Importance Of Frame Selection In Video Quality Assessment

M.T. Qadri*, J. Woods*, K.T. Tan and M. Ghanbari*, Fellow IEEE**
*School of Computer Science and Electronic Engineering, University of Essex, UK.
**School of Electrical and Electronic Engineering Singapore Polytechnic, Singapore/

Abstract

Video sequences contains multiple frames therefore their quality is estimated by determining individual quality metric of each frame then apply the temporal masking affect. However, the integration of each frame's quality metric into one score is very important because each video frame has different spatial features hence have different quality metric. There are several methods available to combine the metric into one score like averaging, linear weighting, worst frames averaging etc. Taking the average of each frame's score is not very useful as humans give more attention to the worst values (most distorted frame) while rating their values. In this paper we evaluated the performance of different integration methods and a different approach is proposed which includes the average of worst selected frames which is discussed in later sections. The work is tested on LIVE video database which consists of 40 video sequences. They have provided the mean opinion scores for each video with the database. The correlation coefficient of 88.21% is achieved when tested with the best model designed.

Keywords- ***Frame Selection, motion vectors, Objective Video Quality Meters, Full Reference meters.***

I. INTRODUCTION

Due to compression of images and videos, the quality degrades and the distortion starts to appear. There are many image and video quality meters [1-8] exists. The technique used for video quality assessment involves the extraction of each frame from a video sequence and quantifying the quality of each frame individually. Then these individual quality metrics are combined into a single quality metric for a complete video sequence before applying the temporal masking affect. The commonly used integration techniques include averaging or weighting. The fact is that the observer gives more importance to the worst incidents and they use their worst experiences while rating the quality.

The quality metric of each frame is first calculated using the frequency domain approach discussed in [9,10]. The full reference method is used and the combined effect of blockiness and blurriness distortion is considered. The very brief introduction of the meter is discussed in section II.

In this work each video sequence consists of more than 250 frames. It is more likely that each frame has different image quality metric as each

Frame has different spatial features and different amount of distortions. The objective video quality also depends upon the nature of motion in the sequence. The nature and intensity of motion also varies in different video sequences therefore the standard deviation of the motion metrics are also used in motion estimation.

The main contribution of this work is to develop the method to combine the quality metric of each frame into a single value for a video sequence. Next section briefly discusses the method for image quality estimation. Section III compares different methods for integrating the quality metric. Section IV highlights the best method of integration with some results and finally section V concludes the paper followed by references used in the work.

II. AN OVERVIEW OF IMAGE QUALITY ESTIMATION

The Full Reference image quality meter which was designed in [10] is used to determine the quality of each frame of a video sequence. Blockiness and blurriness are the main dominant distortions considered in the work. They are estimated in frequency domain. The method includes the edge detection of both reference and coded images to determine the spatial activity of the images. Then edge cancellation process is applied to cancel sharp luminance edges and it leaves only edges due to distortion. Then the frequency domain analysis is applied and the ratio of harmonics to other ac coefficients is calculated for blockiness estimation. For blurriness artifact, the ac coefficients of the coded and reference images are compared as the fact that blurriness reduces the sharpness of a image by eliminating the high frequency coefficients. The meter is briefly explained in figure below.

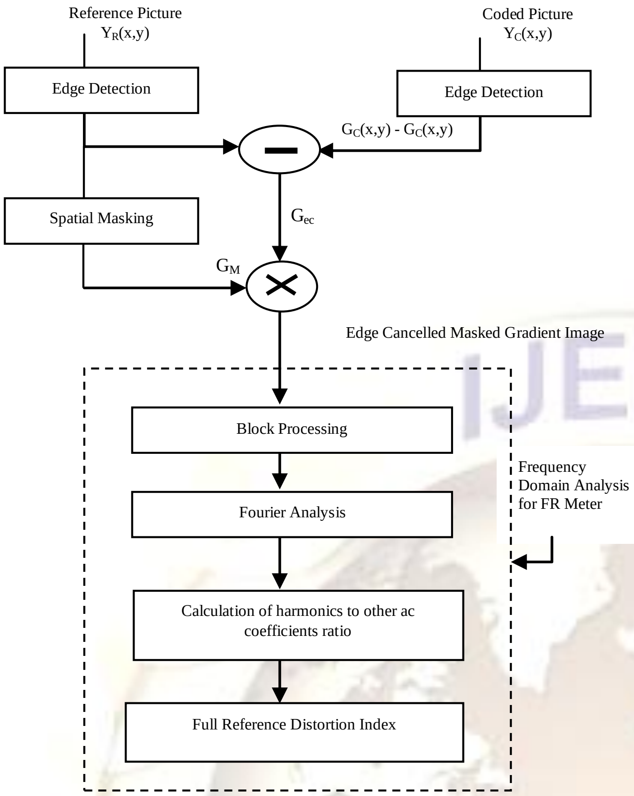

Fig 1. Overview of full reference distortion meter.

