QMET: A New Quality Assessment Metric for No-Reference Video Coding by Using Human Eye Traversal

Pallab Podder[*], Manoranjan Paul[*], and Manzur Murshed[€]
[*]School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW-2795, Australia
[€]School of Information Technology, Federation University,VIC-3842, Australia
{ppodder ; mpaul}@csu.edu.au, manzur.murshed@federation.edu.au

***Abstract*—The *subjective quality assessment* (SQA) is an ever demanding approach due to its in-depth interactivity to the human cognition. The addition of no-reference based scheme could equip the SQA techniques to tackle further challenges. Existing widely used objective metrics- *peak signal-to-noise ratio* (PSNR), *structural similarity index* (SSIM) or the subjective estimator- *mean opinion score* (MOS) requires original image for quality evaluation that limits their uses for the situation having no-reference. In this work, we present a no-reference based SQA technique that could be an impressive substitute to the reference-based approaches for quality evaluation. The *High Efficiency Video Coding* (HEVC) reference test model (HM15.0) is first exploited to generate five different qualities of the HEVC recommended eight class sequences. To assess different aspects of coded video quality, a group of ten participants are employed and their *eye-tracker* (ET) recorded data demonstrate closer correlation among gaze plots for relatively better quality video segments. Therefore, we innovatively calculate the amount of *approximation of smooth eye traversal* (ASET) by using *distance, angle,* and *pupil-size* feature from recorded gaze trajectory data and develop a new- *quality metric based on eye traversal* (QMET). Experimental results show that the quality evaluation carried out by QMET is highly correlated to the HM recommended coding quality. The performance of the QMET is also compared with the PSNR and SSIM metrics to justify the effectiveness of each other.**

Keywords*— *ASET, Eye Traversal, Eye-tracking, HEVC, QMET, Quality Assessment.

I. INTRODUCTION

The demand of *video quality assessment* (VQA) greatly increases due to its broad range of applications in the development and utilization of various video coding algorithms [1]-[3]. In general, the quality estimation is performed in two ways: objective and subjective where the earlier one is more widely used due to its simplicity, ease of use and having real-time applications. A good number of research works have been conducted based on the objective image quality estimation [4]-[6]. However, as human visual system is the ultimate assessor of video quality, the SQA has an eternal demand to the video coding research community. The no-reference SQA methods further deserve an upper-hand as they do not require any ground-truth reference for quality estimation and comparison. Therefore, compared to the full-reference (i.e. original videos as reference) or reduced-reference (i.e. existing of partial signals as reference), the *no-reference* (NR) based approach is more challenging due to the absence of original reference signal to analyze [7]. Moreover, traditional *full-reference* (FR) metrics such as *mean squared error* (MSE), SSIM or the PSNR are not always suitable representative to demonstrate strong correlation with human perceived actual quality. Thus, the authors in [8] carried out human cognition based objective quality assessment system using the eye-tracking technology and evolved more realistic ground truth visual attention data. Finally they exploited the ET aware normal scene saliency to improve their algorithm.

(a) eye-traversal for good quality video contents (b) eye-traversal for poor quality video contents (c) pupil-size vary for good and poor quality

Fig. 1. More concentrated eye-traversing approach is noticed for relatively better quality contents (e.g. image (a)). Using *BQMall* sequence, the opposite is observed in (b) for which the pupil-size sharply increases as shown in (c).

Unlike any objective video quality estimation technique, the subjective one may not be suitable for some real-time uses due to the engagement of human in the process. However, the subjective VQA studies could yield valuable data to evaluate the performance of *objective* methods towards aiming the ultimate goal of matching human perception [9]. 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. To increase the performance of objective model, Jia *et al.* [10] introduce a no-reference VQA model based on the ET data. Using blur and blockiness metric, they try to improve the subjective and objective correlation. Psychological metric based video quality judgment process is introduced by Arndt *et. al.* [11]. They use the pupil dilation feature from eye-tracking data and alpha feature from *electroencephalogram* (EEG) signal to study the video quality perception. Experimental results point that the proportion of alpha activity decreases and the pupil dilation increases once participants watch the degraded video quality. However, they test their scheme only for an arbitrarily selected and degraded portion of a frame that limits their scheme for its further uses. For assessing quality, the authors in [12] effort to bring in a voting process based algorithm by using the eye-tracking technology. It is based on the gaze point weighting process which experimentally acts inversely proportional to the user perceived quality. Using the scan path of eye movements, Tsai *et al.* [13] 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. Using human feedback, eye-tracking and saliency modeling, Podder *et al.* [14], analyze the human engagement behavior with video which may be a pre-requisite of any accurate SQA process. On the other side, the popularly used subjective testing

method- MOS [15][16] is often biased by the testing environment, viewers mode, viewers expertise and many other factors that could limit its uses.

