An Objective No-Reference Perceptual Quality Assessment Metric based on Temporal Complexity and Disparity for Stereoscopic Video

Kwangsung Ha[1], Sung-Ho Bae[2], and Munchurl Kim[2]

[1] Department of Software Platform R&D at LG Electronics / Seoul, Korea kwangsung.ha@lge.com
[2] Department of Electrical Engineering, Korea Advanced Institute of Science and Technology / Daejeon, Korea
s.h.bae@kaist.ac.kr, mkim@ee.kaist.ac.kr

***** Corresponding Author: Munchurl Kim

Received July 14, 2013; Revised July 29, 2013; Accepted August 12, 2013; Published October 31, 2013

** Regular Paper*

***Abstract*:** 3DTV is expected to be a promising next-generation broadcasting service. On the other hand, the visual discomfort/fatigue problems caused by viewing 3D videos have become an important issue. This paper proposes a perceptual quality assessment metric for a stereoscopic video (SV-PQAM). To model the SV-PQAM, this paper presents the following features: temporal variance, disparity variation in intra-frames, disparity variation in inter-frames and disparity distribution of frame boundary areas, which affect the human perception of depth and visual discomfort for stereoscopic views. The four features were combined into the SV-PQAM, which then becomes a no-reference stereoscopic video quality perception model, as an objective quality assessment metric. The proposed SV-PQAM does not require a depth map but instead uses the *disparity* information by a simple estimation. The model parameters were estimated based on linear regression from the mean score opinion values obtained from the subjective perception quality assessments. The experimental results showed that the proposed SV-PQAM exhibits high consistency with subjective perception quality assessment results in terms of the Pearson correlation coefficient value of 0.*808*, and the prediction performance exhibited good consistency with a zero outlier ratio value.

***Keywords*:** No-reference, Stereoscopic video, Subjective quality assessment, Visual fatigue

1. Introduction

According to the increasing demand for realistic broadcasting, which includes presence and realism, 3-Dimensional (3D) video has attracted considerable attention recently. Therefore, 3DTV is expected to be a promising next-generation broadcasting service to maximize the sensation of presence by providing an additional dimension, depth. On the other hand, a number of technical problems need to be solved before a successful 3DTV broadcast service can be achieved. Among these problems, visual discomfort/fatigue problems, such as vomiting, sickness and dizziness caused by viewing 3D video have become important issues. Therefore, 3D video quality and visual comfort evaluation tools are essential for predicting the above problems for the safety of the 3D contents.

Early studies on stereoscopic quality assessments have used the existing video quality metrics for 2D video to stereoscopic video [1, 2]. In [1], the video quality and depth perception were estimated using the 2D quality metrics of PSNR (peak signal-to-noise ratio), SSIM (structural similarity index metric) and VQM (video quality model). The disparity information, however, was not considered, resulting in relatively low performance. Therefore, the stereoscopic video quality model in [2] improved the model in [1] by incorporating the disparity information into conventional 2D quality metrics. On the other hand, the stereoscopic video quality metric in [2] measured the level of degradation of the disparity map using a 2D quality metric in [1], which might not be appropriate for considering the characteristics of the

human visual system (HVS) in stereoscopic vision. Recently, many stereoscopic video quality metrics have been developed using the HVS characteristics of stereoscopic vision [3-5]. Because human visual perception is quite sensitive to edge information and perceived image distortions are strongly dependent on their local features, such as edge/non-edge areas, the model in [3] considered the disparity information based on the local edge characteristics. The evaluation result showed that the prediction accuracy of the model in [3] is not sufficiently accurate, because the model in [3] did not fully consider the effects of a disparity on HVS. In [4], a stereoscopic video quality model was prepared by considering the disparity information with spatial/ temporal complexity and depth position in stereoscopic videos. The visual fatigue model in [4] is difficult to apply when the depth map is unavailable, and it also does not fully consider the various factors that affect the stereoscopic visual quality. Therefore, the model in [4] resulted in low performance, where the Pearson correlation coefficient (PCC) between the measured quality values and estimated quality values was only 0.505.

This study examined some essential features that can affect the perceived depth and visual comfort for video quality assessments on a stereoscopic video by analyzing the subjective quality assessments for various stereoscopic videos. The essential features were then combined to develop a NR SV-PQPM.

This paper is organized as follows: Section 2 explains the experimental method for the subjective quality assessments on stereoscopic video. Section 3 introduces the proposed quality perception model to assess the stereoscopic video in detail. Section 4 presents the experimental results and Section 5 reports the conclusions for the proposed quality perception model for the stereoscopic video, and addresses some issues to further improve the performance of the proposed quality perception model.

2. Subjective Quality Assessment

The performance of an objective quality assessment metric was evaluated against the subjective quality assessment results. Therefore, the experiments for the subjective quality assessment are needed. For this, a subjective quality assessment for stereoscopic video was conducted. This section describes an experimental environment for a subjective quality assessment.

2.1 Experimental Environment

For the subjective quality assessment, for stereoscopic video, a 46” polarized stereoscopic display (Hyundai S465D), which supports stereoscopic video up to a HD resolution of 1,920×1,080 pixels was used. The subjective perceptual quality assessments for the stereoscopic video were carried out on twenty subjects. Each subject was seated in front of the monitor in a viewing distance of approximately 3H (2m). The luminance, angle of viewing, and brightness of the laboratory were set to follow the recommendation by ITU-R BT.500-11 [6] under the experimental condition.

