IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 21, 2019, accepted March 3, 2019, date of publication March 14, 2019, date of current version April 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2904732

EMAN: The Human Visual Feature Based No-Reference Subjective Quality Metric

PALLAB KANTI PODDER[1], (Member, IEEE), MANORANJAN PAUL[2], (Senior Member, IEEE), AND MANZUR MURSHED[3], (Senior Member, IEEE)

[1]Department of Information & Communication Engineering, Pabna University of Science & Technology, Pabna 6600, Bangladesh
[2]School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW 2795, Australia
[3]School of Science, Engineering, and Information Technology, Federation University, Churchill, VIC 3842, Australia

Corresponding author: Manoranjan Paul (mapul@csu.edu.au)

This work was supported in part by the Australian Research Council through the Discovery Projects under Grant DP130103670.

ABSTRACT As the human vision is a definitive assessor of video quality, the expanded interest for no-reference subjective quality assessment (SQA) is focusing on a definitive goal of coordinating with human observation. However, the widely used subjective estimator-mean opinion score (MOS) is often biased by the testing environment, viewers mode, expertise, domain knowledge, and other factors which may influence on actual assessment. In this paper, a no-reference SQA metric is devised by simply exploiting the nature of human eye browsing on videos and analyzing the associated quality correlation features. The high efficiency video coding (HEVC) reference test model is first employed to produce different forms of coded video quality which then displayed to a number of partakers. Their eye-tracker recorded spatiotemporal gaze-data indicate more concentrated eye-traversing approach for relatively better quality. Thus, we calculate the quality assessment related to assorted features such as length pursuit, angle deflection, pupil deviation, and gaze interlude from recorded gaze trajectory. The content and resolution invariant operations are carried out prior to synthesizing them using an adaptive weighted function to develop a new quality metric-eye maneuver (EMAN). Tested results reveal that the quality evaluation carried out by the EMAN is comparatively better than MOS and structural similarity (SSIM) in terms of assessing different aspects of coded video quality for a wide range of single view video contents. For the free viewpoint video (FVV), where the reference frame is not available, the EMAN could also better distinguish different qualities compared to the MOS and SSIM.

INDEX TERMS EMAN, eye-maneuver, gaze trajectory, HEVC, quality evaluation.

I. INTRODUCTION

Video quality estimation is a conspicuous research zone because of its wide scope of utilizations in yielding different improved video compression techniques [1]. Among two basic approaches of quality evaluation, the objective one is simpler and more widely used because of its simplicity and ready-to-use implementations [2]–[4]. The subjective estimator, on the other hand, engages human in the process. The assessment process is further regarded as full-reference [5], [6] (uses the original signal), reduced-reference [7], [8] (uses partial signal) and no-reference [9], [10] schemes. The no-reference method is also called the blind approach which requires no information from the reference signal and more challenging to analyze [11]. Moreover, the existing *full-reference* (FR) approaches such as *mean squared error* (MSE), *peak signal-to-noise ratio* (PSNR) or the *structural similarity index* (SSIM) are applicable merely to the reference oriented situations. This is also partially applicable for the reduced reference-based metrics. These drawbacks mandates the implementation of no-reference based metrics for quality estimation [12], [13].

Using the natural scene statistics, Fang *et al.* [14] propose a blind metric for faded image quality by examining its irregularity attributes and defending the level of deviation from the statistical models. Their experimental results suggest that the natural scene statics models are promising in handling contrast-distorted images except the limitation of capturing some of its specific aspects. The video quality forecast model by utilizing *discrete cosine transform* (DCT)

The associate editor coordinating the review of this manuscript and approving it for publication was Chaker Larabi.

is reported by Zhu *et al*. [15] to discuss about the condensed video recordings. However, this metric is distortion specific and data driven, thus, may incur with lower performance on unknown data and limits its performance on compressed videos. To overcome the evaluation limitation of existing deblocking algorithms for deblocking images, the authors in [16] present a no-reference quality metric for deblocked JPEG images. They first build a deblocked image database with subjective MOS as a ground truth reference and then simultaneously evaluate the blocking and blurring artifacts in smooth and textured regions respectively. This metric could be applied to automatic parameter selection in image deblocking algorithms. By exploiting the human visual model, Li *et al*. [17] introduce a blind image quality estimation metric by exploiting the luminance and structural information. Gu *et al*. [18] incorporate the local and global features of an image for quality estimation using blind metric. This algorithm could amend the performance of general purpose blind quality estimation metrics to a sizable margin. These statistical metrics could signify the image structure analysis but may not be reasonable for some high-quality extent. This because quality observation here is mostly correlated to the visual aspect instead of statistical measurement of the picture [19]. Yet, several features of visual perception are mostly skipped in the literature.

Liu *et al*. [20] carry out the human insight based quality evaluation framework by utilizing eye-tracking technology and develop increasingly sensible ground truth visual saliency model to improve their implementation. The eye-tracker has turned into a non-intrusive, moderate and simple to-utilize device in human attitude analysis as it exactly screens where, when, and what individuals gaze at. The visual perception can also be estimated by employing the software based gaze estimation tool [21].

Unlike objective estimation, the subjective studies could yield valuable data to evaluate the performance of objective methods towards aiming the ultimate goal of matching human perception [22]. To this end, researches in [23] aim at a no-reference objective evaluation metric by utilizing eye-tracker based contrast distortion information. The authors in [24] introduce a model to judge the video quality based on psychological merits such as electroencephalogram signaling, and pupil dilation. Their process was applicable for selectively degraded portion of a video frame which deprived it for further use. Research methodology in [25] use the eye-information based elective process to create a no-reference quality assessment outline. They experimentally prove that for inferior quality contents, partakers consume more time for quality evaluation which is opposite to the approach introduced by Tsai *et al*. [26]. We also observe similar trend in [25] mostly for the still images which however, becomes impractical for videos as the frames move continually. This causes the continuous changes of object positions not to see the same scene even a few frames later. The widely used subjective metric MOS [27], [28] in contrast, is often biased

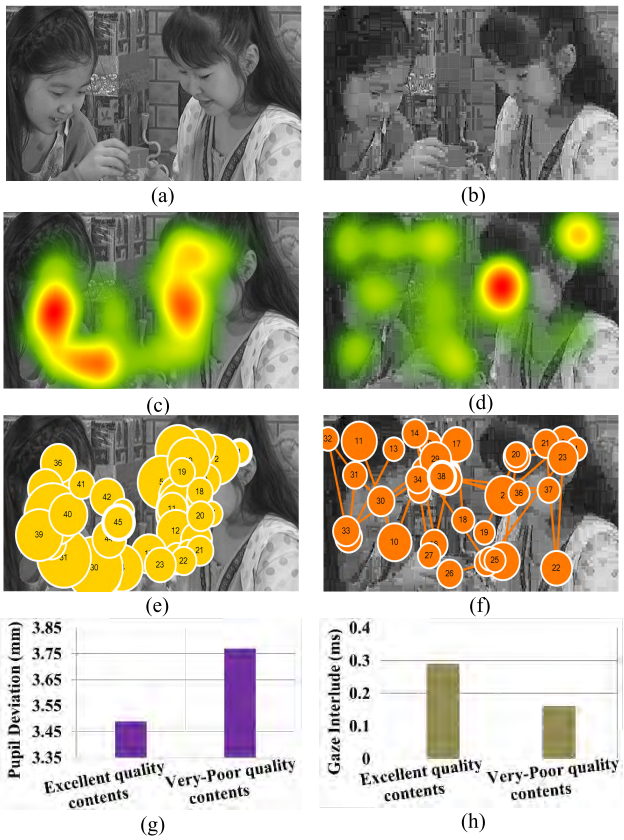


FIGURE 1. Compact eye-browsing is seen for Excellent quality (for example, images in (c) and (e)). The inverse is seen in (d) and (f) for Very-poor quality content that impacted on automatic Pupil deviation increment and Gaze interlude decrement in (g) and (h) respectively.

by many factors and does not show a point-to-point relation between quality and human perception.

