QUALITY ASSESSMENT OF 3D SYNTHESIZED VIEWS WITH DEPTH MAP DISTORTION

Chang-Ting Tsai and Hsueh-Ming Hang

Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan

ABSTRACT

Most existing 3D image quality metrics use 2D image quality assessment (IQA) models to predict the 3D subjective quality. But in a free viewpoint television (FTV) system, the depth map errors often produce object shifting or ghost artifacts on the synthesized pictures due to the use of Depth Image Based Rendering (DIBR) technique. These artifacts are very different from the ordinary 2D distortions such as blur, Gaussian noise, and compression errors. We thus propose a new 3D quality metric to evaluate the quality of stereo images that may contain artifacts introduced by the rendering process due to depth map errors. We first eliminate the consistent pixel shifts inside an object before the usual 2D metric is applied. The experimental results show that the proposed method enhances the correlation of the objective quality score to the 3D subjective scores.

Index Terms—3D image quality assessment, 3D artifacts in view synthesis, depth map induced errors

1. INTRODUCTION

3D videos are becoming more popular recently. The 3D perception is often made by viewing two different views in two eyes, and then they are combined by the Human Visual System (HVS). The ISO/IEC Moving Picture Expert Group (MPEG) is in the process of specifying the 3D video coding (3DVC) standards based on the multiple-view plus depth (MVD) format. Fig 1 shows the framework of ISO/IEC-MPEG 3DVC system. It assumes the input is a 2-view video. Each view has its corresponding depth map, which can be captured by depth sensors or generated by a depth estimation algorithm. Then these data, color images and depth maps, are compressed by a 3D video coder. At the receiver, virtual view images are generated by a view synthesis algorithm. Either transmitted views or synthesized views and their mixtures can be displayed on a 3D monitor. With the popularity of 3D virtual view systems, how to predict the quality of the stereo images with synthesized views becomes an important issue.

There are two quality assessment types: subjective and objective. Subjective quality assessment is performed by human observers, who watch test sequences and give it a quality score. ITU-R BT.500 [1] outlines a few subjective test processes to judge the quality of pictures or videos. The subjective evaluation is costly and time consuming, and cannot implement in a machine. Therefore, the objective assessment is very desirable. The goal of objective quality assessment is to develop a computational model algorithm that can predict the visual quality judged by human.

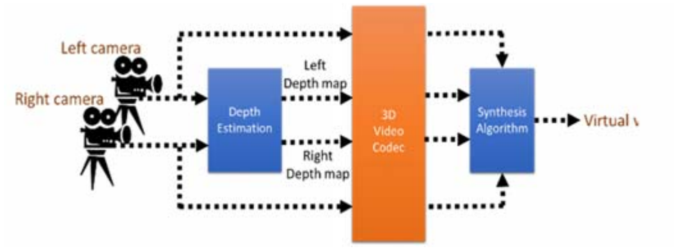


Fig. 1. Framework of 3DVC system.

There are a few existing databases for 3D quality assessment available on the website. Lavoue et al. [2] proposed a 3D computer graphics model database. 88 models between 40K and 50K vertices were generated from 4 reference objects. There are two types of distortion, noise and smoothing. The database proposed by Goldmann et al. [3] contains 10 scenes with various textures and depth structures. Each of the scenes has been captured with different camera baseline in the range 10-50 cm. Benoit et al. [4] proposed database contain three types of distortion (JPEG, JPEG2000, and blur) symmetrically to the stereo pair images. Urvoy et al. [5] distort the stereo pair images based on H.264, JPEG2000, and typical image processing steps such as down sampling and sharpening. IRCCyN/IVC DIBR image database proposed by Bosc et al. [6] contains three MVD sequences. And seven DIBR algorithms use these sequences to generate four new viewpoints for each sequence. The above databases mainly look at the distortions caused by conventional image processing methods such as noise, blurring, and compression applied to images. The distortions due to depth map errors are nearly undiscussed.