To the best of our knowledge, no method in the existing literature has been introduced to develop a NR-SQA metric using eye-tracking that could be an impressive substitute to the reference based approaches. To address this lack, in this work, we innovatively develop a NR-SQA metric by exploiting the gaze trajectory data of human eye traversal. Fig. 1 illustrates the eye traversing approach of a participant with good and poor quality contents. The captured gaze plots of this viewer indicate more concentrated eye traversal and smooth browsing patterns for good quality contents as shown in Fig. 1 (a). Fig. 1 (b) reveals his scattered patterns of browsing for poor quality contents, while (c) indicates an automatic increase of pupil-size during watching the poor quality. As this trend is observed for the whole *n* number of frames in a sequence, we first study the spatio-temporal correlation among gaze points. Now if we determine *angle* and *distance* 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). Since we also discover that the quality variation has an impact on pupil size variation during watching video, three cardinal features *angle*, *distance* and *pupil-size* are calculated for each *potential gaze point* (PGP) from the gaze trajectory data 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). Each *fixation* and *saccade* in the proposed algorithm is called the *vision sensitive potential* (VSP) as we first eliminate all the *unclassified* data. Based on the calculated values of *angle*, *distance* and *pupil-size*, a VSP would turn into an ASET only if it satisfies the predefined *thresholding* (Th) criteria. The ASET is thus calculated for five different *quality segments* (QS) of the same sequence.

The main hypothesis of the proposed algorithm is that the better the quality of a video, the higher percentage of ASETs should be obtained as the viewers could better capture spatio-temporal information from video contents with smooth global browsing. Thus, the QMET score would be higher as well. For the poor contents, viewers perhaps start browsing in a hit and miss manner and continues random switching of eyes for improved quality that could eventually reduce the number of ASETs as well as the QMET score. The experimental reveal a good correlation among the scoring patterns of QMET, PSNR and SSIM. The proposed algorithm could be employed not only for the ET device produced data but also for the software based ET simulator [17]. As the QMET is developed for NR based situations, it could be an impressive alternative to the reference-required approaches (i.e. PSNR, SSIM or MOS) to tackle further challenges of quality estimation.

The rest of the paper is structured as follows. Section II describes the key steps of the proposed technique; Section III explicitly presents the experimental results and discussions; while Section IV concludes the paper.

II. Proposed Technique

The first phase of the proposed quality metric design is the coding quality variation and five different segments preparation which is executed by employing the HEVC [18] reference software HM15.0 [19]. These quality varied videos were then watched by a group of ten participants and their eye-tracking data were captured and analyzed for ASET calculation. Based on the calculated amount of ASETs, finally we develop the new quality metric- QMET to recognize how human perception and response are related to the video quality variation. The key steps are detailed in following sub-sections.

A. Design of the Experiment

All the participants were recruited from the Charles Sturt University by an open invitation disseminated through emails and notice board posters which included a detailed 'Participant Information Sheet' about the project [ethical approval no. 2015/124]. Recruited participants had normal or corrected-to-normal vision and did not suffer from any medical condition to influence our project adversely. A group of 10 people (7 males and 3 females) who were recruited fall within 20-45 age band and were undergraduate/postgraduate students, PhD students, and lecturers of the University. HEVC recommended eight class sequences which were used in this experiment are the representative in the scene ranging from low motion activity to the high motion activity, representing a wide range of contents, different aspects of motion and resolutions. The brief description about the sequences and their working conditions to carry out the experimental process are summarized in TABLE I (more detail about the videos to be found in [20][21]). In order to keep the test away from biasness with color or contrast, initially we design the experiment using their gray scale components only. We generate five different QS of each sequence which include- Excellent (using the *quantization parameter* QP=5), Good (QP=15), Fair (QP=25), Poor (QP=40) and Very-poor (QP=50) and display them to the participants in the same order mentioned here. Each segment of a sequence was 30 seconds long and the segment gap was 3 second.

