Research Article

Objective No-Reference Stereoscopic Image Quality Prediction Based on 2D Image Features and Relative Disparity

Z. M. Parvez Sazzad,[1] Roushain Akhter,[2] J. Baltes,[2] and Y. Horita[1]

[1] *Graduate School of Science and Engineering, University of Toyama, Toyama 930-8555, Japan*
[2] *Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada R3T 2N2*

Correspondence should be addressed to Z. M. Parvez Sazzad, sazzad@univdhaka.edu

Stereoscopic images are widely used to enhance the viewing experience of three-dimensional (3D) imaging and communication system. In this paper, we propose an image feature and disparity dependent quality evaluation metric, which incorporates human visible system characteristics. We believe perceived distortions and disparity of any stereoscopic image are strongly dependent on local features, such as edge (i.e., nonplane areas of an image) and nonedge (i.e., plane areas of an image) areas within the image. Therefore, a no-reference perceptual quality assessment method is developed for JPEG coded stereoscopic images based on segmented local features of distortions and disparity. Local feature information such as edge and non-edge area based relative disparity estimation, as well as the blockiness and the edge distortion within the block of images are evaluated in this method. Subjective stereo image database is used for evaluation of the metric. The subjective experiment results indicate that our metric has sufficient prediction performance.

1. Introduction

Nowadays, three-dimensional (3D) stereo media is becoming immersive media to increase visual experience as natural in various applications ranging from entertainment [1] to more specialized applications such as remote education [2], robot navigation [3], medical applications like body exploration [4], and therapeutic purposes [5]. There are many alternative technologies for 3D image/video display and communication, including holographic, volumetric, and stereoscopic; stereoscopic image/video seems to be the most developed technology at the present [6]. Stereoscopic image consists of two images (left and right views) captured by closely located (approximately the distance between two eyes) two cameras. These views constitute a stereo pair and can be perceived as a virtual view in 3D by human observers with the rendering of corresponding view points. Although the technologies required for 3D image are emerging rapidly, the effect of these technologies as well as image compression on the perceptual quality of 3D viewing has not been thoroughly studied. Therefore, perceptual 3D image quality is an important issue to assess the performance of all 3D imaging

applications. There are several signal processing operations that have been designed for stereoscopic images [7] and some researchers are still working to develop a new standard for efficient multiview image/video coding [8]. They believe the image compression technique that used in 2D image material can also be applied independently on the left and right images of a stereo image pair to save valuable bandwidth and storage capacity. Although subjective assessment is the most accurate method for perceived image quality, it is time consuming, and expensive. Therefore, objective quality evaluation method is required that can automatically predict perceptual image quality.

In the last two decades, a lot of work have been concentrated to develop conventional 2D image/video quality assessment methods. Whereas, still now no comparable effort has been devoted to the quality assessment for 3D/stereoscopic images. A full-reference (FR) quality metric for the assessment of stereoscopic image pairs using the fusion of 2D quality metrics and of the depth information is proposed in [9]. The study evaluated that the FR metric of 2D quality assessment can be used for an extension to 3D with the incorporation of depth information. In [10],

the selection of the rate allocation strategy between views is addressed for scalable multiview video codec to obtain the best rate-distortion performance. In [11], a FR quality metric is proposed for stereoscopic color images. The metric is proposed based on the use of binocular energy contained in the left and right retinal images calculated by complex wavelet transform and bandelet transform. In [12], a FR overall stereoscopic image quality metric has been suggested by combining conventional 2D image quality metric with disparity information. In [13], the quality of 3D videos stored as monoscopic color videos that augmented by pixel depth map and finally this pixel information used for color coding and depth data. In [14], the effect of low pass filtering one channel of a stereo sequence is explored in terms of perceived quality, depth, and sharpness. The result found that the correlation between image quality and perceived depth is low for low pass filtering. A comprehensive analysis of the perceptual requirements for 3D TV is made in [15] along with a description of the main artifacts of stereo TV. In [16], the concept of visual fatigue and its subjective counterpart, visual discomfort in relation to stereoscopic display technology, and image generation is reviewed. To guarantee the visual comfort in consumer applications, such as stereoscopic television, it is recommended to adhere to a limit of "one degree of disparity," which still allows sufficient depth rendering for most applications. In [17], the effects of camera base distance and JPEG coding on overall image quality, perceived depth, perceived sharpness, and perceived eye strain are discussed. The relationship between the perceived overall image quality and the perceived depth are discussed in [18]. In [19], an FR quality assessment model is proposed for stereoscopic color images based on texture features of left image as well as disparity information between left and right images. In [20], a positive relationship between depth and perceived image quality for uncompressed stereoscopic images is described. Subjective ratings of video quality for MPEG-2 coded stereo and nonstereo sequences with different bit rates are investigated in [21]. In [22], a crosstalk prediction metric is proposed for stereoscopic images. The method try to predict level of crosstalk perception based on crosstalk levels, camera baseline, and scene content.

Although perceptual quality of stereoscopic images depends mainly on the factors such as the depth perception, level of crosstalk, and visual discomfort, overall perceptual quality reflects the combined effect of the multidimensional factors [16]. We believe that human visual perception is very sensitive to edge information and perceived image distortions are strongly dependent on the local features such as edge, and nonedge areas and also depth/disparity perception is dependent on the local features of images. Therefore, in this work we propose a no-reference (NR) quality assessment method for stereoscopic images based on segmented local features of distortions and disparity. In many practical applications, the reference image is not available, therefore an NR quality assessment approach is desirable. Here, we limit our work to JPEG coded stereoscopic images only. A similar approach based on three local features such as edge, flat, and texture was made in [23]. The metric used many parameters (thirteen) and local features

(three). Consequently, computational cost of the model was high. Therefore, we consider two local features (edge and nonedge) and less parameters with low computational cost in this paper. A previous instantiation of this approach was made in [24] and promising results on simple tests were achieved. In this paper, we generalize this algorithm, and provide a more extensive set of validation results on a stereo image databases. The rest of the paper is organized as follows: Section 2 describes briefly the subjective database that is used to evaluate our method. The details of our approach is given in Section 3. Results are discussed in Section 4 and finally, the paper is concluded in Section 6.

2. The Subjective Databases

We conducted subjective experiment on 24 bit/pixel RGB color stereoscopic images in the Media Information and Communication Technology (MICT) laboratory, University of Toyama [23]. The database contained JPEG coded symmetric and asymmetric 490 stereoscopic image pairs (70 symmetric, and 420 asymmetric pairs) of size 640×480. Out of all, ten were reference stereo pairs. The seven quality scales (QS: 10, 15, 27, 37, 55, 79, and reference) were selected for the JPEG coder. A double stimulus impairment scale (DSIS) method was used in the subjective experiment. The impairment scale contained five categories marked with adjectives and numbers as follows: "Imperceptible = 5", "Perceptible but not annoying = 4", "Slightly annoying = 3", "Annoying = 2," and "Very annoying = 1". A 10-inch auto stereoscopic, LCD (SANYO) display (resolution: 640×480) was used in this experiment. Twenty-four nonexpert subjects were shown the database; most of them were college/university student. Mean opinion scores (MOSs) were then computed for each stereo image after the screening of postexperiment results according to ITU-R Rec. 500-10 [25]. The details of the experiment were discussed in [24].

3. Proposed Objective Method

The primary function of the human visual system (HVS) is to extract structural or edge information from the viewing field [26]. Therefore, Human visual perception is very sensitive to edge detection, and consequently, perceive distortions should be strongly dependent on local features such as edge, and nonedge. For example, in theory, the visual distortions of an image increase with an increased rate of compression. However, the relationship between the distortions and the level of compressions is not always straight forward. It strongly depends on the texture contents of an image as well. In order to verify the relationship, we analyse the degradation of images which causes visual difficulty, that is, appearance of image distortions at different compression levels for various textures of images. Here, we consider an image (see Figure 1(a)) that contains a variety of textures such as edge and nonedge areas. Out of all edge (nonuniform) and nonedge (uniform) areas in Figure 1(a), we analyse a small portion of uniform and nonuniform areas which are represented by the top-right rectangular box and the

(a) Reference image; Image taken from [27]

