COPYRIGHT NOTICE

A Novel No-reference Subjective Quality Metric for Free Viewpoint Video Using Human Eye Movement

Abstract. The *free viewpoint video* (FVV) allows users to interactively control the viewpoint and generate new views of a dynamic scene from any 3D position for better 3D visual experience with depth perception. Multiview video coding exploits both texture and depth video information from various angles to encode a number of views to facilitate FVV. The usual practice for the single view or multiview quality assessment is characterized by evolving the objective quality assessment metrics due to their simplicity and real time applications such as the *peak signal-to-noise ratio* (PSNR) or the *structural similarity index* (SSIM). However, the PSNR or SSIM requires reference image for quality evaluation and could not be successfully employed in FVV as the new view in FVV does not have any reference view to compare with. Conversely, the widely used subjective estimator- *mean opinion score* (MOS) is often biased by the testing environment, viewers mode, domain knowledge, and many other factors that may actively influence on actual assessment. To address this limitation, in this work, we therefore devise a no-reference subjective quality assessment metric by simply exploiting the pattern of human eye browsing on FVV. Over different quality contents of FVV, the participants' eye-tracker recorded spatio-temporal gaze-data indicate more concentrated eye-traversing approach for relatively better quality. Thus, we calculate the *Length*, *Angle*, *Pupil-size*, and *Gaze-duration* features from the recorded gaze trajectory. The content and resolution invariant operation is carried out prior to synthesizing them using an adaptive weighted function to develop a new *quality metric using eye traversal* (QMET). Tested results reveal that the proposed QMET performs better than the SSIM and MOS in terms of assessing different aspects of coded video quality for a wide range of FVV contents.

Keywords: Eye-traversal, Eye-tracking, Free viewpoint video, Gaze-trajectory, HEVC, QMET, Quality assessment.

1 Introduction

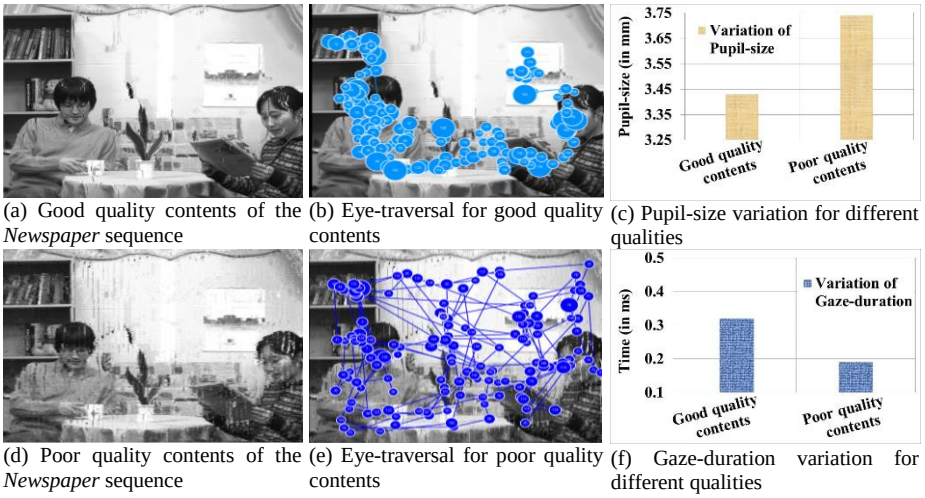
The *video quality evaluation* (VQE) is a promising research area due to its wide range of applications in the development of various video coding algorithms [1][2]. The technical coding areas involved with the FVV are characterized by the view generation using *multiview video coding* (MVC) and the view synthesis. This process first goes through the image warping and then a hole filling technique e.g. the inverse mapping technique or spatial/temporal correlation as simple post processing filtering [3][4]. Since the synthesized view is generated at a virtual position between left and right views, there is no available reference frame for quality estimation of FVV [5]. Usually the quality estimation is performed in two ways: objective and subjective, where the former one is more widely used due to its simplicity, ease of use and having real-time applications. Thus, a good number of citable researches have been conducted based on the objective image quality estimation [6]-[8]. The quality estimation could be further

categorized into full-reference (i.e. original videos as reference), reduced-reference (i.e. existence of partial signals as reference) and no-reference schemes. Among them, the applications of full-reference metrics such as the SSIM or PSNR have been restricted to the reference based situations only and these metrics lose their suitability in estimating different qualities of FVV where the reference frame is not available.

To address the limitations of full-reference metrics, a number of no-reference based research works have recently come into light for quality evaluation [10]-[12]. The introduced statistical metrics may not be suitable to some high quality ranges since the quality perception in these area is mostly due to perceptual *human visual system* (HVS) features, rather than to the statistics of the image [13]. However, different features of the HVS are not actively studied in the existing schemes. The authors in [14] carried out the human cognition based objective quality assessment system using eye-tracking technology and evolved more realistic ground truth visual saliency model to improve their algorithm. Actually, the eye-tracking has become a non-intrusive, affordable, and easy-to-use tool in human behavior research today that allows to measure visual behavior as it objectively monitors where, when, and what people look at. With very few exceptions, anything with a visual component can be eye tracked not necessarily by using the tracking device itself, rather simply employing the software based eye-tracking simulator [15].