To overcome the limitations of MOS, we introduce a *no-reference* subjective testing method using eye-maneuver. Now, let us have a look at FIGURE 1 (a) $\sim$ (d) where the heat map is reproduced for good and poor quality. It is obvious that the heat map is more condensed for better quality as depicted in FIGURE 1 (c) which is also observed for all frames. For better outline, this time, the viewer's eye movement pattern i.e. compact or scattered as shown in FIGURE 1 (e $\sim$ f) is explored. It is then inferred that for good quality, participants investigate video data with smooth eye perusing. Focusing on the viewers' gaze-point distribution, thus, we calculate the spatial features of quality correlation such as *Pupil deviation* (P), and the temporal features of quality correlation such as *Length pursuit* (L), *Angle deflection* (A), and *Gaze interlude* (T) for each *potential gaze point* (PGP) as reported in FIGURE 1. (e $\sim$ f) and (g $\sim$ h). The PGP is a gaze-point that belongs to the *fixation* (i.e. visual concentration on a fixed position) and *saccade* (i.e. abrupt eye movement among various segments of fixation). An operation for making the features (i.e. L, A, P, and T) content and resolution invariant is carried out so

that the metric becomes applicable in different displayable devices with low motion to high motion video contents. Finally, these features are synthesized in the spatiotemporal domain by using a weighted function to construct a new evaluator- the *eye maneuver* (EMAN).

The experimental analysis reveal a good correlation of EMAN with the most existing reference required PSNR, SSIM and the subjective metric MOS in quality test analysis. The result of these metrics are then assessed using two evaluators- *Pearson Linear Correlation Coefficient* (PLCC) and *Spearman Rank-Order Correlation Coefficient* (SRCC) [14]. Tested results show that the proposed EMAN could better distinguish different qualities compared to the MOS and SSIM for a wide range of single view video (where viewpoint is fixed for users) and free view video contents.

The remainder of the paper is organized as follows: Section-II illustrates the correlation analysis on features; Section-III presents the key steps of the proposed EMAN development technique; Section-IV broadly represents the tested results to justify the performance of EMAN both on the *single view video* (SVV) and the FVV, while Section-V concludes the paper.

II. MOTIVATIONAL ANALYSIS

The proposed EMAN metric uses four different features extracted from eye tracker data in order to assess the reconstructed videos encoded with video coding techniques in different qualities. Among four features used in this experiment, the *Length pursuit* (L-unit in pixel) of i^{th} gaze plot is calculated with respect to the $(i+1)^{th}$ plot by means of *Euclidean distance*, while the *Angle deflection*(A in degree) of the i^{th} gaze plot is determined by using the reference of its $(i-1)^{th}$ and $(i+1)^{th}$ plots (here $i = \{1, 2, \ldots, n\}$. The *Pupil deviation* (P-unit in millimeter) is calculated for each i^{th} plot from the average diameter of two pupils. Finally, the *Gaze interlude* (T-in millisecond) is the eye engagement information for each i^{th} plot which are determined by using MATLAB R2012a.

By employing the *High Efficiency Video Coding* (HEVC) [29]–[31] reference Test Model (HM) [32], we reproduce five dissimilar quality segments of a video such as-*Excellent* (using quantization parameter (QP) = 5)), *Good* (using QP = 15), *Fair* (using QP = 25), *Poor* (using QP = 40) and *Very-poor* (using QP = 50) and identify the response of each feature against coding quality variation as presented in FIGURE 2. In summary, the calculated outcomes demonstrate that L, A, P features have a relative, while T has a converse corresponding relationship with video quality deterioration.

Now, using the normalized data (the data normalization process for the features will be discussed in Section III) of the sequences, we determine quality versus features correlation by applying the Q-score (initially justified score of EMAN). The features L, A, P, and T are determined by Q_1, Q_2, Q_3, and Q_4 in equation (1) - (4) respectively and the outcome of

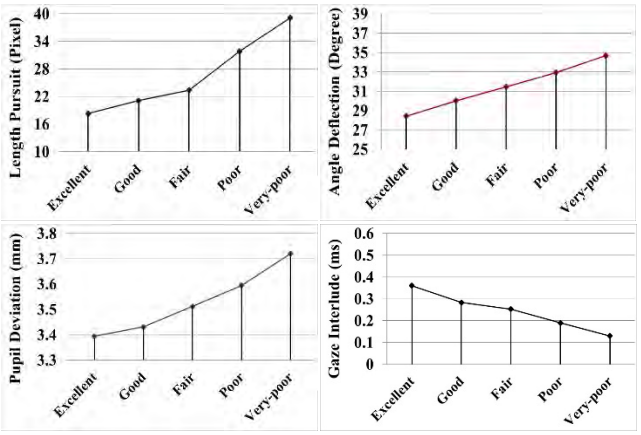


FIGURE 2. Relationship between the features (i.e. Length pursuit, Angle deflection, Pupil deviation, and Gaze interlude) and coded video quality.

which is illustrated in FIGURE 3.

$$Q_1 = L^{\partial L} \tag{1}$$
$$Q_2 = A^{\varphi A} \tag{2}$$
$$Q_3 = (P/2)^{\Im P} \tag{3}$$
$$Q_4 = \sqrt{2T}^{(\aleph/\sqrt{2T})} \tag{4}$$

The symbols ∂, φ, $\Im$, and $\aleph$ are the weighs of L, A, P, and T respectively. The equations are established based on power-law [33]. As the minor value change is noticed due to the change of quality, the features may be best applied by its power notation for an obvious score separation. Unlike power multiplication for the features L, A, and P, the feature T is inversely correlated with Q-score where the power division is applied and shown in FIGURE 3 (d). The reason for applying Q-score is to pre-determine L, A, P, and T features' response to EMAN with respect to quality variation within 0 to 1 range.

The parameter L captures the perceptive distance (i.e. how far), A calculates angular deflection (i.e. how much), P measures pupil-size variation (i.e. how big/small), and T asserts the fixation duration (i.e. how long) between two gaze points. To capture viewers' fixation/saccadic eye movements, these four features played a vital role in perceptive quality analysis. Since L, A, and P works in the spatial domain, these will be merged with T in temporal domain. Thus, we combine them by developing an adaptive weighted function equated as: $Q = L^{\partial L} \times A^{\varphi A} \times (P/2)^{\Im P} \times \sqrt{2T}^{(\aleph/\sqrt{2T})}$. FIGURE 4(a) is the output of this synthesized function which can clearly segregate different qualities. Now, the weight for ∂, φ, $\Im$, and $\aleph$ in the above equation (1)-(4) is fixed with 0.5. This illustrated in FIGURE 4(b) by calculating slopes at each segment and averaging them using various weights. Since the calculated slope using weight 0.5 is sharper than others, it is fixed for the entire experiment. Different weight combination among features may work better; yet produced results reveal a good correlation of EMAN with other existing metrics.

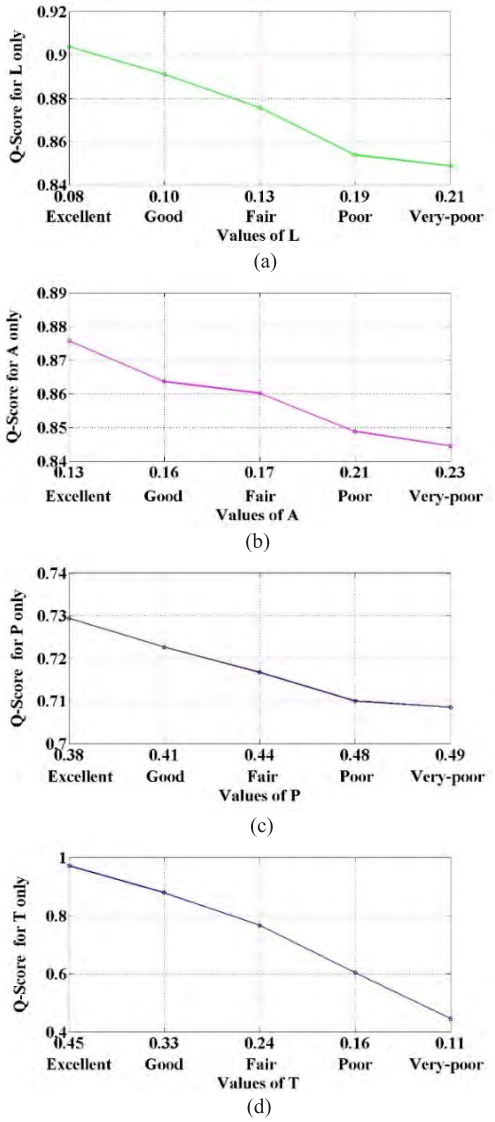


FIGURE 3. The effect of Length pursuit, Angle deflection, Pupil deviation, and Gaze interlude features in isolating diverse quality portions. (a) Q-score for Length pursuit (*L*) by means of equation (1). (b) Q-score for Angle deflection (*A*) by means of equation (2). (c) Q-score for Pupil deviation (*P*) by means of equation (3). (d) Q-score for Gaze interlude (T) by means of equation (4).