(b) Compressed image; QS = 10

(i) QS = 50

(ii) QS = 25

(i) QS = 50

(ii) QS = 25

(iii) QS = 15

(iv) QS = 10

(iii) QS = 15

(iv) QS = 10

(c) A small portion of an uniform area

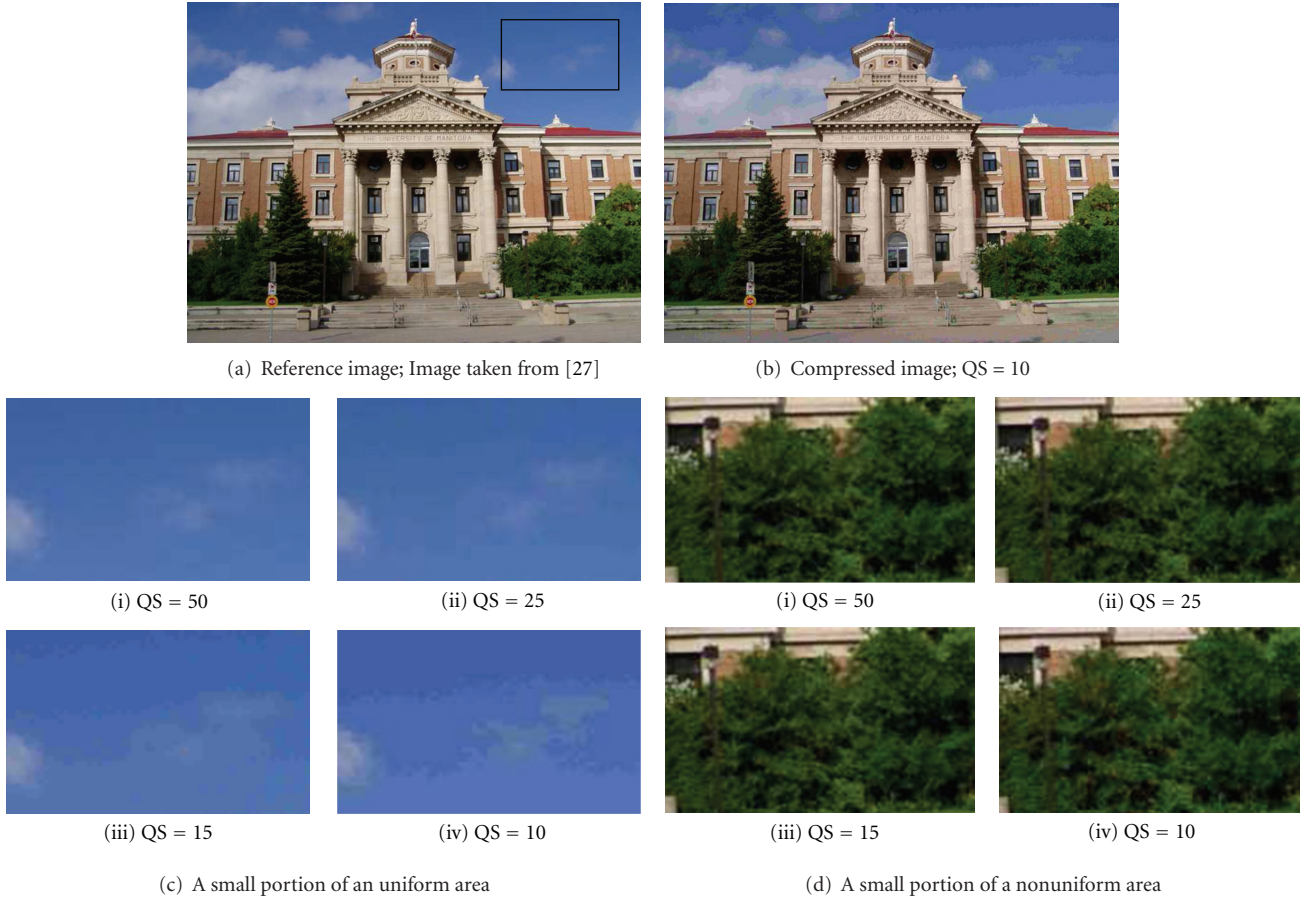(d) A small portion of a nonuniform area

Figure 1: Variation of perceived distortion (uniform and nonuniform areas).

bottom-right rectangular box (dotted line), respectively. A high level of JPEG compression is applied to the image which is shown in Figure 1(b). The result shows the blocking distortions are more visible to uniform areas compared to that the nonuniform areas (see the corresponding areas in the compressed image) even though the level of compression is equal. In order to study the relationship more extensively we apply four levels of compression (QS: 50, 25, 15, and 10) to the image and consider expanded views of the portions of uniform and nonuniform areas (see the rectangular box areas) for each level of compression which are shown in Figures 1(c) and 1(d), respectively. These two figures indicate that perceived distortions for these areas are not similar even though the compression levels are equal. In details, blocking distortions are more visible in uniform areas compared to nonuniform areas (see Figures 1(c)(iii) and 1(d)(iii), and also Figures 1(c)(iv) and 1(d)(iv)). Similarly, the blur distortions are more visible in the nonuniform areas compared to uniform areas (see Figures 1(c)(iii) and 1(d)(iii), and also Figures 1(c)(iv) and 1(d)(iv)). The results indicate that visibility of image distortions are strongly depended on local features such as edge and nonedge areas.

Thus, we also believe that 3D depth perception is strongly dependent on objects, structures, or textures edges of stereo image content. Therefore, an NR perceptual stereoscopic image quality assessment method is proposed based on segmented local features of distortions and disparity in this research. An efficient 2D compression technique, JPEG codec is applied independently on the left and right views of the stereo image pairs. Since JPEG is a block based discrete cosine transform (DCT) coding technique, both blocking and edge distortions may be created during quantization of DCT coefficients in the coded images. Blocking effect occurs due to the discontinuity at block boundaries, which is generated because the quantization in JPEG is block based and the blocks are quantized independently. Here, blockiness of a block is calculated as the average absolute difference around the block boundary. The edge distortion, which makes blurring effect, is mainly due to the loss of high-frequency DCT coefficients, which smooths the image signal within each block. Thus, higher blurring represents a smoother image signal which causes the reduction of signal edge points. Consequently, average edge point detection measures of blocks give more insight into the relative edge distortion in the image. Here, zero-crossing technique is used as an edge detector. Although, the impact of coding distortions on the perceived stereoscopic image quality of an asymmetric image pair depends on the visual appearance of the artifact, where blockiness appears to be much more disturbing than blur [28], we take into account the maximum blockiness and edge distortion measures between the left and right views. Therefore, we consider higher blockiness
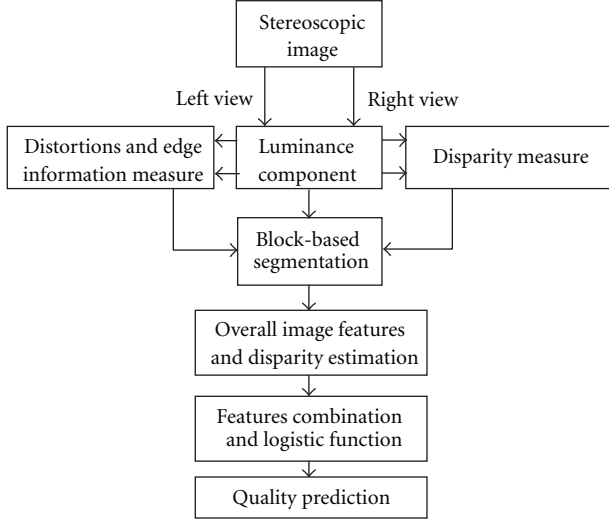
FIGURE 2: Proposed NR quality evaluation method.

and lower zero-crossing values between the two views. For simplicity, only the luminance component is considered to make overall quality prediction of color stereo images. As image distortions as well as disparity are estimated based on segmented local features, a block based segmentation algorithm is applied to identify edge and nonedge areas of an image which is discussed in details in [24]. Subsequently, the distortions and disparity measures are described in the next Sections. The block diagram of the proposed method is shown in Figure 2.

3.1. Image Distortions Measure. We estimate blockiness and zero-crossing to measure JPEG coded image distortions in spatial domain based on segmented local features. Firstly, we calculate blockiness and zero-crossing of each 8×8 block of the stereo image pair separately (left and right images). Secondly, we apply the block (8×8) based segmentation algorithm to the left and right images individually to classify edge, and nonedge blocks in the images [24]. Thirdly, we average each value of blockiness and zero-crossing separately for edge, and nonedge blocks of each image of the stereo pair. Fourthly, the total blockiness and zero-crossing of the stereo image pair is estimated respectively based on the higher blockiness value and lower zero-crossing value between the left and right images distinctly for edge, and nonedge blocks. And finally, we update these blockiness and zero-crossing values by some weighting factors that are optimized by an optimization algorithm. The mathematical features, blockiness and zero-crossing measures within each block of the images are calculated horizontally and then vertically.

For horizontal direction: let the test image signal be $x(m, n)$ for $m \in [1, M]$ and $n \in [1, N]$, a differencing signal along each horizontal line is calculated by

$$
\begin{aligned}
& d_h(m, n) \\
& = x(m, n+1) - x(m, n), \quad n \in [1, N-1], \ m \in [1, M].
\end{aligned}
\tag{1}
$$

Blockiness of a block (8×8) in horizontal direction is estimated by

$$
B_{bh} = \frac{1}{8} \sum_{i=1}^{8} |d_h(i, 8j)|,
\tag{2}
$$

where "i" and "$8j$" are, respectively, number of row and column position, and $j = 1, 2, 3, \ldots (N/8)$.

For horizontal zero-crossing (ZC) we have

$$
d_{h\text{-sign}}(m, n) = \begin{cases} 1 & \text{if } d_h(m, n) > 0, \\ -1 & \text{if } d_h(m, n) < 0, \\ 0 & \text{otherwise,} \end{cases}
\tag{3}
$$

$$
d_{h\text{-mul}}(m, n) = d_{h\text{-sign}}(m, n) \times d_{h\text{-sign}}(m, n+1).
$$

We define for $n \in [1, N-2]$:

$$
z_h(m, n) = \begin{cases} 1 & \text{if } d_{h\text{-mul}}(m, n) < 0, \\ 0 & \text{otherwise,} \end{cases}
\tag{4}
$$

where the size of $z_h(m, n)$ is $M \times (N-2)$. The horizontal zero-crossing of a block (8×8), ZC_{bh}, is calculated as follows:

$$
ZC_{bh} = \sum_{i=1}^{8} \sum_{j=1}^{8} z_h(i, j),
\tag{5}
$$

Thus, we can calculate blockiness and zero-crossing of each available block of the left and right images.

For vertical direction: we can also calculate the differences of signal along each vertical line as follows:

$$
\begin{aligned}
& d_v(m, n) \\
& = x(m+1, n) - x(m, n), \quad n \in [1, N], \ m \in [1, M-1].
\end{aligned}
\tag{6}
$$

Similarly, the vertical features of blockiness (B_{bv}) and zero-crossing (ZC_{bv}) of the block are calculated. Therefore, the overall features B_b and ZC_b per block are given by

$$
B_b = \frac{B_{bh} + B_{bv}}{2}, \quad ZC_b = \frac{ZC_{bh} + ZC_{bv}}{2}.
\tag{7}
$$