Unlike objective quality evaluation, the subjective studies could yield valuable data to evaluate the performance of objective methods towards aiming the ultimate goal of matching human perception [16]. Thus, a number of quality assessment algorithms have been proposed which are closely related to the studies of human visual attention and cognition. The study in [17] proposed a no-reference model using blur and blockiness metric to improve the performance of objective model based on eye-tracker data. The authors in [18] introduced a model to judge the video quality on the basis of psychological merits including- the pupil dilation and electroencephalogram signalling. Albanesi *et al*. [19] used the eye-gaze data to create a voting algorithm to develop a no-reference method. Using the scan path of eye movements, Tsai *et al*. [20] subjectively assessed the perceived image and its colour quality. Conversely, the widely used subjective testing method- MOS [21][22] is often biased by the testing environment, viewers mode, expertise, domain knowledge, age range, and many other factors which may undesirably influence the effectiveness of actual quality assessment process. The authors in [23] although introduced a subjective metric, their initial work is based on the single view video where the viewing angle is fixed for users. Moreover, their introduced approach highly depends on threshold selection for each feature and incur with the lack of proper correlation setting among features. The most importantly, their metric does not perform well in different contents and resolutions of the videos. The proposed method is a significantly extended version of their work where the major amendments include the employment of FVV i.e. in the no reference scenario, increasing number of features, better correlation analysis of features, performing content and resolution invariant operation on features, synthesizing them by an adaptive weighted function, comparing the new metric with PSNR, SSIM, and MOS, and eventually employing two widely used estimators the *Pearson Linear Correlation*

Coefficient (PLCC) and *Spearman Rank-Order Correlation Coefficient* (SRCC) to justify the effectiveness of the proposed QMET for a range of FVV sequences.



(a) Good quality contents of the *Newspaper* sequence

(b) Eye-traversal for good quality contents

(c) Pupil-size variation for different qualities

(d) Poor quality contents of the *Newspaper* sequence

(e) Eye-traversal for poor quality contents

(f) Gaze-duration variation for different qualities

Fig. 1. More concentrated eye-traversing approach is perceived for relatively better quality contents (e.g. *Newspaper* sequence image in (b)). The opposite is noticed in (e) for which the pupil-size sharply increases in (c), while the gaze event duration notably decreases in (f).

Let us first concentrate on Fig. 1 in which (a) and (d) represent a Multiview video sequence namely *Newspaper* encoded as good and poor quality respectively, while (b) and (e) demonstrate the eye traversing approach of a viewer for good and poor quality image contents respectively. The tracked gaze plots indicate more concentrated eye-traversal for relatively better quality contents. Now if we determine *Length* (*L*) and *Angle* (*A*) features for each gaze plot, they could better inform about the viewers nature of browsing (i.e. smooth or random as indicated in Fig. 1 (b) and (e)). Since we also discover that the quality variation has an impact on both the *Pupil-size* (*P*) and *Gaze-duration* (*T*) variation presented in Fig. 1 (c) and (f), hence we calculate four cardinal features- *L*, *A*, *P*, and *T* for each *potential gaze plot* (PGP) from the gaze trajectory of the whole sequence. The PGPs in this test are defined by the fixations (i.e. visual gaze on a single location) and saccades (i.e. quick movement of eyes between two or more phases of fixations). Then we carry out content and resolution invariant operation on the features and adaptively synthesize them using a weighted function to develop the proposed QMET. The higher QMET score promises good quality video as the viewers could better capture its content information with smooth global browsing. Experimental results reveal that the quality evaluation carried out by the QMET could better perform compared to the objective metric SSIM, and the subjective estimator MOS. The proposed QMET is expected to use as an impressive substitute to the MOS in evaluating the objective metrics towards aiming the goal of matching human perception. Since the eye tracker data could be easily captured today by directly employing the software based eye-tracking simulator [24] (i.e. device itself is no longer required), the utility of the QMET could also be more flexible using such simple simulator generated data set.

2 Proposed Method

The first phase of the proposed quality metric design is to conduct the coding quality variation and different segments preparation which is executed by employing the HEVC [25] reference software HM15.0 [26]. These quality varied videos were then watched by a group of ten participants and their eye-tracking data were analyzed using four quality correlation features, i.e. *L*, *A*, *P*, and *T*. The content and resolution invariant operations were performed on the features and then synthesized by an adaptive weighted function to develop a new metric- QMET. The entire process is presented as a process diagram in Fig. 2 and the key steps are detailed in the following sub-sections.

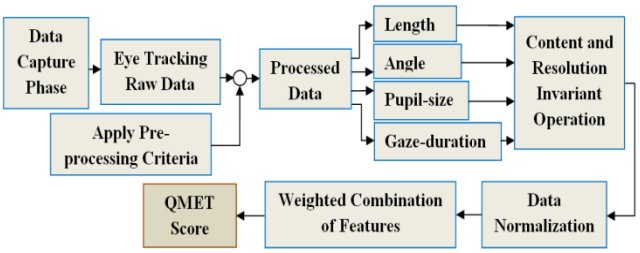


Fig. 2. Process diagram of the proposed QMET development.