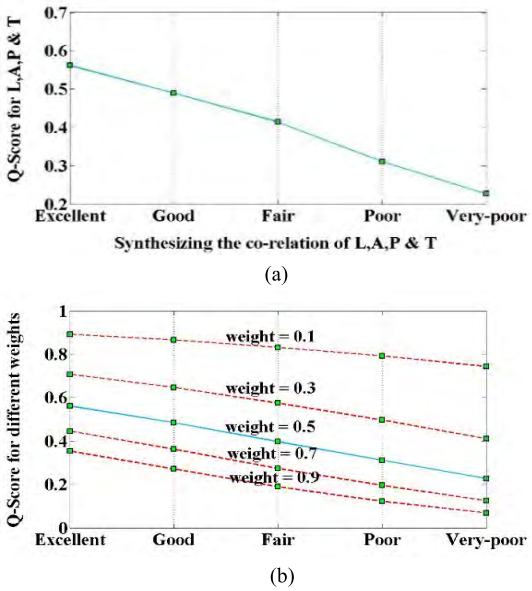


FIGURE 4. Quality segment distinction and relative weight selection by exploiting Q-score. (a) Features *L*, *A*, *P*, and *T* have been combined using a weighted function to distinct different coded quality segments. (b) Weight 0.5 outperforms all other weight combinations and reproduce the graph in (a).

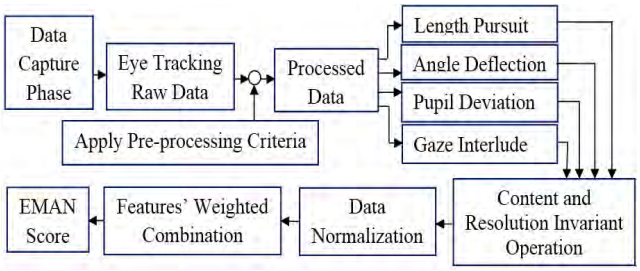


FIGURE 5. Process diagram of the proposed technique.

III. PROPOSED TECHNIQUE

The varied video qualities were generated by HM15.0 and displayed on the screen to capture partakers' eye-tracking raw data. These data were processed mainly by excluding all unclassified parts (that incur with frequent eye blinking of the partakers and found 3% on average for each video) and nullifying the corresponding gaze plots. The content and resolution invariant operation along with data normalization were carried out on four features. The spatiotemporal features were then combined to introduce a new metric EMAN. The whole procedure is exhibited as an outline in FIGURE 5 and the key stages are explained in the following associated sections.

A. DATA CAPTURE AND PRE-PROCESSING

Each partaker joined from the Charles Sturt University (CSU), Australia by an open welcome which was dispersed through messages and notice board notices. The invitation incorporated an itemized 'Member-detail Sheet' about the participation in the project [ethical endorsement no. 2015/124]. The partakers who attended the data capturing session did not suffer from any sort of physical disfunctionalities that could undesirably impact on assessment. A group consisting of fifteen people (including men and women) who were appointed belong to 19-46 age range and were undergraduate, postgraduate students, Ph.D students, and academics of CSU. Experiments were conducted using the *joint collaborative team on video coding* (JCT-VC) recommended eight class sequences including the Class-A, Class-B, Class-C Class-D, and Class-E. The used videos had the versatility of contents, motions and resolutions. The brief detail about videos and test conditions are summarized in TABLE 1 (more specific description about the videos could be found in [34], [35]).

TABLE 1. Sequences for single view video quality evaluation.

Video	Video-Size (W×H)	Class	Quality Types	Segment Duration	Frame Rate
Traffic	2560×1600	A	5	60 seconds	30 fps
Cactus	1920×1080	B			
Tennis	1920×1080	B			
Basketball	832×480	C			
BQMall	832×480	C			
B.Bubble	416×240	D			
Flowervase	416×240	D			
Fourpeople	1280×720	E			

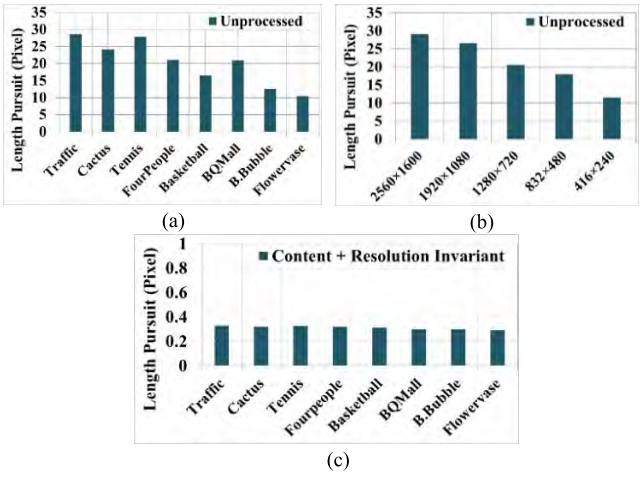


FIGURE 6. Unprocessed Length pursuit (in (a) and (b)) is made content and resolution invariant along with data normalization when processed in (c).

To keep the experiment out of biasness with color, contrast or brightness, merely their gray-scale modules were used. There were five sub-divisions of quality as stated earlier and each of them was displayed a minute with a pause of 3 seconds in between. The necessary calibration using sixteen-point trial run was done for the partakers comfort in the process. The lighting impact was constant for the whole period of time. Upon their agreement, the Tobii eye tracker [36], [37]which was attached to the video display computer came into eye traversal recording operation of the partakers. To finish the whole recording, around 60 minutes were required for each participant. Since eye tracker was operated at 60HZ frequency and the videos were run at 30fps, each frame captured two gaze plots. Hence, a video could capture 18000 gaze plots having 3600 for each segment.

B. INVARIANT OPERATION ON LENGTH PURSUIT

Prior to carrying out the invariant operation on the feature *Length pursuit-L*, let us see its unprocessed values in FIGURE 6(a) and (b). The calculated content based variation using their highest and lowest values go to 61.96% as depicted in (a), while the resolution based variation is 52.57%. Thus, the invariant operation is undertaken on both content and resolution. The first one is comprised of the following phases: (i), we compute the L from PGPs as discussed in Section II; (ii), figure out the average of *potential coordinate* PGP (x, y) which is the pre-determined center C(x, y); (iii), with respect to C(x, y), the*Euclidean distance* of all PGPs is calculated to sort values of L in chronological order. This ordering scheme could prioritize *foveal* fixation and unlike parafoveal, or perifoveal formation, the *foveal* eye fixation is more object centric [38]. (iv), $\mathcal{U}$ number of organized values are taken for calculating radius of motion led region. The selection of $\mathcal{U}$ is justified in the experimental results analysis section; (v), now, the calculated radius is used as a divisor of *length pursuit* for each *potential plot*.

As instance, next to the content invariant operation, the obtained average values of L for the (2560 × 1600), (1280 × 720), and (416 × 240) resolution videos were 0.37, 0.23, and 0.14-pixels respectively where the variation between (2560 × 1600), and (416 × 240) resolutions is 62.16%. Both the video content and display size are responsible for such stunning variation. Once we apply the multipliers reported in TABLE 2 for (2560 × 1600), (1280 × 720), and (416 × 240) resolution based sequences, the post-processed values we obtain are 0.24, 0.23, and 0.21 respectively. The calculated variation between (2560 × 1600) and (416 × 240) resolutions now downs to 12.5%. As the recorded eye-tracking data illustrates a consistent correlation between video resolution and quality, the employed multipliers fitted well in invariant operation. The normalized value now also ranges within 0 to 1. The third and fourth column of TABLE 2 show the earlier and current variation respectively. The current variation notably downs to 10.67% and its implication is revealed in FIGURE 6 (c) which is eventually taken for EMAN scoring.

TABLE 2. A major percentage of variation could be reduced on Length pursuit by performing content plus resolution invariant operation.

Sequence Size	Video size variation neutralization multipliers	Earlier Variation		Current Variation
		Content based (%)	Video-size 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			

C. INVARIANT OPERATION ON ANGLE DEFLECTION

FIGURE 7 (a) and (b) show the content and resolution based unprocessed maximum *Angle deflection-A* values which are 37.49% and 6.25% respectively.