Consequently, the average blockiness value of edge, and nonedge areas of the left image are calculated by:

$$
\begin{aligned}
Bl_e &= \frac{1}{N_e} \sum_{b=1}^{N_e} B_{be}, \\
Bl_n &= \frac{1}{N_n} \sum_{b=1}^{N_n} B_{bn},
\end{aligned}
\tag{8}
$$

where N_e and N_n are, respectively, the number of edge, and nonedge blocks of the image. Similarly, the average blockiness values of Br_e, and Br_n for the right image are calculated.
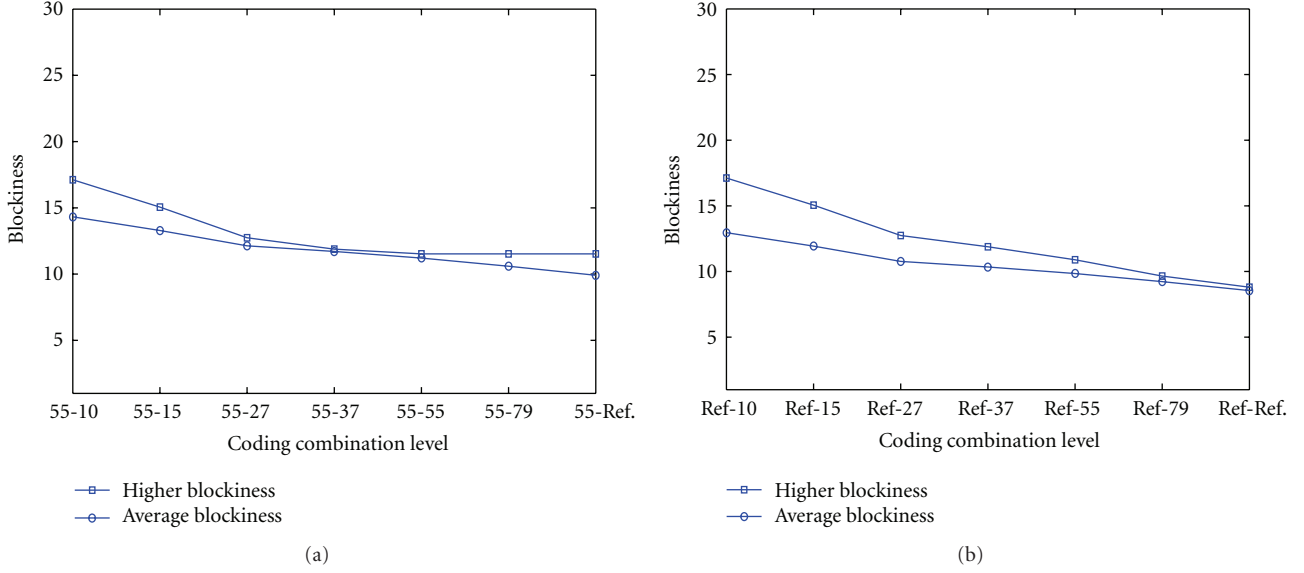
FIGURE 3: Blockiness versus different coding levels for Cattle image pairs (a) L, R: QS55-Seven different coding levels. (b) L, R: Ref-Seven different coding levels.

Accordingly, the average zero-crossing values of ZCl_e, and ZCl_n for the left image are estimated by

$$ZCl_e = \frac{1}{N_e} \sum_{b=1}^{N_e} ZC_{be},$$

$$ZCl_n = \frac{1}{N_n} \sum_{b=1}^{N_n} ZC_{bn}. \quad (9)$$

Similarly, the average zero-crossing values of ZCr_e, and ZCr_n for the right image are calculated. We then calculate the total blockiness and zero-crossing features of edge, and nonedge areas of the stereo image. For the total blockiness features (B_e and B_n) of the stereo image, we consider only the higher values between the left and right images by the following algorithm:

$$B_{e/n}(Bl, Br) = \max(Bl, Br). \quad (10)$$

However for zero-crossing features (ZC_e, and ZC_n), we estimate lower values between the left and right images by the following algorithm:

$$ZC_{e/n}(ZCl, ZCr) = \min(ZCl, ZCr). \quad (11)$$

Finally, the overall blockiness, and zero-crossing of each stereo image pair are calculated by

$$B = B_e^{w_1} \cdot B_n^{w_2}$$
$$Z = ZC_e^{w_3} \cdot ZC_n^{w_4}, \quad (12)$$

where w_1 and w_2 are the weighting factors for the blockiness of edge, and nonedge areas and also w_3 and w_4 are the weighting factors for zero-crossing.

3.1.1. Significance of Considering the Maximum Blockiness of a Stereo Pair.

In this section, we discuss the reason for choosing the maximum blockiness of a stereo pair for our model. The goal is to measure the maximum possible blockiness within a stereo pair so that of the metric can correlate well with human viewers' perception without actual human. Because, blockiness is one of the most annoying artifacts for human eyes. Moreover, the model is developed both for symmetric and asymmetric images. In order to take into count the highest degradation, we consider the maximum blockiness between the left and the right views. To explain the consideration of the maximum blockiness, we took a stereo image "Cattle" (the image from the MICT database [26]). The coding levels versus blockiness of the stereo image are shown in Figure 3. We examine both the highest and average blockiness between the two views. Figure 3 shows variations of blockiness with the increasing of bit rate. The results indicate that the blockiness variation is higher in case of highest of blockiness compared to the average blockiness for increasing of bit rate. The normalized MOS (NMOS) versus blockiness (N-blockiness) with increasing bit rate for two types of stereo images is shown in Figure 4. The coding levels (L, R: Ref-10, Ref-15, Ref-27, Ref-37, Ref-55, Ref-79, and Ref-Ref), and (L, R: 79-10, 79-15, 79-27, 79-37, 79-55, 79-79, and 79-Ref) in the Figure 4 indicate increasing bit rate. Although NMOS scores show an increasing trend with decreasing N-blockiness, the consideration of maximum blockiness (Higher-B) correlates inversely better with NMOS compared to average blockiness (Average-B). Here, NMOS versus the maximum N-blockiness features for edge (i.e, non-plane) and nonedge (i.e., plane) areas along with a wide variety of quality pairs for Car and Cattle images are also shown in Figure 5. The two blockiness features (B_e and B_n) support the similar trend of inverse nature with respect to NMOS. Therefore, the above results suggest that the consideration of the maximum blockiness with the two blockiness features is
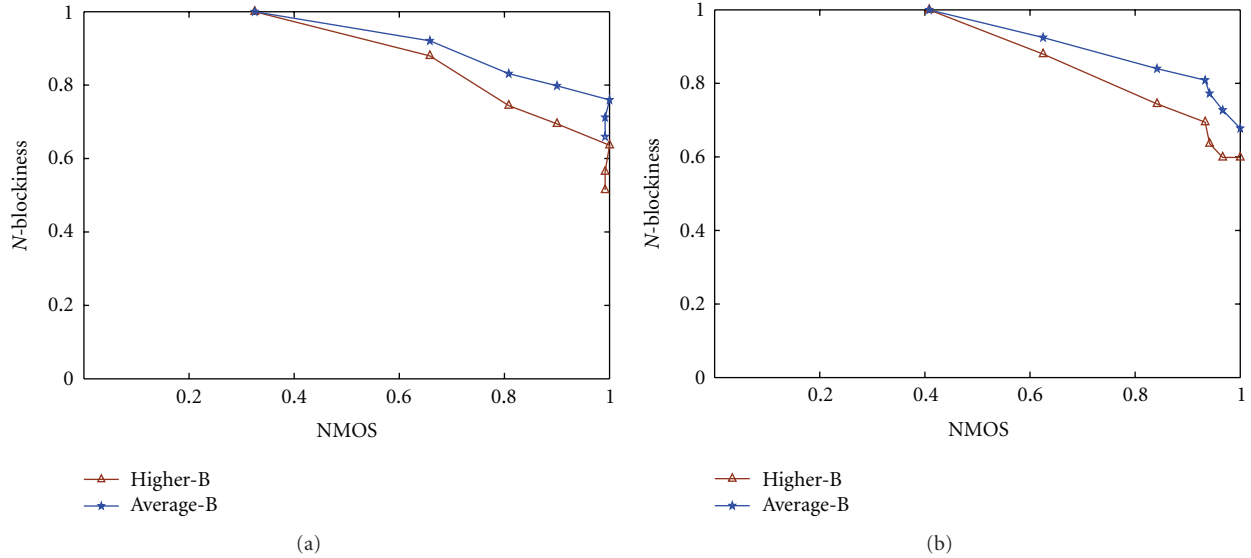
(a)

(b)

FIGURE 4: Normalized MOS versus blockiness for different Cattle image pairs. (a) L, R: Ref-QS10, Ref-QS15, Ref-QS27, Ref-QS37, Ref-QS55, Ref-QS79, and Ref-Ref, (b) L, R: QS79-QS10, QS79-QS15, QS79-QS27, QS79-QS37, QS79- QS55, QS79-QS79, and QS79-Ref.
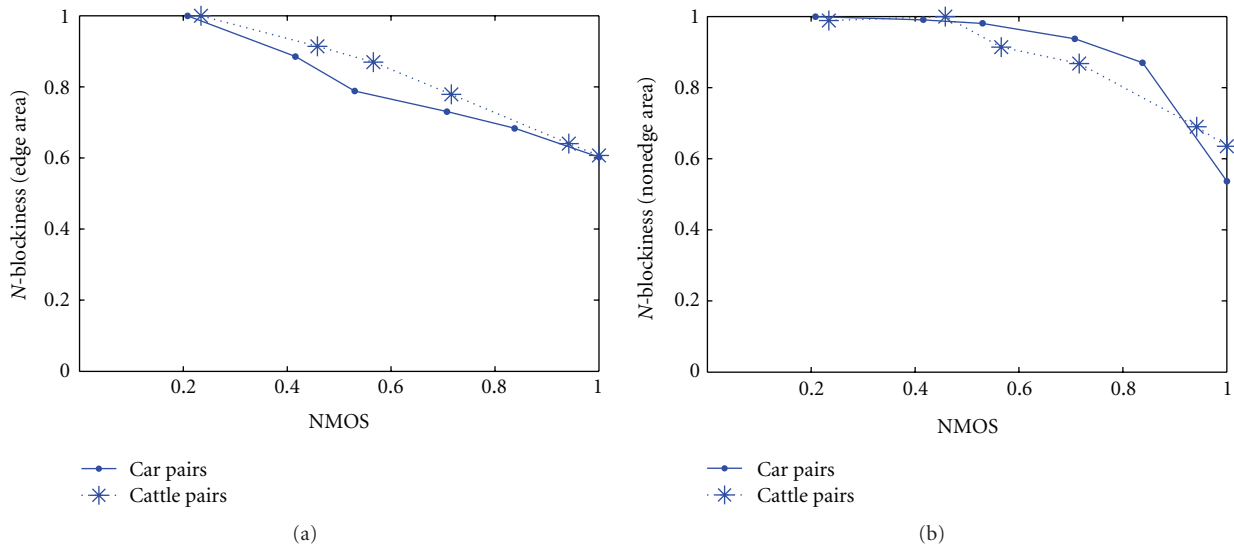

(a)

(b)

FIGURE 5: Normalized MOS versus blockiness for edge and nonedge areas of different Car and Cattle image pairs. (a) NMOS versus N-blockiness (edge areas, B_e), (b) NMOS versus N-blockiness (nonedge areas, B_n).

more justified than the average blockiness for developing of an objective model.