2.1 Data Capture and Pre-processing

The participants (including males and females) who were recruited from the University had normal or corrected-to-normal vision and did not suffer from any medical condition that might be adversely influenced by our project [ethical approval no. 2015/124]. They fall within the 20-45 age band and are undergraduate/postgraduate students, PhD students, and lecturers of the University. A number of multiview sequences which are used in this test comprise the resolution type of 1920×1088 and 1024×768 (detail to be found in [27]). To avoid the biasness with color or contrast, initially we design experiment using the gray scale components only. We generate three different quality types of each video including *Excellent* (using *quantization parameter* QP=5), *Fair* (QP=25), and *Very-poor* (QP=50) and randomly display them to the participants. Calibration and a trial run was performed so that the participants feel comfort about the whole process. Upon their satisfaction, the Tobii eye tracker [28] was employed to record their eye movements. As the device recorded data at 60HZ frequency and allocated frame rate was 30 (fps), each frame could accommodate two gaze points and a single whole video covered 9000 gaze plots having 1800 for each quality segment.

2.2 Correlation Analysis of Features

The *Length* (*L*- in pixel) of the i^{th} *potential gaze plot* is calculated using the two dimensional *Euclidean distance* with respect to the $(i+1)^{th}$ gaze plot, while the *Angle* (*A*- in degree) of the i^{th} plot is calculated by using the reference of its $(i-1)^{th}$ and $(i+1)^{th}$ values (where $i=\{1,2,\ldots,n\}$ and the values of *L* and *A* are not calculated for the 1^{st} and n^{th} plots). The pupil-size (*P*- in mm) and Gaze-duration (*T*- in ms) on the other hand, are determined for each i^{th} plot by averaging the values of left and right pupil size and the eye-tracker recorded timestamp data respectively for all the sequences by

employing MATLAB R2012a (MathWorks Inc, Massachusetts, USA). The overall calculated results indicate that *L*, *A*, *P* features have a proportionate and *T* feature has an inversely proportionate correlation with the video quality degradation as depicted in Fig. 3.

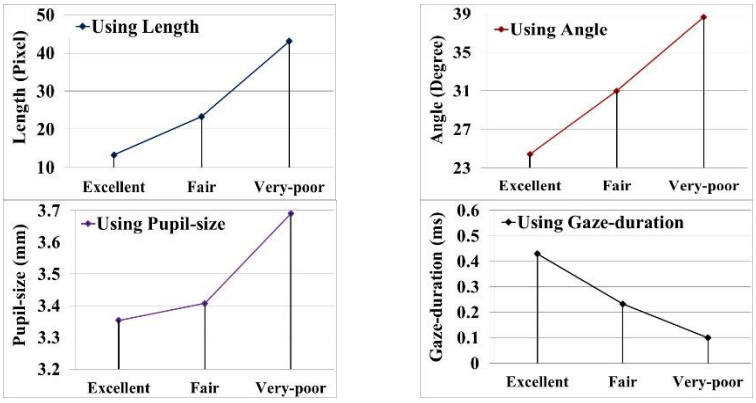


Fig. 3. The *Length* (*L*), *Angle* (*A*), and *Pupil-size* (*P*) features have a proportionate correlation, while the *Gaze-duration* (*T*) feature has an inversely proportionate correlation with quality degradation.

This time, we evaluate the contribution of each individual feature in the context of distinguishing different aspects of coded video quality using dissimilar quality segment and observe that none of them could discretely be the best representative in distinguishing different qualities. We determine the individual Q-score (i.e. calculated pseudo score of the QMET) for each feature by employing the equations (1)-(4), where Q_1, Q_2, Q_3, and Q_4 denote the Q-score for individual *L*, *A*, *P*, and *T* respectively.

$$Q_1 = L^{\partial L} \quad (1)$$

$$Q_2 = A^{\omega A} \quad (2)$$

$$Q_3 = (P/2)^{\delta P} \quad (3)$$

$$Q_4 = \sqrt{2T}^{(\aleph/\sqrt{2T})} \quad (4)$$

here, ∂, ω, δ, and $\aleph$ are the weighting factors of *L*, *A*, *P*, and *T* features respectively. Let us briefly discuss the formation of equations to produce different Q-scores using the power law. A power law is a functional relationship between two quantities, where a relative change in one quantity results in a proportional change in the other quantity, independent of the initial size of those quantities: one quantity varies as a power of another [29]. In our case, the relative value change of the features is unknown, and their corresponding reproduced Q-score is unknown as well, however, whether they have proportionate or inversely proportionate relation is known. For example, lower *L* indicates higher quality and respective higher Q-score, but still, we do not know how much. Since the value change of *L* for each quality segment is not significant (e.g. 0.08 for *Excellent* and 0.12 for Fair and the maximum average does not exceed 0.50), it could be best represented only by its power representation since smaller power with smaller base produces a higher score. Thus, a clear score difference among different quality

segments could be produced. The features A, and P similarly work as L with power-weight multiplication, however, since T has an inversely proportionate relation with Q-score, the power-weight division woks here in the same manner as presented in equations (1)-(4). The rationality of using the Q-score is to predict a better picture of the QMET performance change for various changes of L, A, P, and T within a sizable format that ranges from 0 to 1.