Most existing 3D IQA approached adopt 2D image QA models directly to predict 3D subjective quality [7]-[10]. When virtual-view images are synthesized using distorted depth maps, the depth distortions usually produce object shifting or ghost artifacts on the synthesized images as shown in Fig. 3. The subjective quality of consistently shifted object is still high, but the traditional pixel-by-pixel based 2D IQA model often overly penalize these regions [11].

In this paper, we propose a novel IQA model to assess the visual quality of distorted image synthesized by the depth map distorted by three different types of depth distortions. The proposed metric uses block-matching algorithm to reduce the effect of object shift and the Hausdorff distance to

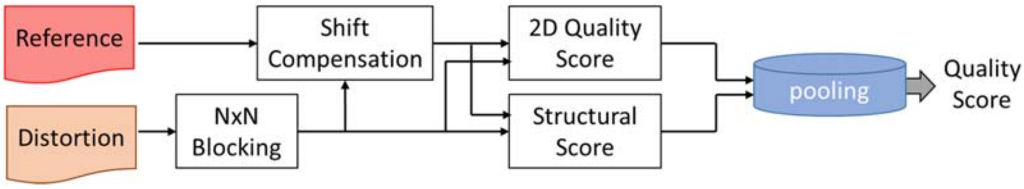


Fig. 2. Flow chart of the proposed model.

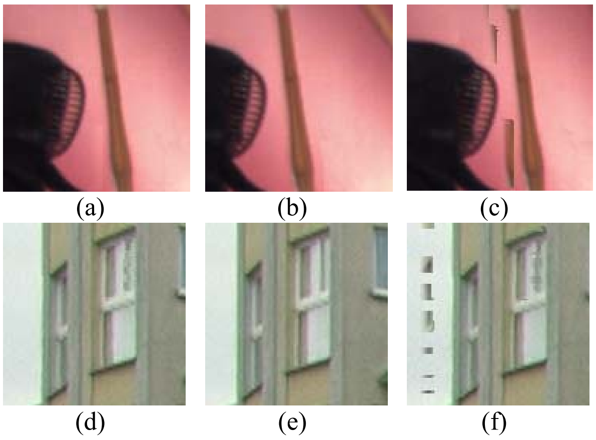


Fig. 3. The artifacts of the synthesized image cause by error depth map. (a) (d) Original; (b) (e) Object shift; (c) (f) Ghost artifact

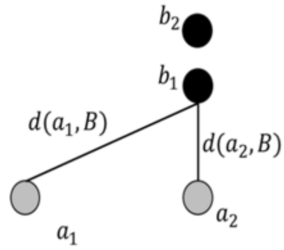


Fig. 4. An illustration of directed Hausdorff distance.

$$h_K(A, B) = {}^{p}K^{th}_{a \in A}\, d(a, B) \tag{4}$$

$$\frac{K}{N_A} = p\% \tag{5}$$

where ${}^{p}K^{th}_{a \in A}$ denotes the *Kth* ranked distance in A and N_A is the number of points in set *A*.

determine the level of a ghost artifact. We test a number 3D images and show that the proposed method improves the correlation of the objective QA model to the subjective score.

The rest of this paper is organized as follows. Section 2 introduces the Hausdorff distance. Section 3 proposes our QA model. Section 4 presents the experimental results. Finally, section 5 concludes this work.

2. THE HAUSDORFF DISTANCE

The Hausdorff distance is used to measure the degree of mismatch between two sets. In the computer vision applications, this distance may refer to the differences between two image patches [12]. The Hausdorff distance of two finite sets, A and B, is defined as follows.

$$H(A, B) = max(h(A, B), h(B, A)) \tag{1}$$

where $h(A, B)$ is called the directed Hausdorff distance from A to B. It is defined as

$$h(A, B) = \max_{a \in A} d(a, B) \text{ and} \tag{2}$$

$$d(a, B) = \min_{b \in B} \|a - b\|. \tag{3}$$

It identifies the longest distance of the points in A to the nearest neighbor in B. For example, in the Fig. 4, let a_1 and a_2 belong to set A, the directed Hausdorff distance $h(A, B)$ equals to $d(a_1, B)$.

The directed Hausdorff distance may be modified to fit a specific application; for example, we consider the Kth ranked point of A.