Twenty subjects (observers) consisting of ten males and ten females (mean age, 25 years; range 20 to 35 years), who participated in the experiments for the subjective perceptual quality assessment on stereoscopic video. All of the subjects had some experience of 3D movie viewing, but they were not directly concerned with the 3D perceptual quality assessment as part of their normal works, and were not experienced stereoscopic quality assessors. Before the subject perceptual quality assessment tests, preliminary subjective tests were performed to evaluate the subject's 3D vision performance using well-made 3D animations. The results showed that all the subjects perceived the depth in the stereoscopic video contents well.

In this study, the aim was not to estimate the quality of the stereoscopic video against their reference video, but to estimate the absolute perceptual qualities of the stereoscopic video under a test. The Single Stimulus (SS) method of ITU-R BT.500-11 was used in these experiments because each evaluation session should not take more than 30 minutes.

In the SS methods, the subjects are asked to assess each video sequence in the stimulus set individually. In the case of a sequence, the subjects provide a score for the entire presentation. A typical assessment trial consists of three displays: a mid-grey adaptation field, a stimulus, and a mid-grey post-exposure field. The duration of these displays are 3, 10 and 15 seconds, respectively, to allow sufficient time for the perceptual quality assessment on the stereoscopic video sequences under the test.

Because each subject uses its own internal scale to construct their judgment, the results of the subjective quality assessment can vary from one person to another. To obtain reliable experimental results, the untrustworthy subjects are screened out according to the guidelines described in ITU-R BT.500-11. For this, the distribution of the test scores is checked to determine if it follows a normal distribution or not by using the Kurtosis coefficient. The subjects are considered untrustworthy if their scores are outside the 95% confidence interval. In this experiment, two of the twenty subjects were rejected as untrustworthy subjects, leaving the results of eighteen subjects for analysis.

2.2 Test Sequences for Stereoscopic Video in Subjective Perceptual Quality Assessments

For ground-wave 3DTV broadcasting, the left and right contents of the stereoscopic video are transmitted and serviced within 6MHz bandwidth of a ground-wave DTV channel in Korea [7]. It is presupposed that the left and right video should be provided at the same level of the existing full HDTV quality [8]. In other words, to ensure backward compatibility, the quality of monoscopic video that the existing DTV can provide as “HDTV” service must be kept. Therefore, the test sequences of stereoscopic video, which are of HD resolution, were used. Table 1 lists the test sequences of the stereoscopic video used in the experiments for subjective perceptual quality assessment.

Table 1. Stereoscopic Video Used for the Experiments.

Sources	Names	Resolutions	Lengths
KBS	Daegu athletics	1,920×1,080/ 30fps	20 min
KBS	Ssireum	1,920×1,080/ 30fps	6 min
Korean Film Council	Stereoscopic video collection	2,560×1,280/ 30fps	15 min
Korean Film Council	Nail	1,920×1,080/ 30fps	12 min
EPFL	Sofa, Bike, Feet, Hallway, Notebook, Car [9]	1,920×1,080/ 25fps	250 frames
GIST	Café [10]	1,920×1,080/ 30fps	300 frames
Nagoya Univ.	Pantomime [11]	1,280×960/ 30fps	500 frames
Poznan	Carpark, Hall1, Hall2, Street [12]	1,920×1,080/ 25fps	300 frames

For the set of stereoscopic video sequences in Table 1, 105 test stereoscopic video sequences of 10 sec were extracted to obtain the test stereoscopic video contents with a range of characteristics. The extracted test sequences of the stereoscopic video clips excluding *Pantomime* and *Stereoscopic video collection* were all converted to a resolution of 1,920×1,080 in YUV420 format. The original version was used for the *Pantomime* stereoscopic sequence. For the *Stereoscopic video collection* sequence, it was cut and down-sampled to the full HD resolution.

2.3 Categorization of Stereoscopic Video Contents

The spatial and temporal characteristics of the video affect the perceived subjective quality of 2D and 3D video. Furthermore, the quality of stereoscopic video is affected significantly by the level of disparity. To prepare a more reliable set of test stereoscopic video sequences, the set of test stereoscopic video sequences were categorized into 27 different classes according to the spatial variance (SV) as spatial complexity, the temporal variance (TV) as temporal complexity, and depth (D). The dynamic range of the values for each feature was divided into three intervals (low, middle and high) giving a total of 27 categories in the combinations of SV, TV and D.

The Edge Histogram Descriptor (EHD) of MPEG-7 [13] was used to calculate the SV values for the stereoscopic video contents, which is defined as

$$SV = \frac{1}{N \times J} \sum_{i} \sum_{j} S_{ij} \qquad (1)$$

where N and J indicate the number of frames and edge types, respectively. S_{ij} is the average histogram value for the j^{th} edge type in i^{th} frame. The threshold values to determine the high, middle, or low SV and TV are experimentally set. The low, middle and high SV categories correspond to the intervals of SV values less than 0.4, between 0.4 and 0.8, and higher than 0.8, respectively. Fig. 1 shows the categories of the three typical stereoscopic video frames for the low, middle and high SV categories.

