COPYRIGHT NOTICE

A Novel Quality Metric Using Spatiotemporal Correlational Data of Human Eye Maneuver

***Abstract*— The popularly used subjective estimator- *mean opinion score* (MOS) is often biased by the testing environment, viewers mode, domain expertise, and many other factors that may actively influence on actual assessment. We therefore, devise a no-reference subjective quality assessment metric by exploiting the nature of human eye browsing on videos. The participants' eye-tracker recorded gaze-data indicate more concentrated eye-traversing approach for relatively better quality. 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 quality evaluation carried out by QMET demonstrates a strong correlation with the most widely used *peak signal-to-noise ratio* (PSNR), *structural similarity index* (SSIM), and the MOS.**

***Keywords*— *Eye-traversal*, *Eye-tracking*, *Gaze-trajectory*, HEVC, MOS, QMET, Quality Assessment.**

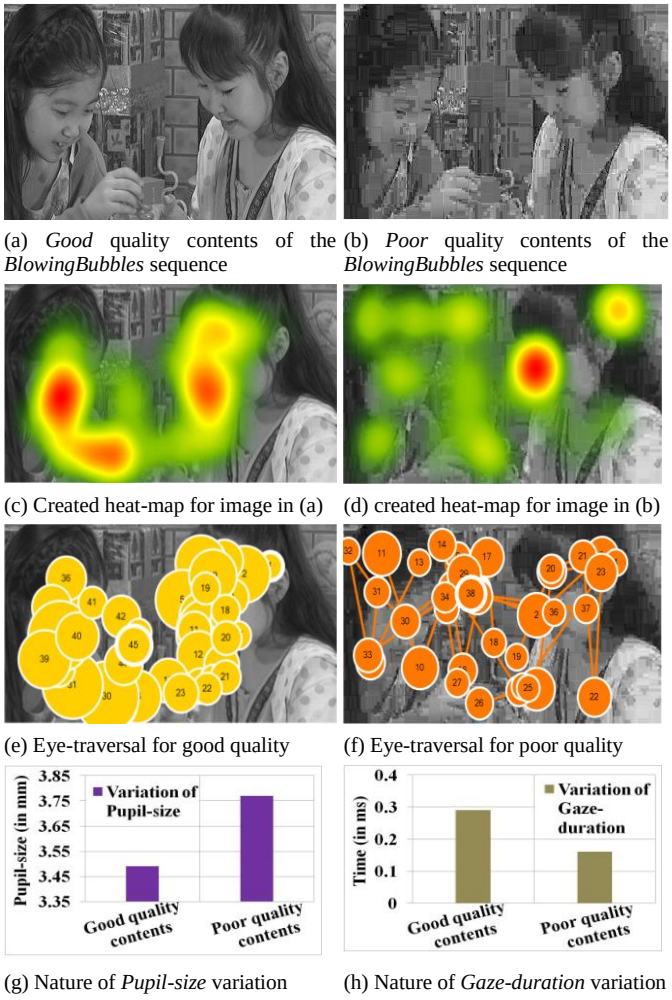
I. INTRODUCTION

Video quality evaluation (VQE) is a prominent research area due to its wide range of applications in the development of various video coding algorithms [1]-[3]. 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 [4]-[7]. The quality estimation could be mainly categorized into full-reference (i.e. original videos as reference), reduced-reference (i.e. existing of partial signals as reference) and no-reference schemes. The no-reference models become more challenging due to the lack of original reference signal to analyze [8]. Moreover, the applications of full-reference metrics such as SSIM or PSNR have been restricted to the reference based situations only. To overcome this limitation, a number of no-reference based research works have recently come into light for quality evaluation. Based on the principle of *natural scene statistics* (NSS), authors in [9] introduce a no-reference quality assessment method of contrast distorted images using unnaturalness characteristics and justifying the degree of deviation from the NSS models. The video quality prediction model using the discrete cosine transform is presented in [10] to analyze the statistics of compressed natural videos. The authors in [11] introduce a no-reference metric for quality assessment of contrast distorted image by analyzing and combining the local and global details of an image. These statistical metrics may not be suitable in some high quality range as quality perception in these area is mostly due to perceptual human visual system (HVS) features, rather than to the statistics of the image [12], however, different features of the HVS is not actively studied in the existing schemes.

The authors in [13] carry out the human cognition based objective quality assessment system using the eye-tracking technology and evolve 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 [14].