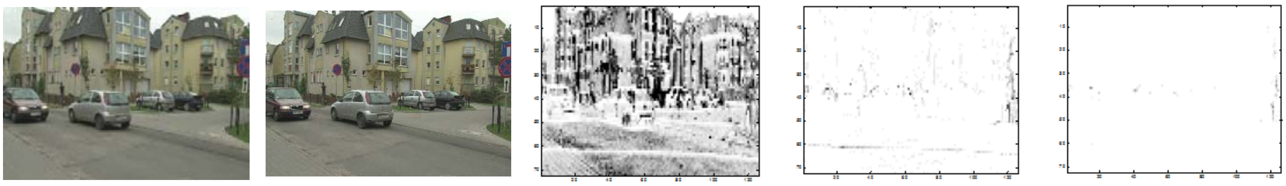
3. PROPOSED MODEL

The proposed QA model is divided into two parts as shown as Fig. 2. One part is computing the traditional 2D quality compensated by object shift. And the other is generating a structural score by the Hausdorff distance.

3.1. 2D Quality Score with Object Shift

In the view-synthesizer, incorrect depth values cause objects to shift horizontally in the synthesized picture. If we look at these "distorted" pictures on a 3D display, the depth perception of objects change somewhat but the subjective image quality shows nearly no difference. Because the traditional 2D metrics are calculated pixel-by-pixel, they are sensitive to object shifts.

We thus like to compensate the "consistent" object shifts before applying the conventional 2D metrics. To find the shifts, a block-matching algorithm is implemented along the x-direction (assuming the pictures are well calibrated). Then, each NxN block in the distorted image finds a best-matched NxN block in the original image. In our experiment, the value N is set 25 as a tradeoff between performance and computational complexity. Fig. 5 illustrates that image (b) synthesized by a quantized depth map has little subjective distortions compared to the reference image (a) synthesized by the original depth map. It produces only object shifts that do not affect the quality judgment by human. Fig. 5(c) and (d) show the SSIM quality maps evaluated on Fig. 5(a) on Fig.



(a) (b) (c) (d) (e)

Fig. 5. (a) original image; (b) image with object shift; (c) SSIM quality map computed based on (a) and (b); (d) SSIM quality map with compensated object shift; and (e) is (d) with extra Gaussian filter.

5(b) without and with shifts compensation, respectively. The darker region in the SSIM map indicates lower quality indices. We will describe the SSIM metric in section 3.2. Fig. 5(c) indicates that the 2D model gives high penalties on the regions with object shifts. After shift compensation, the 2D QA model matches the subjective quality score better.

In the QA map of Fig. 5(d), there are some low quality regions on the road and near the building. They often appear near object boundaries. This is due to the rendering process. Incorrect depth values often produce a synthesized view in which objects pick up the neighboring color pixels in reconstruction. This error can be detected by the pixel-based 2D QA models. However, the subjective quality can still be high. For this reason, we apply a Gaussian filter on the images before they are evaluated by the 2D QA model as shown in Fig. 6. In other words, we now care more on significant structure errors. The result is shown in the Fig. 5(e).

3.2. Structural Similarity (SSIM) index

The full reference 2D QA model, SSIM, was proposed by Wang et al. [13]. It bases on the observation that HVS is highly adapted to the structure information of a scene that pixels have strong inter-dependency. SSIM index is calculated based on three components: luminance, contrast, and structure.

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \quad (6)$$

where x and y are the reference and distorted images, respectively. The luminance, contrast, and structure can be computed from comparing the means, standard deviations, and correlation coefficient of the images. Finally, the overall image quality is evaluated from the average SSIM.

3.3. Structural Score

However, human eyes are sensitive to ghost-type errors or inconsistent object shifts (particularly, along object boundaries). We thus like to design a *structure score* that penalizes these undesired errors. After an NxN block of the distorted image (D_i) matches a corresponding block (R_i) in the reference image, the Canny edge detector is used to pick up the feature points in a block. Then, we compute the Hausdorff distance of the *ith* block $H(D_i)$ based on the feature points in these two blocks.

