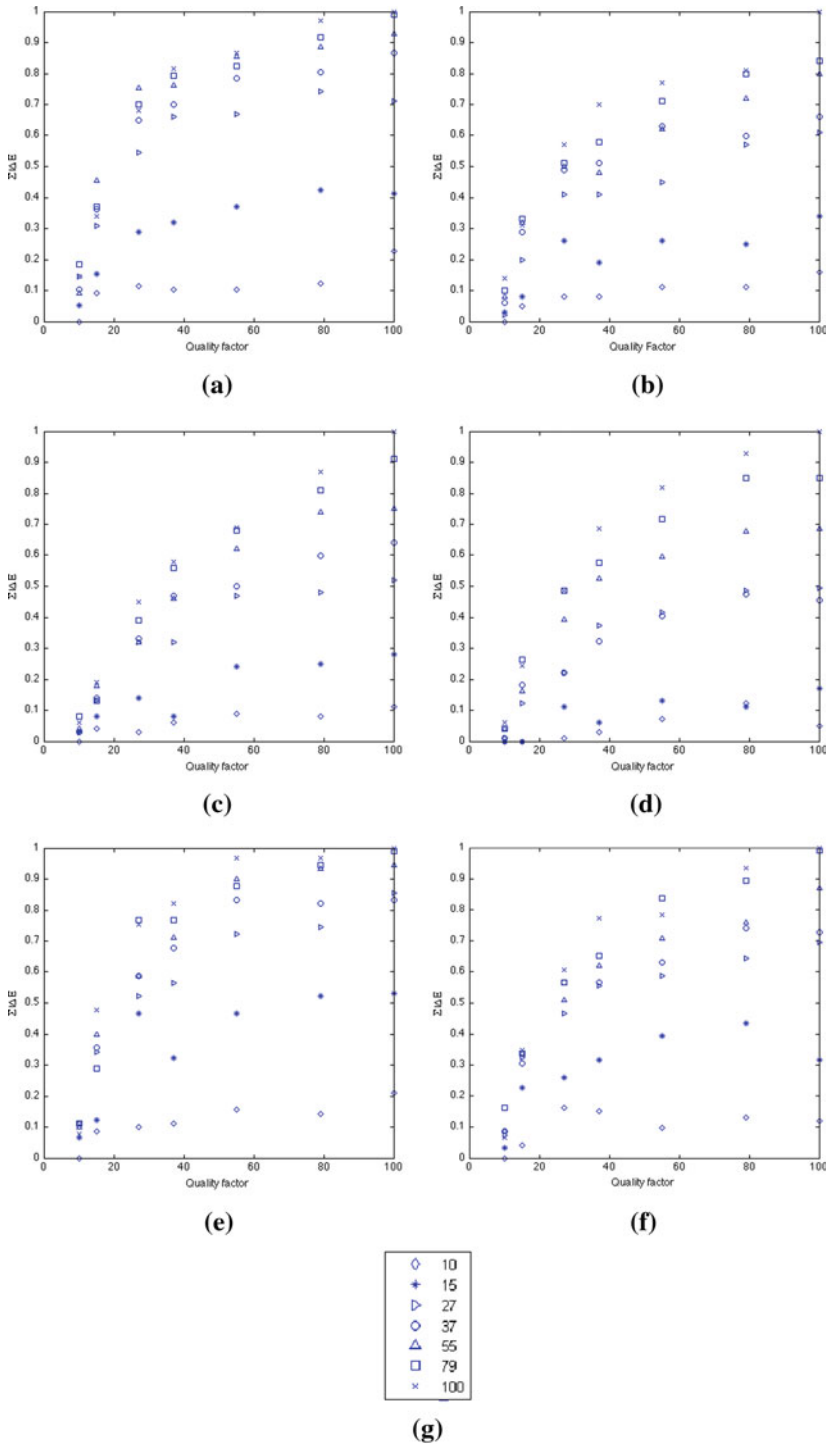


Fig. 21 Variation of the binocular energy difference for JPEG asymmetric coding. **a** Cattle. **b** Goat. **c** Doll. **d** Woman. **e** Doll2. **f** Bird. **g** Legend (quality factor)

