

No-Reference Stereoscopic Image Quality Assessment

Roushain Akhter^a, Z. M. Parvez Sazzad^b, Y. Horita^b, and J. Baltes^a

^a Dept. of Computer Science, University of Manitoba, Winnipeg, MB, Canada;

^b Graduate School of Science and Engineering, University of Toyama, Toyama, Japan

ABSTRACT

Display of stereo images is widely used to enhance the viewing experience of three-dimensional imaging and communication systems. In this paper, we propose a method for estimating the quality of stereoscopic images using segmented image features and disparity. This method is inspired by the human visual system. We believe the perceived distortion and disparity of any stereoscopic display is strongly dependent on local features, such as edge (non-plane) and non-edge (plane) areas. Therefore, a no-reference perceptual quality assessment is developed for JPEG coded stereoscopic images based on segmented local features of artifacts and disparity. Local feature information such as edge and non-edge area based relative disparity estimation, as well as the blockiness and the blur within the block of images are evaluated in this method. Two subjective stereo image databases are used to evaluate the performance of our method. The subjective experiments results indicate our model has sufficient prediction performance.

Keywords: No-reference, Disparity, JPEG, Auto stereoscopic display, Segmentation

1. INTRODUCTION

In recent years, there have been many efforts to enhance the viewing experience of stereo imaging and visual communication systems by incorporating three-dimensional (3D) imaging. This trend has a strong impact on daily life applications ranging from entertainment¹ to more specialized applications such as robot navigation,² remote education,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 have been designed for stereoscopic images⁷ and some researchers are still working to develop a new standard for efficient multi-view 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. There is no doubt that subjective testing is the most accurate method for perceived image quality as it reflects true human perception. However, it is time consuming, expensive, and cannot be done in real time. As a result, the development of objective quality evaluation method (i.e. computational model that can automatically predict perceptual image quality) is getting more attention in the quality assessment field.

Although, several conventional 2D objective quality assessment methods⁹ have been proposed based on peak signal-to-noise ratio (PSNR), mean squared error (MSE), or human visual system (HVS) characteristics in the last two decades for images/ videos, no comparable effort has been devoted to the quality assessment of stereoscopic images. A full-reference (FR) quality metric for the assessment of stereoscopic image pairs using the

Further author information: (Send correspondence to R. Akhter)

R. Akhter: E-mail: rakhter@cs.umanitoba.ca, Tel.: 1-204-474-8838

Z. M. P. Sazzad: E-mail: parvezdu@ctg.u-toyama.ac.jp, Tel.: 81 76 445 6758

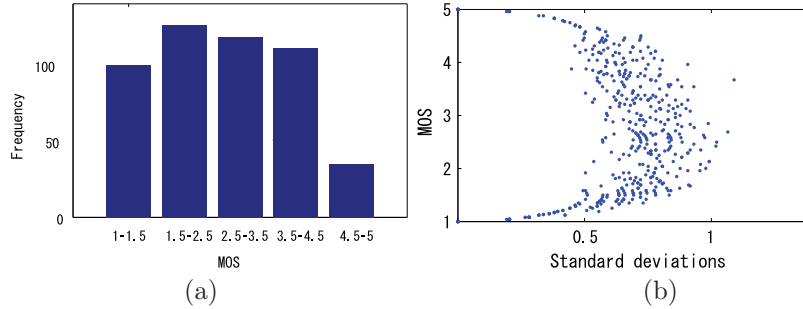


Figure 1. Histogram and standard deviations of MOS scores: (a) Histogram, (b) Standard deviations.

fusion of 2D quality metrics and of the depth information is proposed in.¹⁰ The study is evaluated that the FR metric of 2D quality assessment can be used for an extension to 3D with the incorporation of depth information. In,¹¹ the selection of the rate allocation strategy between views is addressed for scalable multi-view video codec to obtain the best rate-distortion performance. In,¹² 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,¹³ 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 in¹⁴ along with a description of the main artifacts of stereo TV. In,¹⁵ 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,¹⁶ 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.¹⁷ In,¹⁸ 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,¹⁹ 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 non-stereo sequences with different bit rates are investigated in.²⁰