Since L, A, P, and T features could jointly advice about how far, how much, how large, and how long respectively in the spatiotemporal domain, we synthesize them by developing an adaptive weighted cost function as equated by $Q = L^{\partial L} \times A^{\varphi A} \times (P/2)^{\xi P} \times \sqrt{2T}^{(\aleph/\sqrt{2T})}$. The purpose of this multiplication is to keep a persistent relation of L, A, P, and T features with the reproduced Q-score. As the normalized value of the features varies within the range 0 to 1 and their manipulation in equations (1)-(4) also follow this range to yield the quality score, thus, their multiplication could better reproduce the ultimate result within the predefined limit. Note that the weight for ∂, φ, ξ, and $\aleph$ in the equations (1)-(4) is fixed with 0.5 in this test. This is because we further calculate the slope at each point changing the quality (i.e. *Excellent*, Fair, and so on) and determine their average for a number of weights. Since the calculated average using weight 0.5 outperforms the other weight combinations, we fix it for the entire experiment to best distinguish different quality segments which is demonstrated in Fig. 4. The distribution of other combination among features and weights might work better; however, the tested results demonstrate a good correlation of QMET with other metrics.

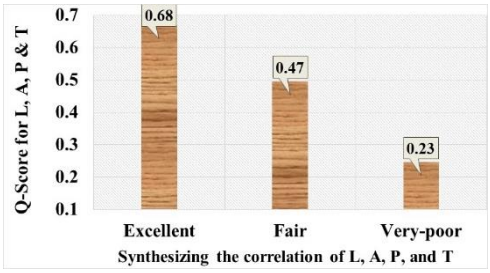


Fig. 4. The synthesizing operation using *Length*, *Angle*, *Pupil-size*, and *Gaze-duration* features could better distinguish different quality segments.

2.3 Invariant Operation on Features

Let us first ponder the content (left in Fig. 5) and resolution (right in Fig. 5) based unprocessed L of two example sequences e.g. *Poznan_Street* and *Newspaper* presented in Fig. 5. The calculated variations between the highest and lowest values are 41.72% and 28.63% according to the contents and resolutions respectively. Since the human vision is not equally susceptible to different video contents and resolutions, we, therefore, carry out the invariant operation on features. The content invariant operation follows a number of steps. First, we calculate the L of the PGPs as mentioned in Section- 2.2; Second, figure out the average of *potential gaze plot* (x) and *potential gaze plot* (y) and entitle them the centre C(x,y); Third, with respect to C(x,y), we estimate the two dimensional *Euclidean distance* of all PGPs and sort the calculated values of length by lowest to the highest order. The rationality of this ordering scheme is due to

prioritizing the foveal central concentration on pixels by partially avoiding the long surrounded parafoveal, or perifoveal fixations [30] that might occur even with attentive eye browsing; Fourth, to determine the object motion area according to the best viewing strategy, we take the average of first μ sorted values (75% in this test as it could help QMET in obtaining the highest score) which is the foreseen radius of captured affective region; Fifth, the radius is then employed as a divisor of calculated lengths for each PGP in the First step.

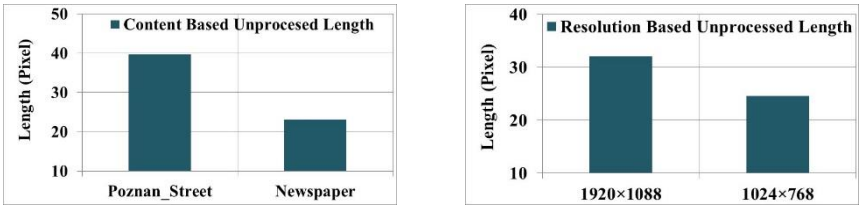


Fig. 5. The video content and resolution based unprocessed *Length*.

Similar to the content based lengths, we also observe a stunning variation of 28.63% for different resolution based lengths in Fig. 5 (right). As a result, we exploit a number of multiplication factors (passively act as compensators) eventually to neutralize the impact of various size video resolutions displayed on the screen. For example, assuming 1024×768 resolution sequence as a reference, the unprocessed lengths of its higher and lower resolution sequences are multiplied by 0.75 and 1.25 respectively. Almost for all the sequences, since the eye-tracker recorded data demonstrates a good correlation among the highest to the lowest resolution videos, the multipliers could perform well in resolution invariant operation. The outcomes then turn into the normalized values ranging within 0 to 1. The resultant effect of content plus resolution invariant operation for L is revealed in the top-left of Fig. 6 which is undertaken for the final QMET scoring. Once the similar operations are performed on the features A, P, and T, the variation effects could be significantly minimized as illustrated in the top-right, bottom-left and bottom-right respectively as demonstrated in Fig. 6.