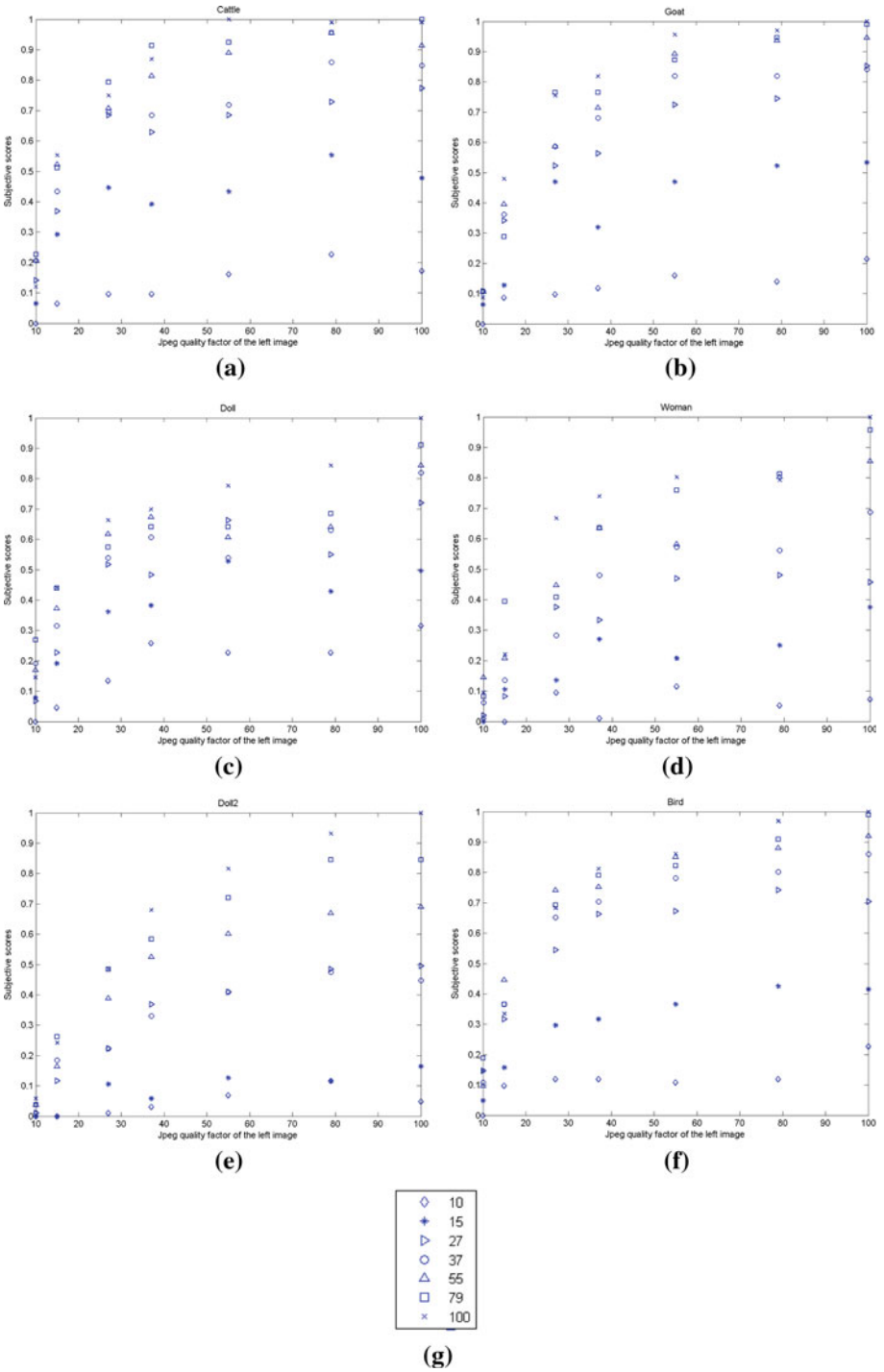


Fig. 22 Subjective scores for JPEG asymmetric coding. **a** Cattle. **b** Goat. **c** Doll. **d** Woman. **e** Doll2. **f** Bird. **g** Legend (quality factor)