TABLE I. Sequences Used in this Experiment

Sequence Name	Resolution (W×H)	Class Types	Quality Types	Segment Duration	Frame Rate
Traffic	2560×1600	A	5	30 sec	30 fps
Cactus	1920×1080	B	5	30 sec	30 fps
Tennis	1920×1080	B	5	30 sec	30 fps
Basketball	832×480	C	5	30 sec	30 fps
BQMall	832×480	C	5	30 sec	30 fps
B.Bubble	416×240	D	5	30 sec	30 fps
Flowervase	416×240	D	5	30 sec	30 fps
Fourpeople	1280×720	E	5	30 sec	30 fps

Calibration and a trial run was performed so that the participants feel comfort about the whole process. For the same reason, the lighting condition was also kept constant for the entire duration of video display. Upon their satisfaction, the Tobii eye tracker [22][23] (attached with the video display computer) was employed to record their eye movements and the completion of whole process was about 30 minutes long for each participant. Since the ET recorded data at 60HZ frequency and videos were run at 30fps, each frame could accommodate two gaze points. Thus, a single whole video contained 9000 gaze plots having 1800 for each segment.

B. ASET Calculation by Data Processing

After the completion of data capturing phase, we start analyzing only using the *fixations* and *saccades* (i.e. VSP) as discussed before. Then the *angle* (in degree) of the i^{th} VSP is

calculated by using the reference of $(i\text{-}1)^{th}$ and $(i\text{+}1)^{th}$ VSP, while the Euclidean *distance* (in pixel) of the i^{th} VSP is calculated with respect to $(i\text{+}1)^{th}$ VSP. For both cases, $i=\{1,2,\ldots,n\}$ and the values of *angle* and *distance* are not calculated for the 1st and n^{th} VSP plots. The *pupil-size* (in millimeter) on the other hand is calculated exactly for each gaze plot by averaging the values of left and right pupil size. Thus the corresponding *angle*, *distance* and *pupil-size* for the total *n* number of plots of a whole sequence are calculated using Matlab[9.0]. The features are then normalized based on the video contents to make them content invariant.

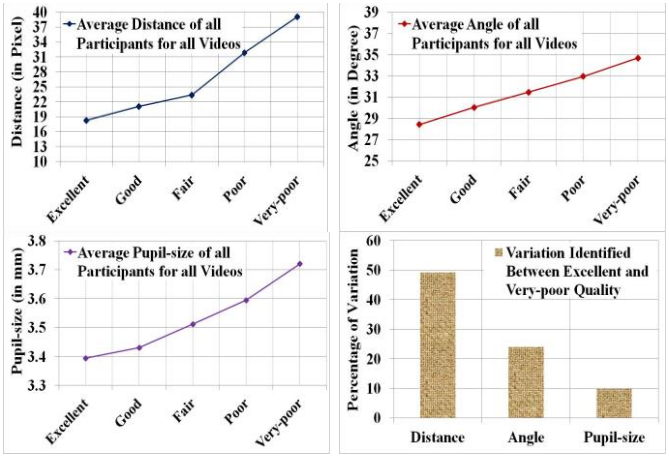


Fig. 2. Reaction principle of *angle*, *distance* and *pupil-size* features with video quality degradation explored in this experiment. It is noticed that all these features have a proportional correlation with quality degradation. The QP=5 to QP=50 sequentially present the Excellent to Very-poor quality segments.

Now let's concentrate to the Fig. 2 which clearly reveals how three features of ASET react with the coding quality variation. Once we calculate the average distance of all participants for all videos at different qualities, we notice a sharp increment of viewing *distance* with respect to the video quality degradation as shown in the top-left portion of Fig. 2. Similarly, we observe a linear increment of *angle* (top-right) and almost exponential increment of *pupil-size* (bottom-left) for video quality degradation. The bottom-right portion of the Figure summarizes how the video quality change could affect these features for calculating ASET. It also presents the percentage of variation for each feature value during experiencing the best to worst quality. The *distance* feature seems more reactive compared to the *angle* or *pupil-size*. However, the overall calculated results confirm that all these features have a proportional correlation with video quality degradation.