Since the feature A does not have a promising impact on the variation of resolution, we merely implement the content invariant operation on it (similar to the L) by the following steps: (i) we determine the *Angle deflection* of the potential gaze points as stated earlier in Section II; (ii) sort the estimated values of A from the lowest to the highest order;

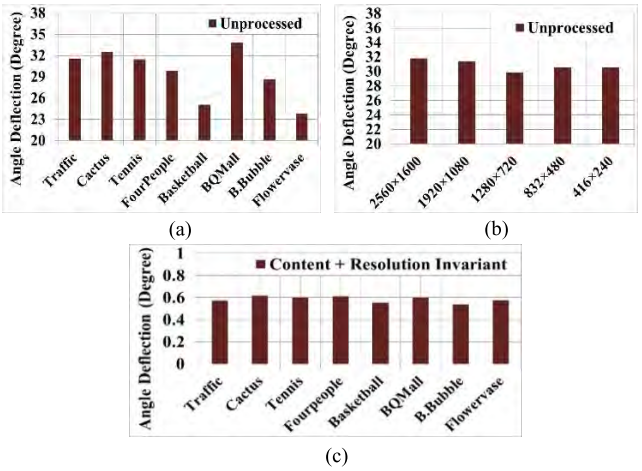


FIGURE 7. The unprocessed Angle deflection presented in (a) is made content invariant since they are already resolution invariant as presented in (b). The outcome in (c) is obtained after performing content plus resolution invariant normalization (ranges from 0 to 1) operation.

(iii), select the standard average values from the range of $\mathcal{U}$ ordered values; (iv) the calculated average is finally used as a divisor of the potential gaze plots determined in the step (i). The normalized value also ranges within 0 to 1. The whole invariant impact on A is revealed in FIGURE 7 (c) with a minor variation of 6.31%.

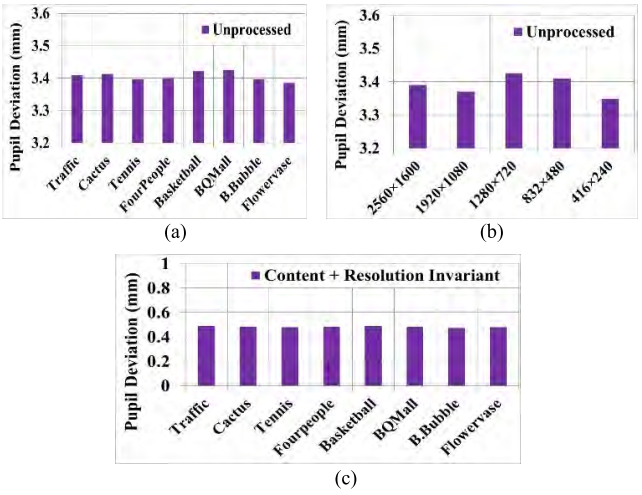


FIGURE 8. The content and resolution based unprocessed Pupil-sizes are shown in (a) and (b) respectively. The outcome in (c) is obtained by only performing the normalization operation (ranges from 0 to 1).

D. NORMALIZATION OF PUPIL DEVIATION

Compared to the features L or A, we detect relatively steady characteristics of the *Pupil deviation* P both on content and resolution as presented in FIGURE 8 (a) and (b) with a tiny variation of 1.37% and 1.68% (less that 10% variation is avoided) respectively. Hence, we skip invariant operations on *Pupil deviation*. However, to normalize the feature P, all *pupil deviation*values are divided by the maximum *Pupil-size* of human (which is taken 6.1 mm from [39]). The invariant

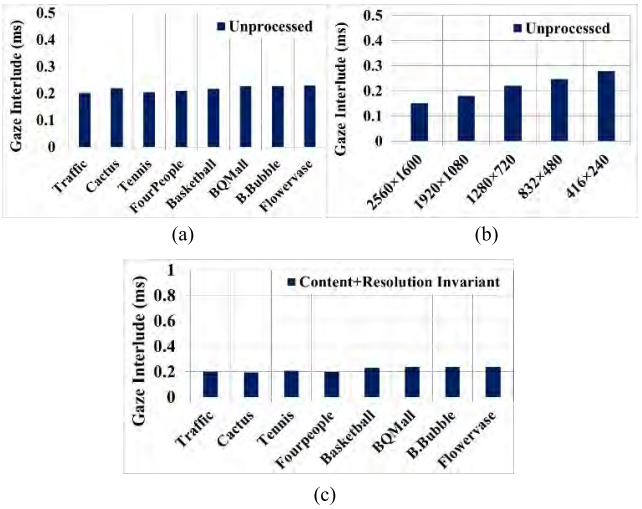


FIGURE 9. Only the resolution based unprocessed Gaze interlude in (b) is made resolution invariant since they are already content invariant as presented in (a). The outcome in (c) is obtained after performing content plus resolution invariant normalization (ranges from 0 to 1) operation.

and normalized P values are displayed in FIGURE 8 (c) that belong to the range 0 to 1 as well.

E. INVARIANT OPERATION ON GAZE INTERLUDE

The calculated *Gaze interlude* presented in FIGURE 9 (a) is already content invariant as the maximum variation is 9.31% among them. However, since the variation among resolutions goes up to 46.78% (in (b)), we make them resolution invariant by exploiting some multipliers which is similar to the *Length pursuit* feature operation which is presented in first and second columns of TABLE 3. The third and fourth column show the earlier and current variation respectively. The variation now downs to 10.23% by operating the multipliers for different resolution types and its implication is shown in FIGURE 9(c).

TABLE 3. A major percentage of variation could be reduced on Gaze interlude by performing the resolution invariant operation.

Sequence Size	Video size variation neutralization multipliers	Earlier Variation		Current Variation
		Content based (%)	Video-size Based (%)	After content plus resolution invariant operation (%)
2560×1600	1.40	9.31	46.78	10.23
1920×1080	1.20			
1280×720	1.00			
832×480	0.90			
416×240	0.80			

F. THE IMPLEMENTATION OF EMAN

In the proposed algorithm, the EMAN score is calculated for every *potential gaze point* as the feature values of L, A, P, and T exist for all these points. Thus, for a single quality segment (i.e. for a QP), more than 3450 EMAN values

are obtained due to having the same number of potential gaze plots (i.e. 3600 plots minus 3% unclassified data). The average EMAN score is then counted for parallel comparison to other metrics. Note that the PSNR or SSIM similarly counts their corresponding value for all frames at a QP (said earlier) and produce their average as a result. From previous statistical analysis, it can be inferred that an automated higher EMAN score is gained if the features L, A, and P are much lower than T. The equation for EMAN is finally derived as:

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

where the associated weighted values of L, A, P and T denoted by ∂, φ, $\bar{\delta}$, and $\aleph$ respectively and the weight is set at 0.5 for the whole experiment (described earlier). An exceptional case is also taken in consideration. If $L = A = 0 \in 30$ continuous frames (as 30 is the frame rate), then the value of $L = 0.1$ and $A = 0.1$ are forced to set out as a penalizing operation. The features L and A are only manipulated as P and T are still $\neq 0$. Note that during the whole experiment, we did not experience such unusual situation. If the EMAN evaluated quality score is close to 1, the video content quality is rated best, while, the opposite happens for the score close to 0.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

To verify the proposed metho's effectiveness, experiments are conducted on the JCT-VC recommended eight class sequences and the FVV sequences. The quality variation of the sequences is carried out by employing the HM15.0. As presented earlier the reproduced *Excellen*t, *Good*,*Fair*, *Poor* and*Very-poor* quality segment has been constructed using QP = 5, 15, 25, 40, and 50 respectively. The performance of the EMAN is then compared with the popularly used objective metrics PSNR, SSIM and subjective estimator MOS. The goal is to exploit the EMAN as an alternative to the MOS by actively incorporating the human cognition.

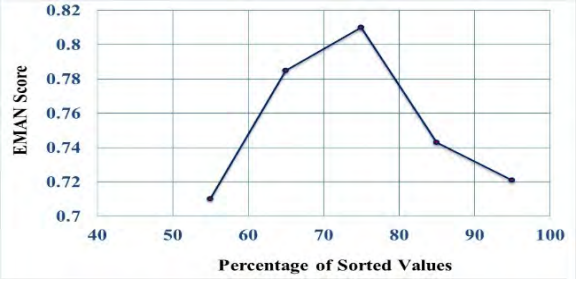


FIGURE 10. Illustration of EMAN scoring by employing the average of $\mathcal{U}$ sorted values of *L* and *A* features. Since using first 75% could yield the highest EMAN score for all sequences, this percentage is used for the whole experiment.

A. EMAN EVALUATED QUALITY SCORE

As discussed in Section-III, we use the average of $\mathcal{U}$ ordered values for determining the fobial radius of moving region which is justified in FIGURE 10. When first 60% values are sorted, a good number of feature values still remain unemployed to boost up EMAN score. In contrast, if much higher percentage is imposed (e.g. 85%), plenty of superfluous values incur with the scoring process that further results in poorer rating of the metric. Since using first 75% could yield the highest EMAN score by further rectifying the visual sensitive plots almost for all sequences, we apply it to calculate the final EMAN score.