Unlike objective quality evaluation, the subjective one is impractical for some applications due to the human engagement in the process. However, it could yield valuable data to evaluate the performance of objective methods towards aiming the ultimate goal of matching human perception [15]. To this end, a number of quality assessment algorithms have been proposed which are closely related to the studies of human visual attention and cognition. Jia *et al*. [16] propose a no-reference model using blur and blockiness metric to improve the performance of objective model based on eye-tracker data. The authors in [17] introduce a model to judge the video quality on the basis of psychological merits including- the pupil dilation and electroencephalogram signalling. Since they test their scheme only for an arbitrarily nominated and degraded portion of a frame which limits their scheme for its further use. Albanesi *et al*. [18] use the eye data to create a voting algorithm to develop a no-reference method. Using the scan path of eye movements, Tsai *et al*. [19] subjectively assess the perceived image and its colour quality. Tested results prove that percipients tend to spend more time in evaluating the image with relatively improved quality. On the other side, the widely used subjective testing method- MOS [20][21] 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 [22] although introduce the QMET, their initial work incur with the limitation of proper feature correlation setting and highly depends on user defined threshold for each feature. This work is a significantly amended version of their previous work where the included major extensions are- the increased number of features, their correlation analysis, performing content and resolution invariant operation, synthesizing them by a weighted function, comparing the new metric with PSNR, SSIM, and MOS, and eventually employing two estimators *Pearson Linear*

Correlation Coefficient (PLCC) and the *Spearman Rank-Order Correlation Coefficient* (SRCC) [9] to justify the effectiveness of QMET for using it as an impressive substitute to the MOS.



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

(b) *Poor* quality contents of the *BlowingBubbles* sequence

(c) Created heat-map for image in (a)

(d) created heat-map for image in (b)

(e) Eye-traversal for good quality

(f) Eye-traversal for poor quality

(g) Nature of *Pupil-size* variation

(h) Nature of *Gaze-duration* variation

Fig. 1. More concentrated eye-traversing approach is perceived for relatively better quality contents (e.g. *BlowingBubbles* sequence image in (c) and (e)). The opposite is noticed in (d) and (f) for which the *Pupil-size* sharply increases in (g), while the *Gaze-duration* notably decreases as shown in (h).

Let us first concentrate on the Fig. 1 in which (a) and (b) represent *BlowingBubbles* sequence having good and poor quality contents respectively, while (c) and (d) indicate their corresponding reproduced heat-maps generated by the eye tracker. As the heat-map inherently indicates the participants' concentration by reproducing deep reddish region in the image, the eye browsing nature in (d) tents to be more scattered compared to the one in (c). This observations trigger us to further calculate each plot of eye tracker recorded spatiotemporal gaze data. Fig. 1 (e) and (f) therefore, demonstrate the eye traversing approach of a viewer for good and poor quality image contents respectively and the tracked gaze plots indicate more concentrated eye-traversal for relatively better quality contents. As the trend is observed for the whole *n* frames of a sequence, the spatiotemporal correlation of gaze plots are analyzed to develop the proposed QMET. The higher QMET score promise good quality video as the viewers could better capture its content information with smooth global browsing.

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 (e~f)). 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 (g~h), therefore, four cardinal features- *L*, *A*, *P*, and T are calculated 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 a new metric-QMET. Experimental results reveal that the quality evaluation carried out by QMET has a good correlation to the HM recommended coding quality and the widely used PSNR, SSIM, and 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 [14] (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.

The remainder of this paper is organized as follows. Section II explicitly presents the key steps of the proposed implementation; Section III evaluates the tested results in detail, while Section IV concludes the paper.

II. PROPOSED TECHNIQUE

The first phase of the proposed quality metric design is to conduct the coding quality variation and five different segments preparation which is executed by employing the *High Efficiency Video Coding* (HEVC) [23] reference test model HM15.0 [24]. These quality varied videos were then watched by a group of ten participants, their eye-tracking data were analyzed using four quality correlation features (i.e. *L*, *A*, *P*, and *T*), performed invariant operation on features, and synthesized them by an adaptive weighted function eventually to develop a new metric-QMET to recognize human perception and response to the video quality variation. The entire process is presented as a process diagram in Fig. 2 and the key steps are detailed in the following sub-sections.

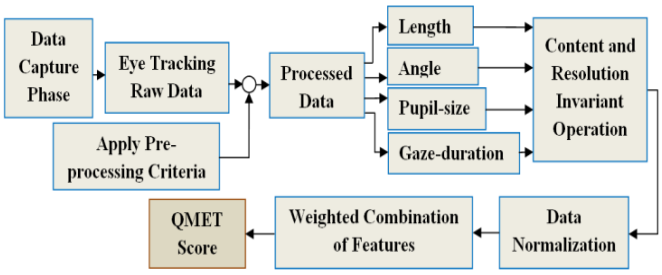


Fig. 2. Process diagram of the proposed technique.