Fig. 3 illustrates the contribution of each individual feature and their combined role for calculating ASETs. The values are calculated for different QS by employing all the participants and all the videos. The Figure shows that *distance* feature (top-left) itself is not well representative as the percentage of ASET counts for Poor QS is higher than that in Fair QS. This statistics contradicts our hypothesis as the quality score using Poor contents could not be higher than Fair contents. Similar attributes could be observed both for the *angle* (top-right) and *pupil-size* (bottom-left) features once they are individually taken into account. However, their combined contributions could significantly segregate different aspects of coding quality as presented in the bottom-right of Fig. 3. Therefore, we exploit all these three features- *angle*, *distance*, *pupil-size* and dispose their obtained values in the following condition: if ((*angle*<=Th_1 && *distance*<=Th_2 && *pupil-size*<=Th_3, "1", i.e. ASET else "0") to determine whether a VSP would be considered as an ASET or not. Based on this calculated percentage of ASETs, the video quality is rated by the proposed QMET. The values of Th_1, Th_2 and Th_3 are directly determined by averaging the normalized values of *angle*, *distance* and *pupil-size* obtained for all participants. Other threshold selection strategy might work better, however, experimental results show that the proposed threshold selection approach provides good results in terms of ASET calculation as well as QMET evaluated quality rating.

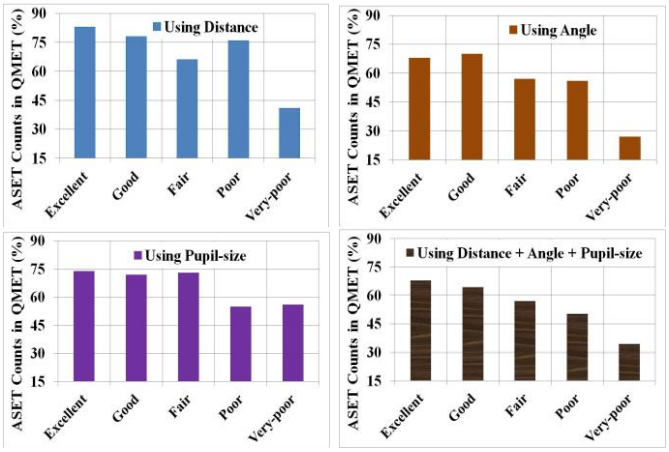


Fig. 3. Individual contribution of *angle*, *distance* and *pupil-size* feature and their combined role (bottom-right) in terms of ASET calculation. The QP=5 to QP=50 sequentially present the Excellent to Very-poor quality segments.

C. Design of Quality Metric

As stated earlier, if a VSP satisfies predefined thresholding criteria, it turns into an ASET, thus we count the percentage of ASETs for each segment (i.e. Excellent, Good, Fair, Poor, Very-poor) of a whole sequence. According to the proposed QMET, higher the amount of ASET is counted for a segment, better the quality is perceived. The quality rating is carried out by the following equation:

$$QMET = \frac{1}{1+\left(\frac{\Gamma-\Theta}{\Gamma}\right)} \qquad (1)$$

where Γ and Θ denote the total number of VSPs and ASETs respectively. The QMET evaluated quality rating ranges from 0.5 to 1 where the score 0.5 and 1 indicate the worst and the best quality respectively. According to the algorithm, higher the ratio obtained from the values of VSP and ASET, better the quality would be determined. In equation (1) for example, if the value of Γ and Θ is the same, the QMET score becomes 1 which is the highest. The opposite happens as the difference between Γ and Θ increases. The rationality of starting the score from 0.5 is due to keep consistency with the HEVC recommended coding, i.e. HEVC allows using QPs from 0 to 51 to indicate various forms of quality. Since even QP=50 could ensure the least standard quality of a video that should not have score '0'. The pseudo-code presented in Fig. 4 summarizes the entire procedure. In the algorithm, the execution $E_t[i]$!= *unclassified* confirms the exclusion of all unclassified data

