DUJASE Vol. 3(1) 47-54, 2015 (January)

Continuous No-Reference Stereoscopic Video Quality Prediction

Z. M. Parvez Sazzad

Dept. of Electrical and Electronic Engineering, University of Dhaka, Dhaka-1000

E-mail: sazzad@univdhaka.edu

Received on 03. 11. 2013. Accepted for publication on 14.07. 2014.

Abstract

In this paper, we propose an objective no-reference continuous video quality assessment method for MPEG-2 MP@ML coded stereoscopic videos based on spatio-temporal segmentation. Segmented local features such as edge and non-edge areas based spatial artifacts, disparity, and temporal features are measured in this method. Blockiness and blur are considered to measure spatial artifacts for each stereo pair frames. A block based different zero-crossing approach is used for disparity measure. In this method, a temporal segmentation approach is considered and each temporal segment is evaluated for artifacts and disparity. Temporal features are calculated separately for left and right video sequences based on segmented local features and sub temporal segment. Different weighting factors are then applied for the two different local features to measure the overall artifacts, disparity, and temporal features of a temporal segment. In order to verify the performance, we conducted subjective experiment on different symmetric and asymmetric coded stereo videos which indicates that our proposed method's prediction performance is quite sufficient.

Keywords: No-reference, Temporal segmentation, Disparity, SSCQE, Auto stereoscopic display.

1. Introduction

Three dimensional (3D) video is getting rapidly more attention for next generation applications, ranging from broadcast television to streaming. This trend towards immersible media is going to have a strong impact on our daily life in different applications such as 3DTV [1], remote education [2], medicine [3], etc. There are many alternative technologies for 3D video display and communication including holographic, volumetric and stereoscopic; stereoscopic video seems to be the most developed technology at the present [4]. Stereoscopic video consists of two videos (left and right views) captured by closely located (approximately the distance between two eyes) two video cameras. These views constitute a stereo pair and can be perceived as a virtual view (i.e., not an actual camera view) in 3D by human observers with the rendering of corresponding view points. Therefore, codec that used in 2D video material can still be applied independently on the left and right views of a stereo video pair to save valuable bandwidth and storage capacity, though MPEG Ad-Hoc group for 3D audio and video is working on a new standard for efficient multi-view video coding [5]. The perceived quality of an image/video is always important to evaluate the performance of all 3D imaging applications and subjective quality assessment is the most accurate method for it. However, it is time consuming and expensive. In addition, this kind of assessment is not suitable for real time monitoring applications. Therefore, objective quality evaluation method is required to assess the quality. In [6], described the quality of 3D videos stored as monoscopic color videos that augmented by pixel depth map and finally this bit information used for color coding and depth data. In [7], the effect of low pass filtering over a 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 when low pass filtering is used. A comprehensive analysis of the perceptual requirements for 3D TV is made in [8] along with a description of the main artifacts which may arise when dealing with stereo TV. In [9], the researchers identified the effect of camera distance and JPEG coding on overall image quality that includes perceived depth, sharpness, and eye strain. A compound FR stereo-video quality metric is proposed based on H.264 coder with composition of two components: monoscopic quality component and stereoscopic quality component in [10]. The relationship between the perceived image quality and the perceived depth are also discussed in [11].

Human visual perception is very sensitive to edge information. Consequently perceive distortions as well as 3D depth perception should be strongly dependent on local features of stereo video content such as edge (non-plane), and non-edge (plane). Spatial contrast sensitivity of the human visual system (HVS) is low when the video contents are high in speed [12]. Perceptual effects of temporal activity (specifically a rapid change of the video content between adjacent frames) of any video content are also dependent on the local features of a scene. In this work, we proposed a no-reference (NR) continuous video quality assessment model for symmetric and asymmetric codec stereoscopic sequences that use the perceived differences of local features of spatial, temporal, and disparity measures. 3D video quality assessment is required to incorporate multidimensional perceptual factors: depth, 3D video impairments, visual comfort, and the combine effect of these factors reflects overall perceptual quality. In this work, we limit our study to MPEG-2 MP@ML codec videos with different bit rates. The subjective experiment results on our stereo videos dataset are used to train and test the model.