Perceived quality of stereoscopic images depends on several factors such as the rendered perception of depth, stereoscopic impairments (keystone distortion, depth plane curvature, puppet theater effect, cross talk, card-board effect, shear distortion, picket fence effect and image flipping) and visual discomfort.¹⁵ However, overall perceptual quality reflects the combined effect of the multidimensional factors. 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 non edge 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 model for stereoscopic images based on segmented local features of artifacts and disparity. In many practical applications, the reference image is not available, therefore a 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.²¹ The metric used too many parameters (thirteen) and local features (three) and used only one subjective database. Therefore, the metric has a chance of over trained on the database. Consequently, it may create an opportunity to falsely improve assessed metric beside the training parameters. Moreover, computational cost of the model was high. Therefore, we consider two local features (edge and non-edge) and eight reduced parameters with low computational cost in this paper. We generalize this algorithm, and provide a more extensive set of validation results on the two databases. The rest of the paper is organized as follows: Section 2 describes briefly the two subjective databases that are used to evaluate our method. The details of our approach is given in Section 3. Results are discussed in Section 4 and finally, Section 5 concludes the paper.

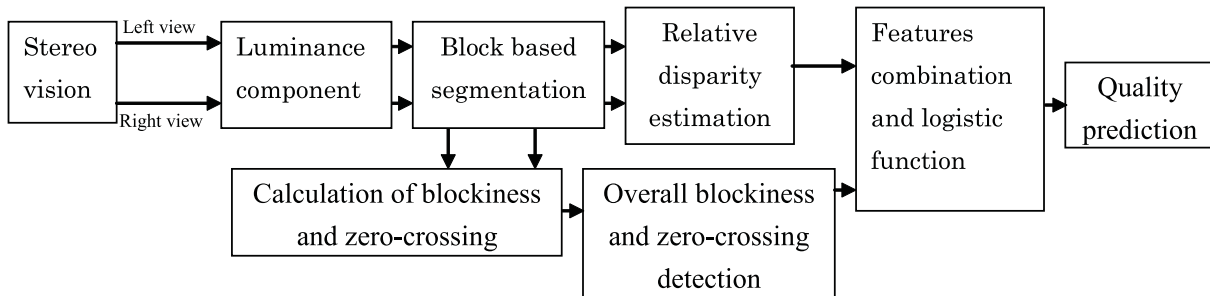


Figure 2. Proposed NR quality evaluation model.

2. THE SUBJECTIVE DATABASES

2.1 Toyama Database

The Media Information and Communication Technology (MICT) lab., University of Toyama conducted subjective experiment on 24 bit/pixel RGB color stereoscopic images.²¹ The database contained JPEG coded symmetric and asymmetric 490 stereoscopic image 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. Both distorted and original images were displayed sequentially. At the end of the presentation, the subject was asked to assess the annoyance he/she felt over all perceptual quality on the distorted stereo image with respect to the reference stereo one. 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”. Twenty-four non-expert subjects (12 males and 12 females, age range: 19-32 years) were shown the database; most of them were college/university student. A 10-inch auto stereoscopic, LCD (SANYO) display (resolution: 640×480 , image splitter technology) was used in this experiment to display the stereoscopic images and the subjects were instructed about the limited horizontal viewing angle to perceive 3D image correctly. Mean opinion scores (MOSs) were then computed for each stereo image after the screening of post-experiment results according to ITU-R Rec. 500-10.²² The MOS histogram and standard deviations of all MOSs of the database are shown in Figure 1.

2.2 IRCCyN/IVC Database

The IRCCyN lab, university of Nantes conducted subjective experiment on 24 bit/pixel RGB color stereoscopic images of size 512×448 .¹⁰ Six reference stereo images and their five degradation levels JPEG and JPEG2000 coded images were used in the database. Total sixty symmetric coded image pairs were consider in the database. JPEG2000 compressions used bit rates ranging from 0.16 bits per pixel (bpp) to 0.71 bpp while JPEG compression involved bit rates ranging from 0.24 bpp to 1.3 bpp. The subjective assessment methodology for video quality (SAMVIQ) method was used in the experiment. The SAMVIQ method has possible to combine quality evaluation capabilities and ability to discriminate similar levels of quality, using an implicit comparison process. The method is based on a random access process to play sequence files. Subjects can start and stop the evaluation process as they wish and can follow their own paces in rating, modifying grades, repeating play out when needed. Each subject used a slider on a continuous scale graded from 0 to 100 defined by 5 linearly quality terms “bad”, “poor”, “fair”, “good”, and “Excellent”. Seventeen subjects, mostly males familiar with subjective tests, with an average age of 28.2 years took part in the test. A 21-inch Samsung SyncMaster 1100MB stereoscopic display (resolution: 1024×768) was uses in the experiment. At the end of the test sessions, the difference mean opinion score (DMOS) is computed as the difference between the MOS for the hidden reference and the MOS one relative to the image. Details of the experiment was discussed in.¹⁰