For the temporal complexity measure, the average magnitude of the motion vectors was used as TV. The motion vectors were calculated based on blocking matching for each block of 16×16. The low, middle, and high TV categories correspond to intervals less than 2.5, ranging from 2.5 to 4.0, and greater than 4.0, respectively. Fig. 2 shows the categories of the three typical stereoscopic video frames for the low, middle and high TV categories.

The disparity between the left and right frames was calculated for each block of 16×16 based on block matching in the horizontal directions. The disparity values were normalized in the range between 0 and 255 with zero disparity of 128. Subsequently, each block was classified

Low (~0.4)

Middle (0.4~0.8)

High (0.8~)

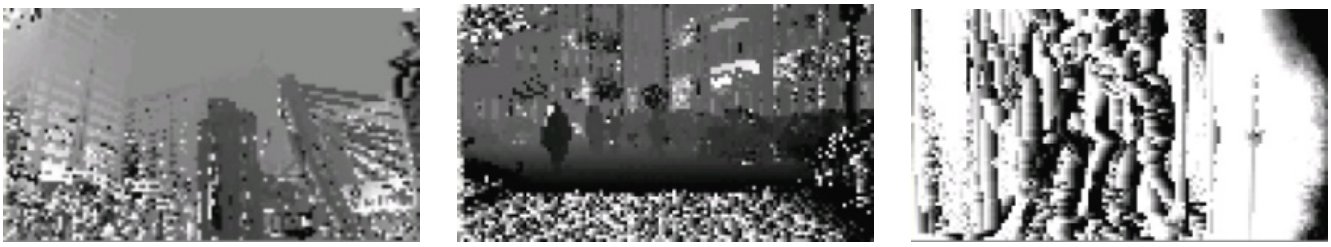
Fig. 1. Categories of the test sequences by SV.

Low (~2.5) Middle (2.5~4.0) High (4.0~)

Fig. 2. Categories of test sequences by TV.

Table 2. Disparity-Based Sequence Categorization.

Disparity Range	Block-level disparity categories	Sequence-level disparity categories		
0~255	0~ 51: Large Negative disparity			
	52~102: Small Negative disparity		Middle	High
	103~153: Zero disparity	Zero		
	154~204: Small Positive disparity			
	205~255: Large Positive disparity			



Zero Middle High

Fig. 3. Categories of the test sequences according to the disparity level.

into one of the five block-level disparity categories – large negative, small negative, zero, small positive and large positive disparities. The test stereoscopic video sequence were categorized into one of the zero, middle and high sequence-level disparity categories based on the most dominant percentage of block-level disparity categories in the sequence. Table 2 lists the disparity-based sequence categorization.

Fig. 3 shows three categories of disparity images estimated for the test stereoscopic video sequences.

Fig. 4 shows the distribution of the 105 test stereoscopic video sequences in 27 categories in the combinations of SV, TV and D. Fig. 5 shows that there are no test sequences that belong to the category of low SV and middle TV due to the absence of a sequence with such SV and TV characteristics in the experiments.

2.4 Analysis on Experimental Results

The disparity information was extracted and compared with the subjective quality scores to analyze the relationship between the features of the disparity and stereoscopic video quality. Section IV describes the

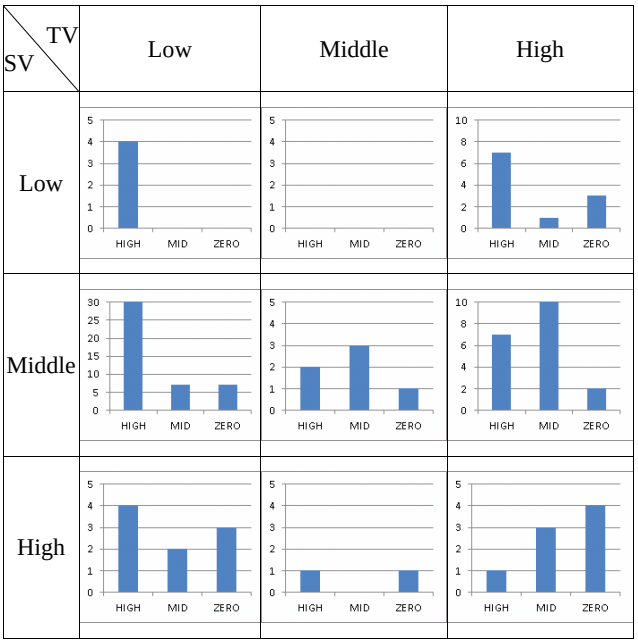


Fig. 4. Distribution of available test sequences.

Fig. 5. Snapshot of stereoscopic video that results in high MOS with smooth change in disparity (ed note: An article is not needed as the first word of a figure/table caption).

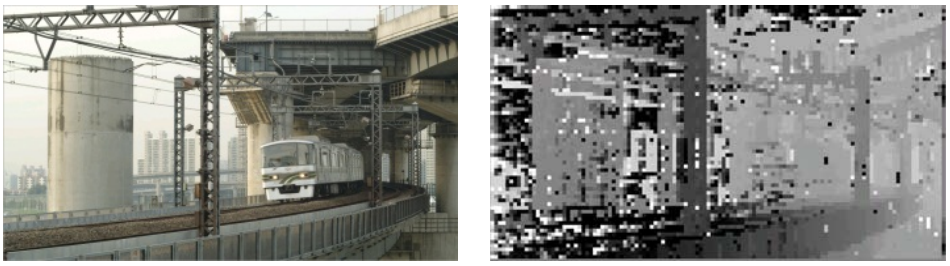


Fig. 6. Snapshot of stereoscopic video that results in high MOS with the objects giving depth perception from near distance to far distance.