2. Subjective Experiments

We conducted subjective experiment in Media Information and Communication Laboratories (MICT), Toyama, Japan. In the experiment, we used single stimulus continuous quality evaluation (SSCQE) method in which a processed video sequence was presented alone without being paired

with its reference version [13]. Fifteen stereo video clips of 15 seconds each with 640 × 480 pixels, 30 fps progressive were used in the experiment. In order to develop long sequence, we combined all clips together and created 3 minutes 45 seconds sequence. Each sequence contents similar symmetric/ asymmetric coded videos. Subsequently seven symmetric/asymmetric stereo video sequences were created by using MPEG-2 MP@ML encoder with four kinds of bit rates 2, 3, 5, and 8 Mbps. The selected bit rates combinations of left (L) and right (R) sequences are (L, R): (5, 5), (8, 2), (3, 2), (3, 3), (5, 3), (8, 5), and (2, 2) Mbps. Video clips order was same in every sequence [14]. Left view (grey scale) of a reference sequence is shown in Figures 1, 2, and 3. All reference clips were produced by NHK, Japan, and made available for research on stereo video. Each reference clip was 15 seconds length with 1920 × 1035 pixels, 30 fps of 24-bit/pixels RGB color space.

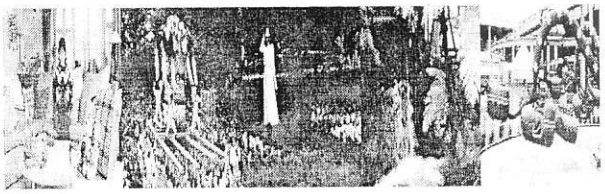

Fig. 1: Left view of reference sequence (0-75) sec

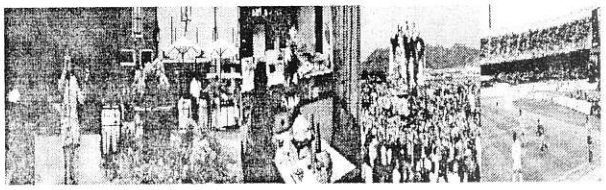

Fig. 2: Left view of reference sequence (76-150) sec

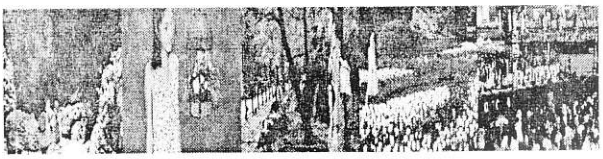

Fig. 3: Left view of reference sequence (151-225) sec

Sixteen non-expert subjects (8 males and 8 females, with average age 23 years) with ages ranging from 20 to 32 are participated in the experiment. Most of them are college/university student and are non-experts in the area of video quality. All subjects are screened prior to participate the session for normal visual acuity with or without glasses, normal color vision, normal stereo depth perception and familiarity with the language. An auto stereoscopic (SANYO) display is used in this experiment to display the stereoscopic video sequences and the subjects are instructed about the limited horizontal viewing angle to perceive 3D video correctly. The subjective test conditions and parameters are summarized in Table 1.

Table 1: Subjective test conditions and parameters

Method	SSCQE
Coder	MPEG-2 MP@ML
Bit Rates	4 kinds (2, 3, 5 and 8 Mbps)
No. of stereo video clips	15 (640×480 pixels) 24-bit/ pixel, RGB
Each clip length	15 sec
No. of test stereo sequences	7 (Each length 3 min 45 s)
Subjects	16 (Non expert, students)
Display	10-inch, LCD 3D Auto stereoscopic
Display resolution	640×480 pixels (LR: 320×480 pixels)
Viewing distance	4H (H = Picture height)
Room illumination	Dark

The subjects were asked to provide their overall perception of quality on a continuous quality scale marked with "Excellent", "Good", "Fair", "Poor", and "Bad". The subjective scores were quantized on a scale of [0...100], 0 being the worst quality and 100 being the best. The slider in the SSCQE test was not a stand-alone hardware device, but a graphical on-screen slider that was steered by moving the mouse up and down, i.e. vertical mouse movements were translated directly into slider shifts. Viewers' familiarity with handling a computer mouse is an additional advantage. SSCQE judgments were given continuously at a sampling rate of 2/s. In order to avoid any recency effects from the previous sequence pair, first clip (Clip-1, 15 sec) voting is rejected in each sequence. Therefore, total 420 samples were collected instead of 450 samples (3 minutes 45 seconds) for each sequence. Mean opinion scores (MOSs) were then computed for each stereo video sequence after the screening of post-experiment results according to ITU-R Rec. 500-10 [13]. Two outlier subjects were detected out of sixteen subjects. Discarding the outliers, the MOS had been computed for each sequence with the 95% confidential interval (CI). Average 95% CI was ± 8.06 for all sequences.