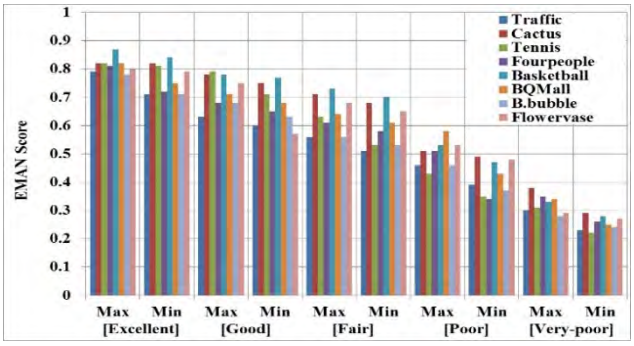


FIGURE 11. The video-wise highest (Max) to the lowest (Min) EMAN score is calculated for all segments using eight videos. The highest score is obtained for the Excellent quality segment and vice-versa. There is also a clear score difference from one segment to another.

TABLE 4. The distinction of Different Segments of Quality Using EMAN Evaluated Quality Score.

Different Segments of Quality	Average of Max Values	Average of Min Values	Average of Max and Min
Excellent (QP=5)	0.80	0.76	0.78
Good (QP=15)	0.72	0.68	0.70
Fair (QP=25)	0.64	0.59	0.61
Poor (QP=40)	0.49	0.42	0.45
Very-poor (QP=50)	0.33	0.28	0.30

The highest and lowest EMAN score for all the segments and all the test videos are presented in FIGURE 11. The score for the *Excellent* fragment is the most noteworthy which gradually decreases with respect to quality degradation and reaches at the least for *Very-poor* portion. Using Max and Min the calculated average of each segment is reported in TABLE 4. In either case, there is a clear decline in scores due to content quality degradation occurring at higher QPs. In the rightmost column, the final average is calculated and the variation between the best and worst quality is recognized by 61.53%.

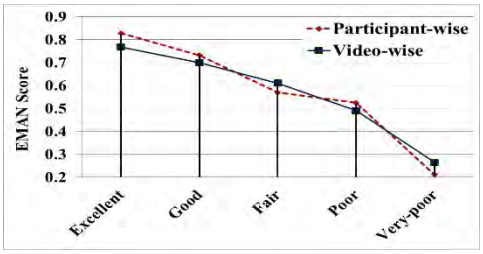


FIGURE 12. The introduced metric EMAN has an inverse proportional correlation with quality degradation.

The resultant score indicates the EMAN's quality segregation proficiency which is further discussed according to participant and video-wise average score in FIGURE 12.

For coded *Excellent* segment, the obtained participant and video-wise EMAN scores are 0.81 and 0.78 respectively which are the highest scores in both cases according to FIGURE 12. In fact, for its best quality, the partakers could better detect and capture information with smooth eye browsing. Conversely, for its *Very-poor* quality segment with lowest scores (i.e. 0.21 and 0.26), partakers get the video by means of hit and miss browsing. The hit and miss means are due to the successful and unsuccessful attempts for pleasant and unpleasant quality respectively. For *Very-poor* contents, the number of hit-and-miss browsing notably rises with time. This leads relatively to the scattered means of browsing. Hence, for the worst quality content, it is highly unlikely to yield upper score by EMAN.

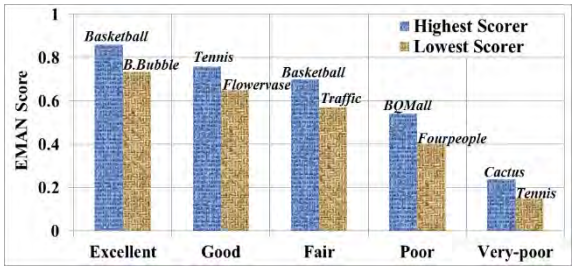


FIGURE 13. For various segments of coded video quality, the Basketball is the highest EMAN scorer, while the Tennis is the lowest one.

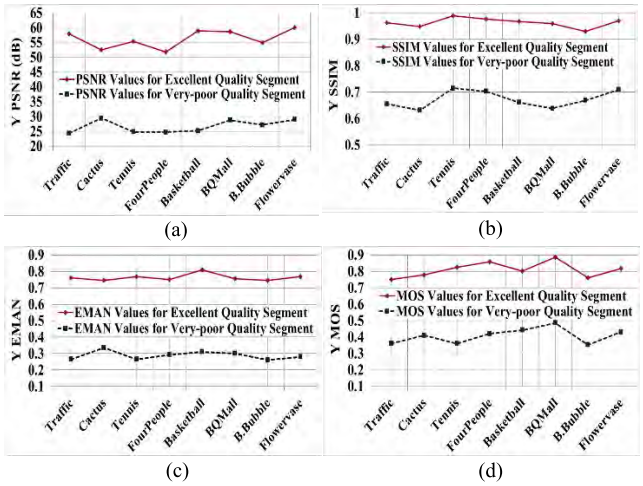
FIGURE 13 illustrates the EMAN estimated quality score for the videos which scores the highest and lowest at various segments. The Figure reveals that for the *Excellent* segment of quality, the *Basketball* sequence performs the best and obtains the score 0.83. The *Tennis* on the other hand, performs the worst for the *Very-poor* quality and scores 0.17. Exactly the inverse happens for the *Excellent* quality with a score of 0.77. The quality corruption therefore, not only influences the partakers' perception but also automatically resists them from better scoring.

Now two cases are observed for the proposed assessment process: (i) for *Basketball*, whether the partakers injected eyes to the vicinity of a specific position to bias the *Length pursuit* and *Angle deflection* feature and obtained the highest score in FIGURE 13; (ii) the reason for which *Tennis* scores the lowest according to the same figure. To justify the answer these questions, let us concentrate on the *Bee-swarm visualization* (i.e. the gaze fixation of all partakers in the form of simultaneous plots that attract attention in a frame) for the 10th frame of both sequences presented in FIGURE 14.

The group-fixations for the 10th image of *Basketball* video at *Excellent* quality is shown in FIGURE 14 (a) where partakers' eye fixation data looks more consistent with object movement. Conversely, the group fixations for the 10th image of *Tennis* at *Very-poor* quality is shown in FIGURE 14 (b) where partakers also located eyes all over the image. However, the entire recorded gaze data pattern appears more scattered (due to hit-and-miss eye movement) and could not satisfy

(a) (b)

FIGURE 14. Eye tracker recorded group visualization data could better inform about the partakers gaze point in an image (taken a single frame from video). (a) Group fixations for 10th image of Basketball video at Excellent quality. (b) Group fixations for 10th image of Tennis at Very-poor quality.



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

FIGURE 15. Calculated four metric values for the Excellent and Very-poor quality. (a) Calculated PSNR for all videos. (b) Calculated SSIM for all videos. (c) Calculated EMAN for all videos. (d) Calculated MOS for all videos.

high scoring criteria and eventually obtained relatively poorer quality score.

B. COMPARISON OF QUALITY EVALUATION

FIGURE 15 illustrates the calculated four metric values for *Excellent* and *Very-poor* quality segment. It is obvious that for the *Excellent* quality, the PSNR and SSIM selected best quality video set is mostly correlated to the EMAN picked video set such as [*Basketball*, *Flowervase*, *Tennis*]. These three metrics reveal the similar characteristics for the *Very-poor* quality as well. However, the MOS does not demonstrate the equivalent video selection pattern in either case. Except *Traffic* video which is evaluated as one the lowest scorers in all metrics, the performance of EMAN is very much similar to PSNR and SSIM, rather than MOS over all videos.