(found 3% on average) that have null values for corresponding gaze plots.

```
Get EYE_TRACKER data Et_{i=1}^{N}
ASET=0;
TH1 = XX; %{Threshold values}
TH2 = YY; %{Threshold values}
TH3 = ZZ; %{Threshold values}
for i=1 to N do %{indicate all VSPs}
  if (Et[i]!=unclassified)
   then Vsp[i]=Et[i]; %{Vsp[n] is a vector of
"fixation" and "saccads" data}
   Vd[i]= distance(Vsp[i], Vsp[i+1]); %{distance()
calculates the Euclidean distance between the
current coordinate and its next coordinate}
   Va[i]= angle(Vsp[i-1], Vsp[i], Vsp[i+1]);
%{angle()calculates the angle of current
coordinate with respect to its previous and next
coordinates assuming the values of i={1,2,...,n}}
   Vp[i]= pupil_size(Vsp[i]); %{pupil_size()
calculates the average pupil size of left and
right eye for the current coordinate position}
   VN1[i]= Norm(Va[i]); % {Normalizes values}
   VN2[i]= Norm(Vd[i]); % {Normalizes values}
   VN3[i]= Norm(Vp[i]); % {Normalizes values}
    if(VN1[i]<=TH1 && VN2[i]<=TH2 && VN3[i]<=TH3)
        then ASET = ASET+1;
    end
  end
end
QMET = 1/(1+((N - ASET)/N))
  if (QMET ≈ 1)
   then QMET Score = High && VQ = good %{quality}
  end
 if (QMET ≈ 0.5)
   then QMET Score = Low && VQ = poor %{quality}
 end
```

Fig. 4. Pseudo-code for the development of QMET and its quality rating.

III. Experimental Results and Analysis

Since the quality rating of the proposed algorithm is highly correlated to the amount of ASET counts, first the average percentage of ASETs for different QS are calculated in two ways: (i) participant-wise ASET selection and (ii) video-wise ASET selection. In the former case, the percentage of ASET count is carried out by exploiting all the participants' eye-tracking data for all the sequences with different qualities. The data in TABLE II present such an example where its average values indicate the highest amount of ASETs (67.72%) to be selected for the Excellent QS. This value gradually downs for rest of the segments and reach 30.56% for Very-poor QS which is difference of 37.16% with the highest value.

TABLE II. Participant-wise ASET selection for Different Qualities.

Participant	ASET Counts (in %) for Different QS				
	Excellent	Good	Fair	Poor	Very-poor
P-1	71.80	65.59	60.57	44.80	42.23
P-2	69.41	62.30	58.71	32.91	27.39
P-3	71.22	63.31	61.32	41.26	40.31
P-4	69.94	53.60	46.12	36.76	28.83
P-5	60.39	56.84	53.45	34.48	25.71
P-6	72.58	48.88	48.82	33.61	26.06
P-7	62.81	49.97	45.58	36.82	33.11
P-8	68.55	58.84	54.86	29.56	21.30
P-9	59.06	51.13	49.45	31.88	27.48
P-10	71.50	64.35	62.14	49.45	33.26
Average	**67.72**	**57.48**	**54.10**	**37.15**	**30.56**

In the next case, the ASET count is carried out by exploiting the eye-tracker recorded data of all the videos for all participants as shown in TABLE III. In this case, the average selected percentage of ASET for the Excellent QS is higher by 33.40% compared to the Very-poor one. In both cases, the algorithm estimated percentage of ASET counts sharply decreases with respect to the quality degradation.

TABLE III. Sequence-wise ASET selection for Different Qualities.