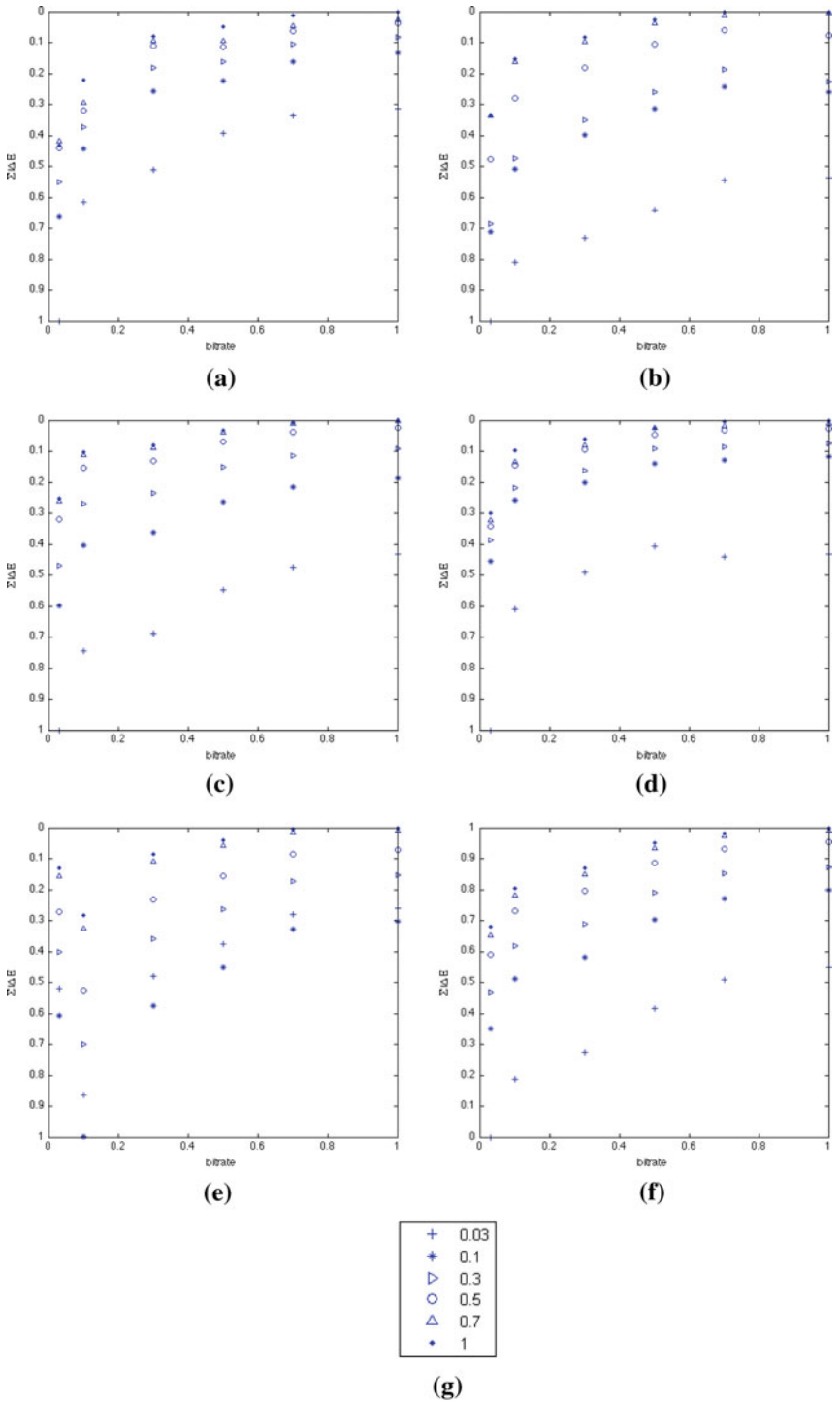


Fig. 23 Variation of the binocular energy difference for JPEG 2000 asymmetric coding. **a** Cattle. **b** Goat. **c** Doll. **d** Woman. **e** Doll2. **f** Bird. **g** Legend (bitrates)