A. Data Capture And Pre-processing

Participants 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]. A total of 15 people (including males and females) who were recruited fall within the 20-45 age band and were undergraduate/postgraduate students, PhD students, and lecturers of the University. The HEVC recommended eight class sequences were used in the test namely: *Traffic* (2560×1600- class A), *Cactus* (1920×1080- class B), *Tennis* (1920×1080- class B), *BasketballDrill* (832×480- class C), BQMall (832×480- class C), *Blowingbubbles* (416×240- class D), *Flowervase* (416×240- class D), and *FourPeople* (1280×720- class E) (detail to be found in [25][26]). To avoid the biasness with color or contrast, initially we design experiment using the gray scale components only. We generate five different quality types of each video including *Excellent* (using *quantization parameter* QP=5), *Good* (QP=15), *Fair* (QP=25), *Poor* (QP=40), and *Very-poor* (QP=50). The video display order randomly varied for sequences. For example, if the visual display for *Traffic* was carried out in the order of *Excellent* to *Very-poor*, for the same participant, the display order for the next video *Cactus* was reversely designed from *Very-poor* to *Excellent* to avoid the participants' biasness with content and quality in the experiment. Each segment was 30 seconds long and the segment gap was 3 second. Calibration and a trial run was performed so that the participants feel comfort about the whole process. Upon their satisfaction, the Tobii eye tracker [27][28] was employed to record their eye movements and the completion of whole process took about 30 minutes for each participant. 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.

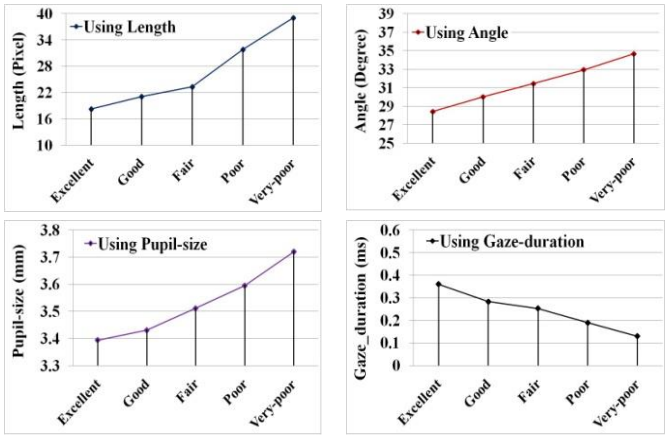


Fig. 3. The observed relationship between the features and the qualities.

B. Correlation Analysis of Features

The *Length* (L- in pixel) of i^{th} PGP 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,…,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 proportional and *T* feature has an inverse proportional correlation with the video quality degradation as depicted in Fig. 3. However, it is observed by the authors in [18] that for relatively poor quality image, the participants tend to spend few more time for assessing quality that contradicts the outcomes presented in the bottom-right of Fig. 3. We also observe this mostly for the still images where people have enough leisure to spend more time to perceive a specific location. This, in turn, becomes impractical for videos since the frames move continuously that causes the continuous changes of object positions and does not allow the viewers to perceive the same scene even a few frames later. Thus, we interestingly notice that for the *Very-poor* quality video, the participants tend to spend rather less time on specific plots which is mostly due to appeared continuing unpleasant quality and the intention of the participants to look for better visibility.

Now, we evaluate the contribution of each individual feature in terms of distinguishing different aspects of coded quality using dissimilar quality segment and observe that none of them discretely could be the best representative in terms of quality distinction. Since the human vision is not equally susceptible to different video contents and resolutions, we, therefore, carry out the content and resolution invariant operation (to be discussed in Section II-C) on the obtained feature values. Using the normalized data set for five sequences (i.e. one from each Class type), then we figure out a relationship of the features with video quality variation which is pictured in Fig. 4. To calculate the Q-score (i.e. the pre-processed score of the QMET) for *L*, *A*, *P*, and *T*, the equations (1)-(4) have been employed respectively where Q_1, Q_2, Q_3, and Q_4 denote the Q-score for individual *L*, *A*, *P*, and *T* respectively which is better illustrated in Fig. 4.

$$Q_1 = L^{\alpha L} \tag{1}$$

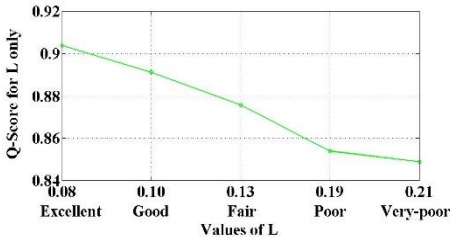
$$Q_2 = A^{\varphi A} \tag{2}$$

$$Q_3 = (P/2)^{\beta P} \tag{3}$$

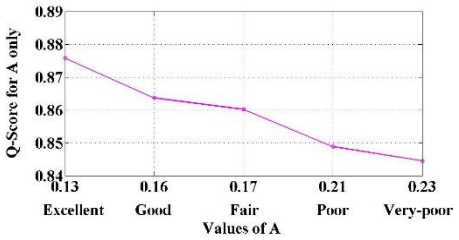
$$Q_4 = \sqrt{2T}^{(\tau/\sqrt{2T})} \tag{4}$$

In the equations, the symbols α, φ, β, and τ denote the weighting factors of *L*, *A*, *P*, and *T* respectively. Let's briefly discuss about the formation of equations to produce different Q-scores using the power law [29]. For example, the value change of *L* for each quality segment is not significant (e.g. 0.08 for *Excellent* and 0.10 for *Good* as shown in Fig. 4 (a) and the maximum average does not exceed 0.60), it could be best represented by its power representation since smaller power with smaller base produces higher score. Thus, a clear score difference among different quality segments could be produced. The features *A*, and *P* also similarly works as *L* with power multiplication, however, since *T* has an inverse relation with Q-score, the power division woks here in the same manner as presented in Fig. 4 (a)-(d). The rationality of using the Q-score is to predict a better picture of the QMET's performance change