3. NR Stereo Video Quality Assessment

In order to develop our objective model, we considered the HVS characteristics and presumed that the perceive distortion, disparity, and temporal activity of two adjacent frames should be strongly dependent on local features such as edge and non- edge areas of a stereo frame content. Therefore, the perceived differences of local features based spatio-temporal segmentation approach are considered for video quality assessment. A previous instantiation of this approach was made in [14]. In this paper, we generalize this algorithm, and provide a more extensive set of validation results. A block diagram of the proposed method is shown in Figure 4. Two video sequences (left and right views) of a stereoscopic video are converted into frames separately and

only the luminance component is used for simplicity. Temporal segmentation is used for videos partition in temporal segments and sub temporal segments as well. Each temporal segment consists of five sub temporal segments and each sub segment has four consecutive frames. Both temporal and sub temporal segments are partially overlapped (only corner frame is overlapped between two successive segments). For spatial segmentation, we use a block (8×8) based segmentation algorithm to classify edge and non-edge areas of a frame/image that was introduced in [15], [16]. All mathematical measures are calculated individually for each temporal segment. The proposed model consists of three measures:

1. Distortions/Artifacts measure
2. Disparity measure
3. Temporal features measure

Subsequently, the measures are described in the next sub Sections.

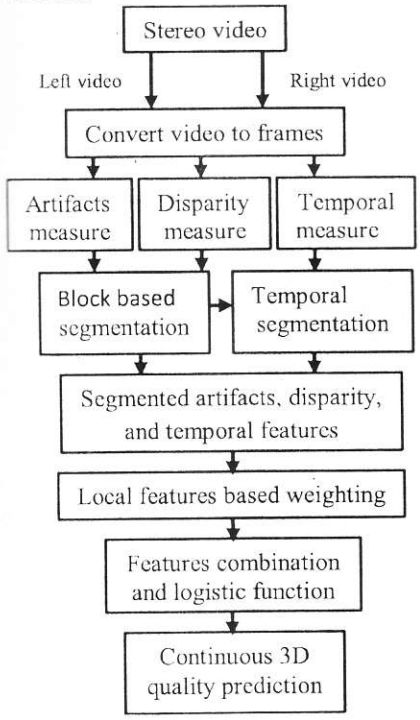

Fig. 4: Proposed NR quality evaluation model.

3.1 Distortions/Artifacts Measure

Since the MPEG-2 coding algorithm is based on the discrete cosine transform (DCT), the degradation/distortion of the frame images are similar to that of JPEG coded images. Consequently, both blocking and blurring artifacts may be created during quantization of DCT coefficients in the coded frame images. Blocking effect occurs due to the discontinuity at block boundaries. 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. Therefore, average edge point detection measures of blocks give more insight into the relative blur in the image. Here, zero-crossing technique is used as an edge detector. Subsequently, local features based blockiness and zero-crossing measures are estimated in this section [17],[18].

Firstly, we calculate blockiness and zero-crossing of each 8×8 block of the stereo frame pair separately for left and right frames. Secondly, we apply the block (8×8) based segmentation algorithm to the left and right frames to classify edge, and non-edge blocks in the frames [15]. Thirdly, we average each value of blockiness and zero-crossing independently for edge, and non-edge blocks of each frame of the stereo pair. And finally, total blockiness and zero crossing for each stereo frame pair is estimated respectively based on the higher blockiness value and lower zero-crossing value between the left and right frames distinctly for edge, and non-edge blocks. The mathematical features, blockiness and zero-crossing measures within each block of the frames are calculated horizontally and then vertically.

For horizontal direction: Let the test frame 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

$$d_h(m,n) = x(m,n+1) - x(m,n) \quad (1)$$
$$n \in [1,N-1] \text{ and } m \in [1,M]$$

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

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

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

For horiz ontal zero-crossing (ZC):

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

$$d_{h-mul}(m,n) = d_{h-sign}(m,n) \times d_{h-sign}(m,n+1) \quad \text{We} \quad (4)$$

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

$$z_h(m,n) = \begin{cases} 1 & if\ d_{h-mul}(m,n) < 0 \\ 0 & otherwise \end{cases} \quad (5)$$

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:

$$ZC_{bh} = \sum_{i=1}^{8}\sum_{j=1}^{8} z_h(i,j) \quad (6)$$

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

For vertical direction: We can also calculate a differencing signal along each vertical line:

$$d_v(m,n) = x(m+1,n) - x(m,n) \quad (7)$$
$$n \in [1, N] \text{ and } m \in [1, M-1]$$

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:

$$B_b = \frac{B_{bh} + B_{bv}}{2}, \; ZC_b = \frac{ZC_{bh} + ZC_{bv}}{2} \quad (8)$$

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

$$Bl_e = \frac{1}{N_e} \sum_{b=1}^{N_e} B_{be} \quad (9)$$

$$Bl_n = \frac{1}{N_n} \sum_{b=1}^{N_n} B_{bn} \quad (10)$$

Where Ne, and Nn are respectively the number of edge, and non-edge blocks of the frame. Similarly, the average blockiness values of Br_e, and Br_n for the right frame are calculated.