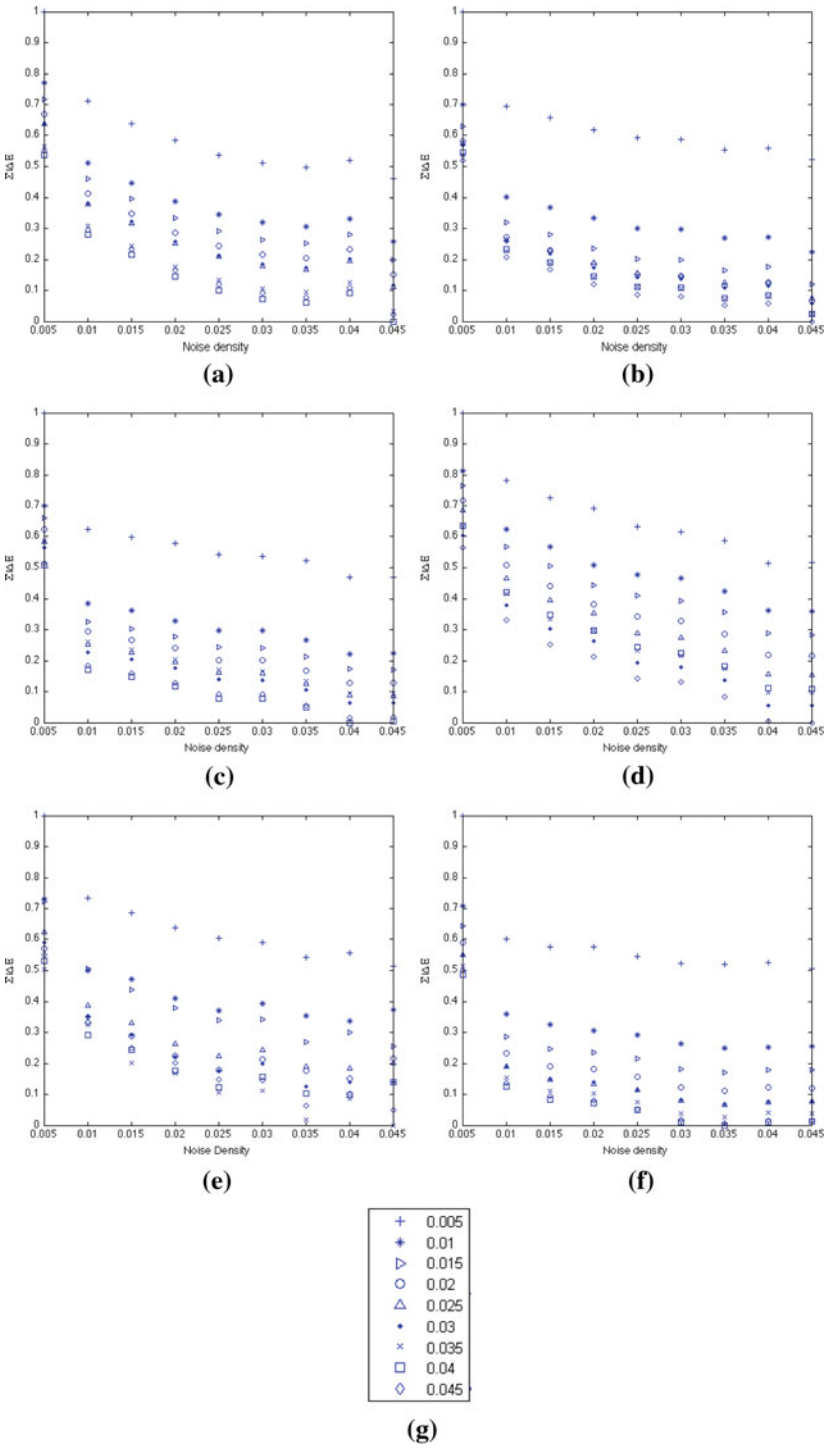


Fig. 24 Variation of the binocular energy difference for salt & pepper noise. **a** Cattle. **b** Goat. **c** Doll. **d** Woman. **e** Doll2. **f** Bird. **g** Legend (noise density)

the quality of reconstructed image is not a linear sum of both qualities but the quality of the binocular signal represented by the amount of binocular energy generated by the pair of stereoscopic images. It allows to conclude that our model reproduces with some fidelity the behavior of the binocular perception.

5.2.2 JPEG 2000

The second impairment concerns the blurring caused by JPEG 2000 because of the quantization of the wavelet coefficients. This artifact creates a reverse reaction than the blocking effect that consists in the depth suppression because of the contours' smoothing. This effect can be explained by the fact that more the contour is smoothed more it is thick and will spread over the ON and OFF regions of the simple cell giving thus a less sensation of depth (Barlow et al. 1967). In the proposed model, the smoothed contours will obtain low coefficients, producing a variation of orientation or even a loss of orientation. The binocular energy of the original and impaired stereo-pairs are different, stating that the perceived quality is different too. From Fig. 23, the same conclusion can be made as for JPEG. The difference of binocular energy computed by our model is very correlated to the strength of the impairment. Also, a gap can be observed for low bitrates of the master image. Except for image *Computer* where the lowest bitrate of the right image provides incoherent results.

5.2.3 Noise

In the case of salt & pepper noise, a similar performance to JPEG is observed. When the noise density is high, the wavelet coefficients of high frequency subbands associated to noisy pixels are high. It implies, as in the previous cases, a variation of the amplitude, phase and orientation leading to a variation of the binocular energy. Figure 24 gathers the results of binocular energy difference for the whole set of images. A low value of noise density means a low impairment et *vice versa*; This is why the curve are inverted in comparison to JPEG and JPEG 2000. The behavior of the binocular energy difference is coherent and shows a high correlation with the strength of noise.

5.2.4 Binocular energy difference versus quality

Starting from the results described above, it seems easy to conclude that there is strong relationship between the difference of binocular energy and perceived quality, independently from the types of impairments that we used in this study. It has been also confirmed by showing the subjective scores for JPEG asymmetric coding that behave similarly to the difference of binocular energy.

This means that the proposed model approximates reliably the binocular energy generated from a pair of stereoscopic images. Also, the difference of binocular energy is an important way to predict the 3D perceived quality. So the BEQM metric based on these aspect might have very good performance in comparison to the state-of-the-art. The next section is dedicated to the performance study.

5.3 Performance evaluation of the BEQM

In this section, the performance of the BEQM (Binocular Energy Quality Metric) (Bensalma and Larabi 2010) is evaluated in comparison to several implementations of 2D metrics. Let

consider (I_lo, I_ro) the original image-pair, (I_ld, I_rd) the impaired image-pair, d_o, d_d respectively original and impaired disparity map and c_o, c_d respectively original and impaired cyclopean images. The used metrics are:

1. PSNR-avg $= (PSNR(I_lo, I_ld) + PSNR(I_ro, I_rd))/2$
2. SSIM-avg $= (SSIM(I_lo, I_ld) + SSIM(I_ro, I_rd))/2$
3. PSNR-disp1 $= sqrt(PSNR\text{-avg} \times PSNR(d_o, d_d))$
4. SSIM-disp1 $= sqrt(SSIM\text{-avg} \times SSIM(d_o, d_d))$
5. PSNR-disp2 $= (PSNR\text{-avg} \times PSNR(d_o, d_d) + 1)$
6. SSIM-disp2 $= (SSIM\text{-avg} \times SSIM(d_o, d_d) + 1)$
7. PSNR-cyclop $= (PSNR(d_o, d_d) + (PSNR(c_o, c_d))/2$
8. SSIM-cyclop $= (SSIM(d_o, d_d) + (SSIM(c_o, c_d))/2$

The evaluation is performed using the Toyama database for JPEG compression because of the availability of subjective scores. The test is made separately for symmetric and asymmetric JPEG coding for 7 bitrates. This lead to a set of image-pairs equal to 637 (13 image pairs $\times 7$ right bitrates $\times 7$ left bitrates).