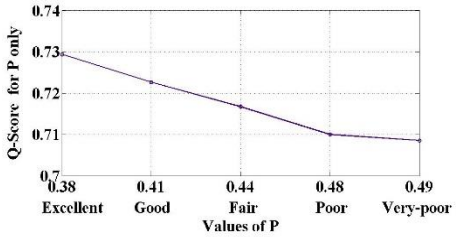
for various changes of *L*, *A*, *P*, and *T* within a sizable format that ranges from 0 to 1. Fig. 4 (a) reveal that *L* itself is not always a good indicator in segregating different quality contents since it produces almost the similar score both for *Poor* and *Very-poor* quality segments. The similar picture could be noticed for *A* and *P* features in Fig. 4 (b) and (c) respectively. The *T* feature in (d) itself is relatively more consistent compared to others, however, if excessive eye blinking incurs with captured data, it indeed loses it suitability.



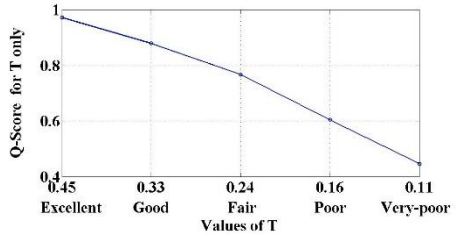
(a) Q-score calculation for different values of *L* using equation (1)



(b) Q-score calculation for different values of *A* using equation (2)



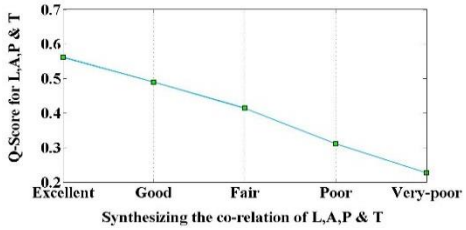
(c) Q-score calculation for different values of *P* using equation (3)



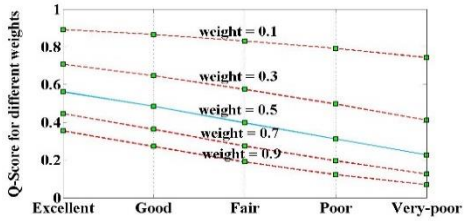
(d) Q-score calculation for different values of *T* using equation (4)

Fig. 4. The individual contribution of *Length*, *Angle*, *Pupil-size*, and *Gaze-duration* features in terms of segregating different quality segments.

Since *L*, *A*, *P*, and *T* features could best advice about how far, how much, how large, and how long respectively both in the spatial and temporal domain, we synthesize them by developing an adaptive weighted cost function as equated by $Q = L^{\partial L} \times A^{\varphi A} \times (P/2)^{\check{s}P} \times \sqrt{2T}^{(\aleph/\sqrt{2T})}$. Using the feature values of Fig. 4, the obtained outcome of the synthesizing operation is demonstrated in Fig. 5 (a) which clearly indicates the distinguishing Q-scores for different qualities. As the normalized value of the features varies within the range 0 to 1 and their manipulation in equation (1)-(4) also follow this range to yield the quality score, thus, their multiplication could better reproduce the ultimate result within the predefined limit. The distribution of other combination among features and weights might work better, however, the experimental results demonstrate a good correlation of QMET with other metrics. Note that the weight for α, φ, β, and τ in the above equations (1)-(4) is fixed with 0.5 in the experiment and the rationality of selecting 0.5 is further validated in Fig. 5 (b). According to the proposed implementation, the quality degradation is inversely proportional to the obtained Q-score. Therefore, we further calculate the slope at each point changing the quality (i.e. *Excellent*, *Good*, and so on) and determine their average for each individual weight used in Fig. 5 (b). Since the calculated average using weight 0.5 outperforms the other weight combinations, we fix it for the entire experiment.



(a) Combined contribution of *L*, *A*, *P*, and *T* features in terms of segregating different quality segments

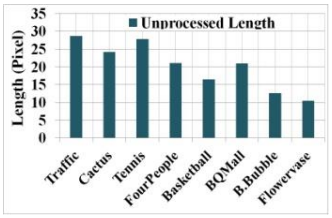


(b) Justification of different weights and selection of 0.5 for the combined feature generated graph in (a)

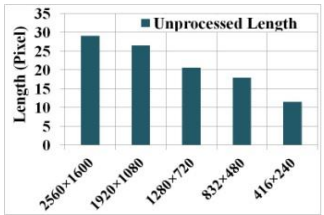
Fig. 5. The effectiveness of synthesizing the correlation of *L*, *A*, *P*, and *T* to best represent the distinguishing quality scores for varied quality segments using the weighting factor 0.5 for which the average slope value is the highest.