Fig. 7. Snapshot of stereoscopic video that results in high MOS with objects of attention in the middle of frames.

extraction of the depth feature.

To examine the effect of the disparity feature characteristics of stereoscopic video on the subjective perception quality assessment results, the estimated disparity images were examined, in which the areas at long distances from the stereoscopic cameras are depicted with positive ('+') disparity, which correspond to the brighter areas, and the areas in close proximity from the stereoscopic camera are indicated as having a negative ('-') disparity, which corresponds to the darker areas.

In the following, some scene characteristics that affect the subjects giving good MOS scores are examined. Firstly, for the stereoscopic video sequences with the positive and negative disparity area connected smoothly, the resulting MOS scores were high. Fig. 5 presents a snapshot of a stereoscopic video sequence that results in high MOS value of 4.5 points. The estimated disparity in the right figure changes smoothly.

Secondly, the subjects tend to give high MOS scores when the objects that help perceive the perspective are positioned at fixed locations from a close distance to a long distance to the stereoscopic cameras. Fig. 6 gives a snapshot of the Stereoscopic video collection sequence, where the power-line towers of trams and the bridge posts are located by giving a good perspective to the content. The resulting MOS score was 4.44 points.

Lastly, high score values tend to be expected when the people or objects that can attract the subject's attention are located in the middle of the frames. Fig. 7 gives a snapshot of the *Daegu athletics* stereoscopic video that results in a high MOS score of 4.22 points, where there is a concentrated object (camera man) in the middle of the frames, which presumably attracts the subject's attention.

In the following, some other scene characteristics that affect the subjects to give low MOS scores were also examined.

First, low MOS scores with an average of 3.0 points or below can be obtained when the difference in the disparity contrast in stereoscopic video is quite large. Fig. 8 presents a snapshot of the *Hallway* stereoscopic video that results in a low MOS score of 3 points, in which the difference in the disparity contrast is quite large. Fig 8 shows the large positive and large negative disparities.

Secondly, low MOS scores with an average of 3.0 or lower are also likely to be obtained when the frame boundary areas in the stereoscopic video frames are perceived to be protruded excessively. This is similar to the paradoxical stereoscopic window affect, which causes

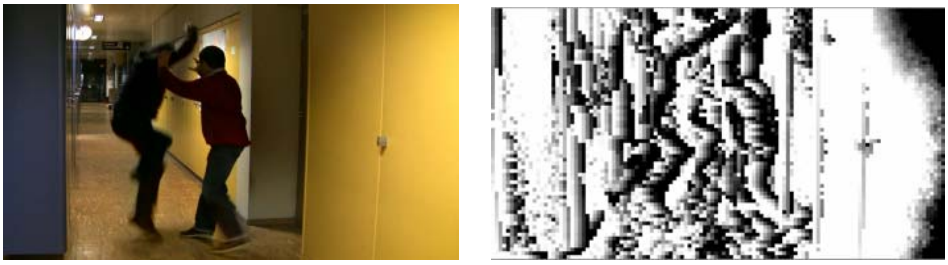


Fig. 8. Snapshot of stereoscopic video that results in low MOS when the difference of the disparity contrast is very large.

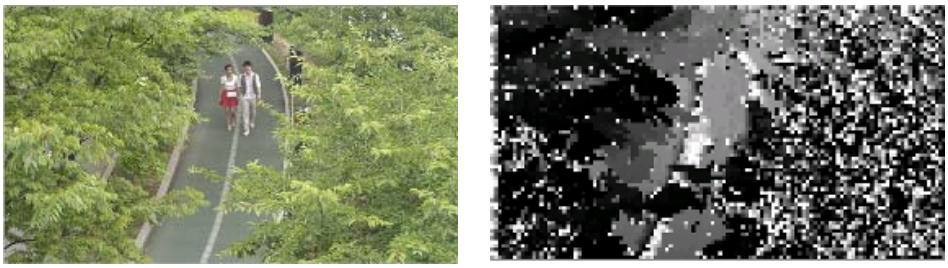


Fig. 9. Snapshot of the stereoscopic video that results in low MOS obtained when the frame boundary areas in the stereoscopic video frames are perceived to be excessively protruded.

Fig. 10. Snapshot of stereoscopic video that results in low MOS when the perceived depth of the video changes drastically in the temporal direction.

visual fatigue due to the collision of occlusion depth cue and parallax depth cue in the frame of 3DTV [14]. Fig. 9 shows a snapshot of a stereoscopic video that results in a low MOS obtained when the frame boundary areas in the stereoscopic video frames are perceived to be protruding excessively.

Visual fatigue occurs when the perceived depth of the stereoscopic video changes drastically in the temporal direction. Fig. 10 presents a snapshot of the *Daegu athletics* stereoscopic video that results in a low MOS score of 2.28 points. In the right figure in Fig. 10, the quarter-sized figures from the top left to the bottom right indicate the disparity estimates in time. Large changes in the disparity estimates in time exist, leading to small MOS scores for a stereoscopic perception quality assessment.