3. OBJECTIVE STEREOSCOPIC IMAGE QUALITY EVALUATION

It has already been established that the primary function of the human visual system (HVS) is to extract structural or edge information from the viewing field, and the HVS is highly adapted for this purpose.²³ Human visual



Figure 3. Block based segmented images

perception is very sensitive to edge detection, consequently, perceive distortions should be strongly dependent on local features such as edge, and non-edge. Thus, we 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 artifacts 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 blurring artifacts 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 difference around the block boundary. The 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 more smooth the image signal which causes the reduction of signal edge points. Consequently, average edge point detection measures of a block gives more insight into the relative blur in the image. Here, zero-crossing technique is used as an edge detector. Although, the impact of coding artifacts 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,²⁴ we take into account maximum blockiness and blur measures between the left and right views. Therefore, we consider higher blockiness and lower zero-crossing values between the two views. The block diagram of the proposed model is shown in Figure 2. For simplicity, only the luminance component is considered to make quality prediction of the color stereo images. As, both features artifacts as well as disparity are estimated based on segmented local features. At first the details of segmentation is discussed in the following sub Section. Subsequently, the artifacts and disparity measures are described in the next sub Sections.

3.1 Block Based Segmentation

In order to classify edge and non-edge blocks of an image, we use a simple block based segmentation algorithm.²¹ First, we establish a simple pixel based segmentation to classify each pixel within the image into either a edge, or non-edge pixel. Initially, standard deviation (STD) of each pixel is estimated within its 3×3 and 5×5 neighborhood pixels. For all corners pixels, we take into account only available pixels for the measures. Let $STD_{3 \times 3}(m, n)$, and $STD_{5 \times 5}(m, n)$ be the standard deviated image of 3×3 and 5×5 neighborhood, respectively. Then we calculate absolute difference, $D_a(m, n)$ by the following equation:

$$D_a(m, n) = |STD_{3 \times 3}(m, n) - STD_{5 \times 5}(m, n)| \quad (1)$$

where $m = 1, 2, \dots, M$ and $N = 1, 2, \dots, N$. Subsequently, we calculate STD of $D_a(m, n)$ by

$$D = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M \times N} (\overline{D_a} - D_{a_i})^2} \quad (2)$$

We then use the following algorithm to classify edge and non-edge pixels of the image.

$$P(m, n) = \begin{cases} 1 & \text{if } D_a(m, n) \geq D \\ 0 & \text{otherwise} \end{cases}$$

where “1” and “0” denote edge and non-edge pixels respectively. Secondly, we classify each block (8×8) of the image into either edge or non-edge block by using the segmentation algorithm.

The block based segmentation algorithm:

$$Sum = n_e + n_n \quad (3)$$

where n_e , and n_n are respectively the number of edge, and non-edge pixels per (8×8) block. Therefore, the “Sum” is the total number of pixels per block.

*if $\left(\frac{n_e}{Sum} > th_n\right)$ then the block is “edge”
else the block is “non – edge”*

The threshold value, $th_n = 0.25$ and it is calculated empirically. The value (0.25) indicates that if more than 25% pixels within a block is edge, the block will be considered as “edge” block. The performance of the block based segmentation algorithm for two reference images are shown in Figure 3. The white, and dark blocks are respectively indicate edge, and non-edge areas in the segmented images. The Figure 3 indicates sufficient segmentation performance. Although, this segmentation is used to predict quality assessment, it is not an exactly accurate segmentation. Because, if any block contains same type of pixels (either edge or non-edge), the block then ideally considers as either edge or non-edge. Otherwise, it is very difficult to identify accurately the type of blocks.