Sequence	ASET Counts (in %) for Different QS.				
	Excellent	Good	Fair	Poor	Very-poor
Traffic	61.15	57.55	54.19	32.99	29.24
Cactus	50.68	43.59	44.69	39.59	24.11
Tennis	66.79	65.16	65.03	41.73	23.10
Basketball	70.74	71.46	51.08	45.49	36.32
BQMall	76.56	68.09	47.43	39.59	31.08
B.Bubble	59.79	61.68	54.60	51.29	30.52
Flowervase	62.76	59.47	49.69	41.67	39.50
Fourpeople	62.01	57.48	44.21	37.97	29.44
Average	**63.81**	**60.56**	**51.36**	**41.29**	**30.41**

From the calculated amount of ASET counts (in %), the QMET evaluated corresponding quality rating is performed which is shown in Fig. 5. Based on participant and video-wise average values, since the highest percentage of ASET is estimated for the segment with Excellent quality, therefore, the highest scores (i.e. 0.83 and 0.80) are also determined by the QMET for this segment. This is because the participants could probably better capture the information from the best quality contents with smooth global browsing. In contrast, for the case of its lowest scores (i.e. 0.52 and 0.50) at Very-poor QS, the participants perhaps watch the video with a trial and error basis; i.e. try to capture information from a portion but fails due do *bad* quality (noisy) and immediately move to the next but still erroneous. As the number of such hit and miss browsing sharply increases with time, both the ASET counts and the quality score decrease. This statistics confirm that any video having really *Poor ~ Bad* quality could never obtain higher percentage of ASET and thus becomes very unlikely to achieve higher quality score using QMET.

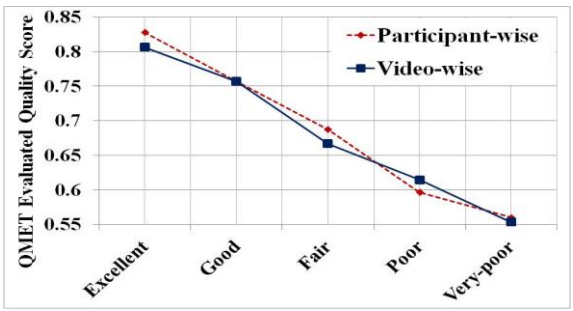


Fig. 5. Once the coding quality (coded by the HM) starts to degrade, the QMET evaluated quality score decreases as well.

Fig. 6 presents the proposed algorithm evaluated score for the sequences which performs best and worst at different QS (i.e. at different QPs). The Figure reveals that once we calculate their average obtained score for Excellent and Very-poor QS, the *BQMall* and *Tennis* performs the best and worst respectively. However, *Tennis* was the third highest scorer at Excellent QS that is reported in TABLE IV. This information also indicate that the video quality degradation could radically affect the participants' cognition at any time in scoring regardless of considering its contents or types. Now two questions arise: (i) For the *BQMall*, did people fix their eyes to the proximity of a particular location to obtain the highest score? (ii) why did *Tennis* score the lowest according to the

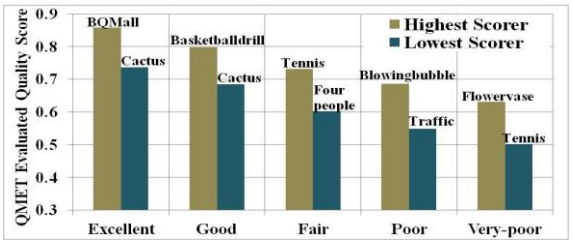
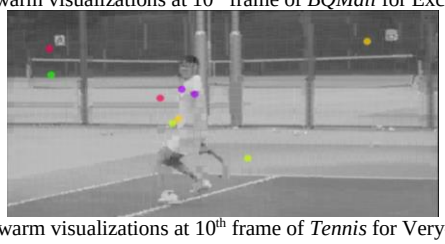


Fig. 6. Different sequences that score highest and lowest for different QS.

Fig. 6? To answer these questions, let us concentrate to the Fig. 7. For the *BQMall* sequence (in (a)), participants' gaze locations (colored dots of ET generated Bee swarm visualizations) confirm a global browsing over the frame and the supporting recorded gaze data of its entire duration also confirm more concentrated pattern of browsing. Therefore the amount of ASET counts (76.56% according to TABLE III) as well as the quality score (0.82 according to TABLE IV) reach the highest. In contrast, for the *Tennis* sequence (in (b)), participants also located eyes globally but overall in such a haphazard manner (being affected by unsuccessful attempts due to poor image quality) that could not meet the QMET thresholding criteria for most cases and also scored the lowest.