3. Proposed Stereoscopic Video Perceptual Quality Assessment Model (SV-PQAM)

In this section, an NR quality assessment model for stereoscopic video is proposed and explained in detail. The proposed stereoscopic video quality perception model (SV-PQAM) predicts the quality of the original stereoscopic image without considering the 2D artifacts. The methods for extracting the features that affect the stereoscopic video quality are explained based on an analysis of the results of the subjective quality assessment experiments.

3.1 Feature extractions for SV-PQAM

Disparity estimation

When depth information is unavailable in a stereoscopic video, the estimated disparity can be used for quality perception modeling. A horizontal block matching algorithm [15] between the left and right video was used to extract the disparity information in a stereoscopic video. The extracted disparity information was normalized from 0 to 255. A value of 128 indicates zero disparity. The search range was from -31 to +31 consisting of 63 levels. The disparity can be estimated by

$$disp^n(x,y) = \min_{i \in SR} MSE^n(x,y,i) \quad (2)$$

where (*x*,*y*) is a block location, *i* indicates the displacement in the horizontal direction and SR is the search range. $MSE^n(x,y,i)$ is the mean square error between the left block and right block located in (*x*,*y*), and is given by the following:

$$MSE^n(x,y,i) = \frac{1}{M \times N} \cdot \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \left\{ f_L^n(x+p, y+q) - f_R^n(x+p+i, y+q) \right\}^2 \quad (3)$$

where $f_L^n(x,y)$ and $f_R^n(x,y)$ indicates the intensity pixel values at (*x*,*y*) in the nth left and right images, respectively. To obtain the estimated disparity in the range between 0 and 255, the disparity in (2) was normalized by

$$f_{disparity}^n(x,y) = disp^n(x,y) \times \frac{255}{DB} + 128 \quad (4)$$

where *DB* indicates the depth budget that defines the search range (in this paper, *DB* is 63 because the search range is ±31). Each disparity image was divided into 32×32 sub images (blocks), and the average disparity for each block was calculated. The average disparity for each block in the nth image is given by the following:

$$f_{avg}^n(p,q) = \frac{1}{W \times H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} f_{disparity}^n(i,j) \quad (5)$$

where *W* and *H* are the width and height of the *W*×*H* sub-image. (*p*,*q*) is a *W*×*H* sub-image location. i and j indicate the coordinates of the mean disparity intensities.

Temporal variance

The perception of depth becomes more difficult for a stereoscopic video with a larger amount of motion activity [16]. The amount of motion activity needs to be considered in the design of objective quality assessment metrics because the high motion activity entails a large disparity variation. The average motion vector magnitude was calculated as a motion activity feature in the stereoscopic video. The average motion vector magnitude as the temporal variance feature is given by the following:

$$tv = \frac{1}{N \times B} \sum_{n=0}^{N-1} \sum_{b=0}^{B-1} \sqrt{\left(MV_{b,x}^n\right)^2 + \left(MV_{b,y}^n\right)^2} \quad (5)$$

where n is the frame number, N is the total number of frames, *b* indicates the *b*-th block in the frame, and *B* is the total number of blocks. $MV_{b,x}^n$ and $MV_{b,y}^n$ are the *x*- and *y*-component of the motion vector, MV_b^n, for the *b*-th block in the *n*-th frame. Here, the motion search range was set to 64×64. The motion activity in (6) was categorized into five scales such as

$$TV = \begin{cases} 5, & if\ tv > 4.5 \\ 4, & if\ 3.5 < tv \le 4.5 \\ 3, & if\ 2.5 < tv \le 3.5 \\ 2, & if\ 1.5 < tv \le 2.5 \\ 1, & if\ tv \le 1.5 \end{cases} \quad (7)$$

where *TV* is the quantized temporal variance that has one of the five levels.

Intra-frame disparity variation

When the disparity variation in the spatial domain is large, the resulting MOS values tend to be low from the subjective quality assessments. Relatively high MOS values are obtained if the disparity variation occurs smoothly in the spatial domain, even though the spatial variation in disparity is large. Therefore, intra-frame disparity variation is defined as the difference in the disparity values between the neighboring blocks.

By using the mean disparity of each block in (5), the disparity difference between the neighboring blocks were calculated, and the average intra-frame disparity variation DV_s can be expressed as

$$DV_s = \frac{1}{N} \sum_{n=0}^{N-1} \sqrt{\frac{1}{K \times L} \sum_{p=0}^{K-1} \sum_{q=0}^{L-1} \left\{ DV_s^n(p,q) \right\}^2} \quad (8)$$

where *K* and *L* indicate the number of sub-images in the row and column in the nth image. In this paper, *K* and *L* were set to 32, and *N* is the total number of frames. Fig. 11 shows the computation of the intra-frame disparity variation for a block against its neighboring blocks. In (8), $DV_s^n(p,q)$ is the intra-frame disparity variation for a block located at (*p*,*q*), and is given by

$$DV_s^n(p,q) = \frac{1}{8} \sum_{i=-1}^{1} \sum_{j=-1}^{1} \left| f_{avg}^n(p,q) - f_{avg}^n(p+i, q+j) \right| \quad (9)$$

Inter-frame disparity variation

Both the intra-frame disparity variation and the disparity variation in the temporal direction also affect the subjective quality for a stereoscopic video. The perception of depth becomes difficult when the disparity changes in the temporal direction occur rapidly, even though the object moves slowly. Therefore, the inter-frame disparity variation DV_t was considered:

$$DV_t = \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{K \times L} \sum_{p=0}^{K-1} \sum_{q=0}^{L-1} \left| f_{avg}^n(p,q) - f_{avg}^{n-1}(p,q) \right| \quad (10)$$