Subsequently, the average zero-crossing values of ZCl_e, and ZCl_n for the left frame are estimated by:

$$ZCl_e = \frac{1}{N_e} \sum_{b=1}^{N_e} ZC_{be} \quad (11)$$

$$ZCl_n = \frac{1}{N_n} \sum_{b=1}^{N_n} ZC_{bn} \quad (12)$$

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

$$B_{e/n}(Bl, Br) = \max(Bl, Br) \quad (13)$$

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

$$ZC_{e/n}(ZCl, ZCr) = \min(ZCl, ZCr) \quad (14)$$

Let Be_f and ZCe_f be the total blockiness and zero-crossing of edge areas of a stereo frame pair, respectively. Then we calculate these two features for each sub-temporal segment by the following equations:

$$Be_s = \frac{1}{F} \sum_{f=1}^{F} Be_f \quad (15)$$

$$ZCe_s = \frac{1}{F} \sum_{f=1}^{F} ZCe_f \quad (16)$$

where Be_s and ZCe_s represent blockiness and zero-crossing for a sub-temporal segment of edge areas. And "F" denotes number of frames in a sub segment, here F = 4. Subsequently, we estimate the blockiness (Be_t) and zero-crossing (ZCe_t) of a temporal segment of edge areas by the following equations:

$$Be_t = \frac{1}{S} \sum_{s=1}^{S} Be_s \quad (17)$$

$$ZCe_t = \frac{1}{S} \sum_{s=1}^{S} ZCe_s \quad (18)$$

where "S" represent number of sub-temporal segment in a temporal segment. Here S = 5. Similarly, the blockiness (Bn_t) and zero-crossing (ZCn_t) features for non-edge areas are computed. Lastly, the overall blockiness (B_t) and zero-crossing (Z_t) for each temporal segment of stereo frame pairs are calculated by

$$B_t = Be_t^{w1} \cdot Bn_t^{w2} \quad (19)$$

$$Z_t = ZCe_t^{w3} \cdot ZCn_t^{w4} \quad (20)$$

where w1, and w2 are the weighting factors for the blockiness of edge, and non-edge areas and also w3, and w4 are the weighting factors for zero-crossing.

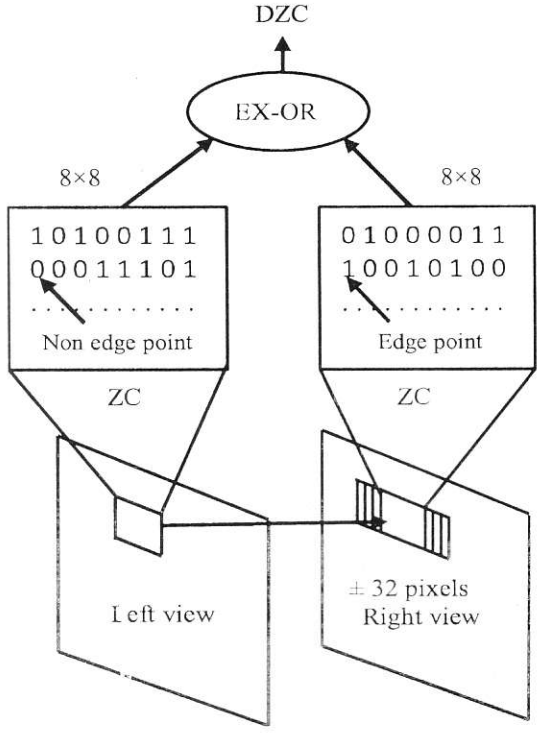

Fig. 5: Disparity estimation approach.

3.2 Disparity Measure

In this section, a block-based edge difference approach is used for disparity estimation. Although, many features based approaches are used for stereo matching, 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 frame into non overlapping 8×8 blocks with classification of edge and non-edge blocks. For each block of the left frame, 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 frame.

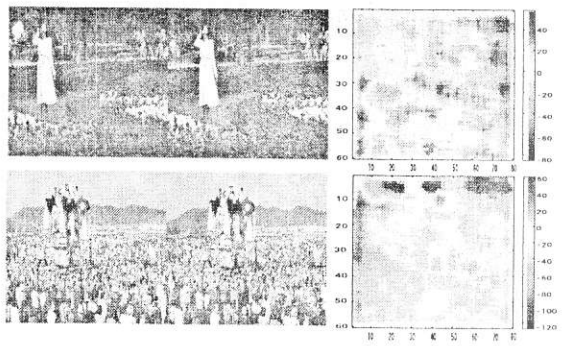

Fig. 6: Stereo frame pairs and its depth map.