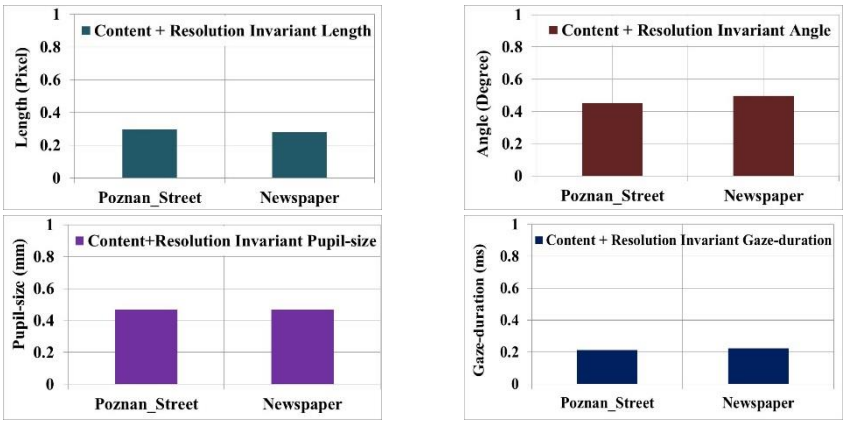


Fig. 6. The obtained values of L, A, P, and T (normalized) after performing the content and resolution invariant operation.

2.4 The Development of QMET

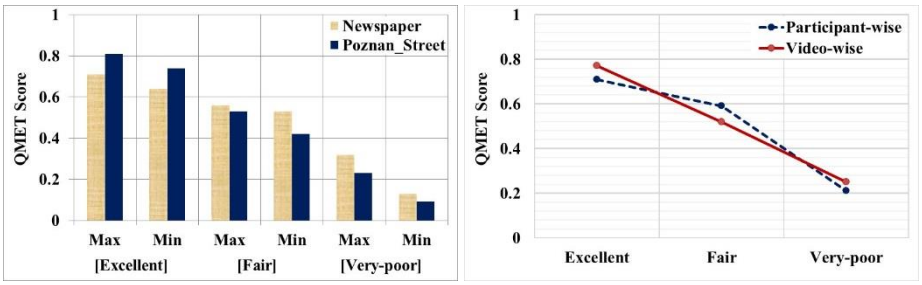
According to the hypothesis of the proposed algorithm, if relatively lower values of L, A, and P, and higher values of T belong to a PGP, it should produce higher QMET score. Thus, the QMET is calculated for all PGPs of each segment (i.e. *Excellent*, *Fair*, and *Very-poor*) of a sequence by adaptively synthesizing the features as follows:

$$Q_{MET} = L^{\partial L} \times A^{\varphi A} \times (P/2)^{\delta P} \times \sqrt{2T}^{(\aleph/\sqrt{2T})}. \quad (5)$$

Where the weight for ∂, φ, δ, and $\aleph$ is fixed with 0.5 as stated earlier. In an unusual case, if the normalized values of L and A become 0 for 30 consecutive frames (as the frame rate is kept 30 in this test), then a mimicking operation is performed. The rationality of allocating such operation is due to handling the consecutive 0s that may incur with the intentional eye fixation of participants to a certain PGP. Thus, the user data which have got stack over the frames are forcefully panelized by arbitrarily setting the value of L=0.1 and A =0.1. This operation is applicable only for the features L and A since P and T are still !=0 then. Note that during this test, we did not experience such unusual situation and carried out no such operation.

3 Experimental Outcomes

The QMET evaluated maximum and minimum scores for each quality segment using two example sequences are presented in Fig. 7 (a). For both sequences, the obtained score for the *Excellent* quality segment is the highest which gradually decreases with respect to the quality degradation and reaches its lowest for the *Very-poor* segment of quality. Compared to the *Newspaper*, the QMET score sharply decreases for the *Poznan_Street* sequence. This is because compared to its *Excellent* quality segment, the recorded supporting gaze data for the *Very-poor* quality incur with recurrent unsuitable feature values and produce a lower QMET score. Once we calculate the average score of each Max and Min for the individual quality segment, we notice that the average recognition of variation between the best and worst quality becomes 72.35% which indicate a clear quality distinguishing capability of the QMET.

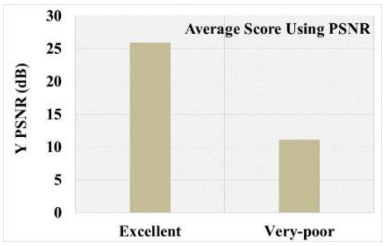


(a) Maximum (Max) to minimum (Min) QMET score at each quality segment using two test sequences

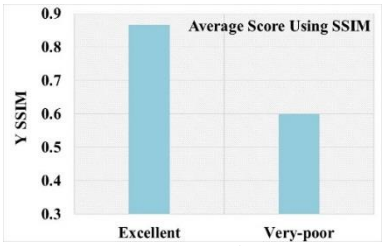
(b) The QMET score has a proportionate correlation to the coded video quality (person and video-basis)

Fig. 7. Different scoring orientations of QMET for a wide range of qualities (both the participant and video-basis).