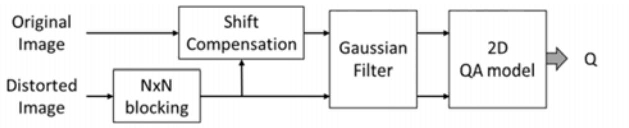


Fig. 6. Flow chart of the part evaluates quality score.

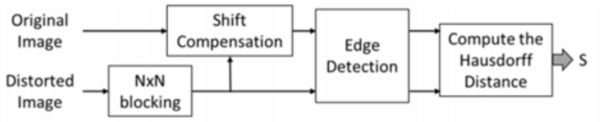


Fig. 7. Flow chart of the part evaluates structural score.

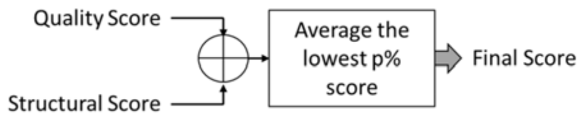


Fig. 8. Flow chart of pooling stage.

$$H(D_i) = max(h_K(D_i, R_i), h_K(D_i, R_i)) \quad (7)$$

A modified directed Hausdorff distance is adopted by our model, which is defined by

$$h_K(A,B) = {}^{70}K^{th}_{a \in A}\, d(a,B) \quad (8)$$

The parameter 70% is obtained empirically. Then, we normalize the distance $H(D_i)$ value between 0 and 1.

$$H_{normalize}(D_i) = \frac{H(D_i)}{2N} \quad (9)$$

Because the structural score is opposite to the distance: smaller distance indicates the ghost artifact is light and thus, the structural score S should be higher, and vice versa. We therefore define the structural score as

$$S(D_i) = 1 - H_{normalize}(D_i) \quad (10)$$

3.4. Pooling

After obtaining the above two scores, we add them to form the final score.

$$F(D_i) = \alpha \cdot Q(D_i) + (1-\alpha) \cdot S(D_i) \quad (11)$$

where $F(D_i), Q(D_i)$ and $S(D_i)$ are the final score, the 2D quality score with shift, and structural score, respectively, for the *ith* block in the distorted image; and α is a parameter used to adjust the relative importance of these two components. In our experiment, we set α to 0.5.

Typical 2D QA metrics calculate the average of all the pixels or blocks of an image to produce the final image quality index. But in a synthesized image, the object shift and the ghost artifacts appear in specific regions due to the depth-based rendering process. So, we use the lowest p% of quality scores instead of using all scores in calculating the final index [14].

$$final\_score = \frac{1}{N_p} \sum_{n}^{N_p} F(D_n) \quad (12)$$

where N_p is the lowest p% quality scores $F(D_n)$.

4. EXPERIMENTAL RESULTS

4.1. Test Sequences

Six multiview plus depth sequences provided by MPEG 3DVC for the 3DVC contest [15] are used in our experiment. Fig 9 shows the sequences *Poznan Hall2*, *Kendo*, *Balloons*, *Poznan Street*, *Lovebird*, and *Newspaper1* from left to right and top to bottom. The frame number, input camera views, and displayed stereo pair (views) are given in Table 1. The view synthesis algorithm producing virtual views is *"VSRS-1D-Fast"* implemented in HTM version 3.1, which is an HEVC based reference software developed by the ITU/MPEG 3DV group.

The distortions applied to the depth maps are (1) *Offset*, (2) *Quantization*, and (3) *Gaussian_noise*. For the *Offset* type noise, we add a constant value to all the pixels of the depth map.

$$d_{offset}(x,y) = d_{original}(x,y) + offset\_value \quad (13)$$

where $d_{original}(x,y)$ and $d_{offset}(x,y)$ denote the depth values at coordinate (*x,y*) of the original depth map and the distorted depth map, respectively; and $offset\_value$ is a given integer.