3.1.2. Significance of Considering the Minimum Zero-Crossing of a Stereo Pair. An analysis of choosing the minimum zero-crossing value between the left and the right views of a stereo pair is given in this section. In [29], it has been discussed that the average edge point detection within image blocks gives better insight of edge distortion measurement within an image. Consequently, the zero-crossing values show a decreasing (i.e., increasing edge distortion) trend with the increasing compression level. Therefore, there is a relationship with the transition of zero-crossing and the overall edge distortion within an image. In order to study the

relationship, we take a stereo image pair, Cattle. Normalized MOS (NMOS) versus zero-crossing (N-zero crossing) of the stereo image is shown in Figure 6. We consider both the minimum (Lower-ZC) and the average zero-crossing (Average-ZC) value of the stereo pair. The Figure 6 shows that the minimum zero-crossing measure is correlated better to the NMOS score compared to that of the average zero-crossing. In addition, the N-zero crossing values show an increasing trend for increasing bit rate. Subsequently, the NMOS versus the minimum N-Zero crossing features for edge and nonedge areas over a variety of quality pairs for Car and Cattle images are shown in Figure 7. The two zero crossing features follow the similar trend of correlation with respect to NMOS. Therefore, the results indicate that
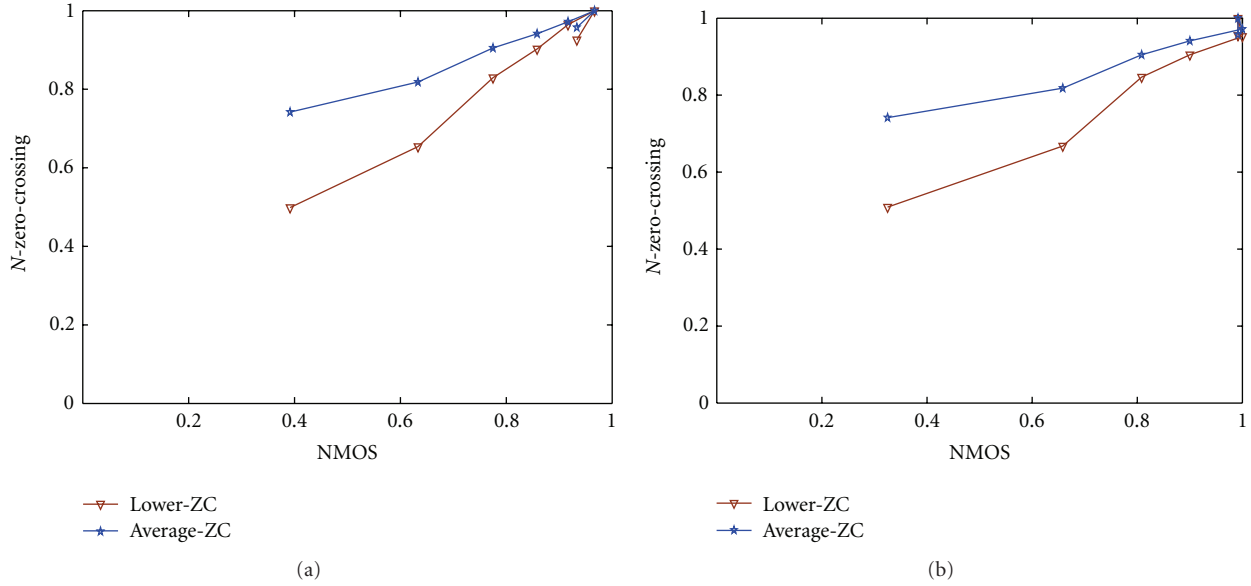
(a)

(b)

FIGURE 6: Normalized MOS versus zero-crossing for different Cattle image pairs. (a) L, R: QS55-QS10, QS55-QS15, QS55-QS27, QS55-QS37, QS55-QS55, QS55-QS79, and QS55-Ref, (b) L, R: Ref-QS10, Ref-QS15, Ref-QS27, Ref-QS37, Ref-QS55, Ref-QS79, and Ref-Ref.
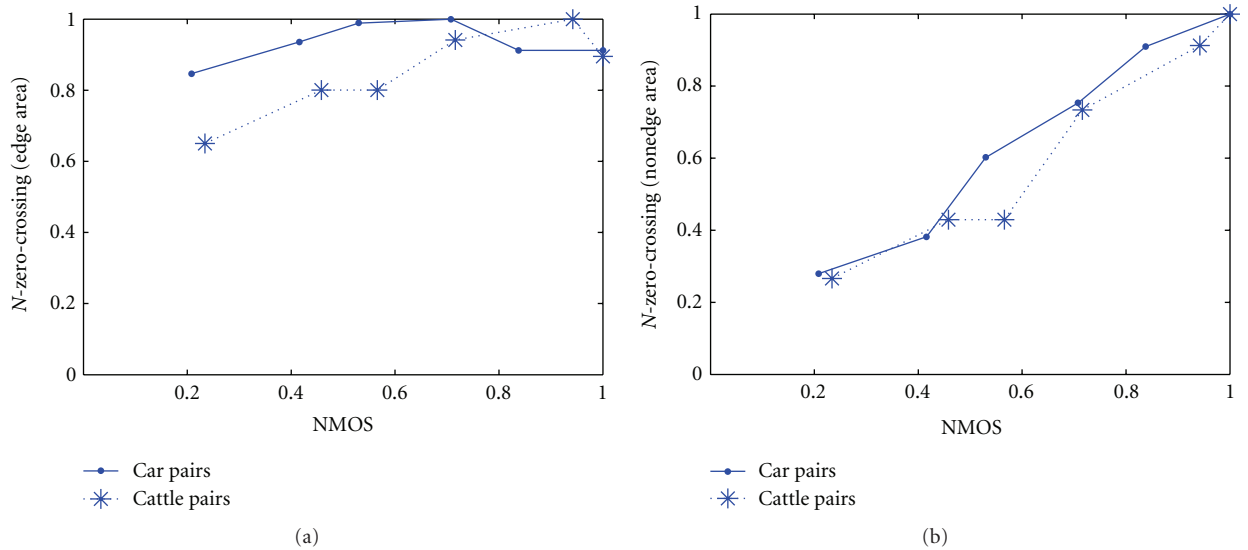

(a)

(b)

FIGURE 7: Normalized MOS versus zero-crossing for edge and nonedge areas of different Car and Cattle image pairs. (a) NMOS versus N-Zero crossing (edge areas, ZC_e), (b) NMOS versus N-Zero crossing (nonedge areas, ZC_n).

the two zero crossing features (ZC_e and ZC_n) measures along with the minimum zero-crossing are more justified than the average zero-crossing to develop the quality prediction metric.

3.2. Relative Disparity Measure. To measure disparity, we use a simple feature-based block matching approach. Many feature-based approaches are applied for stereo matching/disparity estimation which are discussed in [30]. Here, a fixed block based difference zero-crossing (DZC) approach is employed in this work. The principal of the disparity estimation is to divide the left image into nonoverlapping 8×8 blocks with classification of edge and nonedge blocks.

For each 8×8 block of the left image, stereo correspondence searching is conducted based on minimum difference zero-crossing (MDZC) rate between the corresponding block and up to ± 128 pixels of the right image. The disparity estimation approach is shown in Figure 8. Here, zero-crossing (horizontal and vertical) of a block is estimated according to Section 3.1. "1," and "0" indicate zero-crossing (edge) and nonzero-crossing (nonedge) points, respectively. In order to reduce computational cost, we restricted the correspondence search to 1D only (i.e., horizontally) and within ± 128 pixels. Moreover, the stereoscopic images database that we consider in this research are epipolar rectified images. Therefore, the displacement between the left and right view of a stereo
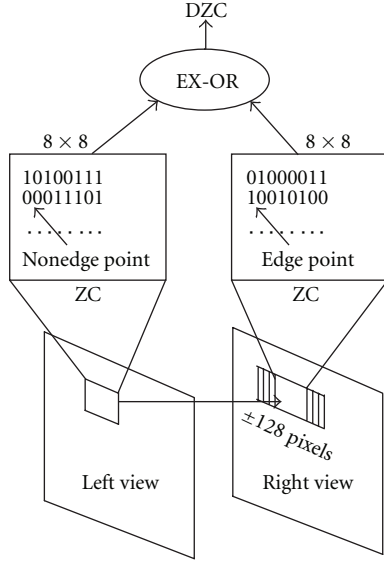
FIGURE 8: Disparity estimation approach.

pair is restricted in horizontal direction only. The depth maps of the two sample stereo image pairs for block size 4×4, 8×8, and 16×16 with searching area ± 128 pixels are shown in Figure 9. Colors in the depth maps that are indicated by vertical color bars in right are estimated depths of the image pairs. Subsequently, depth maps of different symmetric and asymmetric Cattle images are shown in Figure 10. Figures 9 and 10 show that the performance of the disparity algorithm is adequate for the block size 8×8 with searching areas of ± 128 pixels. The effect of different block size and searching areas on this disparity estimation are discussed in details in [29]. Although disparity is a measure of position displacement between the left and right images, an intensity based DZC rate is determined between the block of a left image and the corresponding searching block in the right image as relative disparity in this work.