(a) Bee swarm visualizations at 10th frame of *BQMall* for Excellent QS



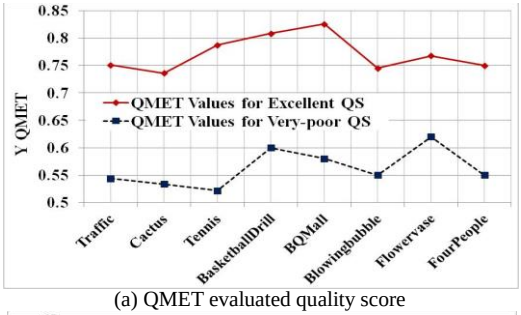
(b) Bee swarm visualizations at 10th frame of *Tennis* for Very-poor QS

Fig. 7. Bee swarm visualizations captured from ET that could determine the participants gaze locations in a frame.

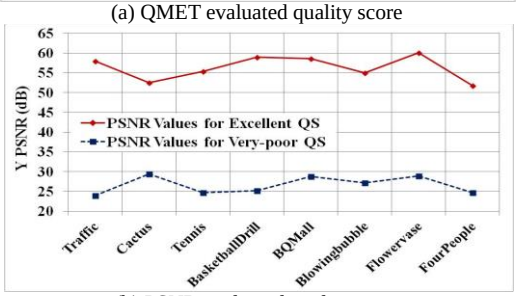
TABLE IV. THE QMET, PSNR AND SSIM EVALUATED QUALITY SCORE (USING ALL SEQUENCES) FOR EXCELLENT AND VERY-POOR QS.

Sequence	Excellent QS			Very-poor QS		
	QMET	PSNR (dB)	SSIM	QMET	PSNR (dB)	SSIM
Traffic	0.75	57.99	96.19	0.54	23.93	65.5
Cactus	0.73	52.45	94.85	0.53	29.38	62.09
Tennis	0.78	55.38	98.88	0.52	24.73	73.44
Basketball	0.80	58.95	96.71	0.60	25.13	66.08
BQMall	0.82	58.59	95.92	0.58	28.74	63.80
B.Bubble	0.74	54.98	92.91	0.55	27.14	66.82
Flowervase	0.76	60.04	96.88	0.62	28.92	73.57
Fourpeople	0.75	51.75	97.68	0.55	24.71	72.02

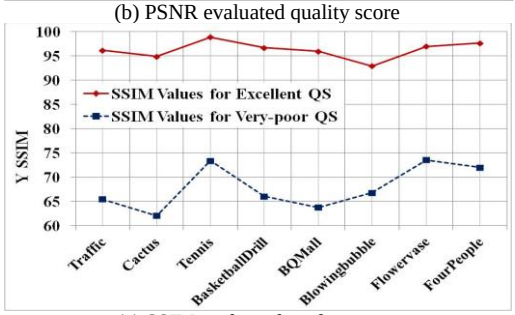
Now we evaluate the calculated score of the QMET with two other popularly used objective metrics PSNR and SSIM. The values of PSNR and SSIM are generated by employing HM15.0. From the whole range of segments, we just mention the results obtained from the Excellent and Very-poor one and present them in TABLE IV. For fair comparison among three metrics, these values are further graphically presented in Fig. 8 in which (a) (b) and (c) indicate the QMET, PSNR and SSIM evaluated plots respectively. The reason of demonstrating only two segments results is due to make a clear graphical difference between the best and worst quality assessed by the three techniques. The QMET evaluated best quality video-set (i.e. highest, second highest and third highest scorer) for the Excellent QS (i.e. using QP=5) include {*BQMall*, *Basketball*, *Tennis*} while for the same QS, PSNR and SSIM selected sets include the {*Flowervase*, *Basketball*, *BQMall*} and {*Tennis*, *Flowervase*, *FourPeople*} respectively. *Flowervase* sequence which is common in the highest scoring list of both PSNR and SSIM also obtains the second highest score in the QMET's evaluation criteria. On the other hand, for the Very-poor QS (i.e. using QP=5), the most visible dissimilarity could be found for the *Tennis* as it obtains the lowest score for both QMET and PSNR's assessment criteria. However, the SSIM scores it highest although its quality is not satisfactory as shown in Fig. 7 (b). This is perhaps the SSIM is a perception-based model that considers degradation in an image mainly by recognizing change in structural information. However, similarity among three metrics could be noticed for the *Traffic* sequence as it is assessed one of the lower scorers by the three metrics, while *Cactus* is assessed lower by QMET and SSIM.