The disparity estimation approach is shown in Figure 5. Here zero-crossing (horizontal and vertical) of a block is estimated according to Section 3.1. "1", and "0" indicate zero-crossing (edge) and non zero-crossing (non-edge) points, respectively. 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 two sample stereo frame pairs are shown in Figure 6. Colors in the depth maps that are indicated by vertical color bars in right are estimated depths of the frames pairs. Although disparity means a measure of position displacement between the left and right frames, a difference zero-crossing rate is determined between the block of left frame and the corresponding searching block in the right frame as relative disparity. The zero-crossing rate values are then averaged separately for edge and non-edge areas of a stereo frame pair. Let ZCl_b, and ZCr_b be the zero-crossing of a block of left frame and the corresponding searching block of right frame, respectively. The difference zero-crossings of the block can be estimated by the following equation:

$$DZC_b = ZCl_b \oplus ZCr_b \quad (21)$$

Thus, we can calculate difference zero-crossing rate of the 8×8 block by

$$DZ_b = \frac{1}{8\times 8}\sum DZC_b \quad (22)$$

Subsequently, the average difference zero-crossing rates of edge, and non-edge areas of the left frame are calculated by

$$DZ_e = \frac{1}{N_e}\sum_{b=1}^{N_e} DZ_{be} \quad (23)$$

$$DZ_n = \frac{1}{N_n}\sum_{b=1}^{N_n} DZ_{bn} \quad (24)$$

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

Thus, we calculate these disparity features for a sub-temporal segment by the following:

$$DZe_s = \frac{1}{F}\sum_{f=1}^{F} DZe_f \quad (25)$$

$$DZn_s = \frac{1}{F}\sum_{f=1}^{F} DZn_f \quad (26)$$

where DZe_s and DZn_s are respectively relative disparity of edge and non-edge areas of a sub-temporal segment. Here F = 4 (number of frame per sub-temporal segment). Accordingly, we estimate the disparity features (DZe_t) of edge areas for a temporal segment by using the highest sub-temporal feature:

$$DZe_t = \max(DZe_s); \quad s = 1, 2, 3,..... \quad (27)$$

where "s = 5" denotes number of sub-temporal segment of a temporal segment. Similarly, we can calculate DZn_t for non-edge area. Finally, the overall disparity feature per temporal segment is estimated by:

$$DZ_t = DZe_t^{w5} \cdot DZn_t^{w6} \quad (28)$$

where w5, and w6 are respectively the weighting factors of the disparity features of edge, and non-edge areas.

3.3 Temporal Features Measure

In order to measure temporal features, segmented local features based temporal information of sub temporal segment are estimated separately for left and right videos. Block (8×8) based segmentation algorithm is applied only for the first frame of each sub temporal segment to classify the edge and the non-edge blocks. The temporal feature of edge area (TIe_s) of a sub-temporal segment is calculated by:

$$TIe_s(m,n,t_k) = \frac{1}{N_e}\sqrt{\frac{1}{64}\sum\left|x_e(m,n,t_f) - x_e(m,n,t_{f+k})\right|^2} \quad (29)$$

$$TIe_s = \sum_{n=1}^{3} TIe_s(m,n,t_k) \quad (30)$$

where t_f and t_{f+k} denote first and successive frames of a sub temporal segment, respectively and k = 1, 2, 3. Ne is the number of edge blocks of the first frame and the frame signal of edge area is denoted by $x_e(m,n)$. Similarly we can calculate TIn_s for non-edge area. Subsequently, we compute the temporal features (TIe_t) of edge areas for a temporal segment by using the highest sub-temporal feature:

$$TIe_t = \max(TIe_s); \quad s = 1, 2,.....5 \quad (31)$$

Similarly, we can calculate TIn_t for non-edge area. As we consider the highest sub-temporal feature, therefore if scene cut exists in a temporal segment the feature can easily detect the high variation of scene content/motion between two adjacent clips.

Let $LTIe_t$, $RTIe_t$ and $LTIn_t$, $RTIn_t$ be the temporal segment features of a temporal segment of left and right videos for edge and non-edge areas, respectively. Thus, the temporal segment features for edge and non-edge areas of a stereo video are calculated by:

$$TIt_e = \frac{LTIe_t + RTIe_t}{2} \tag{32}$$

$$TIt_n = \frac{LTIn_t + RTIn_t}{2} \tag{33}$$

where TIt_e and TIt_n are the temporal segment features for edge and non-edge areas. Finally, the overall temporal feature per temporal segment is estimated by:

$$TI_t = TIt_e^{w7} \cdot TIt_n^{w8} \tag{34}$$

where w7, and w8 are respectively the weighting factors of the temporal features of edge, and non-edge areas.