Disparity distribution of frame boundary area

Severe visual fatigue is caused when the frame boundary in a stereoscopic video frame contains extremely large or small disparity due to collision of the occlusion depth cue and parallax depth cue. Therefore, the disparity

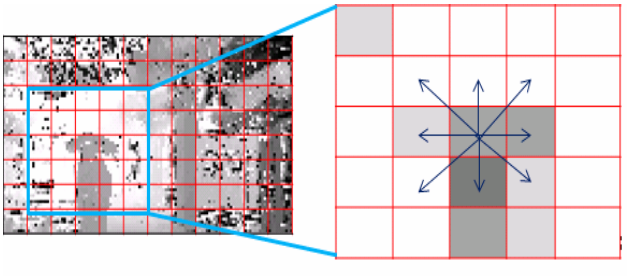


Fig. 11. Illustration of computing disparity variation in spatial domain for a block against its neighboring blocks.

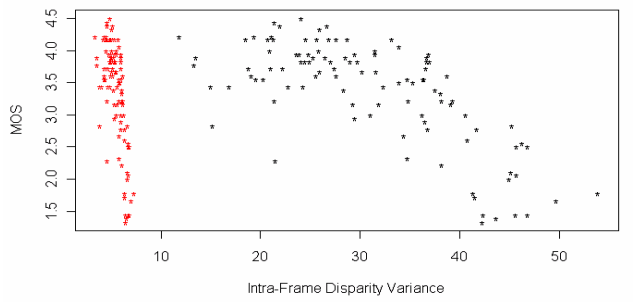


Fig. 12. Relationship between the intra-frame disparity variation and MOS (black: DV_s, red: $\sqrt{DV_s}$).

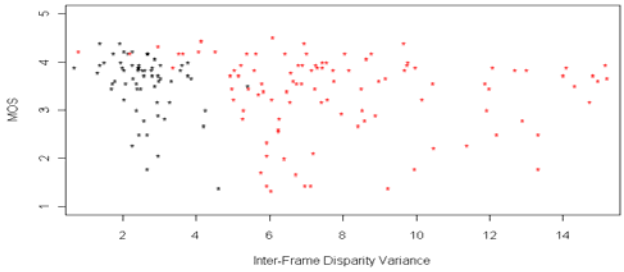


Fig. 13. Relationship between the inter-frame disparity variation and MOS (red: DV_t, black: $\frac{DV_t}{TV}$).

distribution in the area along the frame boundary to the design of our SV-PQAM was considered.

Similarly, after calculating the mean disparity for each frame according to (5), the disparity distribution D_b in the frame boundary area was obtained, in which its width in the left and right boundary areas and its height in the bottom and top areas are defined as one-eighth the width and height of the frame, respectively. The disparity distribution in the frame boundary area was calculated using the following equation:

$$D_b = \frac{1}{N}\sum_{n=0}^{N-1}\sqrt{\frac{1}{BC}\sum_{i\in B}\sum_{j\in B}\left\{f_{avg}^{n}(p,q)-128\right\}^2} \qquad (11)$$

where BC represents the number of blocks in the frame boundary area (448 $W\times H$ blocks), and B is the frame boundary area. In (10), the disparity distribution was calculated against the zero disparity by subtracting 128 from the disparity values of the boundary blocks.

3.2 Design of SV-PQAM with visual quality features

To design SV-PQAM with the features of TV, DV_s, DV_t and D_b, a linear regression method was used by combining the visual features. Although the linear regression model is a simple method that models the linear relationship through a statistical observation, it is difficult to model accurately the human visual system, which is a multi-dimensional complex system. Therefore, some transformed terms were added to increase the model accuracy in addition to the extracted features, such as $\sqrt{DV_s}$, $\sqrt{DV_t/TV}$, DV_s^2, and $\log TV$.

Fig. 12 shows the relationship between the disparity variation and the MOS values obtained by the subjective quality assessments. The distribution of the MOS values (black dots) exhibits a logarithmic relation over the DV_s with larger dispersion. The distribution (red dots) of MOS values shows a more linear relationship with less variation over $\sqrt{DV_s}$. Therefore, $\sqrt{DV_s}$ was considered in this SV-PQAM.

Fig. 13 shows that distribution of the MOS values against DV_t. The MOS values (red dots) appear to have no correlation with DV_t. This is because the high motion activity in stereoscopic video results in large disparity variation. For the distribution of MOS values (black dots) against DV_t/TV, an approximate linear relationship was observed.

Therefore, the TV-normalized inter-frame disparity variation DV_t/TV was added to SV-PQAM. The final perceptual quality model, SV-PQAM, is given by

$$SV-QPAM = w_0 + w_1\times\log TV + w_2\times DV_s^2 + w_3\times\sqrt{DV_s} + w_4\times\sqrt{\frac{DV_t}{TV}} + w_5\times D_b + w_6\times D_b^2 \qquad (12)$$

4. Analysis of the Experimental Results

To confirm the effectiveness of the proposed SV-PQAM, the subjective quality assessment experiments were separated into training and test sets. Sixty stereoscopic video sequences were selected randomly for the training data, and the remaining 45 sequences were used as the test data for verification. The effectiveness of the features was combined into a simple linear model without transformed features as follows:

$$SLM = w_0 + w_1\times TV + w_2\times DV_s + w_3\times DV_t + w_4\times D_b \qquad (13)$$

Table 3 and Fig. 14 show the performance of the simple linear model in (13), and the determination

Table 3. Performance of a Simple Linear Model.