The scatter plots of the objective scores obtained by BEQM and the subjective scores are shown in Fig. 25a for asymmetric JPEG coding and Fig. 25b for symmetric JPEG coding. Figures 25c–f show respectively scatter plots for Average of PSNR and SSIM on left and right views of the stereo pair and PSNR and SSIM using the disparity map. We can observe that the proposed metric gives very coherent results in comparison to subjective scores. However the PSNR shows a very bad behavior that was foreseeable while SSIM is relatively sound. The consideration of the disparity map improves significantly the results.

Three figures suggested by the VQEG were used to evaluate the performance of the proposed stereo quality metric by comparing the subjective scores (MOS) and the predicted scores (MOS_p). That is, Pearson correlation coefficient (PCC) and Mean Absolute Error (MAE) for prediction accuracy, Spearman rank order correlation coefficient (SCC) for prediction monotonicity and Root Mean Square Error (RMSE) for prediction consistency. For the interpretation of the results, MAE and RMSE are the best when values are close to 0 and PCC and RCC are the best when values are close to 1.

Tables 1, 2 and 3 give respectively the results for asymmetric JPEG coding, symmetric and the whole JPEG set using the performance tools described above. It is easy to notice that BEQM provides the best results on the three tables. One can observe that 3D features improve significantly the performance of 2D metrics. Here can be raised the problem of disparity map extraction because the performance of the metric lies on it.

Regarding the computational performance of the proposed metric, it is difficult to discuss it between 2D metrics and real 3D ones because the latter are of course more complex due to their construction. Also, the execution time of BEQM depends on the complexity of the stereo-pair and can vary from one pair to another. Nevertheless, Table 4 gives the average complexity ratio between the proposed BEQM and, PSNR and SSIM when used with the disparity information. We can notice that the disparity calculation increases significantly the complexity of PSNR and SSIM.

6 Conclusion and future works

In this paper we proposed a novel perceptual metric for stereoscopic image pairs aiming to judge the quality as the HVS does by taking into account the binocular fusion process. The novelty lies in modeling the 3D perception of the HVS in opposition to the existing

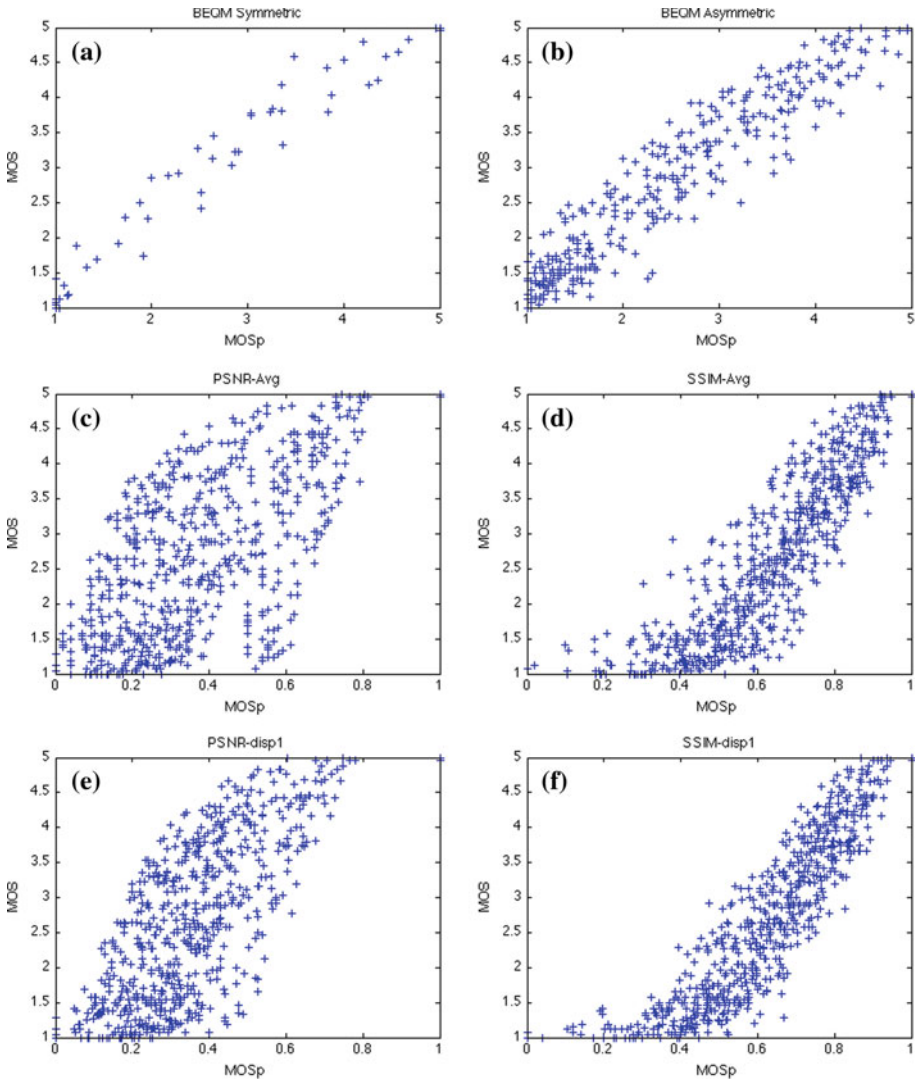


Fig. 25 Scatter plots of objective scores vs. subjective scores for **a** BEQM with symmetric coding, **b** BEQM with asymmetric coding, **c** average of PSNR, **d** average of SSIM, **e** PSNR with disparity map and **f** SSIM with disparity map

metrics that try to assess a 3D content either by using 2D metrics or by exploiting the depth information.