FIGURE 16 (a∼d) illustrate four metrics assessed average score for the *Excellent* and *Very-poor* quality by exploiting all the videos. The calculated percentage of variation between the best and worst quality obtained by PSNR, SSIM, EMAN, and MOS are 55.08, 32.98, 57.95, and 49.53 respectively. This data have been pictorially demonstrated further in FIGURE 16(e) where the EMAN performs better

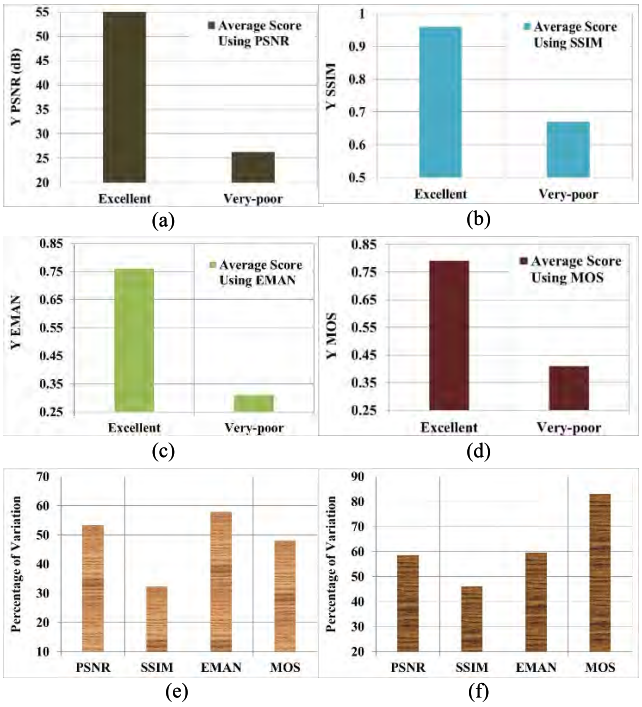


FIGURE 16. The Excellent and Very-poor quality segment separation capability of four metrics where the EMAN performs better than others according to (e). The MOS in contrast could detect the highest variation by experiencing a very wide range of scores (sometimes inconsistent compared to other metrics) from the partakers as shown in (f).

than others. The calculated highest variation for four metrics are also reported in FIGURE 16(f). The MOS tends to produce higher score in FIGURE 16(f) which is mostly due to providing an arbitrary scoring opportunity to the partakers during assessing quality.

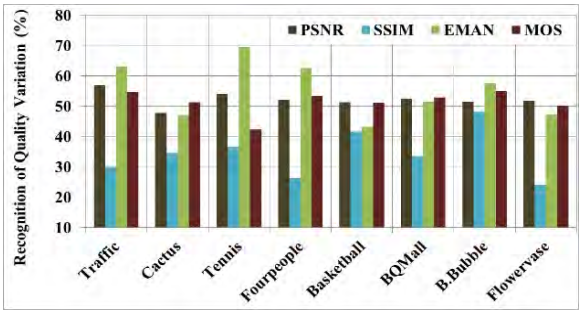


FIGURE 17. Four metric estimated level of variation detection using all class videos.

Now we reproduce the video-wise performance for four metrics with respect to maximum quality distinction capacity and display it in FIGURE 17. For *Tennis* video, the most noticeable difference is captured by the EMAN due to perceiving the best quality of object movement with smooth *Length pursuit*, *Angle deflection*, minimum *Pupil deviation* and maximum *Gaze interlude* to get better EMAN score. Surprisingly, the worst quality of the same video attributed inversely proportional to the characteristics of four features to get inferior EMAN score, thereby creating such stunning difference. The similar statistics is also obtained for *Traffic*, *Fourpeople*, and *Blowingbubbles* video. It is also obvious that unlike SSIM, the PSNR, EMAN, and MOS similarly work over all videos reported in FIGURE 17.

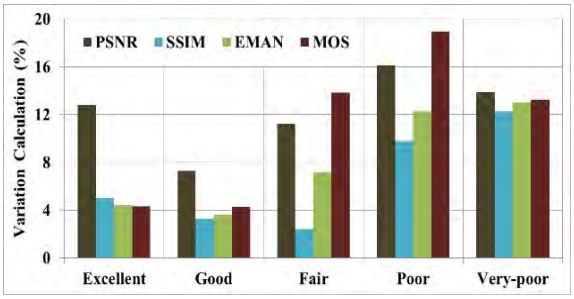


FIGURE 18. Four metrics obtained quality score for each segment where lower percentage of variation indicate better metric performance.

Now, two noticeable points: (i) if distinctive video substance are coded utilizing the equivalent QP (for example 5 for *Excellent*), the created score ought not to have much variations. In any case, the PSNR could not pursue this pattern and goes at the apex as uncovered in FIGURE 18. Although the EMAN performs better than the PSNR, and MOS almost in all cases, the SSIM appears most stable in this regard. Observation (ii): even though the same video is coded with versatile quality, the employed metric should produce various scores accordingly. The EMAN reveals similar result to PSNR, however, it performs much better than the SSIM and MOS almost in all cases according to FIGURE 19.

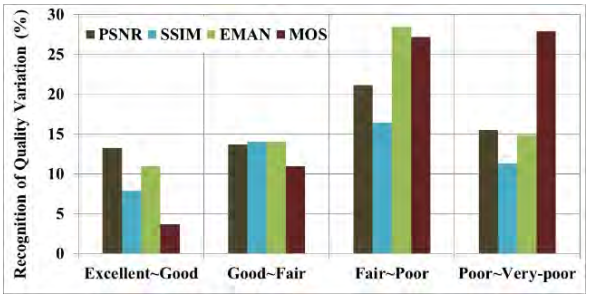


FIGURE 19. Four metrics obtained quality score for each segment where higher percentage of variation indicate better metric performance.

The PLCC and SRCC are two widely used time utilized execution estimators which have been employed for further performance justification of EMAN, PSNR, SSIM, and MOS. The efficiency of a metric is justified by depending on its obtained higher values of PLCC and SRCC [14]. Hence, for all videos, the PLCC of each metric is first calculated and their average value reported in TABLE 5 indicate that the EMAN outperforms all the metrics except PSNR. This mostly due to the use of different coded video qualities since the sensitivity of PSNR from *Excellent* to *Very-poor* quality is very high. This is also applied in quality restoration of lossy compression.

TABLE 5. The PLCC evaluated performance of four metrics over eight class sequences.

Sequences	PLCC			
	PSNR	SSIM	EMAN	MOS
Traffic	0.72	0.63	0.58	0.57
Cactus	0.57	0.58	0.6	0.61
Tennis	0.73	0.66	0.67	0.62
Fourpeople	0.58	0.68	0.64	0.67
Basketball	0.78	0.63	0.72	0.64
BQMall	0.77	0.57	0.68	0.63
B.Bubble	0.68	0.51	0.59	0.62
Flowervase	0.74	0.69	0.67	0.66
Average	0.70	0.62	0.65	0.63

TABLE 6. Position of four metrics in terms of PLCC and SRCC's assessment principle.

Used Metrics	PLCC	SRCC	Average of PLCC and SRCC
PSNR	0.70	0.72	0.71
SSIM	0.62	0.59	0.60
EMAN	0.65	0.61	0.63
MOS	0.63	0.61	0.62

Similar to the PLCC, the SRCC of each metric is calculated and their average value is summarized in TABLE 6. From the data of the table it can be inferred that in both cases, the EMAN is relatively a better estimator compared to SSIM and MOS although the PSNR after all, is the highest scorer. The average of the PLCC and SRCC for each metric is also reported in TABLE 6 which also infer the similar result. However, it should be noted that unlike PSNR, no reference is available once the EMAN based assessment is carried out. Since the scoring example of these measurements are closely comparable regarding recognizing distinctive quality substance as outlined in FIGURE 17, FIGURE 18, FIGURE 19, TABLE 5 and TABLE 6, the newly introduced EMAN could be spoken as another individual from of metric in this family. Therefore, it can be utilized effectively as a decent option in contrast to the emotional estimator- MOS.

C. FREE VIEWPOINT VIDEO (FVV) QUALITY ASSESSMENT USING EMAN

The FVV generation is comprised of image warping and then hole filling procedure such as inverse mapping or post filtering [40]. The synthesized view is generated at a simulated position between left and right views where there is no reference outline for quality estimation of such video [41]. The *depth image based rendering* (DIBR) is a practical way to reduce storage and transmission bandwidth for synthesized videos from texture and its corresponding depth map. However, in the DIBR technique, portions of regions are not visible in the virtual position due to the front objects termed as occlusion, which create some holes in the view synthesis. Therefore, unlike *single view video* (SVV), such crack like artifacts in the coded FVVs are more visible to the users and become more obvious at lower bit-rates (such as

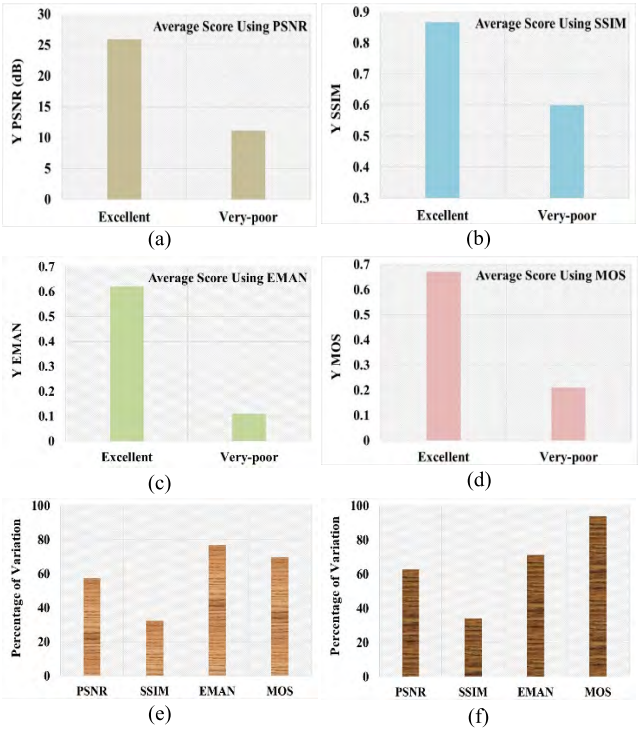


FIGURE 20. The Excellent and Very-poor quality segment separation capability of four metrics in a FFV where EMAN performs better than others according to (e). The MOS in contrast detects the highest variation by experiencing a very wide range of scores (sometimes inconsistent compared to other metrics) from the partakers as shown in (f).

using QP = 50) by reproducing the worst quality. Due to this correlation with quality variation, it is anticipated that EMAN should provide higher PLCC and SRCC values than MOS for FVV.