C. Invariant Operation on Features

Let's first ponder the unprocessed *L* in Fig. 6 (a~b) and calculate 61.96% and 52.57% variations (using highest and lowest values) according to the contents and resolutions respectively. The content invariant operation is thus first performed to neutralize the impact of contents to human vision that follows a number of steps:



(a) Content based



(b) Resolution based

Fig. 6. The video content and resolution based unprocessed *Length*.

First, we calculate the *L* of PGPs as mentioned in Section- II-B; **Second,** figure out the average of PGP(*x*) and PGP(*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; **Fourth,** to determine the object motion area according to the best viewing strategy, we take the average of first 75% sorted values which is the foreseen radius of captured affective region; **Fifth,** the radius is then employed as a divisor of the calculated lengths for each PGP in the First step.

TABLE I. EFFECT OF CONTENT + RESOLUTION INVARIANT OPERATION ON LENGTH WHICH IS EMPLOYED FOR EVENTUAL QMET SCORING.

Resolution	Multiplication factor for resolution invariant operation	**Previous variations**		**Present variation**
		Content based	Resolution based	After content plus resolution invariant operation
2560×1600	0.65	61.96 %	52.57 %	10.67 %
1920×1080	0.75			
1280×720	1.00			
832×480	1.25			
416×240	1.75			

Since we also observe its stunning variations for different resolutions in Fig. 6 (b), therefore, to execute the resolution invariant operation, we exploit a number of multiplication factors as inserted in the second column of TABLE I. The rationality of employing such multiplication factors is to best neutralize the impact of various size display resolutions appeared on the screen. For example, before the resolution invariant operation, the variation between (2560×1600), and (416×240) resolution based sequences was 62.16%, however, applying the multipliers in TABLE I, variation now downs to 12.5%. Moreover, almost for all the sequences since the eye tracker recorded data shows 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 third column of TABLE I shows the previous variations for all sequences, while the fourth column reveals that the content plus resolution invariant operation could reduce the average variation down to only 10.67%. The resultant effect is revealed in the top-left of Fig. 7 for the feature *L* and 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. 7.

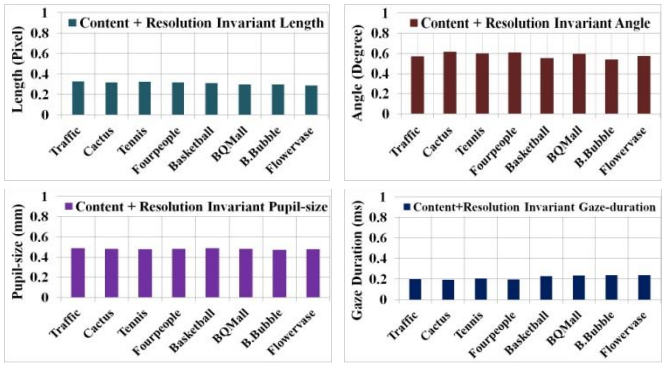


Fig. 7. The content plus resolution invariant normalized *L*, *A*, *P*, and *T* features.

D. The Development of QMET

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

$$Q_{MET} = L^{\alpha L} \times A^{\varphi A} \times (P/2)^{\beta P} \times \sqrt{2T}^{(\tau/\sqrt{2T})} \quad (5)$$

where the associated weighted values of *L*, *A*, *P*, and *T* denoted by α, φ, β, and τ respectively are fixed with 0.5 in this experiment. The rationality of this multiplication is to keep a consistent relation of *L*, *A*, *P*, and *T* features with the previously reproduced Q-score. As the normalized value of the features varies within the range 0 to 1 and their manipulation in equation (1) - (4) also follow this range to yield the quality score, thus, their multiplication could better reproduce the ultimate result within the predefined limit. The distribution of other combination among features and weights might work better, however, the experimental results demonstrate a good correlation of QMET with other metrics. 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 reason 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 the whole experiment, we did not experience such unusual situation. If the QMET evaluated quality scores are closer to 1, the video content quality is rated best, while, the opposite happens for scores closer to 0.

III. EXPERIMENTAL RESULTS

The QMET evaluated minimum and maximum scores for each segment of quality using all videos are presented in Fig. 8 while, their counted average score indicate it's quality segregation proficiency for varied qualities. The calculated score for the *Excellent* quality segment is 0.79 which gradually decreases with quality degradation and reach 0.27 for *Very-poor* quality contents once we calculate the average score of Max and Min for each quality segment. Thus we find a clear score declining pattern from *Excellent* to *Very-poor* quality contents.