In order to measure the relative disparity, firstly, the segmentation algorithm is applied to left image only to classify edge and nonedge blocks. Secondly, block-based DZC is estimated in the two corresponding blocks between the left and right images. Thirdly, we average the DZC rate values separately for edge and nonedge blocks. Finally, the values are updated with some weighting factors. If ZCl, and ZCr be the zero-crossing of a block of left image and the corresponding searching block of right image, respectively. The DZC of the block can be estimated by the following equation:

$$\text{DZC} = \text{ZCl} \oplus \text{ZCr}, \tag{13}$$

where the symbol, "$\oplus$" indicates a logical Exclusive-OR operation. Subsequently, DZC rate (DZCR) is calculated by

$$\text{DZCR} = \frac{1}{8 \times 8} \sum \text{DZC}. \tag{14}$$

For horizontal direction: let ZCl_h, and ZCr_h be the zero-crossing of a block of left image and the corresponding

searching block of right image in horizontal direction, respectively. The DZC_h of the block are estimated by the following equation:

$$\text{DZC}_h = \text{ZCl}_h \oplus \text{ZCr}_h, \tag{15}$$

Thus, we can calculate DZC_h rate (DZCR_h) of the 8×8 block by

$$\text{DZCR}_h = \frac{1}{8 \times 8} \sum \text{DZC}_h. \tag{16}$$

Therefore, the average DZCR_h (AZC_h) for edge, and nonedge blocks of the left image are calculated by

$$\text{AZC}_{h_e} = \frac{1}{N_e} \sum_{e=1}^{N_e} \text{DZCR}_{h_e}, \tag{17}$$

$$\text{AZC}_{h_n} = \frac{1}{N_n} \sum_{e=1}^{N_n} \text{DZCR}_{h_n}, \tag{18}$$

where N_e and N_n are, respectively, the number of edge, and nonedge blocks of the left image.

For vertical direction: similarly, we can calculate AZC_{v_e} and AZC_{v_n}. Subsequently, the total relative disparity features for edge, AZC_e and nonedge, AZC_n areas are estimated by the following equation:

$$\text{AZC}_e = \frac{\text{AZC}_{h_e} + \text{AZC}_{v_e}}{2}, \qquad \text{AZC}_n = \frac{\text{AZC}_{h_n} + \text{AZC}_{v_n}}{2}. \tag{19}$$

Finally, the overall relative disparity feature is estimated by

$$\text{DZ} = \text{AZC}_e^{w_5} \cdot \text{AZC}_n^{w_6} \tag{20}$$

where w_5 and w_6 are, respectively, the weighting factors of the disparity features for edge, and nonedge areas. In order to verify the estimation of the two disparity features (AZC_e and AZC_n) the normalized MOS versus the disparity features for edge and nonedge areas over the different quality pairs for Car and Cattle images are shown in Figure 11. The two disparity features also maintained the similar increasing trend of correlation nature with respect to NMOS. Therefore, it is indicated that the two disparity features measures are also justified to develop the prediction metric. Although 3D depth perception is a complex process, we believe it has a strong correlation with objects/structural information of a scene content that is near to the viewers. In order to verify this statement, we compare three stereoscopic images of similar scene contents and noticed that the distance of the near objects/structures to the viewers in second and third images is decreasing in comparison with the first image that is shown in Figure 12. Consequently, the depth perceptions are increasing from the images one to third according to the viewer's perception. Eventually, the proposed disparity feature (DZ) measure is shown in Figure 13 for edge and nonedge areas within the images. The figure shows the normalized DZ features for the two different areas of
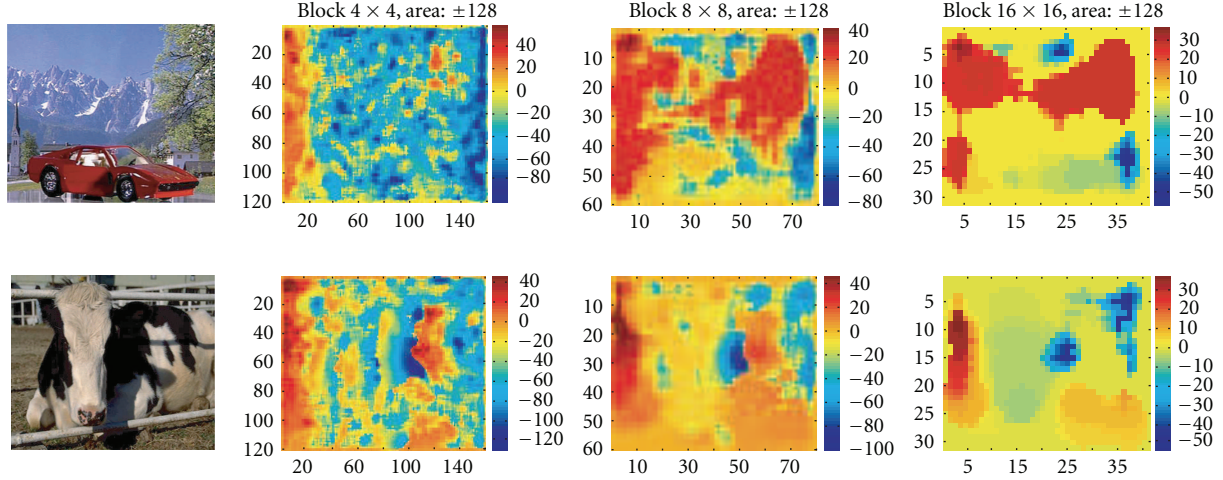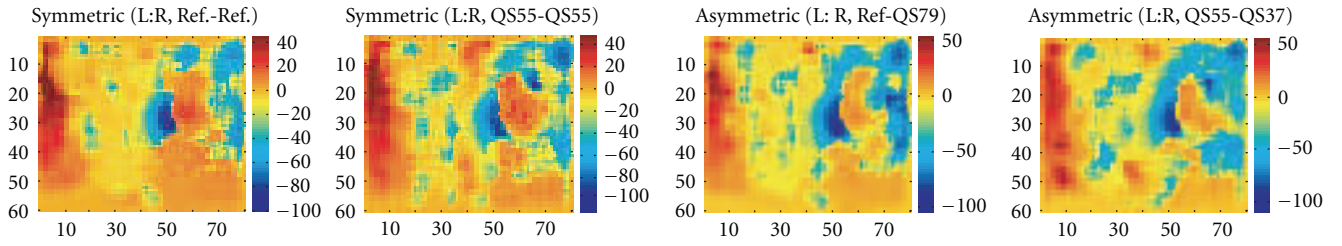
FIGURE 9: Car and Cattle images and its depth maps with different block sizes and searching area ±128.

FIGURE 10: Depth maps of distinct symmetric and asymmetric pairs for Cattle images.

the images. The DZ values for edge areas in Figure 13(a) indicate that the first image's depth is lower than the second and similarly, the DZ value of second image is lower than the third image. Therefore, the increasing trend of DZ features for edge areas on similar scene contents confirms the human visual depth perception of the images. Although the DZ features for edge areas support the depth perception, we also consider the DZ features for nonedge areas to measure the relative depth perception of other objects/structures of scene contents in this algorithm.

3.3. Features Combination. We can combine the artifacts and disparity features to develop a stereo quality assessment metric in different way. In order to investigate the best suitable features combination equation, we studied the following equations:

Case 1.

$$S = \alpha(DZ) \cdot B \cdot Z \qquad (21)$$

Case 2.

$$S = \alpha + \beta(DZ) \cdot B \cdot Z \qquad (22)$$

Case 3.

$$S = \alpha(DZ) + \beta(B) + \gamma(Z) \qquad (23)$$

Case 4.

$$S = \alpha(DZ) + \beta B \cdot Z, \qquad (24)$$

where α, β, and γ are the method parameters. The method parameters and weighting factors (w_1 to w_6) are must be estimated by an optimization algorithm with the subjective test data. The proposed method performance is also studied without disparity by the following equation:

$$S = \alpha + \beta B \cdot Z. \qquad (25)$$

We consider a logistic function as the nonlinearity property between the human perception and the physical features. Finally, the obtained MOS prediction, MOS_p, is derived by the following equation [31]:

$$MOS_p = \frac{4}{1 + \exp[-1.0217(S - 3)]} + 1. \qquad (26)$$

Here, Particle Swarm Optimization (PSO) algorithm is used for optimization [32].

4. Results

In order to verify the performance of our method we consider the MICT stereo image database (see Section 2). To use the database, we divide the database into two parts for training and testing. The training database consists of five randomly selected reference stereo pairs (from the total ten) and all of their different combinations of symmetric/asymmetric coded stereo images (245 stereo pairs). The testing database consists of the other five reference stereo pairs and their symmetric/asymmetric coded versions (245 stereo pairs), and also there is no overlapping
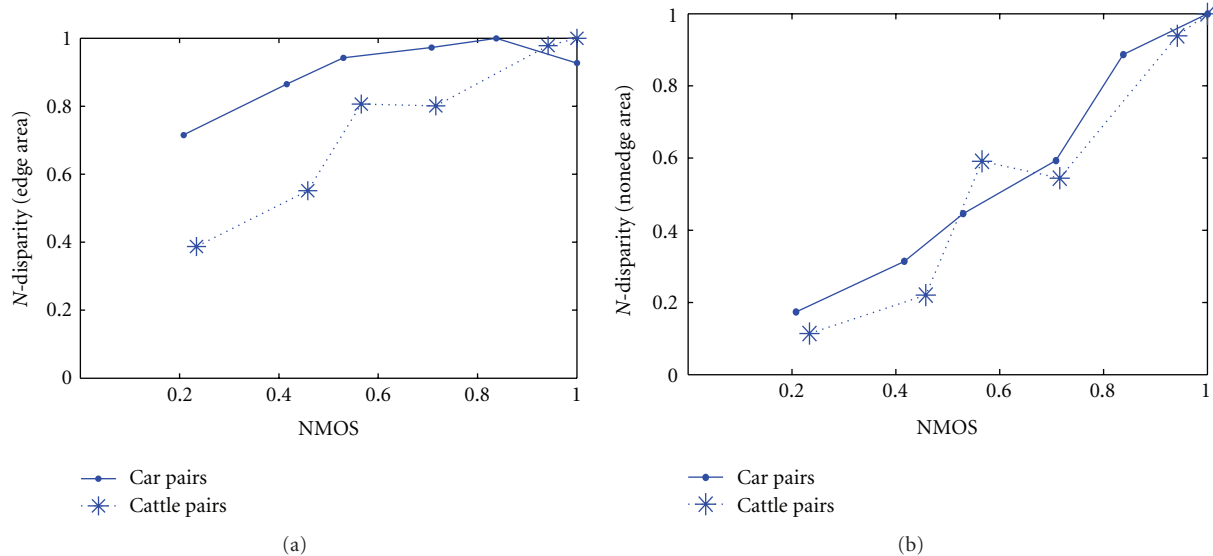
(a)

(b)

FIGURE 11: Normalized MOS versus Disparity for edge and nonedge areas of different Car and Cattle image pairs: (a) NMOS versus N-Disparity (edge areas, AZC_e) (b) NMOS versus N-Disparity (nonedge areas, AZC_n).