Since the HVS remains the reference in terms of stereoscopic vision, we focused on modeling some of its properties. Specifically, the binocular fusion of which the 3D perceptual image is the result. First, the behavior of simple cells, whose are the first to receive the retinal information in the visual cortex, has been studied and modeled using spatial-frequency transforms like the Discrete Wavelet Transform (DWT), Complex Wavelet Transform (CWT) and the bandelet transform. This choice has been made because a simple cell is characterized by its amplitude, orientation, phase and size and analogically the dyadic square can be

Table 1 Performance comparison of the objective and subjective scores of different metrics for JPEG asymmetric coding (best results are given in bold)

Metrics	PCC	SCC	RMSE	MAE
PSNR-avg	0.6241	0.5913	0.8741	0.734
SSIM-avg	0.8909	0.8814	0.5082	0.3920
PSNR-disp1	0.7260	0.6973	0.7693	0.6334
SSIM-disp1	0.9182	0.9154	0.4431	0.3504
PSNR-disp2	0.6774	0.6366	0.8230	0.6857
SSIM-disp2	0.9203	0.9147	0.4378	0.3385
PSNR-cyclop	0.7775	0.7539	0.7035	0.5889
SSIM-cyclop	0.8260	0.8122	0.6306	0.5112
BEQM	**0.9490**	**0.9423**	**0.3663**	**0.2877**

Table 2 Performance comparison of the objective and subjective scores of different metrics for JPEG symmetric coding (best results are given in bold)

Metrics	PCC	SCC	RMSE	MAE
PSNR-avg	0.8518	0.8412	0.7229	0.5710
SSIM-avg	0.9561	0.9533	0.4045	0.2831
PSNR-disp1	0.8863	0.8819	0.6393	0.5068
SSIM-disp1	0.9621	0.9633	0.3763	0.2814
PSNR-disp2	0.8691	0.8606	0.6827	0.5415
SSIM-disp2	0.9640	0.9624	0.3671	0.2579
PSNR-cyclop	0.8905	0.8887	0.6279	0.4892
SSIM-cyclop	0.9045	0.9083	0.5885	0.4588
BEQM	**0.9835**	**0.9816**	**0.2499**	**0.1862**

Table 3 Performance comparison of the objective and subjective scores of different metrics for the whole JPEG images set (best results are given in bold)

Metrics	PCC	SCC	RMSE	MAE
PSNR-avg	0.6590	0.6250	0.8782	0.7317
SSIM-avg	0.9030	0.8954	0.5016	0.3826
PSNR-disp1	0.7526	0.7252	0.7688	0.6315
SSIM-disp1	0.9275	0.9247	0.4365	0.3427
PSNR-disp2	0.7033	0.6679	0.8300	0.6924
SSIM-disp2	0.9289	0.9244	0.4323	0.3308
PSNR-cyclop	0.8023	0.7804	0.6970	0.5784
SSIM-cyclop	0.8441	0.8298	0.6260	0.5052
BEQM	**0.9560**	**0.9513**	**0.3535**	**0.2756**

Table 4 Average complexity ratio with regards to BEQM

Quality metric	BEQM	PSNR-disp	SSIM-disp
Average complexity ratio	1	0.814	0.819

characterized in the same manner. The binocular fusion is performed by the complex cell from the output of simple cells. This process consists in finding for each simple cell of the dominant eye (right image in our case) a correspondent in the other (left image) that maximizes the binocular energy. The proposed metric BEQM (Binocular Energy Quality Metric) is formulated as the difference of binocular energy between the original pair and the impaired pairs. The experimental results show a high correlation between the binocular energy and the strength of impairments leading to a high performance of the metric in comparison to exiting ones.

Several future works are envisaged as for example the construction of a comprehensive database that will allow to have a solid reference for metrics comparison for stereoscopic imaging. Another important work is the extension of BEQM to 3D video assessment by taking into account the temporal properties of the HVS.

Acknowledgments This work has been supported by the project QuIAVU funded by the French Research Agency.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, 2*(2), 284–299.

Akhter, R., Sazzad, Z. M. P., Horita, Y., & Baltes, J. (2010). No reference stereoscopic image quality assessment. In *Image quality and system performance* (vol. 7524, pp. 17–21). San Jose, California, USA.

Ates, H. F., & Orchard, M. T. (2003). A nonlinear image representation in wavelet domain using complex signals with single quadrant spectrum. *Asilomar Conference on Signals, Systems, Computers, 2*, 1966–1970.