3.2 Image Artifacts Measure

To measure JPEG coded stereo image artifacts, we estimate blockiness and zero-crossing in spatial domain based on segmented local features. Firstly, we calculate blockiness (B) and zero-crossing (ZC) 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 (see Section 3.1) to the left and right images individually to classify edge, and non-edge blocks in the images. Thirdly, we average each value of B and ZC separately for edge, and non-edge 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 non-edge 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 is $x(m, n)$ for $m \in [1, M]$ and $n \in [1, N]$, a differencing signal along each horizontal line is calculated by

$$d_h(m, n) = x(m, n + 1) - x(m, n), \quad (4)$$

$$n \in [1, N-1] \quad \text{and} \quad m \in [1, M]$$

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

$$B_{bh} = \frac{1}{8} \sum_{i=1}^8 |d_h(i, 8j)| \quad (5)$$

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

For horizontal zero crossing (ZC):

$$z_h(m, n) = \begin{cases} 1 & \text{if horizontal ZC at } d_h(m, n) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

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:

$$ZC_{bh} = \sum_{i=1}^8 \sum_{j=1}^8 z_h(i, j) \quad (7)$$

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 a differencing signal along each vertical line:

$$d_v(m, n) = x(m + 1, n) - x(m, n), \quad (8)$$

$$n \in [1, N] \quad \text{and} \quad m \in [1, M-1]$$

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:

$$B_b = \frac{B_{bh} + B_{bv}}{2}, \quad ZC_b = \frac{ZC_{bh} + ZC_{bv}}{2} \quad (9)$$

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

$$Bl_e = \frac{1}{N_e} \sum_{b=1}^{N_e} B_{be} \quad (10)$$

$$Bl_n = \frac{1}{N_n} \sum_{b=1}^{N_n} B_{bn} \quad (11)$$

where N_e , and N_n are respectively the number of edge, and non-edge blocks of the image. Similarly, the average blockiness values of Br_e , and Br_n for the right image are calculated. The average zero crossing values of ZCl_e , and ZCl_n for the left image are estimated by:

$$ZCl_e = \frac{1}{N_e} \sum_{b=1}^{N_e} ZC_{be} \quad (12)$$

$$ZCl_n = \frac{1}{N_n} \sum_{b=1}^{N_n} ZC_{bn} \quad (13)$$

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 non-edge 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:

$$B_{e/n}(Bl, Br) = \begin{cases} Bl & \text{if } Bl \geq Br \\ Br & \text{otherwise} \end{cases}$$

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

$$ZC_{e/n}(ZCl, ZCr) = \begin{cases} ZCl & \text{if } ZCl \leq ZCr \\ ZCr & \text{otherwise} \end{cases}$$

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

$$B = B_e^{w_1} \cdot B_n^{w_2} \quad (14)$$

$$Z = ZC_e^{w_3} \cdot ZC_n^{w_4} \quad (15)$$

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

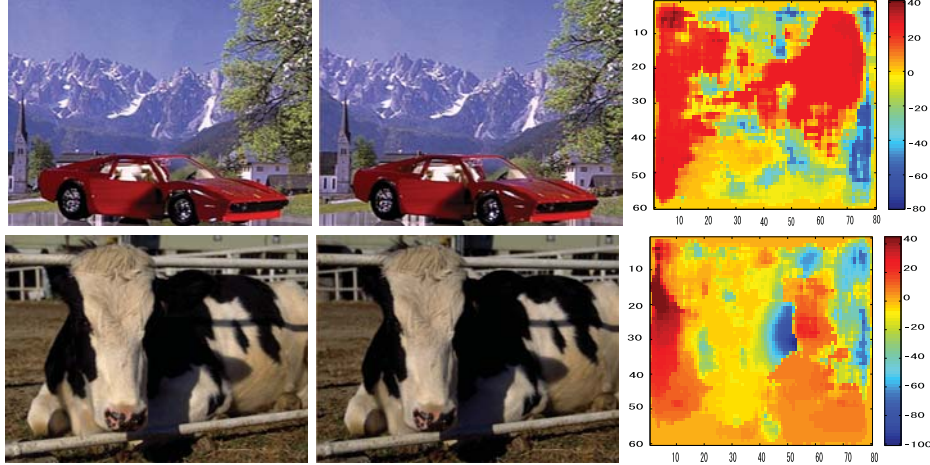


Figure 4. Stereo pairs and its depth maps.