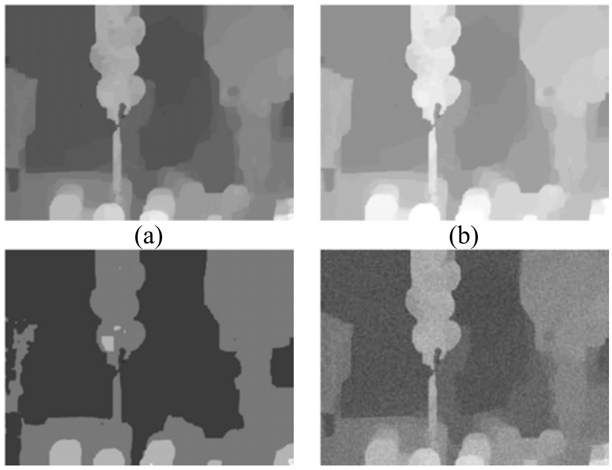
The *Quantization* type error is to quantize the depth map pixel values into specific levels.

$$d_{quantization}(x,y) = \left\lfloor \frac{d_{original}(x,y)}{q\_step} \right\rfloor \cdot q\_step \quad (14)$$

where $d_{quantization}(x,y)$ is the distorted depth value at coordinates (*x,y*); q_step is the quantization step.

The *Gaussian_noise* adds white Gaussian noise to the depth map with different variance, $\sigma_{gaussian}$. We set *offset_value=60* and *100*; *q_step*=60 and 80; and $\sigma_{gaussian}$ = 0.01 and 0.05 in ingenerating our synthesized distorted images. In total, there are 6 distorted pictures and 1 reference picture for each sequence. So, there are totally 42 (7x6) stereo images in our database.

Fig. 9. The sequences used in our experiment.



(a) (b) (c) (d)

Fig. 10. Three types of depth map distortion.
(a) Original (b) *Offset* (c) *Quantization*
(d) *Gaussian_noise*

Table 1. Test sequences details

Sequence Name	frame	Input Views	Output Stereo Pair
Poznan Hall2	90	7 - 6	6.5 - 6
Poznan Street	30	4 - 3	3.5 - 3
Kendo	32	3 - 5	4 - 5
Balloons	1	3 - 5	4 - 5
Lovebird1	80	6 - 8	7 - 8
Newspaper	100	4 - 6	5 - 6

4.2. Subjective Test

The Toshiba 47TL515U 47-inch 3D television is used to display the stereo pictures. Twenty-two observers with an average age of 23.8 participated in our subjective evaluation. The stereo pairs use in our experiment is shown as Fig. 11. The right view picture displayed on the stereoscopic monitor is the original and the left view picture is the synthesized virtual view. In the 42 tests, there are distorted and hidden reference stereo pairs. We follow the single stimulus (SS) testing procedure described in the document ITU-R BT.500 [1]. For each test image, observers are asked to rate a quality score (5: *Excellent*; 4: *Good*; 3: *Fair*; 2: *Poor*; 1: *Bad*), which are the so-called opinion score. All the observers' opinion

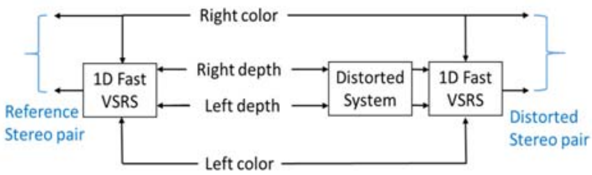


Fig. 11. Illustration of stimulus generation systems.

Fig. 12. Artifact introduced by Gaussian_noise.

Table 2. Performance comparison of Case 1.

Metrics	PLCC	RMSE	OR
PSNR	0.58	0.70	0.042
SSIM	0.37	0.80	0.083
MSSIM	0.51	0.74	0.083
UQI	0.40	0.79	0.167
VIF	0.53	0.73	0.083
VSNR	0.35	0.81	0.167
Proposed	**0.90**	**0.38**	**0**

Table 3. Performance comparison of Case 2.

Metrics	PLCC	RMSE	OR
PSNR	0.52	0.71	0.086
SSIM	0.42	0.75	0.057
MSSIM	0.42	0.75	0.057
UQI	0.36	0.77	0.114
VIF	0.56	0.69	0.029
VSNR	0.25	0.80	0.200
Proposed	**0.77**	**0.53**	**0.029**

scores are averaged to compute the mean opinion score (MOS). The difference between the MOS of a distorted picture and the corresponding reference image is called difference mean opinion score (DMOS).