The above full reference quality meter is used to determine the image quality of each frame of a video sequence. The next section discusses different methods of integrating quality metric.

III. METHODS TO INTERGADE IMAGE QUALITY METRIC

Each video sequence has many frames which need to be processed individually for quality estimation of each frame. The quality metric for each frame is then integrated in the end for single quality metric for that video sequence. The integration of quality metric of each frame into a single value is very important and can be done by many ways. This section discusses different approaches which can be used for integration.

It is more likely that most of the frames in any video sequence has different spatial features and therefore has different distortion levels which results in different quality metric for each frame. Few different methods to integrate these scores are explored in this work which are discussed below.

A. Averaging

It is the easiest method to integrate the image quality metrics by simply taking the average of all individual frame values as described in equation below.

$$IQM_{Av} = \frac{1}{N}\sum_{n=0}^{n=N-1} IQM_{each\ frame} \quad \text{\_\_\_\_A}$$

Where, 'N' is the total number of frames in each video, 'n' is the individual frame number and $IQM_{each\ frame}$ is the individual image quality score of each frame. Due to the fact that every frame might have different distortion levels, therefore their simple average might not be so useful. It is fact that viewers give more importance to the worst incidents or sometimes they rate the quality by watching recent scenes because the video may consists of many frames. Therefore simple averaging technique may not be efficient for longer video sequences. The method is tested on LIVE database [11] for video sequence which consists of 40 different video sequences with 250 to 500 frames in each video. They have provided the Mean Opinion Scores (MOS) of each video sequence. The correlation coefficient of 67% is obtained with using averaging method.

B. Linear Weighting

Normally video sequences have many frames (250-500 in our case), it is observed that the users give more importance to the frames they watched recently while marking their observations which is called recency effect. Considering this human behavior, the objective scores are weighted using a linear function. We gave more importance to the recent frames as compared with the ones appeared earlier. The linear weighting function is described as below.

$$IQM_{Lin.Weighted} = \frac{1}{K1}\sum_{n=0}^{n=N-1} W_n(n).IQM_{each\ frame} \quad \text{\_\_B}$$

Where $W_n(n) = \frac{(1-x)}{(N-1)}n + x$, is the weighting function, which is controlled by the parameter 'x'. The recency effect decreases by increasing the value of 'x'. The value '$K1$' is the scaling factor to keep the weighting factor under 1.0, its value is equal to$k_1 = \frac{(N-1)(x+1)}{2}$. The method is tested on LIVE video database [11] and the correlation coefficient of 70.58% is obtained.

C. Minkowski Summation

The linear weighting depends upon the temporal location of the frame in video sequence therefore it may not be strong enough to manipulate the recency effect. Form experiments, it is observed that location of frame is not much important as compared with the peak intensity of the distortion. The observers are more influenced by few strong stimuli during the rating of their subjective scores. For this purpose Minkowski summation is used which enhances the significance of outstanding events. The degree of enhancement is controlled by the parameter 'x'. For larger value of 'x', the strong stimuli become more dominant in the final score. The equation for Minkowski summation is given below.

$$IQM_{Mink.\ sum} = \left\{\frac{1}{N}\sum_{n=0}^{n=N-1}(IQM_{each\ frame})^x\right\}^{1/x} \_\_C$$

The correlation coefficient of 73.68% is achieved after testing the model on LIVE video database.

D. Worst Samples Averaging

This is another method to enhance the significant events by only considering the worst values of the objective scores within the video sequence. All objective scores of each frame are arranged in descending order (by considering the distortion level) and the average of first 'x' frames is taken as the quality metric. This is explained is equation below.

$$IQM_{worst} = \frac{1}{x}\sum_{n=0}^{n=N-1} IQM_{sorted} \_\_\_\_D$$

Where, IQM_{sorted} are the sorted objective scores of each frame. They are arranged in descending order and its first value will be the worst quality frame of the sequence. The correlation coefficient of 84.98% is achieved using the worst sample averaging.