3.3 Relative Disparity Measure

Although, many features based approaches are used for stereo matching/disparity estimation,²⁵ a simple block based difference zero-crossing (DZC) rate approach is used in this work. The principal of the disparity estimation is to divide the left image into non overlapping 8×8 blocks with classification of edge and non-edge 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 same corresponding block and up to ± 32 pixels of the right image. In order to reduce computational cost, we restricted the correspondence search to 1D (i.e. horizontally) only and within ± 32 pixels. The depth maps of the two sample stereo image pairs are shown in Figure 4. Colors in the depth maps that are indicated by vertical color bars in right are estimated depths of the images pairs. Although disparity means 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. Two type of relative disparities are considered in this method.

d_1 : DZC rate between only the same corresponding block of the two views (like a measure of difference).

d_2 : DZC rate between a block of a left image and the correspondence searching block (MDZC rate based) within ± 32 pixels of the right image.

In order to measure disparity, firstly, the segmentation algorithm is applied to left image only to classify edge and non-edge blocks. Secondly, block based DZC is estimated in the two corresponding blocks (according to d_1 or d_2) between the left and right images. Thirdly, we average the DZC rate values separately for edge and non-edge 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:

$$DZC = ZCl \oplus ZCr \quad (16)$$

$$DZCR = \frac{1}{8 \times 8} \sum DZC \quad (17)$$

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:

$$DZC_h = ZCl_h \oplus ZCr_h \quad (18)$$

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

$$DZCR_h = \frac{1}{8 \times 8} \sum DZC_h \quad (19)$$

Subsequently, the average $DZCR_h$ (AZC_h) for edge, and non-edge blocks of the left image are calculated by

$$AZC_{h_e} = \frac{1}{N_e} \sum_{e=1}^{N_e} DZCR_{h_e} \quad (20)$$

$$AZC_{h_n} = \frac{1}{N_n} \sum_{e=1}^{N_n} DZCR_{h_n} \quad (21)$$

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

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

$$AZC_e = \frac{AZC_{h_e} + AZC_{v_e}}{2}, AZC_n = \frac{AZC_{h_n} + AZC_{v_n}}{2} \quad (22)$$

Finally, the over all disparity feature is estimated by

$$DZ = AZC_e^{w_5} \cdot AZC_n^{w_6} \quad (23)$$

where w_5 , and w_6 are respectively the weighting factors of the disparity features for edge, and non-edge areas.

3.4 Features Combination

In order to combine artifacts and disparity features to develop a stereo quality assessment model. We consider the following features combined equation:

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

where α , and β are the model parameters. The model parameters and weighting factors (w_1 to w_6) are must be estimated by an optimization algorithm with the subjective test data. 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.⁹

$$MOS_p = \frac{4}{1 + \exp[-1.0217(S - 3)]} + 1 \quad (25)$$

Here, Particle Swarm Optimization (PSO) algorithm is used for optimization.²⁶

Table 1. Model parameters and weighting factors for quality scale, 1-5

$\alpha = -88.8009$	$\beta = 95.0422$	
$w_1 = 0.0264$	$w_2 = -0.0241$	$w_3 = -0.0202$
$w_4 = -0.0044$	$w_5 = 0.00086$	$w_6 = 0.0129$

Table 2. Model parameters and weighting factors for quality scale, 0-100

$\alpha = -80.0253$	$\beta = 133.0728$	
$w_1 = -0.0093$	$w_2 = -0.0269$	$w_3 = 0.0024$
$w_4 = 0.0184$	$w_5 = 0.0137$	$w_6 = -0.0379$

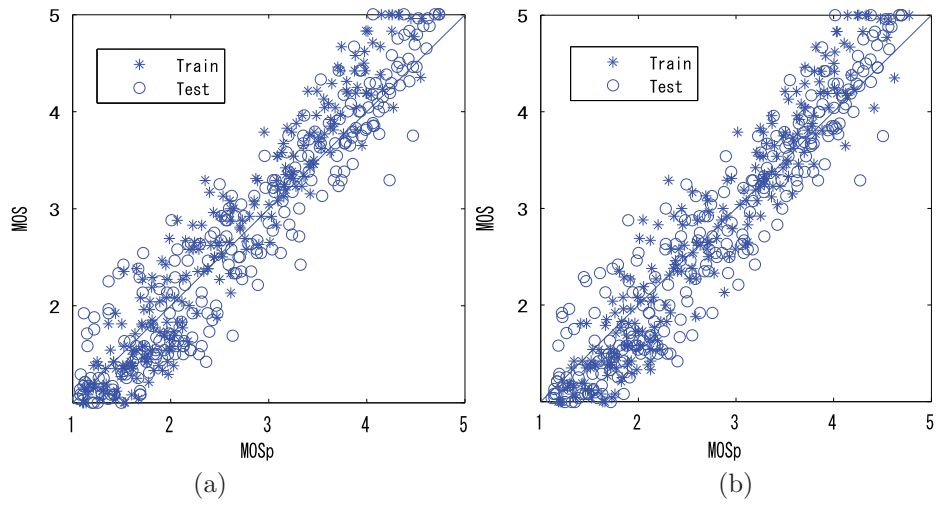


Figure 5. MOS versus MOSp of Toyama database: (a) With disparity, d_1 (b) With disparity, d_2