Trials	60 training sequences			45 test sequences			R^2
	PCC	RMSE	OR	PCC	RMSE	OR	
1	0.693	0.428	0.033	0.701	0.169	0	0.4958
2	0.671	0.460	0.033	0.697	0.165	0	0.5511
3	0.694	0.429	0.017	0.705	0.184	0	0.6367
4	0.672	0.450	0.017	0.679	0.173	0	0.5169
5	0.693	0.425	0.033	0.685	0.200	0	0.6075
Avg.	0.685	0.438	0.026	0.693	0.178	0	0.5616

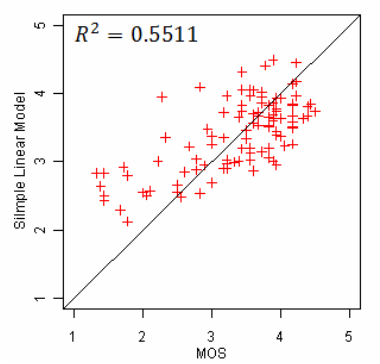


Fig. 14. Simple linear model versus MOS.

Table 4. Results of Verification.

Trials	60 training sequences			45 test sequences			R2
	PCC	RMSE	OR	PCC	RMSE	OR	
1	0.839	0.247	0.016	0.778	0.188	0	0.7719
2	0.868	0.231	0.033	0.804	0.118	0	0.7593
3	0.864	0.217	0.017	0.825	0.108	0	0.7983
4	0.874	0.199	0.000	0.807	0.147	0	0.7513
5	0.862	0.216	0.000	0.826	0.114	0	0.7836
Avg.	0.861	0.222	0.013	0.808	0.135	0	0.7729

coefficient values between the MOS and SLM. Both Table 3 and Fig. 14 show that the model accuracy is rather low.

Table 4 lists the performance of the proposed final SV-PQAM model in (12) for the experiments of five times by randomly selecting 60 training sequences and 45 testing sequences. The coefficients of determination in Table 1 were 0.75 or higher, maintaining the proposed SV-PQAM being a reliable model. The R-system [17] was used to estimate the regression parameters of the linear regression model.

As shown in Table 4, the proposed SV-PQAM accurately predicted the MOS values of the objective quality assessment in terms of the Pearson Correlation Coefficients (PCCs) of 0.861 and 0.808 for the training and test data, respectively. The prediction consistency is good with a zero outlier ratio (OR) value for the test data. In the perspective of the estimation accuracy, the Root Mean Square Errors (RMSEs) were 0.222 and 0.135 out of a total of 5 points for the training and test data, respectively, which shows the sufficient estimation accuracy with only 4.4% and 2.7% prediction errors for the

Table 5. Estimated Values of Regression Parameters.

$w_0 = -2.276$	$w_1 = -0.298$	$w_2 = -0.002$	$w_3 = 1.253$
$w_4 = -0.730$	$w_5 = 1.983$	$w_6 = -0.316$	

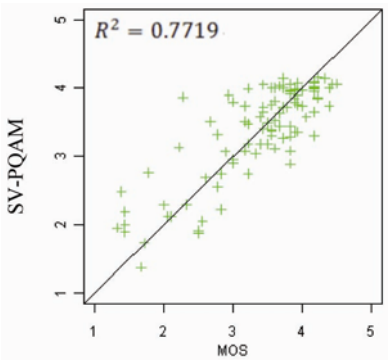


Fig. 15. SV-PQAM versus MOS.

training and test data, respectively. Table 5 summarizes the regression parameter values of the proposed SV-QPAM. Fig. 15 shows the relationship between the SV-PQAM versus MOS, where SV-PQAM is in a good agreement with MOS.

5. Conclusion

This study examined the factors that affect the human perception of depth and visual comfort from stereoscopic video. For this, subjective quality assessments were conducted using the SS method. After analyzing the results, four factors were extracted: (i) Mean magnitude of the motion vector; (ii) Disparity variation in the intra-frames; (iii) disparity variation in the inter-frames; and (iv) disparity distribution of frame boundary areas. Finally, these four factors were combined to propose a NR SV-PQAM, which does not require precise depth map data. The model parameters were estimated using a linear regression model based on the results of the subjective quality assessment.

The experimental results showed that the proposed model exhibits high consistency with a subjective quality assessment results, with a PCC value of 0.808, and the prediction consistency was good with a zero OR value for the test data.

Despite its satisfactory performance, there is room for further improvement. First, due to the incomplete disparity estimation algorithm, there is a problem of increasing the estimation error when the disparity in the contents is excessively large, or when the correspondence points are different due to noise. The proposed SV-PQAM only considers the temporal variations of the disparity magnitudes but not the direction of the disparity change, which has been analyzed as an important feature that affects the perceptual quality assessment. A certain level of disparity change can cause visual fatigue when it is excessive. On the other hand, the subjects appear to experience less visual fatigue when the disparity change in

direction is small enough.