Avcibas, I., Sankur, B., & Sayood, K. (2002). Statistical evaluation of image quality measures. *Journal of Electronic Imaging, 11*(2), 206–223.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology, 193*, 327–342.

Bensalma, R., & Larabi, M. C. (2010). Stereo image coding based on binocular energy modeling. In *International conference image processing (ICIP)* (pp. 2989–2992).

Bensalma, R., & Larabi, M. C. (2010). Towards a perceptual quality metric for color stereo images. In *International conference image processing (ICIP)* (pp. 4037–4040). Hong Kong.

Blake, R., & Wilson, H. R. (1991). Neural models of stereoscopic vision. Trends in neurosciences. *International Journal of Computer Vision, 14*, 445–452.

Boev, A., Gotchev, A., Egiazarian, K. O., Aksay, A., & Akar, G. B. (2010). Towards compound stereo-video quality metric: A specific encoder-based framework. In *IEEE SSIAI* (pp. 218–222). Denver, Colorado, USA.

Campbell, F. W., Cooper, G. F., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology, 203*, 223–235.

Campisi, P., LeCallet, P., & Marini, E. (2007). Stereoscopic images quality assessment. In *European signal processing conference*. Poznan, Poland.

Candès, E., ans Demanet, L., Donoho, D., & Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modelling and Simulation, 5*(3), 861–899.

Cheng, I., & Boulanger, P. (2005). A 3D perceptual metric using just-noticeable-difference. In: *In Eurographics Short Presentations* (pp. 97–100).

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

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Journal on Nature, 352*, 156–195.

Do, M., & Vetterli, M. (2001). Pyramidal directional filter banks and curvelets. *IEEE proceedings on international conference image processing*.

Donghyun, K., Dongbo, M., Juhyun, O., Kwanghoon, S., & Seonggyu, J. (2009). Depth map quality metric for three-dimensional video. *Image Quality and System Performance, 7237*(29), 723719.

Ellinas, J. N., & Sangriotis, M. S. (2004). Stereo image compression using wavelet coefficients morphology. *Image and Vision Computing, 22*(4), 281–290.
Field, D. J., & Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London, 228*, 379–400.
Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy model, position shifts and phase shifts. *Vision Research, 36*(12), 1839–1857.
Foster, K. H., Gaska, J. P., Marcelja, S., & Pollen, D. A. (1983). Phase relationships between adjacent simple cells in the feline visual cortex. *Journal of Physiology, 345*, 22.
Goldmann, L., & Ebrahimi, T. (2010) 3D quality is more than just the sum of 2D and depth. In *IEEE International workshop on hot topics in 3D*.
Gorley, P., & Holliman, N. (2008). Stereoscopic image quality metrics and compression. In *Image quality and system performance*, vol. 6803 (pp. 1–11). San Jose, California, USA.
Hewage, C., & Martini, M.G. (2010). Reduced-reference quality metric for 3D depth map transmission. In *IEEE 3DTV conference* (pp. 1–4). Tampere, Finland.
Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.
Hubel, D. H., & Wiesel, T. N. (1970). Stereoscopic vision in macaque monkey. cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Journal of Nature, 225*, 41–42.
Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*(6), 1233–1258.
Kaptein, R. G., Kuijsters, A., Lambooij, M. T. M., IJsselsteijn, W. A., & Heynderickx, I. (2008). Performance evaluation of 3D-TV systems. In *image quality and system performance* (pp. 1–11). San Jose, California, USA.
Keita, T., & Takeshi, N. (2005). Unstructured light field rendering using on-the-fiy focus measurements. In *IEEE international conference on multimedia and expo*.
Kingsbury, N. (1997). Image processing with complex wavelets. *Philosophical Transactions on Royal Society London A, 357*, 2543–2560.
Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Physiology, 16*, 37–68.
Lavoué, G., Gelasca, E. D., Dupont, F., Baskurt, A., & Ebrahimi, T. (2006). Perceptually driven 3D distance metrics with application to watermarking. *Image Quality and System Performance, 6312*(29), 63120L.
Le Pennec, E., & Mallat, S. (2005). Bandelet image approximation and compression. *SIAM Multiscale Modeling and Simulation, 4*(3), 992–1039.
Liu, A., Gaska, J. P., Jacobson, L. D., & Pollen, D. A. (1992). Interneuronal interaction between members of quadrature phase and anti-phase pairs in the cat's visual cortex. *Vision Research, 32*, 1193–1198.
Mallat, S. (1989). A theory for multiresolution signal decomposition : the wavelet represntation. *IEEE, PAMI, 11*(7), 674–693.
Mallat, S., & Peyré, G. (2006). Orthogonal bandelet bases for geometric image approximation. *Communications on Pure and Applied Mathematics, 61*(9), 1173–1212.
Media Information and Communication Technology (MICT) Lab-oratory. (2011). Mict image quality evaluation database. http://mict.eng.u-toyama.ac.jp/mict/index2.html. Accessed 24 Sept 2011.
Meesters, L. M. J., IJsselsteijn, W. A., & Seuntiens, P. J. H. (2004). A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(3), 381–391.
Miyahara, M., Kotani, K., & Algazi, V. R. (1998). Objective picture quality scale (PQS) for image coding. *IEEE Transaction Communications, 46*(9), 1215–1225.
Nath, S. K., & Dubois, E. (2006). An improved, wavelet-based, stereoscopic image sequence codec with SNR and spatial scalability. *Signal Processing Image Communication, 21*(3), 181–199.
Ohzawa, I., & Freeman, R. D. (1986). The binocular organization of simple cells in the cat's visual cortex. *Journal of Neurophysiology, 56*, 221–242.
Ohzawa, I., & Freeman, R. D. (1986). The binocular organization of complex cells in the cat's visual cortex. *Journal of Neurophysiology, 56*, 243–259.
Olsson, R., & Sjostrom, M. A. (2007). Depth dependent quality metric for evaluation of coded integral imaging based 3D-images. In *IEEE 3DTV*. Kos, Greece.
Palmer, L. A., & Davis, T. L. (1981). Receptive-field structure in cat striate cortex. *Vision Research, 46*, 260–276.
Peyre, G. (2005). *Geometrie multi-échelles pour les images et les textures*. Ph.D. thesis, Ecole Polytechnique.
Pollen, D. A., & Ronner, S. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science, 212*, 1409–1411.