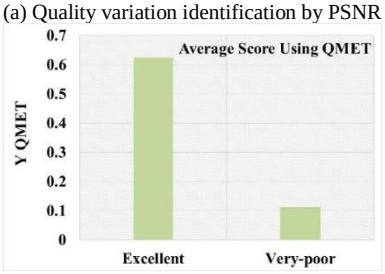
Fig. 7 (b) demonstrates the participant-wise and video-wise average QMET score for three different quality segments. The proposed QMET could obtain the highest score i.e. 0.78 and 0.71 for the *Excellent* quality segment according to both the video and participant as presented in Fig. 7 (b). This is because the participants could better capture information from the best quality contents with smooth global browsing. Conversely, for its lowest scores i.e. 0.25 and 0.21 at *Very-poor* segment, participants perhaps watch the video with a trial and error basis; i.e. try to capture content information but do not succeed due do its unpleasant quality and then immediately move to the next but still erroneous. As the number of such hits and miss browsing sharply increases with time, the quality score also decreases as plenty of inappropriate feature values incur with the scoring process. Therefore, for a sequence having really *Poor~Very-poor* quality, it becomes very unlikely to acquire higher quality score using the proposed QMET.



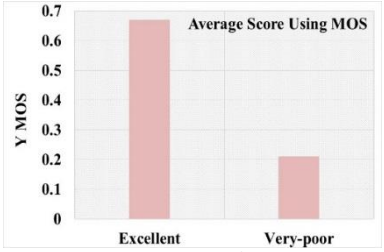
(a) Quality variation identification by PSNR



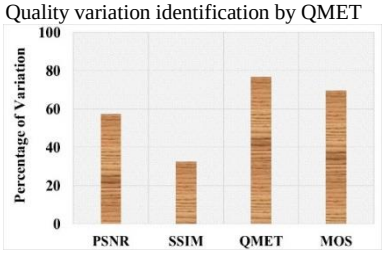
(b) Quality variation identification by SSIM



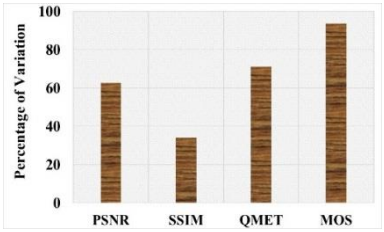
(c) Quality variation identification by QMET



(d) Quality variation identification by MOS



(e) Four metrics estimated average variations



(f)Four metrics assessed maximum variations

Fig. 8. In the Figure, (a~d) reveal the average quality variation identification carried out by the PSNR, SSIM, QMET, and MOS for the *Excellent* and *Very-poor* quality segments of free viewpoint videos which is more explicitly presented in (e), while (f) indicates the maximum achievable difference (e.g. the difference between the highest score of *Excellent* quality and the lowest score of *Very-poor* quality segment) obtained by four metrics.

This time, for better justifying the performance of QMET against the PSNR, SSIM, and the MOS using the FVV, two different quality segments (i.e. *Excellent* and *Very-poor*) have been taken into account. The calculated average score of four metrics for these segments are reported in Fig. 8 (a)-(d). The obtained percentages of variations between the highest score (for *Excellent* quality segment) and the lowest score (for *Very-poor* quality segment) using PSNR, SSIM, QMET, and MOS are 57.39, 32.49, 78.51, and 69.71 as represented in Fig. 8 (e). The outcomes indicate that the QMET estimated average quality segregation score outperforms the rest of the metrics. This is because viewers could better capture good quality synthesized video content with smooth global browsing. Conversely, the poorly reconstructed synthesized views incur with the localized edge reconstruction and crack like artifacts. Thus, the recorded gaze data of poor contents indicate participants' haphazard means of browsing (being affected by unsuccessful attempts due to unpleasant quality) that could not meet the balanced feature correlation criteria and generate lower QMET score. Fig. 8 (f) indicates the maximum achievable difference (e.g. the difference between the highest score of *Excellent* quality and the lowest score of *Very-poor* quality segment) picked out by the four metrics where the MOS could outperform the other metrics. The *Very-poor* quality segment of some synthesized video (e.g. *Newspaper*) incur with an arbitrarily nominated lower score such as 0.05 (out of 1.0) which lead to such stunning variations. The calculated results for free viewpoint videos in Fig. 8 indicate that the improvement using the subjective assessment such as MOS could perform better than those of the objective metrics PSNR and SSIM. This is mostly due to the PSNR and SSIM do not find an available reference image to calculate the score in this regard. However, according to Fig. 8 (e), the human visual perception based QMET could demonstrate relatively improved performance compared to the MOS in terms of segregating different aspects of coded video quality.