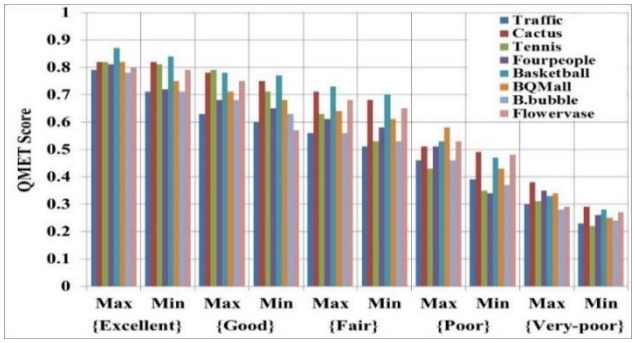
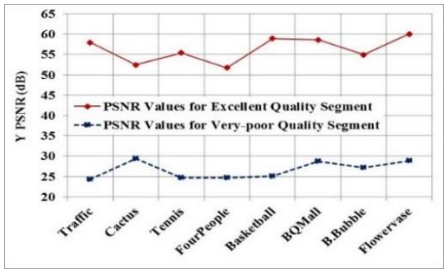
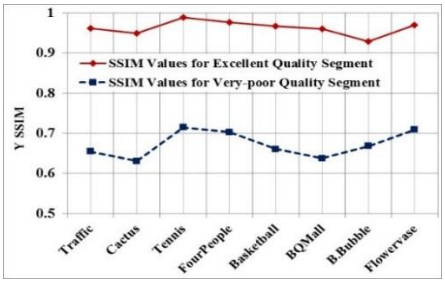


Fig. 8. The maximum (Max) to minimum (Min) QMET score for each quality segment showing the score degradation from *Excellent* to *Very-poor*.

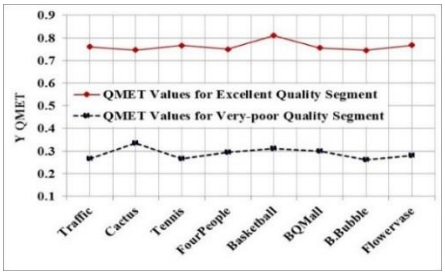
In Fig. 9, the PSNR and SSIM evaluated best quality video-set (i.e. highest, second-highest and third-highest scorer) for the *Excellent* segment include [*Flowervase*, *Basketball*, *BQMall*] and [*Tennis*, *FourPeople*, *Flowervase*] respectively. Conversely, for the same segment, the QMET and MOS picked sets include [*Basketball*, *Flowervase*, *Tennis*], and [*BQMall*, *Fourpeople*, *Flowervase*] respectively. The *Flowervase* sequence is common in the highest scoring list of all the metrics.



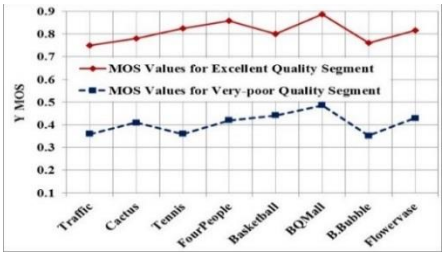
(a) PSNR evaluated quality score



(b) SSIM evaluated quality score



(c) QMET evaluated quality score



(d) MOS evaluated quality score

Fig. 9. The PSNR, SSIM, QMET, and MOS scores for *Excellent* and *Very-poor*. These two graphs could best segregate the best and worst quality.

In contrast, for the *Very-poor* quality segment, major dissimilarity could be found for the *Tennis* as it obtains the lowest score according to the PSNR, QMET, and MOS's assessment criteria, however, the SSIM scores it highest. This is possibly the SSIM is a perception-based model that considers degradation in an image mainly by recognizing the change in structural information. Interestingly, similarity among four metrics could be noticed for the *Traffic* sequence as it is assessed one of the lower scorers by all these metrics. The proposed QMET could obtain the highest and lowest score 0.82 and 0.26 for the *Excellent* and *Very-poor* quality segment respectively. This stunning difference is because the participants could better capture information from the best quality contents with smooth global browsing. However, for *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. Thus, a sequence having really *Poor~Very-poor* quality, it is very unlikely to acquire higher score using QMET.

Fig. 10 (a~d) illustrate the PSNR, SSIM, QMET, and MOS evaluated average scores obtained for all videos using two quality segments (i.e. *Excellent* and *Very-poor*). The calculated percentage of variation between the highest and lowest score using four metrics are 54.89, 33.07, 58.16, and 49.83 respectively which indicate the QMET could best segregate the best and worst quality contents as shown in Fig. 10 (e). The four metrics estimated maximum variations i.e. the calculated difference between the highest score of *Excellent* quality and the lowest score of *Very-poor* quality are further shown in Fig. 10 (f) in which the MOS tends to obtaining the highest score which is mostly because of inserting the arbitrary score of the participants for both *Excellent* and *Very-poor* quality contents.