The FVV does not have any available reference frame, and the EMAN also does not require any reference frame for quality evaluation. Thus, the EMAN could also be helpful eventually to compare different scoring patterns of reference-based metrics such as PSNR or SSIM. Hence, we further employ the EMAN on FVV coding performance analysis. The goal is to evaluate the view synthesis algorithms using EMAN. To assess the performance of four metrics on FVV, the *Excellent* and *Very-poor* segments are considered in FIGURE 20(a $\sim$ d) as we have done earlier in FIGURE 16. The score difference between these two segments obtained by EMAN is 78.51 which is much higher than rest of the metrics taken as reported in FIGURE 20(e). The MOS in contrast could detect the highest variation by experiencing a very wide range of scores (which appears inconsistent compared to other metrics) from the partakers as shown in (f). For some synthesized video such as *Newspaper*, the partakers rank them very low as in the depth image based rendering, numerous holes made the quality too poor. As a result, some of them started scoring from 0.05 (out of 1.0) which lead to such notable variations.

The calculated result for FVV in FIGURE 20 has been compared with the result presented in FIGURE 16 for SVV and further demonstrated in FIGURE 21 for better judgment.

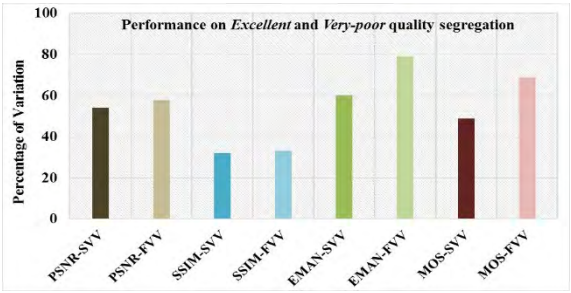


FIGURE 21. The Excellent and Very-poor quality distinguishing capability of the PSNR, SSIM, EMAN, and MOS both on SVV and FVV. Higher percentage of variation indicates better quality differentiating capability of a metric.

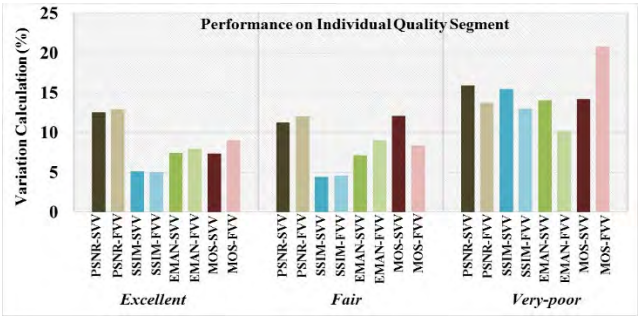


FIGURE 22. The Excellent and Very-poor quality distinguishing capability of the PSNR, SSIM, EMAN, and MOS both on SVV and FVV. The higher percentage of variation indicates better quality differentiating capability of a metric.

In the case of FVV, the quality distinguishing capability of all the metrics has been relatively improved compared to the SVV. The objective estimators PSNR-FVV and SSIM-FVV could not stabilize any ground truth reference upon which they depend. The subjective estimators such as MOS-FVV or EMAN-FVV for the same reason, could do better as it does not need any reference frame for scoring. Similar to the FIGURE 18, the score of four metrics for each segment is calculated using the FVV and also compared with SVV results in FIGURE 22. The performance of the metrics in FIGURE 22 is assessed based on the proposition that the score variations should not be prominent if different videos are coded with the same quality. Both for the *Excellent* and *Fair* segment, the variation calculated by the PSNR-FVV and EMAN-FVV is relatively higher than those of the PSNR-SVV and EMAN-SVV respectively and perform almost in a similar fashion. Unlike the *Excellent* quality segment, the MOS-FVV remains more stable compared to the MOS-SVV for *Fair* quality segment but the SSIM-FVV remains most stable in this aspect. In contrast, for *Very-poor* quality segment, the performance improvement of three metrics is more obvious in FVV in terms of lowering the variations except for the MOS. However, according to the overall calculated result of FIGURE 22, the EMAN-FVV performs relatively better compared to PSNR-FVV as well as MOS-FVV in most cases although the SSIM-FVV outperforms all other metrics by limiting its variations.

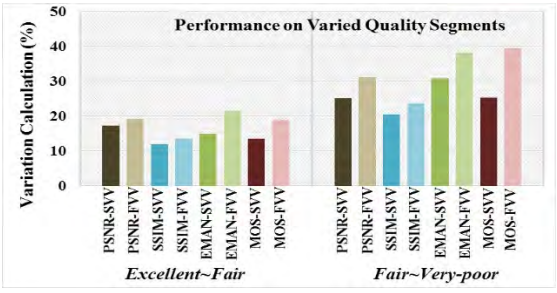


FIGURE 23. The Excellent and Very-poor quality distinguishing capability of the PSNR, SSIM, EMAN, and MOS both on SVV and FVV. The higher percentage of variation indicates better quality differentiating capability of a metric.

To verify proposition-2, i.e. if the same video is coded with various qualities, for each of the variations, the quality score should be different. This has been justified using (i) *Excellent* ∼ *Fair* and (ii) *Fair* ∼ *Very-poor* and reported in Fig. 23. For case-(i), all the metrics perform better than SVV approach where EMAN-FVV and PSNR-FVV are much better than others. The SSIM-FVV is the least responsive metric in this regard. For the second range of segment, the FVV based performance of four metrics appear much better than the SVV based outcomes, however, this time the EMAN-FVV and the MOS-FVV perform their best in quality separation. Over all segments of quality, the performance of the subjective metrics is superior to the objective ones.

TABLE 7. Average performance of four metrics according to the assessment requirement of PLCC and SRCC.

Performance Estimators	PSNR-SVV	PSNR-FVV	SSIM-SVV	SSIM-FVV	EMAN-SVV	EMAN-FVV	MOS-SVV	MOS-FVV
PLCC	0.69	0.68	0.61	0.63	0.64	0.69	0.62	0.68
SRCC	0.71	0.70	0.58	0.62	0.60	0.71	0.60	0.68

For additional analysis, the calculated results for SVV and FVV are reported in TABLE 7 by implementing both the PLCC and SRCC's assessment requirement. Although PSNR-SVV is the best metric according to TABLE 7, the score obtained from the FVV indicate a very closer correlation among the PSNR, EMAN, and the MOS while, the EMAN performs relatively better than the MOS and SSIM according to both performance estimators- PLCC and SRCC. Therefore, in addition to the objective metrics, the outcomes of the free view synthesized videos could also be assessed by organizing the subjective estimators and incorporating their scores for more realistic evaluation.

V. CONCLUSION

The widely used subjective estimator- *mean opinion score* (MOS) is often biased by the testing environment, viewers mode, expertise, domain knowledge, and many other factors which may influence on actual assessment. In this work, we, therefore, introduce a no-reference subjective metric by statistically analyzing four quality correlation features form Human vision and synthesize them using an adaptive cost

function to develop a new metric- *eye maneuver* (EMAN). Tested results reveal that the quality evaluation carried out by the EMAN performs relatively better than the MOS and *structural similarity index* (SSIM) in terms of assessing different aspects of coded video quality for a wide range of single view video contents. For the *free viewpoint video* (FVV) where the reference frame is not available, the EMAN could also better distinguish different qualities compared to the MOS and SSIM. Since the human visual stimuli could also be captured by directly employing the software based eye tracking simulators (where device itself is no longer needed), the utility of EMAN could be more flexible over an extended areas of applications.