Finally, an NR SV-PQAM can be constructed easily because the features used in the proposed SV-PQAM can be integrated easily in the bit stream domains of stereoscopic video codecs.

Acknowledgement

This work was supported by the IT Research and Development Program of MKE/KEIT under Grant 10039199 (A study on core technologies of perceptual quality based scalable 3-D video codecs)

References

[1] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1969–1976, 2008. Article (CrossRef Link)

[2] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," *Proc. IEEE Int. Conf. Image Processing*, Oct. 2008, San Diego, CA. Article (CrossRef Link)

[3] Z. M. P. Sazzad, S. Yamanaka, and Y. Horita, "Continuous stereoscopic video quality evaluation," *in Proc. SPIE*, vol. 7524, Jan. 18-21, San Jose, USA, 2010. Article (CrossRef Link)

[4] J. Choi, D Kim, Bumsub Ham, Sunghwan Choi and Kwanghoon Sohn, "Visual Fatigue Evaluation and Enhancement for 2D-Plus-Depth Video", *in Proc. ICIP*, Sep. 2010, Hong Kong. Article (CrossRef Link)

[5] Z. M. Parvez Sazzad, M. Sato, Y. Kawayoke, and Y.Horita, "No-reference Image Quality Evaluation based on Local Features and Segmentation," The Journal of the Institute of Image Electronics Engineers of Japan(IIEEJ), Vol. 37, no.3, pp.335-345, May 2008.

[6] Recommendation ITU-R BT.500-11, "Methodology for the Subjective Assessment of the Quality of Television Pictures"

[7] W. Baek, S. Lee, J. Kim and S.-I. Park "Korean terrestrial 3DTV broadcasting system: Current status", *Proc. IEEE Int. Conf. Consumer Electron*., pp.889 - 890 2011. Article (CrossRef Link)

[8] Y. Chang and M. Kim, "A Joint Rate Control Scheme in a Hybrid Stereoscopic Video Codec System for 3DTV Broadcasting," *IEEE Transactions on Broadcasting*, vol. 59, no. 2, pp. 265-280, June 2013. Article (CrossRef Link)

[9] L. Goldmann, F. D. Simone, T. Ebrahimi: "A Comprehensive Database and Subjective Evaluation Methodology for Quality of Experience in Stereoscopic Video", *Electronic Imaging (EI), 3D Image Processing (3DIP) and Applications*, San Jose, USA, 2010. Article (CrossRef Link)

[10] Y. Kang, E. Lee, J. Jung, J. Lee, I. Shin, Y. Ho, GIST, 3D Video Test Sequence and Camera Parameters. ISO/IEC JTC1/SC29/WG11 MPEG 2009/m16949, 2009.

[11] Article (CrossRef Link)

[12] M. Domanski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K.Wegner. Poznan. Multiview video test sequences and camera parameters. ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, 2009.

[13] C. Won, D. Park, and S. Park, "Efficient Use of MPEG-7 Edge Histogram Descriptor", *ETRI Journal*, Volume 24, Number 1, February 2002. Article (CrossRef Link)

[14] http://www.3dtv.fr/NAB09_3D-Tutorial_BernardMendiburu.pdf

[15] Z. Sazzad, S. Yamanaka, and Y. Horita, "Continuous stereoscopic video quality evaluation," *in Proc. SPIE*, vol. 7524, Jan. 18-21, San Jose, USA, 2010. Article (CrossRef Link)

[16] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, "Effect of Disparity and Motion on Visual Comfort of Stereoscopic Images", *Proceedings of SPIE-IS&T Electronic Imaging*, vol. 6055, 60550B (2006). Article (CrossRef Link)

[17] http://www.r-project.org/

Kwang-sung Ha received the B.S degree in Computer Science and Engineering from Korea University of Technology and Education (KUT), Cheonan, Korea in 2009, and his M.S degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea in 2011. He is currently a research engineer in the department of Software Platform R&D at LG Electronics. He is working on developing WebOS TV Platform and selected as a Software Coding Expert in 2012.

Sung-Ho Bae graduated summa cum laude from Kyung-Hee University, Suwon, Korea, receiving the double BS degrees in computer engineering and electrical engineering in 2011, and the M.S degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea in 2012. He is currently pursuing the Ph.D degree in Electrical Engineering at KAIST. His research interests perceptual image and video processing, pattern recognition and machine learning.

Munchurl Kim received the B.E. degree in electronics from Kyungpook National University, Daegu, Korea, in 1989, and the M.E. and Ph.D. degrees in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 1992 and 1996, respectively. After his graduation, he joined the Electronics and Telecommunications Research Institute, Daejeon, Korea, as a Senior Research Staff Member, where he led the Realistic Broadcasting Media Research Team. He was an Assistant Professor from Feb. 2001 to Aug. 2005, and an Associate Professor from Sept. 2005 to Feb. 209 at the School of Engineering, Information and Communications University (ICU), Daejeon. He was an Associate Professor from Mar. 2009 to Aug. 2013, and is now a Full Professor from Sept. 2013 at the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon. He has been involved with scalable video coding and high efficiency video coding in JCT-VC standardization activities of ITU-T VCEG and ISO/IEC MPEG. His current research interests include video coding, visual quality assessments on 3-D/UHD video, visual information processing, pattern recognition and machine learning.

Copyrights © 2013 The Institute of Electronics Engineers of Korea