Fig. 9. The performance comparison of PSNR, SSIM, QMET, and MOS metrics on the *Excellent*, *Fair*, and *Very-poor* quality segment using FVV. Lower the calculated variation for a segment better the metric performance is presumed.

Now, two interesting observations: first, if different video contents are coded using the same quality (e.g. QP=5 for *Excellent*), the reproduced scores should not have stunning variations. However, the PSNR could not follow this trend and for most of the quality segments, its variation goes the highest as revealed in Fig. 9. Thus, it might lose its

suitability for a wide range of free view video sequences. On the other side, for the *Very-Poor* quality segment, the participants perhaps provide some unusually perceived arbitrary score for which the MOS reaches its apex and its proficiency drops down in this regard. This is also an example that mandates the development of another subjective metric other than MOS that could opt for relatively fairer scoring. Although the QMET performs better than PSNR and MOS, the SSIM appears most stable in this regard. This is because the SSIM is a perception-based model that considers degradation in an image mainly by recognizing the change in structural information.

To justify the second observation, i.e. even the same sequence is coded with a range of qualities, the recognition of quality variation should be prominent which has been verified by employing two ranges of variations (*Excellent ~ Fair* and *Fair ~ Very-poor*) and reported in Fig. 10. For the first range of segments, all the metrics with free view video although perform in a similar manner, the QMET appears the most responsive in differentiating the range of qualities. The SSIM tends to be the least responsive metric in this regard. For the second range of segments (i.e. *Fair ~ Very-poor*), the QMET and the MOS reach their apex to indicate their best performance in the context of quality segregation. Interestingly, for both range of segments, the subjective estimators perform relatively better compared to the objective ones.

Fig. 10. The PSNR, SSIM, QMET, and MOS metrics recognized percentage of quality variation for a range of quality segment differences. Higher the calculated percentage of variation detection in segments [X~Y], better the metric performance is presumed.

For further performance estimation of four metrics, the calculated results using entire videos used in this test are reported in Table 1 by implementing both the PLCC and SRCC's evaluation criteria. A good quality metric is expected to achieve higher values in both PLCC and SRCC [10]. According to both PLCC and SRCC's judgement, the QMET reveal the similar performance compared to the PSNR, however, it could obtain relatively higher score compared to the SSIM and MOS. In fact, the obtained results of the proposed metric are promising given the fact that no information about the reference image is available to the QMET for evaluating quality. Since the scoring pattern of four metrics are approximately similar in terms of distinguishing different quality contents as illustrated in Fig. 9, Fig. 10, and Table 1, the proposed QMET could be well represented as a new member of the quality metric family and successfully employed as an impressive alternative to the subjective estimator MOS. It could also be employed to evaluate the effectiveness of using the objective metrics PSNR and SSIM since the QMET does not require any ground-truth reference for quality estimation.

Table 1. Average performance of four metrics according to both PLCC and SRCC's evaluation criteria.

Performance Estimators	PSNR	SSIM	QMET	MOS
PLCC	0.68	0.63	0.69	0.68
SRCC	0.71	0.62	0.71	0.68

The potential application of QMET could be the evaluation of synthesized views (images) reproduced by different FVV generating algorithms. A good number of contributions could be found in the literature which claim about the image quality improvement mostly depending on the objective metric PSNR, SSIM or the subjective estimator MOS. However, it is presented earlier that the subjective estimator MOS performs better than the objective metrics in most cases during evaluating the FVV quality. Since the proposed QMET is mostly correlated to the proximity of human cognition, its assessment process is presumed to be more neutral compared to the MOS for assessing different aspects of coded video quality. Moreover, since the view synthesis algorithms go through some post-processing phases such as inverse mapping or inpainting for crack filling, it is highly anticipated to obtain higher quality evaluation score using QMET especially for those algorithms successfully overcoming the crack filling artifacts.

4 Conclusion

In this work, a no-reference video quality assessment metric has been developed based on the free view video. The newly developed metric QMET could be an impressive substitute to the popularly used subjective estimator MOS for quality evaluation and comparison. In the metric generation process, the human perceptual eye- traversing nature on videos is exploited and discovered the patterns of *Length*, *Angle*, *Pupil-size*, and *Gaze-duration* features from the recorded gaze trajectory for varied video qualities. The content and resolution invariant operations are carried out prior to synthesizing them using an adaptive weighted function to develop the QMET. The experimental analysis reveal that the quality evaluation carried out by the QMET is mostly similar to the MOS and the reference required PSNR and SSIM in terms of assessing different aspects of quality contents. Eventually, the outcomes of four metrics have further been tested using the Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank-Order Correlation Coefficient's (SRCC) evaluation criteria which indicate that the QMET could relatively better perform compared to the MOS and the SSIM for a wide range of free viewpoint video contents. Since the eye-tracker data could be easily captured nowadays by directly employing the software based eye-tracking simulator (i.e. device itself is no longer required), the utility of the QMET could also be more flexible using such simple simulator generated data set. Work is undergoing for the project "View synthesis using Gaussian mixture modelling of images from adjacent

views for free viewpoint and multiview video with eye-tracker-based quality assessment" where the newly developed QMET would be applied in a broader context such as increasing the number of free viewpoint videos and quality segments using the colour image components.