(a) QMET evaluated quality score



(b) PSNR evaluated quality score

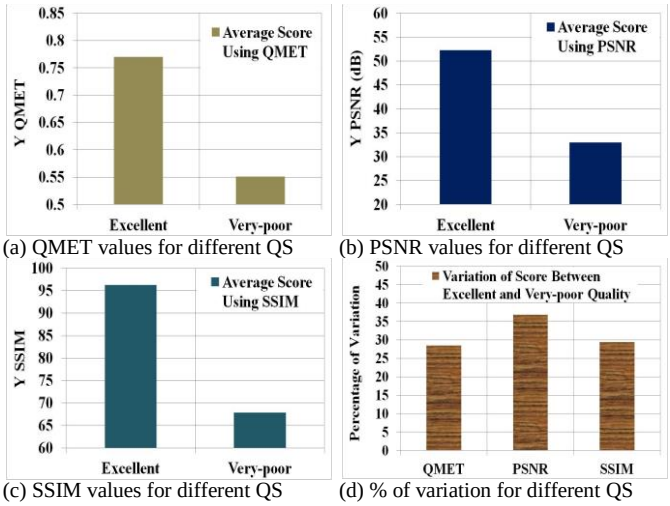


(c) SSIM evaluated quality score

Fig. 8. QMET, PSNR and SSIM score for Excellent and Very-poor QS.

Fig. 9 (a~c) presents the QMET, PSNR and SSIM evaluated average scores obtained for all videos using two QS mentioned in Fig. 8. This score difference for the QMET, PSNR and SSIM are 0.22, 19.28dB and 28.35 respectively. However, once we calculate the percentage of variation between the highest and the lowest score, the QMET could calculate 28.43% difference

while these values are 36.89% and 29.46% for the PSNR and SSIM respectively. This means the PSNR could best segregate the best and worst quality contents, while the QMET and the SSIM perform almost in a similar fashion. Although the scoring patterns of three metrics are roughly similar in terms of distinguishing the best and worst quality as shown in Fig. 9 (d), the proposed QMET could be employed as an impressive alternative to those of the objective metrics. This is because unlike PSNR or SSIM, the QMET does not require any ground-truth reference for quality estimation. Since the eye tracker data could be easily captured today by directly employing the software based eye-tracking simulator (i.e. ET device is not needed), the utility of the QMET could also be made more flexible using such simulator collected data sets. For its further performance improvement, work is undergoing to determine the least value of QMET that could differentiate two closer segments of video quality.



(a) QMET values for different QS (b) PSNR values for different QS

(c) SSIM values for different QS (d) % of variation for different QS

Fig. 9. In the Figure, (a~c) reveal the QMET, PSNR and SSIM induced average values for Excellent and Very-poor QS, while (d) indicates the three metrics estimated percentage of variations between the best and worst quality.

IV. Conclusions

In this work, we present a no-reference based subjective quality assessment technique that could be an impressive substitute to the reference-required approaches for quality estimation and comparison. The amount of approximation of smooth eye traversal (ASET) is innovatively calculated from recorded gaze data by employing the *angle*, *distance* and *pupil-size* features. Eeventually we develop human eye traversal based new quality metric- QMET and compare its performance with the popularly used PSNR and SSIM metrics. Experimental results demonstrate a good similarity among these three metrics in terms of distinguishing the best and worst quality video contents. The proposed algorithm could be applied both on the eye-tracker device and software based eye-tracking simulator recorded data. Thus, the QMET could be a suitable alternative for the reference based metrics to tackle further challenges of video quality evaluation.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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, 2010.

[8] 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.

[9] 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, 2010.

[10] 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.

[11] 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.

[12] 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, 2014.

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

[14] P. K. Podder, M. Paul, T. Debnath, and M. Murshed, "An Analysis of Human Engagement Behaviour Using Descriptors from Human Feedback, Eye Tracking, and Saliency Modelling" *IEEE International Conference on Digital Image Computing: Techniques and Application*, pp. 1-8, November 2015.

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

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

[21] 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.

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

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