4.3. Results and Analysis

In the full-reference quality assessment, we need a reference image for evaluating the distorted image. We chose the virtual image synthesized from the original depth map as the reference. We use the following function for fitting the QA scores to the subjective DMOS.

$$DMOS_p = \frac{b_1}{1+ exp(-b_2(score-b_3))} \qquad (15)$$

where $score$ is the quality score obtained from an objective QA model; $DMOS_p$ is the fitted DMOS value produced by the $score$ with the optimally selected parameters, $b_1, b_2, and\ b_3$. Parameters $b_1, b_2, and\ b_3$ are are obtained through the regression step to minimize the error between $DMOS$ and $DMOS_p$.

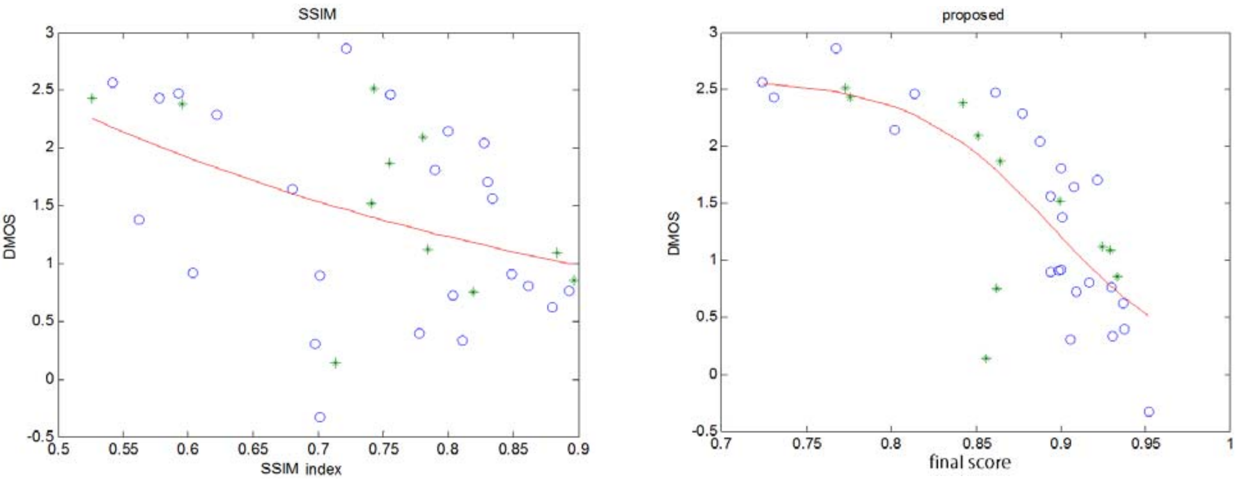
Three criterions are computed between $DMOS$ and $DMOS_p$ to evaluate the performance of a QA model: (1) Pearson Linear Correlation Coefficient (PLCC), (2) Root Mean Square Error (RMSE), and (3) Outlier Ratio (OR). We compare the proposed QA metric with several commonly used 2D models, which are PSNR, SSIM, MSSIM, UQI, VIF, and VSNR provided by the MeTriX MuX Visual Quality Assessment Package [16].

In data analysis, we consider two cases. Case 1 contains *Offset* and *Quantization* distortions only, because they cause similar artifacts, object shift and ghost artifact. Case 2 includes all types of distortions. Tables 1 and 2 show the performance comparison of different models in terms of PLCC, RMSE, and OL. Our proposed model has the highest correlation with the subjective test (0.9), and the minimum RMSE (0.38) and OL (0) in Case 1 when $\alpha = 0.3$. And the highest correlation (0.77) and the minimum RMSE (0.53) and OL (0.029) in Case 2 when $\alpha = 0.5$. Fig 13 shows the scatter plot of the quality scores obtained by the objective model against the DMOS.