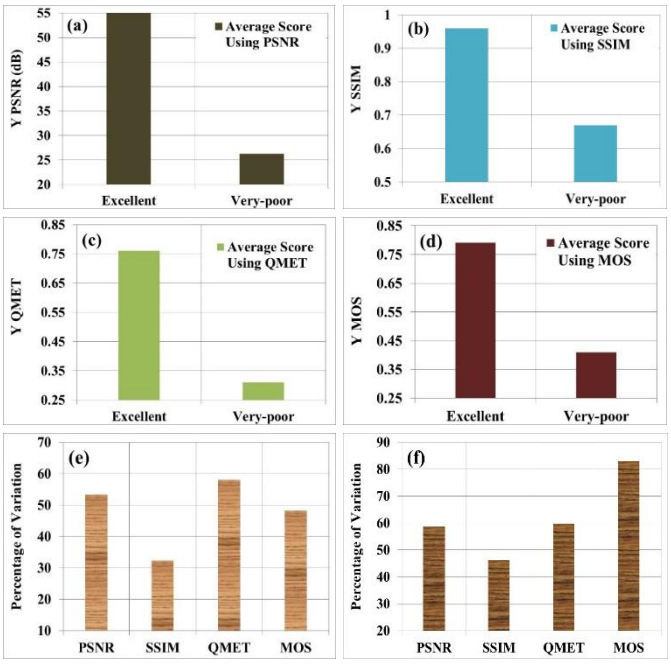


Fig. 10. In the Figure, (a~d) reveal the PSNR, SSIM, QMET, and MOS induced average values for the *Excellent* and *Very-poor* quality segment; (e) indicates the four metrics estimated average percentage of variation between the best and worst quality; while (f) points to the maximum achievable difference (e.g. the difference between the highest score of *Excellent* quality and the lowest score of *Very-poor* quality) obtained by the metrics.

The quality variation recognition score for all individual sequence is further discussed in Fig. 11 as we use a wide range of video contents and resolutions. We observe in Fig. 11 that the QMET could recognize its maximum quality variation for the *Tennis* sequence. This is because the participants could better capture the information from its *Excellent* quality segment with smooth global browsing. However, for its *Very-poor* quality segment, the participants' distorted pattern of browsing produce imperfect feature values and device poor score that eventually results in a higher score difference. The opposite happens for the *Basketball* or *Flowervase* because of relatively higher and continuing correlation among the aforementioned quality segments. Unlike SSIM, PSNR, QMET, the MOS perform in a similar fashion over all sequences shown in Fig. 11.

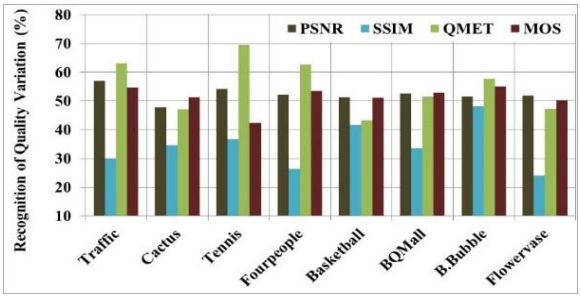


Fig. 11. The PSNR, SSIM, QMET, and MOS estimated percentage of quality variation using all the Class sequences used in this experiment.

Now, two interesting observations: first, if different video contents are coded using the same QP (e.g. 5 for *Excellent*), the produced scores should not have a 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. 12. Interestingly, for the *Fair* and *Poor* quality, the participants perhaps provide some unusually perceived arbitrary score for which the MOS loses its suitability in this regard. This is also an example that mandates the development of another human perception based metric that could opt for relatively fairer scoring. Although the QMET performs better than PSNR and MOS, the SSIM appears most stable in this regard.

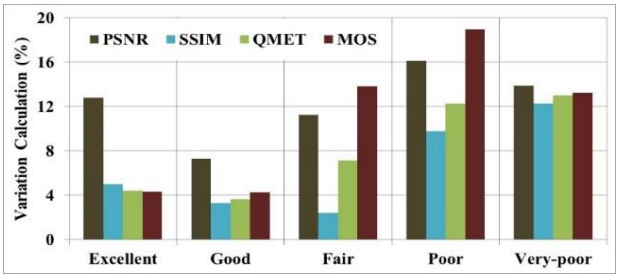


Fig. 12. The obtained score using the PSNR, SSIM, QMET, and MOS when videos are coded for the same segment of quality. For each segment, lower the percentage of variation, better the metric is presumed.

About the second observation, while using the same sequence coded with a range of QPs, due to its different quality variations, the score variations should be prominent as well. Although the QMET proves analogous results with the PSNR, it outperforms the SSIM and MOS in most cases as presented in Fig. 13. The results in the Figure come from averaging the calculated score of all the Class sequences used for experiment.

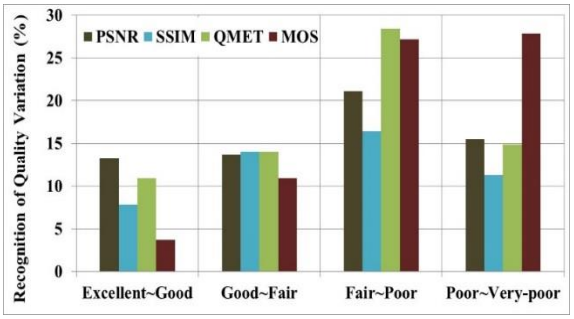


Fig. 13. The PSNR, SSIM, QMET, and MOS recognized percentage of quality variation due to the varied ranges of qualities. The higher differences obtained at segments [X~Y] indicate relatively better recognition capability of a metric.