Features Combination

In order to combine disparity, distortions, and temporal features to constitute a stereo video quality assessment model we consider the following equation:

$$S = \alpha(DZ_t) + \beta B_t Z_t + \gamma(TI_t) \tag{35}$$

where α, β, and γ are the method parameters. The model parameters and the weighting factors (w1 to w8) are must be estimated by an optimization algorithm with the subjective test data. Here, Particle Swarm Optimization (PSO) algorithm is used for optimization [19]. The proposed model performance is also studied without disparity by the following features combine equation:

$$S = \beta B_t Z_t + \gamma(TI_t) \tag{36}$$

We consider a logistic function as the nonlinearity property between the human perception and the physical features. Finally, the obtained MOS prediction, MOSp, per a temporal segment is derived by the following equation [20].

$$MOS_p = \frac{99}{1 + \exp[-1.0217(S - 50)]} + 1 \tag{37}$$

Table 2: Model parameters and weighting factors

α = -30.4057	β = 14.3223	γ = 68.6298
w_1 = -0.0013	w_2 = -0.0105	w_3 = -0.0113
w_4 = -0.0352	w_5 = -0.0284	w_6 = 0.0367
w_7 = -0.0118	w_8 = 0.0037	

Results

To verify the performance of our proposed model, we consider our stereo video dataset (SSCQE MOS scale, 0-100, see Section 2) and divide the dataset into two parts for training and testing. The training dataset consists of four (Seq-1, Seq-2, Seq-3, and Seq-4) symmetric/asymmetric coded stereo video sequences (from the total seven). The sequences' bit rates combinations are LR: (5, 5), (8, 2), (3, 2), and (3, 3) Mbps, respectively. The testing dataset consists of the others three symmetric/asymmetric coded stereo sequences (Seq-5, Seq-6, and Seq-7) and also there is no overlapping between training and testing. The testing sequences' bit rates combinations are LR: (5, 3), (8, 5), and (2, 2) Mbps, respectively. The model's parameters and weighting factors are obtained by the PSO algorithm with all of our training sequences are shown in Table 2.

Table 3: Evaluation results for all sequences

Model	Training
	OR
With disparity (WD)	0.0268
Without Disparity (WOD)	0.0393
	Testing
With disparity (WD	0.0547
Without Disparity (WOD)	0.0563

Table 4: Evaluation results for training and testing

Seqs.	Bit rate (Mbps)	Ave. 95% CI	Training	
			WD	WOD
			OR	OR
Seq-1	LR(5,5)	±8.875	0.0476	0.0381
Seq-2	LR(8,2)	±7.933	0.0048	0.0571
Seq-3	LR(3,2)	±7.879	0.0381	0.0452
Seq-4	LR(3,3)	±8.705	0.0167	0.0167
			Testing	
Seq-5	LR(5,3)	±7.404	0.0214	0.0214
Seq-6	LR(8,5)	±8.482	0.0048	0.0095
Seq-7	LR(2,2)	±7.126	0.1381	0.1381

Although Pearson linear Correlation Coefficient is one of the standard performance evaluation procedures for quality assessment, it is not a suitable criteria for continuous quality prediction. Because every individual human observer's response time is slightly different that makes slight temporal variation of their opinion of picture quality in SSCQE method. However, the objective prediction samples are very particular to the temporal segments of a stereo video sequence. Therefore, point to point correlation is not an appropriate measure between subjective and objective samples specifically in continuous quality prediction. Subsequently, we believe that outlier ratio (OR) is the most important evaluation criteria for continuous quality prediction. Here, we follow OR as a standard performance

evaluation metric between objective (MOSp) and subjective (MOS) scores [21]. The evaluation results for training and testing sequences are summarized in Table 3. It has been observed from Table 3 that the evaluation metric, OR is quite sufficient. Specifically, our proposed model provides sufficient prediction consistency (lower OR). It has also been observed from Table 3 that our model performance with disparity (WD) is slightly better compared to without disparity (WOD). The evaluation results for every individual sequences are summarized in Table 4. The continuous MOS prediction (MOSp) for every sequence with 95%CI and MOS are shown in Figures 7 to 13. Figures 8 to 14 and Table 4 indicate that the model's continuous prediction consistency is sufficient except the test sequence, Seq-7, LR: (2, 2) Mbps in Figure 14. In the Figure, the two major miss prediction areas are marked by a circle. The circles' corresponding clips are "Amusement park" and "Football". Both video clips are in high motion (i.e., video content changes of adjacent frames in the clips is too much) and camera work of these two videos are also high. Noise increases and decreases rapidly within a very short time because of low encoding bit rates. Therefore, subject could not identify the high quality frames. Similarly, in our model the calculated TI features within the sub-temporal segments could not follow the fast changing TI features within the sub-temporal segments. In order to take into account the temporal dependency between the sub-temporal segments, a weighting function can be considered to improve the performance.