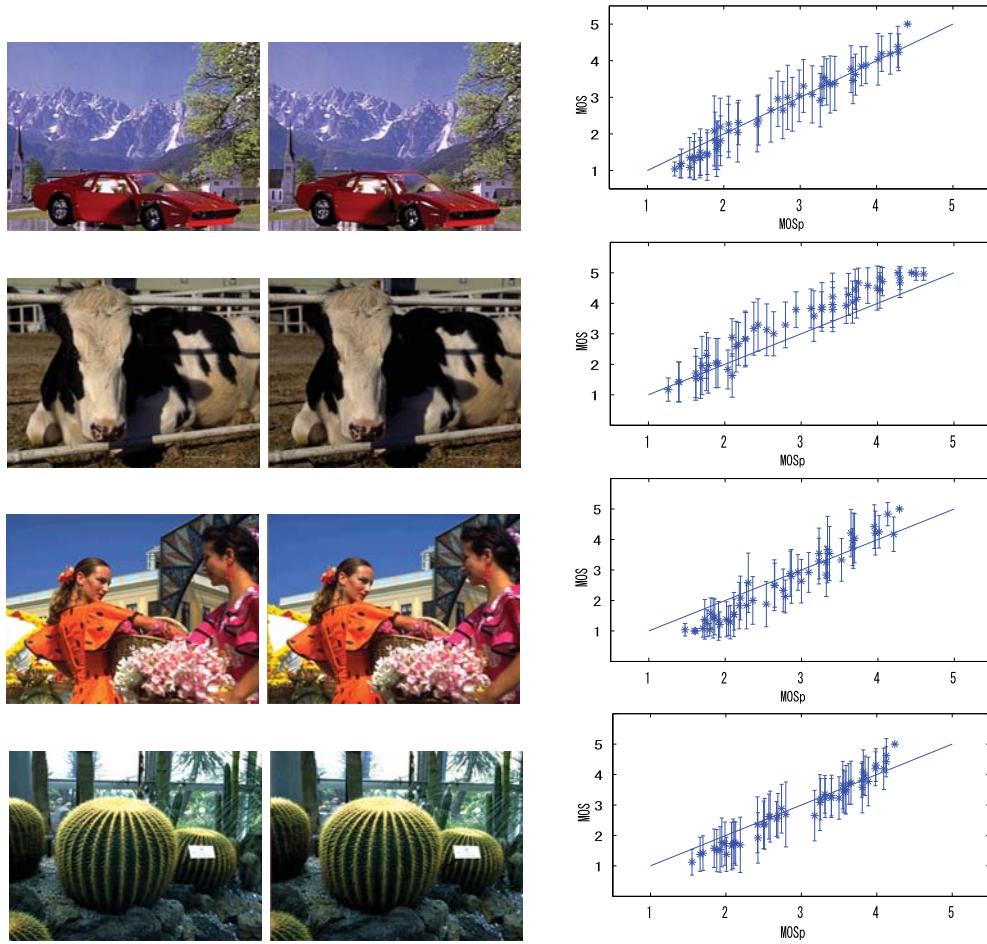


Figure 6. The MOSp performances on texture variety of stereo pairs over the quality range. The predictions points * and ± 2 standard deviation intervals are shown for each stereo pair.

4. RESULTS

In order to verify the permanence of our method extensively we considered two databases: Toyama and IRCCyN,s (see Section 2.1, and 2.2). Since our method is using two different quality scale related (Toyama: MOS scale, 1-5, and IRCCyN, DMOS scale, 0-100) databases, it is difficult to develop a common mathematical relationship between these two scales. Therefore, Pinson and Wolf²⁷ presented a mapping method to convert one subject scale to another, however the performance was not good at all for all subjective data sets. As a result, we do not combine the two different subjective experiment data, but the real subjective scores of both databases are considered. We also estimate the method's parameters and weighting factors for the two databases individually. We consider only the JPEG coded stereo images from the IRCCyN database. The logistic function of quality scale is 0-100 in the work⁹ is also used in our approach.