References

1. S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 4, pp. 516-529, April 2012.
2. K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Transactions on Cybernetics*, vol. 46, no.1, pp. 284-297, 2016.
3. D. Rahman, and M. Paul, "Adaptive weighting between warped and learned foregrounds for view synthesize" *IEEE International Conference on Multimedia and Expo*, 2017.
4. C. Zhu, and S. Li, "Depth image based view synthesis: New insights and perspectives on hole generation and filling" *IEEE Transactions on Broadcasting*, vol. 62, no. 2, pp. 82-93.
5. F. Battisti, E. Bosc, M. Carli, P. L. Callet, and S. Perugia, "Objective image quality assessment of 3D synthesized views," *Signal Processing, Image Communications*, vol. 30, pp. 78-88, January 2015.
6. M. Xu, J. Zhang, Y. Ma, and Z. Wang, "A novel objective quality assessment method for perceptual video coding in conversational scenarios," *IEEE Visual Communications and Image Processing Conference*, pp. 29-32, December, 2014.
7. J. You, T. Ebrahimi, and A. Perkis, "Attention driven foveated video quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 200-213, January 2014.
8. K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, " Quality assessment considering viewing distance and image resolution," *IEEE Transactions on Broadcasting*, vol. 61, no. 3, pp. 520-531, September 2015.
9. H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artefacts in images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 529-539, April 2010.
10. Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-Reference Quality Assessment of Contrast-Distorted Images Based on Natural Scene Statistics" *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838-842, July 2015.
11. K. Zhu, C. Li, V. Asari, and D. Saupe, "No-Reference Video Quality Assessment Based on Artifact Measurement and Statistical Analysis" *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 4, pp. 533-545, April 2015.
12. K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization" *IEEE Transactions on Cybernetics*, June 2016. DOI: 10.1109/TCYB.2016.2575544.
13. S. Tourancheau, F. Autrusseau, Z. M. P. Sazzad, and Y. Horita, "Impact of Subjective Dataset on the performance of image quality metrics," *International Conference on Image Processing*, pp. 365-368, 2008.
14. H. Liu, and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971-982, 2011.
15. M. Bohme, M. Dorr, M. Graw, T. Martinetz, and E. Barth, "A software framework for simulating eye trackers" *ACM Symposium on eye tracking research and applications*, pp. 251-258, 2008.

16. K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427-1440, June 2010.
17. L. Jia, X. Zhong, and Y. Tu, "No-reference video quality assessment model based on eye tracking data" *International conference on Information*, *Electronics*, *and Computer,* pp. 97-100, 2014.
18. S. Arndt, J. Radun, J. N. Antons, S. Moller, "Using eye-tracking and correlates of brain activity to predict quality scores," *IEEE International Workshop on Quality of Multimedia Experience*, pp. 281-285, 2014.
19. M. G. Albanesi, and R. Amadeo, "A new algorithm for objective video quality assessment on eye tracking data" *IEEE International Conference on Computer Vision Theory and Applications*, pp. 462-469, January 2014.
20. C. M. Tsai, S. S. Guan, and W. C. Tasi, "Eye movements on assessing perceptual image quality" *Springer International Publishing*, pp. 378-388, 2016.
21. F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation" *IEEE International Conference on Image Processing*, pp. 3097-3100, September, 2011.
22. R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives" *Multimedia Systems*, vol. 22, no. 2, pp. 213-227, 2016.
23. P. Podder, M. Paul, and M. Murshed, "QMET: A new quality assessment metric for no-reference video coding by using human eye traversal" *Image and vision computing New Zealand*, 2016.
24. M. Bohme, M. Dorr, M. Graw, T. Martinetz, and E. Barth, "A software framework for simulating eye trackers" *ACM Symposium on eye tracking research and applications*, pp. 251-258, 2008.
25. Bross, Han, W. J. Ohm, J.R. Sullivan, and G. J. Wiegand, "High Efficiency Video Coding Text Specification Draft 8," JTCVC- J1003, Sweden 2012.
26. Joint Collaborative Team on Video Coding (JCT-VC), HM Software Manual, CVS server at: (http://hevc.kw.bbc.co.uk/svn/jctvc-hm/), date of exploration December 2016.
27. P. Podder, M. Paul, D.M. Rahaman, and M. Murshed, "Improved depth coding for HEVC focusing on depth edge approximation," *Signal Processing: Image Communications*, vol. 55, pp. 80-92, July 2017.
28. An Exploration of Safety Issues in EyeTracking" http://www.academia.edu/245642/An_Exploration_of_Safety_Issues_in_Eye_Tracking , retrieve date April, 2015.
29. The basics of Power law. https://en.wikipedia.org/wiki/Power_law, date of exploration: December 2016.
30. M.M. Salehin, and M. Paul, "Human visual field based saliency prediction method using Eye Tracker data for video summarization," *IEEE International Conference on Multimedia & Expo*, July 2016.