(I)   (II)   (III)

FIGURE 12: Images with different depth perception based on variation of near objects' distance: Decreasing distance of near objects (from images I to III) creates a significant increase of depth perception.

between training and testing. In order to provide quantitative measures on the performance of the proposed method, we follow the standard performance evaluation procedures employed in the video quality experts group (VQEG) FR-TV Phase II test [33], where mainly pearson linear correlation coefficient (CC), average absolute prediction error (AAE), root mean square prediction error (RMSE), and outlier ratio (OR) between objective (predicted), and subjective scores were used for evaluation. The evaluation result along with all above mentioned features combination equations are shown in Table 1. The table indicates that out of all the combined equations, (24) (Case 4) provides the highest prediction performance among others. Consequently, the proposed method considers (24). The method's parameters and weighting factors are obtained by the PSO optimization algorithm with all of the training images are shown in Table 2. To measure the performance as well as justification of the estimated image features of our proposed method we also consider the following prediction performances:

(1) Methods with disparity:

  (i) proposed model (i.e., considering blockiness, zero-crossing, and disparity) using the features combining Equation (24);

TABLE 1: Evaluation results on different features combined equations with disparity.

Methods	Training			
	CC	AAE	RMSE	OR
Case 1	0.916	0.907	1.044	0.298
Case 2	0.953	0.332	0.401	0.086
Case 3	0.961	0.286	0.348	0.065
Case 4 (considered)	0.964	0.276	0.336	0.061
	Testing			
Case 1	0.865	0.918	1.064	0.314
Case 2	0.942	0.339	0.406	0.053
Case 3	0.931	0.348	0.431	0.037
Case 4 (considered)	0.940	0.339	0.413	0.040

TABLE 2: Method parameters and weighting factors (MOS scale, 1 to 5).

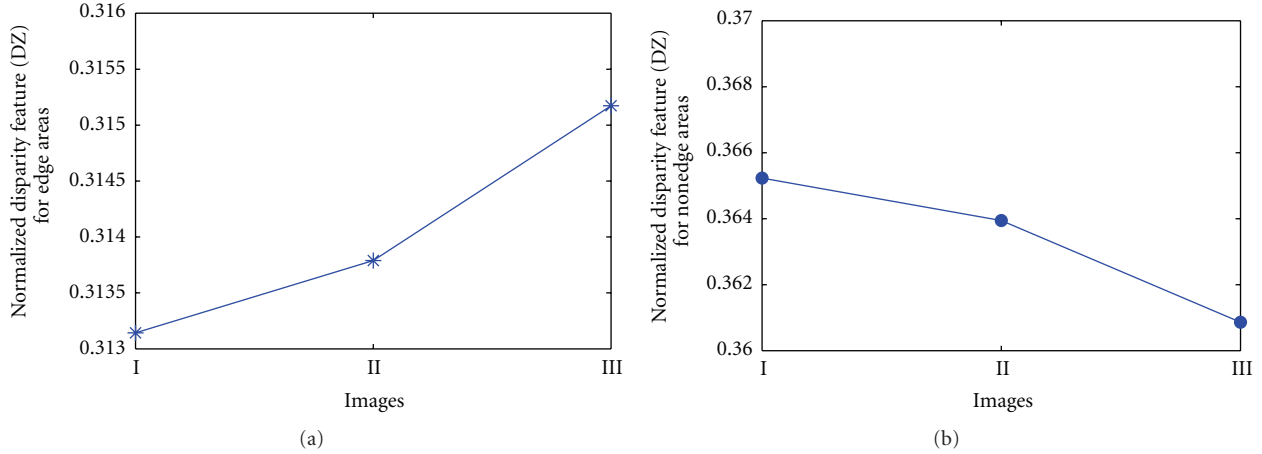|		
$\alpha = 58.064452$	$\beta = -51.026118$	
$w_1 = 0.036062$	$w_2 = 0.00513$	$w_3 = 0.010634$
$w_4 = -0.026979$	$w_5 = -0.017522$	$w_6 = 0.013169$

FIGURE 13: Normalized disparity of the images (I, II, and III) for edge and nonedge areas. (a) Edge areas, (b) nonedge areas.

(ii) method considering only blockiness and disparity using the following features combined equation:

$$S = \alpha(DZ) + \beta B. \qquad (27)$$

(iii) method considering only zero-crossing and disparity using the following features combined equation:

$$S = \alpha(DZ) + \beta Z. \qquad (28)$$

(iv) conventional method with disparity (i.e., consider blockiness, zero-crossing, and disparity without segmentation) using the features combining Equation (24).

(2) Methods without disparity:

(i) method considering blockiness, and zero-crossing using the features combine Equation (25).

(ii) method considering only blockiness by using the following equation:

$$S = \alpha + \beta B. \qquad (29)$$

(iii) method considering only zero-crossing using the following equation:

$$S = \alpha + \beta Z. \qquad (30)$$

(iv) conventional method considering blockiness, and zero-crossing using (25) without segmentation.

(3) Another method:

(i) method considering the blockiness and zero-crossing distinctly for the two views of a stereo pair and measure the quality score of the left

and the right views independently using the features combining Equation (25), and average them without disparity, "2D quality mean" [18].

The evaluation results of all the above mentioned methods are summarized in Tables 3, 4, and 5. Table 3 shows that the proposed method's performances for every one of the evaluation metrics are quite sufficient both for the training and the testing datasets. It has also been observed from the Table 3 that the proposed method provides sufficient prediction accuracy (higher CC), and sufficient prediction consistency (lower OR). The result in Table 3 also prove that the proposed method (i.e., incorporation of the perceptual difference of image distortions and disparity) demonstrates superior quality prediction performance compare to the conventional method with disparity. Tables 3 and 4 also show that the method performances are superior compared to the without disparity. Whereas, 2D quality mean performance is not sufficient even compared to without disparity approach (i.e., considering only blockiness and zero-crossing) (see Tables 4 and 5). Although, the incorporation of disparities measure to the FR stereo image quality assessment method [9] indicates poor results, our proposed method (with relative disparity) indicates better result compared to without disparity (i.e., considering only blockiness and zero-crossing). It is clear from Tables 3 and 4 that all methods performances with disparity are superior compared to without disparity. Therefore, the relative disparity measure which is considered in our proposed method can be a significant measure for 3D quality prediction. In order to understand the significance of estimated image features (i.e., blockiness and zero-crossing), we consider the above mentioned methods which used both features, blockiness and zero-crossing individually with and without disparity. It is clear from Tables 3 and 4 that the performance of the method considering only zero-crossing is better compared to the method considering only blockiness both for with and without disparity. Therefore, zero-crossing feature is more significant compared to blockiness feature for quality prediction. Proposed method's weighting factors also show

Table 3: Methods' evaluation results for training and testing with disparity.

Methods	Training			
	CC	AAE	RMSE	OR
Proposed method	0.964	0.276	0.336	0.061
Only blockiness with disp.	0.867	0.529	0.664	0.086
Only zero-crossing with disp.	0.897	0.484	0.578	0.110
Conventional method with disp.	0.903	0.479	0.597	0.114
	Testing			
Proposed method	0.940	0.339	0.413	0.040
Only blockiness with disp.	0.833	0.537	0.724	0.106
Only zero-crossing with disp.	0.804	0.578	0.707	0.102
Conventional method with disp.	0.881	0.473	0.577	0.127

Table 4: Methods' evaluation results for training and testing without (wo) disparity.

Methods	Training			
	CC	AAE	RMSE	OR
Blockiness and zero-crossing wo disp.	0.953	0.322	0.401	0.074
Only blockiness wo disp.	0.705	1.037	1.199	0.367
Only zero-crossing wo disp.	0.883	0.515	0.610	0.110
Conventional method wo disp.	0.904	0.530	0.646	0.131
	Testing			
Blockiness and zero-crossing wo disp.	0.932	0.349	0.432	0.053
Only blockiness wo disp.	0.705	1.041	1.196	0.359
Only zero-crossing wo disp.	0.814	0.595	0.719	0.114
Conventional method wo disp.	0.854	0.548	0.649	0.159

Table 5: Another method's evaluation results for training and testing.

Method	Training			
	CC	AAE	RMSE	OR
2D quality mean	0.912	0.432	0.55	0.078
	Testing			
2D quality mean	0.89	0.40	0.534	0.057

the deviance. Weighting factors (w_3 and w_4) of zero-crossing are higher compared to weighting factors (w_1 and w_2) of blockiness (see Table 2).

The MOS versus MOS_p of our proposed method for training and testing images are respectively shown in Figures 14(a), and 14(b). The symbols "∗" and "+," respectively, indicate MOS_p points for the databases of training and testing. Figure 14 confirms that the proposed method's overall quality prediction performance is sufficient not only on known dataset but also on unknown dataset. The MOS versus MOS_p performance of the proposed method is also shown in Figure 15 distinctly for symmetric and asymmetric images. Figure 15 shows that the overall prediction performance is almost equally well for both symmetric and asymmetric coded pairs. However, the performance

Table 6: Evaluation results comparison on MICT database.

Method	Training			
	CC	AAE	RMSE	OR
Proposed, NR	0.964	0.276	0.336	0.061
Method, NR [23]	0.966	0.292	0.367	0.069
Method, FR [9]	0.945	0.310	0.381	0.065
2D quality mean, FR [34]	0.779	0.715	0.846	0.261
	Testing			
Proposed, NR	0.940	0.339	0.393	0.037
Method, NR [23]	0.935	0.350	0.421	0.065
Method, FR [9]	0.929	0.370	0.441	0.082
2D quality mean, FR [34]	0.758	0.722	0.844	0.208

trend is slightly inferior for symmetric pairs compared to asymmetric pairs. Because, the proposed method takes into account the highest visual artifacts between the two views. Subsequently, the highest visual artifacts measures are not significant in those symmetric pairs who are very low levels of compression or close to reference pairs. The MOS_p points "∗" and the error bars of ±2 standard deviation intervals of four different stereo images are shown in Figure 16. Error bars show the ±2 standard deviation interval of the MOS. The figure indicates the predictions consistently performed well in almost similar nature on variety of image contents. Although, the incorporation of disparities measure to the FR stereo image quality metrics [9] indicate poor results, our method with the relative disparity indicates better results compared to without disparity. Therefore, the local features-based relative disparity and distortions can be a significant measure for overall stereoscopic image quality prediction. In order to estimate computational cost of the proposed algorithm, we calculate the computing time of the algorithm on an Intel (R) Core (TM) i3 processor with 2.53 GHz clock speed and 2 GB RAM accompanied with Windows 32-bit operating system. Figure 17 shows the average computing time of stereo images with different resolutions. The average computational cost, specifically for 640 × 480 pixels stereo image, of our proposed algorithm is approximately 52 sec which is sufficient to perform the computation on the machine configuration.