E. Worst Samples Averaging with Standard Deviation

The quality metrics also depend upon the contents of the video sequence. Some video sequences have very large motion vectors with less standard deviation like camera moving across trees or in other detailed areas. On other side, if a sequence has moving background with static objects in foreground like train travelling in a field, then there will be less variations in motion vectors as compared with their standard deviation. Therefore, the uniformity of motion in video sequence is also very important for quality estimation. The figures below discuss the motion estimation of two different types of video sequences.

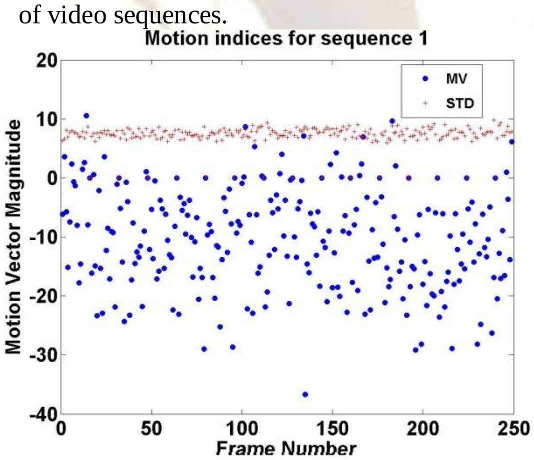

Fig 2. Sequence 1 with camera panning across trees.

It can be seen from figure 2 that the sequence has large variations in motion vectors of each frame but has very little variations in standard deviation because the video sequence contains the movement of camera across trees. On other side, if the sequence has moving background like train passing through the field, then there will be more variations in the standard deviation of its motion vectors as shown in figure below.

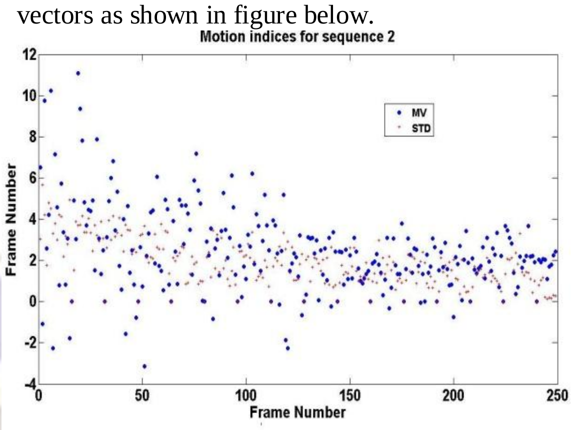

Fig 3. Sequence 2, train travelling across field

The above two figures conclude that the average of worst values of motion vectors and the standard deviation of all motion vectors in each frame both are very useful in estimation the video quality accurately. The ratio of mean motion vectors to the standard deviation is used in this research to determine the video quality metrics. The parameter is tested on LIVE video database [11] and the correlation coefficient of 88.21% is found.

The table below summarizes the results obtained so far using above techniques.

Table1 Results of different integrating functions.

Where, Q_i is the integrated score of each sequence and q_n is the quality score of each frame. The parameters W_n , x and Std are explained in above sections.

Integrating Function	**Integrated Scores**	**Correlation Coefficient**
Averaging	$Q_i = \frac{1}{N}\sum_{n=0}^{n=N-1} q_n$	67%
Linear Weighting	$Q_i = \frac{1}{K1}\sum_{n=0}^{n=N-1} W_n(n).q_n$	70.58%
Minkowski Summation	$Q_i = \left\{\frac{1}{N}\sum_{n=0}^{n=N-1}(q_n)^x\right\}^{1/x}$	73.68%
Worst Samples Averaging	$Q_{i*} = \frac{1}{x}\sum_{n=0}^{n=N-1} q_n$, *sort in descending order first.*	84.98%
Worst Samples with Standard Deviation	$Q_i = Q_{i*}./Std_{frame}$	88.21%

The below section discusses how to select window size for worst motion vectors frames.