Rittermann, M. A. (2004). A proposal for the quality assessment of 3D video objects. In *International workshop on image analysis for multimedia interactiveservices*. Lisboa, Portugal.
Sarnoff Corporation (2003) JND metrix technology. evaluation (2003) Version available http://www.sarnoff.com/productsservices/videovision/jndmetrix/downloads.asp.
Schanda, J. (2007). *Colorimetry: Understanding the CIE System*. Hoboken, NJ, USA: Wiley.
Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision, 47*(1/2/3), 7–42.
Scharstein, D., & Szeliski, R. (2002). Middlebury stereo vision page. vision middlebury data base (2002) Version available http://vision.middlebury.edu/stereo/.
Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. G. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine, 22*(6), 123–151.
Sheikh, H. R., Bovik, A. C., & deVeciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing, 14*(12), 2117–2128.
Sheikh, H. R., & Bovik, A. C. (2006). Image information and visual quality. *IEEE Transactions on Image Processing, 15*(2), 430–444.
Tikanmäki, A., Gotchev, A., Smolic, A., & Müller, K. (2008). Quality assessment of 3D video in rate allocation experiments. In *IEEE international symposium on consumer electronics*. Algarve, Portugal.
VQEG (2008). Final report from the video quality experts group on the validation of objective models of multimedia quality assessement. Technical Report on PHASE I 2008, VQEG.
Wang, Z., Lu, L., & Bovik, A. C. (2004). Video quality assessment based on structural distortion measurement. *Signal Processing Image Communication, 19*(2), 121–132.
Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003). Multi-scale structural similarity for image quality assessment. In *IEEE asilomar conference on signals, systems and computers* (pp. 1398–1402). Pacific Grove, CA.
Watson, A. B. (1993). Dctune: A technique for visual optimization of dct quantization matrices for individual images. *Society for Information Display Digest of Technical Papers, 1*, 946–949.
Weken, D. V., Nachtegael, M., & Kerre, E. E. (2004). Using similarity measures and homogeneity for the comparison of images. *Image and Vision Computing, 22*(9), 695–702.
Woo, W., Ortega, A., & Iwadate, Y. (1999). Stereo image coding based on binocular energy modeling. In *International conference image processing (ICIP)* (pp. 467–471).
Xing, L., You, J., Ebrahimi, T., & Perkis, A. (2010). A perceptual quality metric for stereoscopic crosstalk perception. In *IEEE international conference on image processing* (pp. 4033–4036). Hong Kong, China.
You, J., Xing, L., Perkis, A., & Wang, X. (2010). Assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In *International workshop on video processing and quality metrics*. Scottsdale, Arizona, USA.

Author Biographies

Rafik Bensalma graduated in computer science engineering from the high-school of computer science, Algeria, in 2005. In 2007, he received master degree research in image processing from the University of Poitiers. In same year, he joined the Signal, Image and Communication Laboratory (XLIM-SIC) at the University of Poitiers, France, as a Ph.D. candidate. He received the Ph.D. degree in signal and image processing from the University of Poitiers in 2011. Since 2012, he has been a Research Engineer in the same University. His research activities are focused on stereoscopic coding and 3D quality assessment.

Mohamed-Chaker Larabi received his Ph.D. from the University of Poitiers (2002). He is currently the associate professor in charge of the perception, color and quality activity at the same university. His actual scientific interests deal with image and video coding and optimization, 2D and 3D image and video quality assessment, and user experience. He works on Human Visual System modeling (spatial, temporal and spatio-temporal, binocular) for the enhancement of several algorithms such as compression, digital cinema, etc. Chaker Larabi is a member of the French National Body for the ISO JPEG committee (since 2000)/MPEG and chair of the Advanced Image Coding group. He serves as a member of divisions 1 and 8 of CIE, is a member of IS&T, and a senior member of IEEE. He is involved in many local, regional, national and international projects.