5. Performance Comparison

In this section, we compare the performance of my proposed method against our recently published NR model [23]. The method uses three local features (edge, flat, and texture) and the MICT database. Our proposed method's evaluation results on the same database are shown in Table 6. The table shows that the performance of our proposed method is superior compared to the published method both for the training and testing databases. As a comparison, we can also compare the performance of my proposed method against the currently published FR method presented in [9]. We evaluate the performance of the method on the same database (MICT database). Table 6 shows that the performance of our proposed model is better even compared
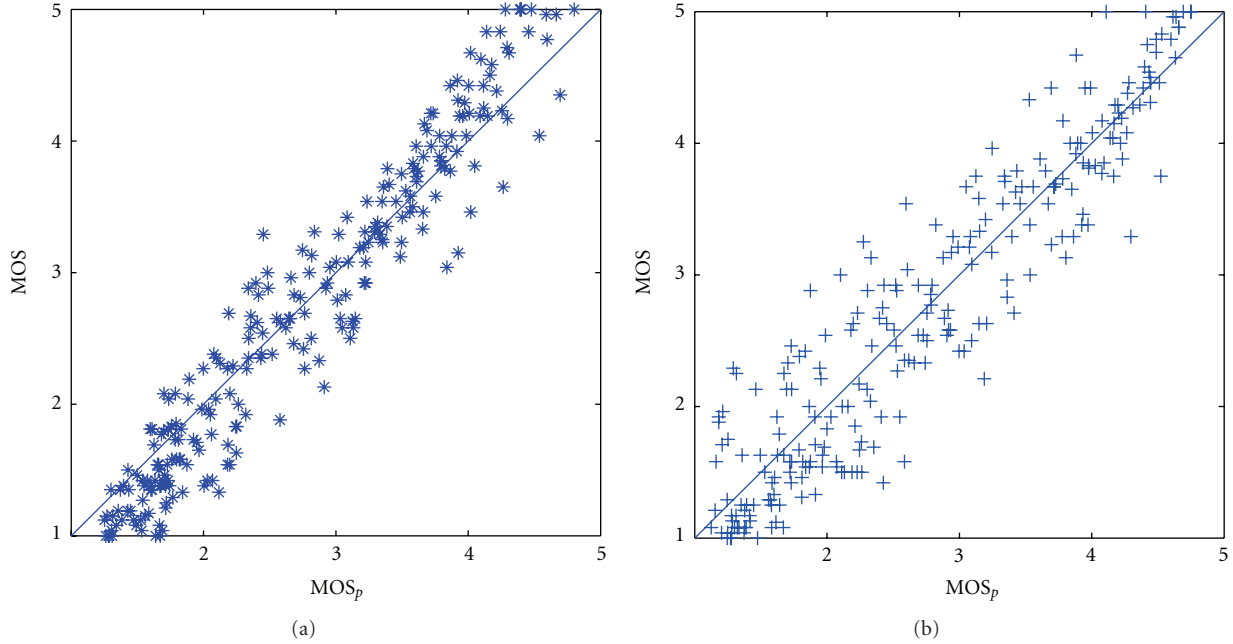
FIGURE 14: Proposed method's MOS versus MOS$_p$. (a) Training, (b) Testing.

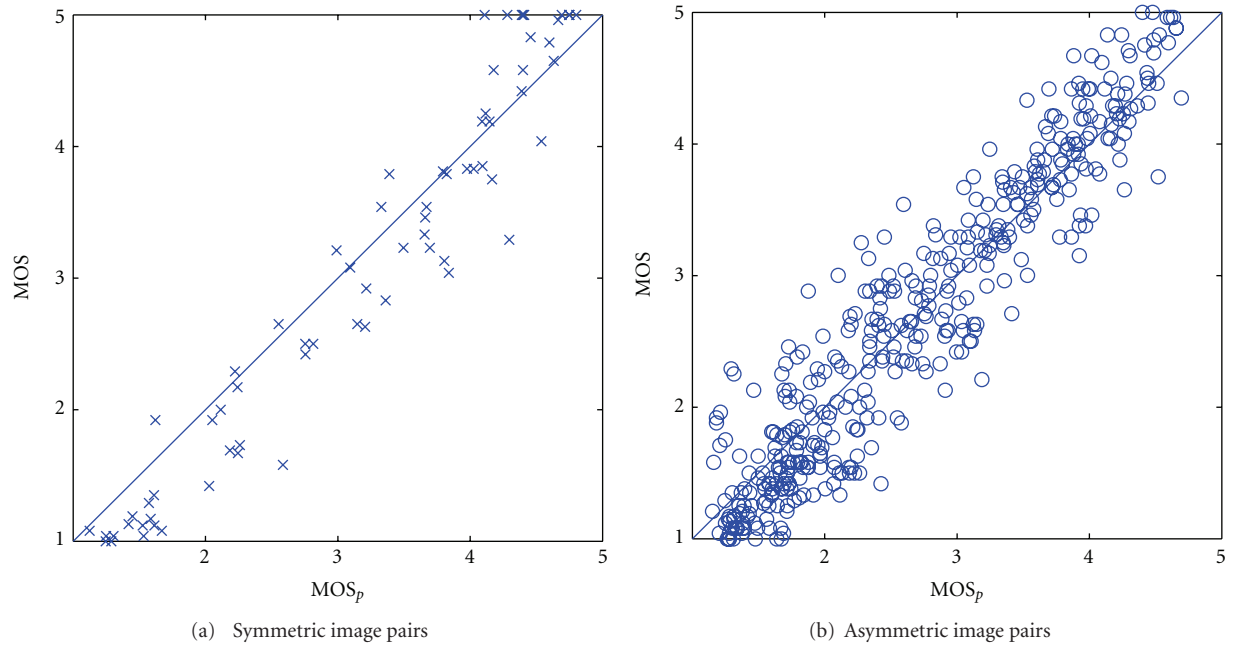(a) Symmetric image pairs

(b) Asymmetric image pairs

FIGURE 15: Proposed method's MOS versus MOS$_p$ for symmetric and asymmetric coded images.

to the FR method [9]. We want to make another comparison according to the idea of some researches. Some researchers claim 2D image quality metric can be used for 3D or stereoscopic image quality prediction by averaging the 2D quality metric for the left and the right views without the disparity features estimation [18]. We want to point out simple 2D averaging technique is not suitable for stereoscopic image quality prediction even if a good quality 2D FR quality metric is used for quality prediction. According to this idea, we compare the performance of our proposed method against

the popular FR objective method for 2D quality assessment [34]. We also evaluate the performance of the method on the same database. Table 6 shows that the performance of our proposed model is more better compared to the averaging method of 2D quality. It is apparent from this result that the 2D quality mean approach is not enough for 3D quality prediction. The proposed method's performance can also be compared with another recently published FR stereo image quality assessment [11]. The method is also used the same MICT database. The FR method's reported CC on the MICT
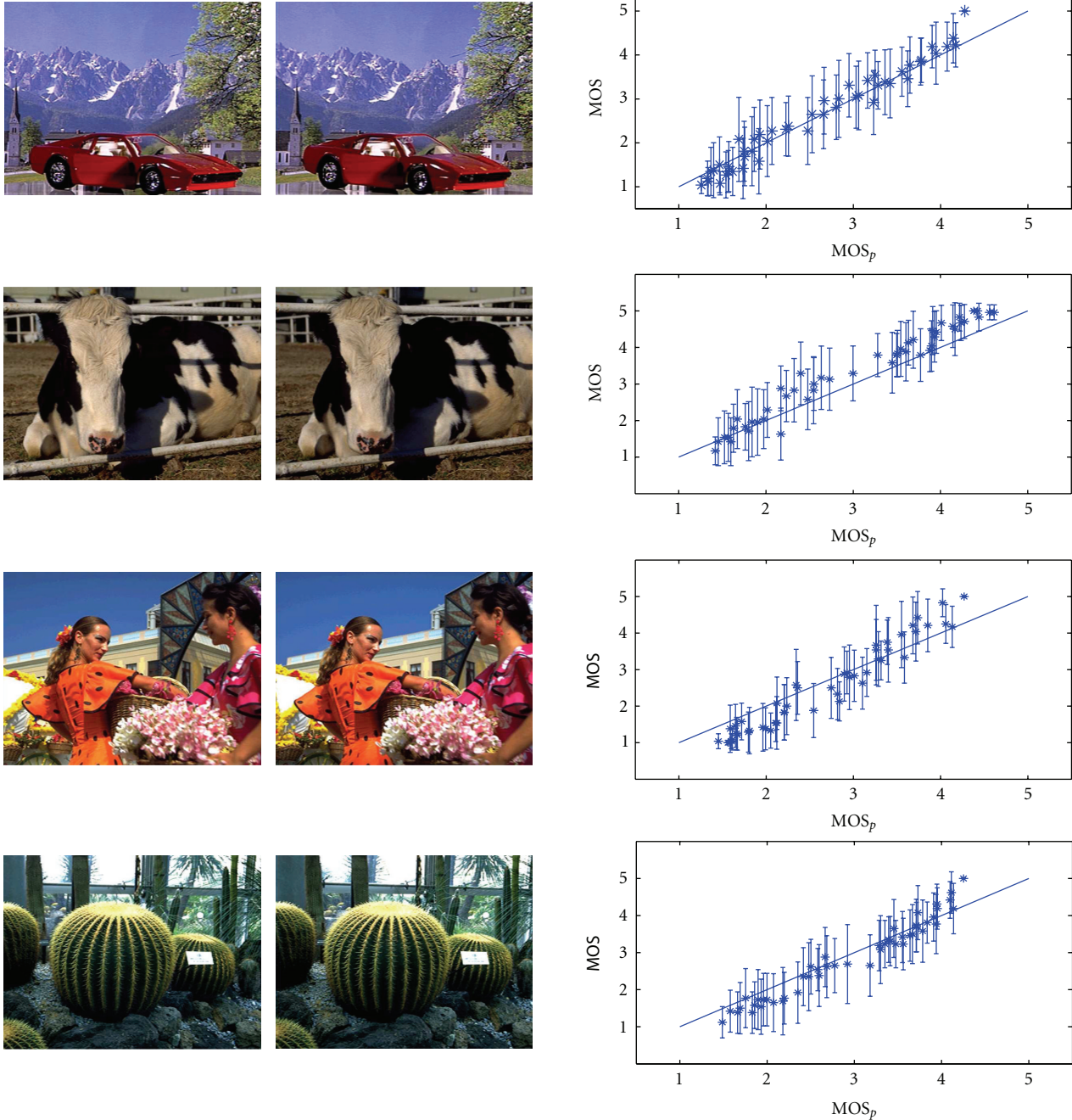
FIGURE 16: The MOS_p performances on texture variety of stereo pairs over the quality range. The predictions points ∗ and ±2 standard deviation intervals are shown for each stereo pair.