In order to use the Toyama 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 between training and testing. Similarly, for the IRCCyN database, the training dataset consists of ten randomly detected stereo pairs and the testing consists of the other ten stereo pairs with no overlapping between training and testing. The model's parameters and weighting factors are obtained for the two quality scales by the PSO optimization algorithm with all of the training images are shown in Tables 1 and 2. In order to provide quantitative measures on the performance of the proposed method, we will follow the standard performance evaluation procedures employed in the video quality experts group (VQEG) FR-TV Phase II test,²⁸ 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. We consider four approaches for evaluations; proposed method with relative disparity d_1 , d_2 , without disparity, and 2D quality mean (i.e. proposed approach is used as a fusion of 2D quality metrics). The evaluation results are summarized in Tables 3, and 4 for the Toyama and IRCCyN databases, respectively. The Tables shown that the proposed model performances with disparities for every one of the evaluation metrics are quite sufficient for the both databases. It has also been observed from the Tables 3, and 4 that the proposed model provides sufficient prediction accuracy (higher CC), and sufficient prediction consistency (lower OR). The Tables also shows the method performance with the two relative disparities are closed and superior compared to without disparity. Whereas, "2D quality mean" approaches performance is not sufficient even compared to without disparity approach. Although, the incorporation of real disparities measure to the FR stereo quality metrics¹⁰ indicate poor results, our method with relative disparities indicate better results compared to without disparity. Therefore, the relative disparity can be a significant measure for 3D quality prediction. The MOS versus MOS predictions (MOSp) of our proposed method with disparities for Toyama databases is shown in Figure 5. The symbols * and o respectively indicate MOSp points for the databases of training and testing. The MOSp points * and the error bars of ± 2 standard deviation (STD) intervals of four different stereo images are shown in Figure 6. Error bars show the ± 2 STD interval of the MOS. The Figure indicates the predictions consistently performed well in almost similar nature on variety of image content. In addition, we compare the performance of our proposed method against the recently published method.²¹ The method was used the Toyama database only. Our proposed method's evaluation results on the same database are shown in Table 5. Table 5 shows that though both methods' evaluation performance are all most same for the training database the performance of our proposed method is superior compared to the published method on the testing database.

5. CONCLUSION

We propose a NR stereoscopic image quality assessment method for JPEG coded symmetric/asymmetric images which used the perceptual differences of local features such as edge and non-edge. Local features based artifacts and disparity measures are estimated in this approach. Two subjective databases are used to verify the performance of the method. The results show that the model 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 by irrespective of coders.

Table 3. Evaluation results for training and testing (Scale, 1-5)

<i>Proposed model</i>	<i>Training (245 stereo pairs)</i>			
	CC	AAE	RMSE	OR
With disparity (d_1)	0.962	0.303	0.379	0.065
With disparity (d_2)	0.957	0.307	0.381	0.078
Without disparity	0.957	0.324	0.395	0.078
Using 2D mean	0.916	0.406	0.513	0.074
<i>Testing (245 stereo pairs)</i>				
With disparity (d_1)	0.943	0.320	0.391	0.049
With disparity (d_2)	0.939	0.336	0.408	0.049
Without disparity	0.930	0.359	0.438	0.061
Using 2D mean	0.883	0.438	0.559	0.061

Table 4. Evaluation results (training + testing) on IRCCyN's dataset (Scale, 0-100)

<i>Model</i>	<i>JPEG coded stereo pairs only</i>			
	CC	AAE	RMSE	OR
With disparity (d_1)	0.94	6.67	7.99	0.05
With disparity (d_2)	0.94	6.38	7.82	0.05
Without disparity	0.91	8.27	10.24	0.15
Using 2D mean	0.90	8.32	10.28	0.15

ACKNOWLEDGMENTS

Author would like to thank MICT lab, Japan and IRCCyN lab, France for providing their stereo image databases.