The performance of the proposed QMET is further compared with the PSNR, SSIM, and MOS by employing two frequently used performance estimators: PLCC and the SRCC. A good quality is expected to achieve high values in both PLCC and SRCC [9]. First, the PLCC score of four metrics are calculated according to the sequences and the corresponding scores obtained for each one are reported in TABLE II. Almost in all cases, the QMET demonstrates inferior performance compared to the PSNR, however, it could outperform both the SSIM and MOS for most of the sequences reported in in TABLE II.

TABLE II. AVERAGE PERFORMANCE OF FOUR METRICS ACCORDING TO THE EVALUATION CRITERIA OF PLCC.

Sequences	**PLCC**			
	PSNR	SSIM	QMET	MOS
Traffic	0.73	0.61	0.57	0.56
Cactus	0.56	0.59	0.61	0.60
Tennis	0.71	0.67	0.66	0.61
Fourpeople	0.59	0.69	0.65	0.68
Basketball	0.77	0.62	0.71	0.63
BQMall	0.79	0.55	0.69	0.64
B.Bubble	0.69	0.53	0.58	0.61
Flowervase	0.73	0.68	0.68	0.65
Average	**0.69**	**0.61**	**0.64**	**0.62**

Similarly, we also calculate the result using SRCC and the produced overall average gain is summarized in TABLE III. In terms of both PLCC and SRCC's assessment criteria, the QMET reveal relatively improved performance compared to the SSIM and MOS, however, the PSNR is a clear winner as it obtains the highest score in both cases. In fact, the obtained results of the proposed approach 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. 11, Fig. 12, Fig. 13, TABLE II, and TABLE III, 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. Since the proposed QMET is entirely the human cognition based metric and the participants have almost no scope to actively manipulate the score, its assessment process is relatively more neutral compared to the MOS.

TABLE III. AVERAGE PERFORMANCE OF FOUR METRICS ACCORDING TO BOTH PLCC AND SRCC'S EVALUATION CRITERIA.

Quality metrics	**PLCC**	**SRCC**
PSNR	0.69	0.71
SSIM	0.61	0.58
QMET	0.64	0.61
MOS	0.62	0.60

IV. CONCLUSION

In this work, we introduce a novel no-reference subjective quality assessment metric that could be an impressive substitute to the popularly used subjective estimator MOS for quality evaluation and comparison. We simply exploit the human eye traversal on videos and discover the patterns of *Length*, *Angle*, *Pupil-size*, and the *Gaze-duration* features from the recorded gaze trajectory. The content and resolution invariant operation is carried out prior to synthesizing them using weighted function to develop a new quality metric- QMET. Tested analysis reveal a good correlation of QMET with the widely used PSNR and SSIM, while in most cases it performs relatively better than the MOS in terms assessing different aspects of coded video quality. Since the eye tracker data could be easily captured today by directly employing the software based eye-tracking simulator (i.e. device itself is no longer required), the utility of the newly developed metric could also be more flexible on such simple simulator generated data set. Other than the video coding applications, the QMET could be also applied in classroom education such as written program evaluations. For instance, a highly organized program (written in any programming language) could be anticipated having higher QMET score compared to the poorly organized one.

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, January 2016.

[3] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695-4708, December 2012.

[4] 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.

[5] 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.

[6] 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.

[7] W. Zhang, A. Borji, Z. Wang, P. L. Callet, and H. Liu, "The applications of visual saliency models in objective image quality assessment: a statistical evaluation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1266-1278, June 2016.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] C. M. Tsai, S. S. Guan, and W. C. Tasi, "Eye movements on assessing perceptual image quality" *Springer International Publishing*, pp. 378-388, 2016.

[20] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation" *IEEE International Conference on Image Processing*, pp. 3097-3100, September, 2011.

[21] 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.

[22] P. Podder, M. Paul, and M. Murshed, "QMET: A new quality assessment metric for no-reference video coding by using human eye traversal" *IEEE International conference on Image and vision computing New Zealand*, November, 2016.

[23] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, December 2012.

[24] Joint Collaborative Team on Video Coding (JCT-VC), HM Software Manual, CVS server at: (http://hevc.kw.bbc.co.uk/svn/jctvc-hm/).

[25] P. K. Podder, M. Paul, and M. Murshed. "Fast mode decision in the HEVC video coding standard by exploiting region with dominated motion and saliency features," *PLOS One*, vol. 11, no. 3, March 2016.

[26] S. Ahn, B. Lee, and M. Kim, "A novel fast CU encoding scheme based on spatiotemporal encoding parameters for HEVC inter-coding" *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 422-435, March 2015.

[27] An Exploration of Safety Issues in EyeTracking" http://www.academia.edu/245642/An_Exploration_of_Safety_Issues_in_Eye_Tracking, retrieve date April, 2015.

[28] Tobii Eye Tracker Manual, Tobii Studio TM 2.2, September 2010.

[29] The basics of Power law. https://en.wikipedia.org/wiki/Power_law, date of exploration: December 2016.