REFERENCES

[1] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

[2] J. You, T. Ebrahimi, and A. Perkis, "Attention driven foveated video quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 200–213, Jan. 2014.

[3] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 520–531, Sep. 2015.

[4] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu, "The application of visual saliency models in objective image quality assessment: A statistical evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1266–1278, Jun. 2016.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[6] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[7] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.

[8] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[9] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu, "No-reference noticeable blockiness estimation in images," *Signal Process., Image Commun.*, vol. 23, no. 6, pp. 417–432, 2008.

[10] R. Hassen, Z. Wang, and M. Salama, "No-reference image sharpness assessment based on local phase coherence measurement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 2434–2437.

[11] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529–539, Apr. 2010.

[12] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[13] Y. Xue, B. Erkin, and Y. Wang, "A novel no-reference video quality metric for evaluating temporal jerkiness due to frame freezing," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 134–139, Jan. 2015.

[14] 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 Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.

[15] K. Zhu, C. Li, V. Asari, and D. Saupe, "No-reference video quality assessment based on artifact measurement and statistical analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 4, pp. 533–546, Apr. 2015.

[16] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, and B. Chen, "No-reference quality assessment of deblocked images," *Neurocomputing*, vol. 177, pp. 572–584, Feb. 2016.

[17] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, Dec. 2016.

[18] 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 Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017. doi: 10.1109/TCYB.2016.2575544.

[19] S. Tourancheau, F. Autrusseau, Z. M. P. Sazzad, and Y. Horita, "Impact of Subjective Dataset on the performance of image quality metrics," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 365–368.

[20] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, Jul. 2011.

[21] M. Böhme, M. Dorr, M. Graw, T. Martinetz, and E. Barth, "A software framework for simulating eye trackers," in *Proc. Symp. Eye Tracking Res. Appl.*, 2008, pp. 251–258.

[22] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[23] L. Jia, X. Zhong, and Y. Tu, "No reference video quality assessment model based on eye tracking datas," in *Proc. 2nd Int. Conf. Inf., Electron. Comput.*, 2014, pp. 97–100.

[24] S. Arndt, J. Radun, J.-N. Antons, and S. Möller, "Using eye-tracking and correlates of brain activity to predict quality scores," in *Proc. 6th Int. Workshop Qual. Multimedia Exper.*, Sep. 2014, pp. 281–285.

[25] M. G. Albanesi and R. Amadeo, "A new algorithm for objective video quality assessment on eye tracking data," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, Jan. 2014, pp. 462–469.

[26] C.-M. Tsai, S.-S. Guan, and W.-C. Tasi, *Eye Movements on Assessing Perceptual Image Quality*. Springer, 2016, pp. 378–388.

[27] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3097–3100.

[28] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives," *Multimedia Syst.*, vol. 22, no. 2, pp. 213–227, Mar. 2016.

[29] P. K. Podder, M. Paul, and M. Murshed, "QMET: A new quality assessment metric for no-reference video coding by using Human eye traversal," in *Proc. Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Nov. 2016, pp. 1–6.

[30] *High Efficiency Video Coding*, document ITU-T Rec. H.265 and ISO/IEC 23008-2 (HEVC), ITU-T and ISO/IEC, Apr. 2013.

[31] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[32] CVS. *Joint Collaborative Team on Video Coding (JCT-VC), HM Software Manual*. Accessed: Dec. 12, 2017. [Online]. Available: http://hevc.kw.bbc.co.uk/svn/jctvc-hm/

[33] (Dec. 2016). *The Basics of Power Law*. [Online]. Available: https://en.wikipedia.org/wiki/Power_law

[34] 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, Mar. 2016, Art. no. e0150673.

[35] S. Ahn, B. Lee, and M. Kim, "A novel fast CU encoding scheme based on spatiotemporal encoding parameters for HEVC inter coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 422–435, Mar. 2015.

[36] (Apr. 2015). *An Exploration of Safety Issues in EyeTracking*. [Online]. Available: http://www.academia.edu/245642/An_Exploration_of_Safety_Issues_in_Eye_Tracking

[37] *Tobii Eye Tracker Manual, Tobii Studio 2.2*, Sep. 2010. Accessed: Dec. 12, 2017. [Online]. Available: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf

[38] M. M. Salehin and M. Paul, "Human visual field based saliency prediction method using Eye Tracker data for video summarization," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Jul. 2016, pp. 1–6.

[39] W. J. Donnelly and A. Roorda, "Optimal pupil size in the human eye for axial resolution," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 20, no. 11, pp. 2010–2015, 2003.

[40] D. M. M. Rahaman and M. Paul, "Adaptive weighting between warped and learned foregrounds for view synthesize," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Jul. 2017, pp. 49–54.

[41] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, "Objective image quality assessment of 3D synthesized views," *Signal Process., Image Commun.*, vol. 30, pp. 78–88, Jan. 2015.

PALLAB KANTI PODDER (S'14–M'18) received the B.Sc. (Hons.) and M.Sc. degrees from the Department of Information and Communication Engineering (ICE), in 2008 and 2010, respectively, and the Ph.D. degree from the CSU Machine Learning Research Unit, Charles Sturt University, NSW, Australia, in 2017. After his M.Sc. degree, he joined as a Lecturer with the Computer Science and Engineering Department, Bangladesh University, Dhaka, Bangladesh. Then, he joined as a Lecturer and promoted to Assistant Professor with the Department of Information and Communication Engineering, Pabna University of Science and Technology (PUST), Pabna, which is one of the renowned universities in Bangladesh. He is currently an Associate Professor, Advisor IEEE (PUST branch), and the Director of the Software Engineering Laboratory in the same department. He has research collaboration with the CSU Machine Learning Research Unit, Charles Sturt University, Australia. He has published more than 30 journal articles and international conference proceedings in the areas of communication, image processing, video compression, and video quality assessment. He is a member of the Bangladesh Computer Council (BCC) and Australian Computer Society (ACS).

MANORANJAN PAUL (M'03–SM'13) received the Ph.D. degree from Monash University, in 2005. He was a Research Fellow with the University of New South Wales, Monash University, and Nanyang Technological University. He is currently an Associate Professor with the School of Computing and Mathematics, Charles Sturt University. He has published more than 160 refereed papers in international journals and conferences. He regularly published journal articles in the IEEE TRANSACTIONS. His major research interests include image/video coding, EEG signal analysis, and computer vision. He was a Keynote Speaker in the IEEE Digital Image Computing, Techniques, and Applications (DICTA-17), WoWMoM-14 Workshop, DICTA-13, and ICCIT-10. He is a Senior Member of ACS (Australian Computer Society). He received $15M competitive grant including the Australian Research Council (ARC) Discovery Project and Australian Government's CRC Project grant. He is the ICT Researcher of the Year 2017 by the Australian Computer Society (ACS). He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the *EURASIP Journal on Advances in Signal Processing*. He has also served as a Guest Editor for the *Journal of Multimedia* and the *Journal of Computers* for five special issues and the Program Chair of PSIVT 2017 and Publicity Chair-DICTA 2016.

MANZUR MURSHED received the B.Sc. Engg. (Hons.) degree in computer science and engineering from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1994, and the Ph.D. degree in computer science from Australian National University (ANU), Canberra, Australia, in 1999. He also completed his Postgraduate Certificate in Graduate Teaching from ANU, in 1997. He is a Robert HT Smith Professor with Federation University, Australia, and the Personal Chair (formerly Monash University, Gippsland Campus) and was one of the founding directors of the Centre for Multimedia Computing, Communications, and Artificial Intelligence Research (MCCAIR). He has so far published more than 200 refereed research papers and received more than $1 million nationally competitive research funding, including three Australian Research Council Discovery Projects grants, in 2006, 2010, and 2013 on video coding and communications, and a large industry grant, in 2011 on secured video conferencing. To date, he has successfully supervised 19 Ph.D. students. His major research interests include video technology, information theory, wireless communications, distributed computing, and security and privacy. He received a University Gold Medal from BUET, in 1994, the inaugural Early Career Research Excellence Award from the Faculty of Information Technology, Monash University, in 2006, and the Vice-Chancellor's Knowledge Transfer Award (commendation) from the University of Melbourne, in 2007. He is also the Emeritus Professor with the School of Engineering and Information Technology, Faculty of Science and Technology, Federation University, Australia. He is an Editor of the *International Journal of Digital Multimedia Broadcasting* and has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, in 2012, and as a Guest Editor of special issues for the *Journal of Multimedia*, from 2009 to 2012.