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, Nov. 2007.
- [2] J. Baltés, S. McCann, and J. Anderson "Humanoid Robots: Abarenbou and DaoDan," RoboCup 2006 - Humanoid League Team Description Paper.
- [3] A. M. William, and D. L. Bailey, "Stereoscopic visualization of scientific and medical content," in Proc. SIGGRAPH'06, Boston, Mass, USA, Jul.-Aug. 2006.
- [4] C. F. Westin, "Extraction 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. Dodgson, "Auto stereoscopic 3-D displays," IEEE Computer, vol. 38, no. 8, pp. 31-36, Aug. 2005.
- [7] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," IEEE Trans. on PAMI vol. 25, pp. 993-1008, 2003.
- [8] A. Smolic, P. Kauff, "Interactive 3-D video representation and coding technology," in Proc IEEE, Special Issue on Advances in Video Coding and Delivery, vol. 93, no. 1, pp. 98-110, Jan. 2005.
- [9] Z. M. P. Sazzad, and Y. Horita, "Local region-based image quality assessment independent of JPEG and JPEG2000 coded color images," J. of Electronic Imaging, vol. 17(3), pp. 033002-1-17, Jul-Sep 2008.
- [10] A. Benoit, P. Le. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," EURASIP Journal on Image and Video Processing, vol. 2008, pp.1-13, 2008.
- [11] N. Ozbek, A. M. Tekalp, and E. T. Tunali, "Rate allocation between views in scalable stereo video coding using objective stereo video quality measure," in Proc. IEEE ICASSP, Hawaii, USA, Apr. 15-20, 2007.
- [12] A. Tikanmaki, and A. Gotchev, "Quality assessment of 3D video in rate allocation experiments," in Proc. IEEE ISCE, Algarve, Portugal, Apr., 14-16, 2008.
- [13] L. Stelmach, W. J. Tam, D. Meegan, and A. Vincent, "Stereo image quality: Effects of mixed spatio-temporal resolution," IEEE Trans. on Circuits and Systems for Video Technology, vol. 10, no. 2, pp. 188-193, March 2000.

Table 5. Evaluation results comparison (Scale, 1-5)

<i>Method</i>	<i>Training (245 stereo pairs)</i>			
	CC	AAE	RMSE	OR
Proposed	0.962	0.303	0.379	0.065
Method ²¹	0.966	0.292	0.367	0.069
<i>Testing (245 stereo pairs)</i>				
Proposed	0.943	0.320	0.391	0.049
Method ²¹	0.935	0.350	0.421	0.065

- [14] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntjens, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381-391, March 2004.
- [15] M. T. M. Lambooi, W. A. IJsselsteijn, and I. Heynderickx, "Visual Discomfort in Stereoscopic displays: A Review," *SPIE IS&T*, vol. 6490, Jan. 2007.
- [16] P. Seuntjens, L. Meesters, and W. IJsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *IEEE ACM Trans. on Applied perception*, vol. 3, no. 2, pp. 95-109, April 2009 (online).
- [17] 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 Letter*, vol. 44, no. 6, pp. 963-965, July 2008.
- [18] Y. Horita, Y. Kawai, Y. Minami, and T. Murai, "Quality evaluation model of coded stereoscopic color image," in *Proc. SPIE* vol. 4067, pp. 389-398, 2000.
- [19] W. A. IJsselstrijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: effects of camera parameters and display duration," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 225-233, March 2000.
- [20] W. J. Tam, and L. B. Stelmach, "Perceived image quality of MPEG-2 stereoscopic sequences," in *Proc. SPIE*, vol. 3016, pp. 296-301, Feb. 10-13, San Jose, CA, USA, 1997.
- [21] Z. M. P. Sazzad, S. Yamanaka, Y. Kawayoke, and Y. Horita, "Stereoscopic image quality prediction," in *Proc. IEEE QoMEX*, San Diego, CA, USA, July 29-31, 2009.
- [22] ITU-R BT.500-10. Methodology for the Subjective Assessment of the Quality of Television Pictures.
- [23] Z. Wang, "Rate scalable foveated image and video communications," PhD thesis, Dept. of ECE, The University of Texas at Austin, Dec. 2003.
- [24] 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," *J. Exp. Psychol.: Appl.*, vol. 7, pp. 143-153, 2001.
- [25] Brian P. McKinnon, "Point, Line Segment, and Region-Based Stereo Matching for Mobile Robotics," Masters thesis, Dept. of Computer Science, University of Manitoba, August 2009.
- [26] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *Proc. IEEE ICNN*, Perth, Australia, pp. 1942-1948, Nov. 1995.
- [27] M. Pinson, and S. Wolf, "An objective method for combining multiple subjective data sets," *Proc. of SPIE Video Communications and Image Processing*, Lugano, Switzerland, July 2003.
- [28] 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/>.