5. Conclusion

A continuous NR objective quality assessment model is proposed for MPEG-2 MP@ML coded stereoscopic videos based on spatio-temporal segmentation that use the perceptual differences of local features such as edge and non-edge. Spatial distortions and disparity measures of a stereoscopic pair frame are calculated based on aforementioned features. Local and temporal segmentation based temporal features have been estimated for the left and right video sequences. Distinct weighting factors for each of those local features are then applied to measure the total distortion, disparity, and temporal features for each temporal segment. We verify the performance of our proposed model on our subjective stereo dataset, which indicates sufficient quality prediction performance. Future research can be extended by generalizing the approach irrespective of encoders as well as developing a suitable weight function for each features of temporal segment.

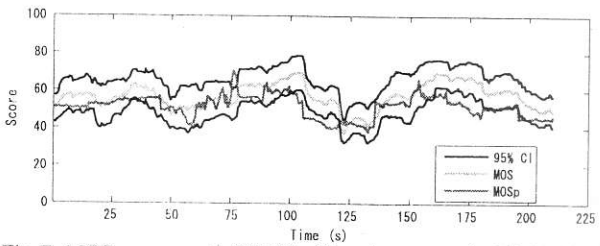

Fig. 7: MOSp scores with 95%CI of Seq-1, symmetric: LR (5, 5) Mbps

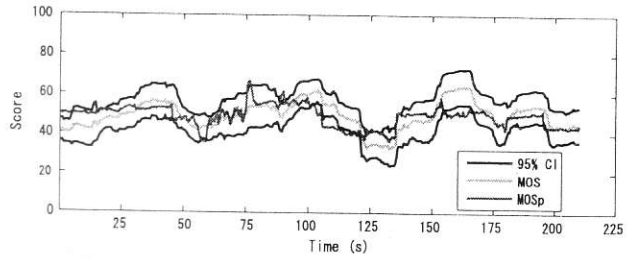

Fig. 8: MOSp scores with 95%CI of Seq-2, symmetric: LR (8, 2) Mbps

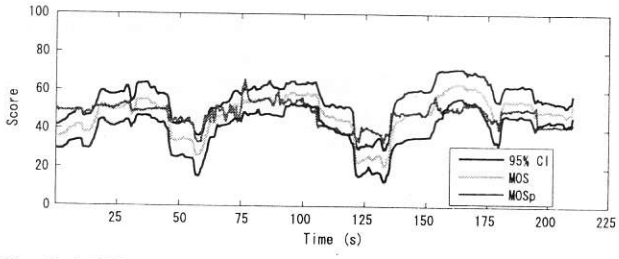

Fig. 9: MOSp scores with 95%CI of Seq-3, symmetric: LR (3, 2) Mbps

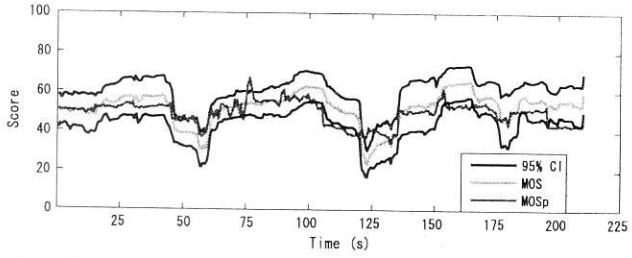

Fig. 10: MOSp scores with 95%CI of Seq-4, symmetric: LR (3, 3) Mbps

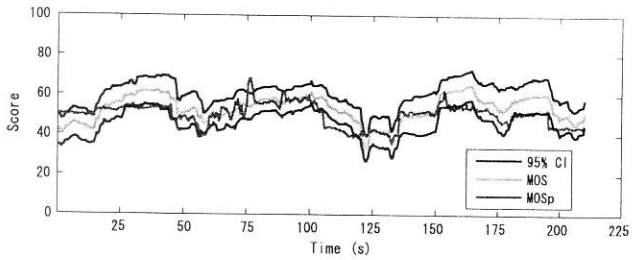

Fig. 11: MOSp scores with 95%CI of Seq-5, symmetric: LR (5, 3) Mbps

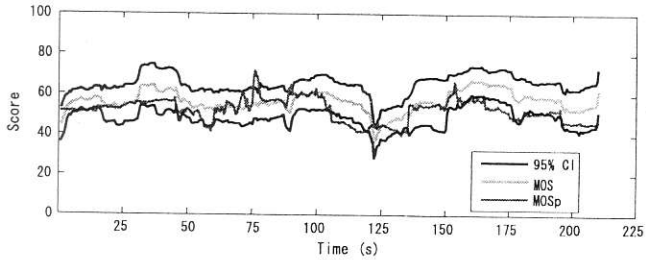

Fig. 12: MOSp scores with 95%CI of Seq-6, symmetric: LR (8, 5) Mbps

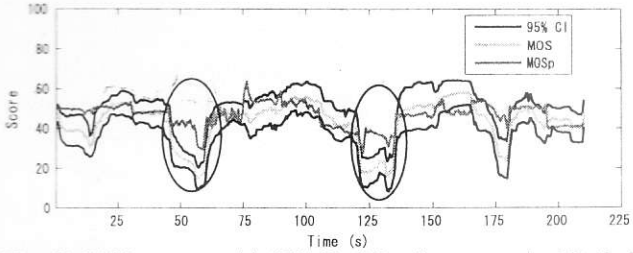

Fig. 13: MOSp scores with 95%CI of Seq-7, symmetric: LR (2, 2) Mbps

Acknowledgment

Author would like to thank NHK, Japan for providing the reference stereo video clips.

References

1. 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. F. Westin, "Extraction brain connectivity from diffusion MRI," IEEE Signal processing magazine, vol. 24, no. 6, pp. 124-152, 2007.
3. M. William, and D. L. Bailey, "Stereoscopic visualization of scientific and medical content," in Proc. SIGGRAPH'06, Boston, Mass, USA, Jul.-Aug. 2006.
4. N. Dodgson, "Auto stereoscopic 3-D displays," IEEE Computer, vol. 38, no. 8, pp. 31-36, Aug. 2005.