We notice that the PLCC of Case 2 is lower than that of Case 1 for our proposed model, when the *Gaussian_noise* distortion is included. This is because its artifact is a blur-like distortion as shown in Fig. 12. In the subjective test, this blur-like distortion is masked by the right-view original image in a stereo pair [17]. Thus, our model gives a too low score on some low DMOS *Gaussian_noise* images.

5. CONCLUSIONS

In this paper, we propose a computational quality assessment model to estimate the quality of distorted image synthesized by a distorted depth map. Three different types of distortions, *Offset*, *Quantization* and *Gaussin_noise* are tested. We compensate the consistent object shift by using a shift-compensation mechanism. We also use the Hausdorff distance to identify the degree of the ghost-type artifacts at object boundaries. Finally, these two scores are combined into a final score. The proposed model not only inherits the properties of the conventional 2D quality metric but also considers the new artifacts that introduced by the synthesis process due to depth map error. In comparison with the popular QA models, the experimental results show the proposed method has higher PLCC and lower RMSE and OR in matching the subjective scores.



(a) (b)

Fig. 13. Scatter plots of the objective quality scores against DMOS. (a) SSIM (b) Proposed

(Blue ◦ : *Offset* and *Quantization*. Green *: *Gaussian_noise*. Red line: Regression line.)

6. ACKNOWLEDGEMENT

This work was supported in part by the NSC, Taiwan under Grant NSC 101-2221-E-009-136, NSC 102-2218-E-009-003 and by the Aim for the Top University Project of National Chiao Tung University, Taiwan.

7. REFERENCES

[1] ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” Tech. Rep. BT.500-11, ITU-R, 2002.

[2] G. Lavoue, et al., “Perceptually driven 3D distance metrics with application to watermarking,” In Proceedings of SPIE, vol. 6312, 2006.

[3] L. Goldmann, et al., “Impact of Acquisition Distortions on the Quality of Stereoscopic Images,” *5th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Scottsdale, USA, 2010.

[4] A. Benoit, et al., “Quality Assessment of Stereoscopic Images,” *EURASIP Journal on Image and Video Processing*, 2008.

[5] M. Urvoy, et al, “NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences,” *Fourth International Workshop on Quality of Multimedia Experience*, Yarra Valley, July 2012.

[6] E. Bosc, et al., “Towards a New Quality Metric for 3-D Synthesized View Assessment,” *IEEE Journal on Selected Topics in Signal Processing*, 2011.

[7] P. Hanhart, F. De Simone, and T. Ebrahimi, “Quality assessment of asymmetric stereo pair formed from decoded and synthesized views,” *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pp.236,241, 5-7 July 2012.

[8] J. You, et al., “Perceptual Quality Assessment for Stereoscopic Images Based on 2D Image Quality Metrics and Disparity Analysis,” *International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM*, Scottsdale, AZ, USA, 2010.

[9] Y. Zhang, et al., “A multiview video quality assessment method based on disparity and SSIM,” , *2010 IEEE 10th International Conference on Signal Processing* (ICSP), vol., no., pp.1044,1047, 24-28 Oct. 2010.

[10] K. Wang, et al., “Stereoscopic 3D video coding quality evaluation with 2D objective metrics,” *Proc. SPIE 8648, Stereoscopic Displays and Applications XXIV*, 86481L, March 12, 2013.

[11] E. Bosc, et al., “Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?” *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp.2597,2600, 11-14 Sept. 2011.

[12] D. P. Huttenlocher, et al., “Comparing images using the Hausdorff distance,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.15, no.9, pp.850,863, Sep. 1993.

[13] Z. Wang, et al., “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[14] A. K. Moorthy and A. C. Bovik “Visual importance pooling for image quality assessment,” *IEEE J. Special Topics in Signal*, vol.3, no.2, pp.193-201, Apr. 2009.

[15] ISO/IEC JTC1/SC29/WG11, “Description of Core Experiments in HEVC based 3D Video Coding,”, N12354, Dec. 2011.

[16] MeTriX MuX Visual Quality Assessment Package, http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

[17] F. Lu, et al., “Quality assessment of 3D asymmetric view coding using spatial frequency dominance model,” *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp.1-4, 4-6 May 2009.