database is 0.97, whereas our proposed NR method CC on the same database is 0.96. It indicates that even though our method is NR the prediction performance is very close to the FR method [11]. Moreover, the FR method converted the MOS scale 1–5 linearly to the MOS scale 0-1, which is not truly mapped the subjective scores between the two scales [35].

In order to extensively verify the performance of the proposed method, we consider another stereo image database. The database was created by IVC and IRCCyN laboratory, University of Nantes, France. As the proposed method is designed for JPEG coded stereo images, we use only the JPEG coded images from the database. In the database, there are thirty JPEG coded stereo images for six different reference images. The images were coded at a wide range of bit rates ranging from 0.24 bpp to 1.3 bpp. The details of the database are discussed in [9]. As the database used difference mean opinion score (DMOS) with different scale (DMOS scale, 0 to 100), it is very difficult to develop a mathematical relationship between the two different scales (MOS scale: 1 to 5, and DMOS scale: 0 to 100). Although Pinson and Walf presented a mapping method to convert one subjective
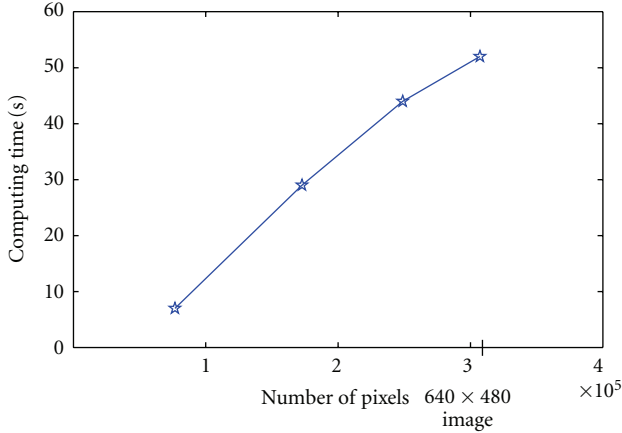
FIGURE 17: Computation time of the proposed algorithm.

TABLE 7: Method parameters and weighting factors (continuous DMOS scale, 0 to 100).

$\alpha = 140.62388$	$\beta = -85.924017$	
$w_1 = -0.017834$	$w_2 = -0.000617$	$w_3 = 0.019968$
$w_4 = 0.017483$	$w_5 = 0.000042$	$w_6 = 0.00717$

scale to another, the performance was not sufficient for all subjective data sets [35]. Consequently, we estimate the suitable optimized model parameters and weighting factors for DMOS scale, 0 to 100 by using the same equations with different logistic function as follows:

$$\text{DMOS}_{p(100)} = \frac{99}{1 + \exp[-1.0217(S - 50)]} + 1. \quad (31)$$

Therefore, in order to use the database we randomly divide the database into two parts for training and testing and also there is no overlapping between training and testing. The method's parameters and weighting factors with the training images are shown in Table 7 for DMOS scale, 0 to 100. The proposed method's CCs for the training and testing images are, respectively, 0.93 and 0.91. Subsequently, the proposed method's performance can again be compared with the FR method (e.g., C4 d2: considering better performance disparity algorithm, "bp Vision") [9]. The prediction performance for all JPEG coded stereo images is shown in Table 8. The table shows that proposed NR method's performance is almost better for the evaluation metrics even compared to the FR method. It is clear from the table that our proposed NR method performance is sufficient and better compared to the published FR method. Therefore, the Tables 6 and 8 confirm that our proposed method performance is sufficient and better compared to the others recently published method.

6. Conclusion

In this paper, we propose an NR stereoscopic image quality assessment method for JPEG coded symmetric/asymmetric images which used the perceptual differences of local

TABLE 8: Evaluation results of the comparison on IVC-IRCCYN laboratory's database.

Method	JPEG stereo images			
	CC	AAE	RMSE	OR
Proposed, NR	0.925	7.689	9.612	0.033
Method (C4 d_2), FR [9]	0.927	8.610	10.141	0.033

features such as edge and nonedge. Local features based distortions and relative disparity measures are estimated in this approach. A popular subjective database is used to verify the performance of the method. The result shows that the method performs quite well over wide range of stereo image content and distortion levels. Although the approach is used only for JPEG coded stereo images, future research can be extended to generalize the approach irrespective of any coded stereoscopic images.

References

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 10–21, 2007.

[2] A. M. William and D. L. Bailey, "Stereoscopic visualization of scientific and medical content," in *Proceedings of the ACM SIGGRAPH 2006 Educators Program–International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '06)*, Boston, Mass, USA, August 2006.

[3] J. Baltes, S. McCann, and J. Anderson, "Humanoid Robots: Abarenbou and DaoDan," RoboCup 2006—Humanoid League Team Description Paper.

[4] C. F. Westin, "Extracting brain connectivity from diffusion MRI," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 124–152, 2007.

[5] Y. A. W. De Kort and W. A. Ijsselsteijn, "Reality check: the role of realism in stress reduction using media technology," *Cyberpsychology and Behavior*, vol. 9, no. 2, pp. 230–233, 2006.

[6] N. A. Dodgson, "Autostereoscopic 3D displays," *Computer*, vol. 38, no. 8, pp. 31–36, 2005.

[7] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[8] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technology," *IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, pp. 98–110, 2005.

[9] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP Journal on Image and Video Processing*, vol. 2008, Article ID 659024, 2008.

[10] N. Qzbek, A. M. Tekalp, and E. T. Tunali, "Rate allocation between views in scalable stereo video coding using an objective stereo video quality measure," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. I1045–I1048, Honolulu, Hawaii, USA, April 2007.

[11] R. Bensalma and M. C. Larabi, "Towards a perceptual quality metric for color stereo images," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 4037–4040, Hong Kong, September 2010.

[12] 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 *Proceedings of the International Workshop on Video Processing and Quality Metrics*

for Consumer Electronics (VPQM '01), Scottsdale, Ariz, USA, 2010.

[13] A. Tikanmaki and A. Gotchev, "Quality assessment of 3D video in rate allocation experiments," in *Proceedings of the IEEE International Symposium on Consumer Electronics (ISCE '08)*, Algarve, Portugal, April 2008.

[14] L. Stelmach, W. J. Tam, D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 188–193, 2000.

[15] L. M. J. Meesters, W. A. Ijsselsteijn, and P. J. H. Seuntiëns, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381–391, 2004.

[16] M. T. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: a review," in *Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490 of *Proceedings of the SPIE*, January 2007.

[17] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation," *IEEE ACM Transactions on Applied Perception*, vol. 3, no. 2, pp. 95–109, 2009.

[18] 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.

[19] Y. Horita, Y. Kawai, Y. Minami, and T. Murai, "Quality evaluation model of coded stereoscopic color image," in *Visual Communications and Image Processing*, vol. 4067 of *Proceedings of the SPIE*, pp. 389–398, June 2000.

[20] W. A. Ijsselsteijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: effects of camera parameters and display duration," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 225–233, 2000.

[21] W. J. Tam and L. B. Stelmach, "Perceived image quality of MPEG-2 stereoscopic sequences," in *Human Vision and Electronic Imaging II*, vol. 3016 of *Proceedings of the SPIE*, pp. 296–301, San Jose, Calif, USA, February 1997.

[22] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "A perceptual quality metric for stereoscopic crosstalk perception," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 4033–4036, Hong Kong, September 2010.

[23] Z. M. P. Sazzad, S. Yamanaka, Y. Kawayoke, and Y. Horita, "Stereoscopic image quality prediction," in *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEx '09)*, pp. 180–185, San Diego, CA, USA, July 2009.

[24] R. Akhter, Z. M. Parvez Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," in *Stereoscopic Displays and Applications XXI*, vol. 7524 of *Proceedings of the SPIE*, San Jose, CA, USA, January 2010.

[25] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. BT.500-10, Geneva, Switzerland, 2000.

[26] Z. Wang, *Rate scalable foveated image and video communications [Ph.D. thesis]*, Department of ECE, The University of Texas at Austin, 2003.

[27] University of Manitoba, http://umanitoba.ca/.

[28] D. V. Meegan, L. B. Stelmach, and W. J. Tam, "Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery," *Journal of Experimental Psychology: Applied*, vol. 7, no. 2, pp. 143–153, 2001.

[29] R. Akhter, *Perceptual image quality for stereoscopic vision [M.S. thesis]*, Department of Computer Science, University of Manitoba, 2011.

[30] B. P. McKinnon, *Point, line segment, and region-based stero matching for mobile robotics [M.S. thesis]*, Department of Computer Science, University of Manitoba, 2009.

[31] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 257–268, 2008.

[32] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Perth, Australia, December 1995.

[33] VQEG, "Final Report from the video quality experts group on the validation of objective models of video quality assessment, FR-TV Phase II (August 2003)," http://www.vqeg.org/.

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

[35] M. Pinson and S. Wolf, "An objective method for combining multiple subjective data sets," in *Proceedings of the SPIE Video Communications and Image Processing*, Lugano, Switzerland, July 2003.