5. 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.
6. Tikanmaki, and A. Gotchev, "Quality assessment of 3D video in rate allocation experiments," in Proc. IEEE ISCE, Algarve, Portugal, Apr., 14-16, 2008.
7. 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.
8. L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens, "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.
9. 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 Trans. on Applied perception, vol. 3, no. 2, pp. 95-109, April 2009.
10. 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.
11. A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G. B. Akar, "Towards compound stereo-video quality metric: a specific encoded-based framework," in Proc. IEEE SSIAI, Denver, USA, Mar. 26-28, 2006.
12. F. Yang, S. Wan, Y. Chang, and H. R. Wu, "A novel objective no-reference metric for digital video quality assessment," IEEE Signal Processing Letter, vol. 12, no. 10, pp. 685-688, Oct. 2005.
13. ITU-R BT.500-10. Methodology for the Subjective Assessment of the Quality of Television Pictures.
14. Z. M. Parvez Sazzad, S. Yamanaka, and Y. Horita, and J. Baltes, "Continuous stereoscopic video quality evaluation,"in Proc. SPIE, vol. 7524, Jan. 18-21, San Jose, USA, 2010.
15. Z. M. Parvez Sazzad, S. Yamanaka, Y. Kawayoke, and Y. Horita, "Stereoscopic image quality prediction," in Proc. IEEE QoMEX, San Diego, CA, USA, July 29-31, 2009.
16. Z. M. Parvez Sazzad, S. Yamanaka, and Y. Horita, "Spatio-Temporal Segmentation Based Continuous No-Reference Stereoscopic Video Quality Prediction,", Proc IEEE QoMEX, Trondheim, Norway, June 21-23, 2010.
17. Z. M. Parvez Sazzad, Rafik Bensalma, Mohamed Chaker Larabi, "Feature based no-reference continuous video quality prediction model for coded stereo video," Proc. CGIV 2012: Sixth European Conference on Colour in Graphics, Imaging, and Vision, Amsterdam, Netherlands, May 6-9, 2012.
18. Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-Reference perceptual quality assessment of JPEG compressed images," in Proc. IEEE ICIP, pp. I-477-480, Sept. 2002.
19. J. Kennedy and R. Eberhart, "Particle Swarm Optimization," Proc. IEEE ICNN, Perth, Australia, pp. 1942-1948, Nov. 1995.
20. Z. M. Parvez 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.
21. 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/.