IV. WINDOW SELECTION OF WORST QUALITY FRAMES

While rating the quality of video sequences, the users give more importance to the worst quality frames. The worst frames can be selected by first arranging the quality metric of each sequence in descending order then take the average of 'x' worst frame values. In this work each video sequence consists of 250 or more frames. The selection window always start with the worst value of quality metric which will be the first value because it is arranged in descending order. As we increase the window size, the correlation coefficient decreases because we come nearer to the average value of quality metric. The affect of increasing window size for selecting worst number of frames is explained in figure below.

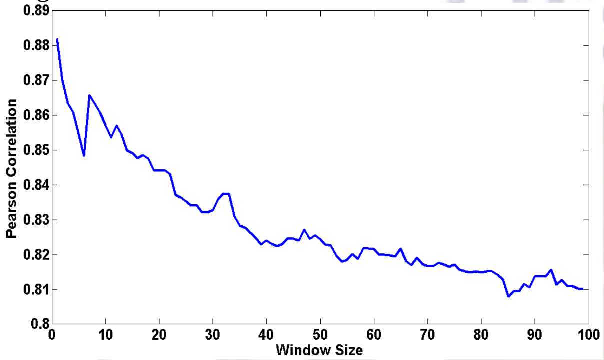

Fig 4. Impact of increasing window size for worst frame selection.

From above figure, it can be observed that the video quality metric mainly depends upon the worst quality metrics. It is also observed that the frames with good scores doesn't play role in quality estimation. As the window size is increased, its average value becomes smaller because of the inclusion of better quality frame values and therefore the quality of distortion meter is decreased as can be seen in figure 4.

CONCLUSION

This paper highlights different methods to integrate the quality scores of each frame in a video sequence. Since video sequences consists of multiple frames and each frame has different quality metric. The simplest method includes averaging of each frame's metric but as humans give more importance to the worst values, as seen from figure 3, therefore the frames with good quality metric can be ignored. Therefore the quality of video sequence can be estimated by only considering worst quality frames. Another important result for this work is that the motion vectors are itself not enough for quantifying the quality metric as different sequences have different intensity and uniformity of motion. For this purpose the standard deviation of the motion vectors are also used for the motion estimation. The meter is tested on LIVE video database [11] which consists on 40 different video sequences with their MOS provided. The Pearsons correlation of 88.21% is achieved using the above quality meter approach.

REFERENCES

[1] Z. Wang and A.C. Bovik, "A universal image quality index", IEEE Signal Processing Letters, 9(3):81–84, March 2002.

[2] A.A. Webster, C.T. Jones, M.H. Pinson, S.D. Voran, S. Wolf, "An objective video quality assessment system based on human perception," SPIE Human Vision, Visual Processing, and Digital Display IV, San Jose, CA, February 1993, pp. 15–26.

[3] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," IEEE Signal Processing Letters, Vol. 4, Nov. 1997, pp. 317-320.

[4] R. Barland and A. Saadane, "A new reference free approach for the quality assessment of mpeg coded videos," in 7th Int. Conf. Advanced Concepts for Intelligent Vision Systems, Sep. 2005, vol. 3708, pp. 364–371.

[5] G.A.D Punchihewa, D. G. Bailey and R. M. Hodgson, "Objective Quality Assessment of Coded Images:The development of New Quality Metrics", 2005.

[6] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, "No-reference image quality assessment for jpeg2000 based on spatial features," Signal Processing: Image Communication, vol. 23, no. 4, pp. 257–268, April 2008.

[7] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video", IEEE Transactions on Image Processing, vol.19, no.6, pp.1427-1441, June 2010.

[8] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "A Subjective Study to Evaluate Video Quality Assessment Algorithms", SPIE Proceedings Human Vision and Electronic Imaging, Jan. 2010.

[9] M.T.Qadri, K.T.Tan and M.Ghanbari, "Frequency Domain Blockiness Measurement for Image Quality Assessment", International Conference on Computer Technology and Development (ICCTD 2010), pp. 282-285, Cairo, Egypt, 2010

[10] M.T.Qadri, K.T.Tan and M.Ghanbari, "Frequency Domain Blockiness and Blurriness Meter for Image Quality Assessment, International Journal of Image Processing (IJIP), Volume (5) Issue (3), July 2011.

[11] LIVE video database for subjective video scores, http://live.ece.utexas.edu/research/quality/live_video.html