OPEN ACCESS

University of BRISTOL

Edge Dissimilarity Reduced-Reference Quality Metric with Low Overhead Bitrate

Farah Diyana Abdul Rahman*[1], Dimitris Agrafiotis[2], Ahmad Imran Ibrahim[3]
[1,2]Department of Electrical and Computer Engineering, Department of Mechatronics Engineering,
International Islamic University Malaysia
[3]Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK

Article Info	Abstract
	In multimedia transmission, it is important to rely on an objective quality metric which accurately represents the subjective quality of processed images and video sequences. Reduced-reference metrics make use of side-information that is transmitted to the receiver for estimating the quality of the received sequence with low complexity. In this paper, an Edge-based Dissimilarity Reduced-Reference video quality metric with low overhead bitrate is proposed. The metric is evaluated by finding the dissimilarity between the edge information of original and distorted sequences. The edge degradation can be detected in this manner as perceived video quality is highly associated with edge structural. Due to the high overhead using the Soergel distance, it is pertinent to find a way to reduce the overhead while maintaining the edge information that can convey the quality measure of the sequences. The effects of different edge detection operator, video resolution and file compressor are investigated. The aim of this paper is to significantly reduce the bitrate required in order to transmit the side information overhead as the reduced reference video quality metric. From the results obtained, the side information extracted using Sobel edge detector maintained consistency throughout the reduction of spatial and temporal down-sample. *Copyright © 2018 Institute of Advanced Engineering and Science. All rights reserved.*

Corresponding Author:

Farah Diyana Abdul Rahman,
Department of Electrical and Computer Engineering,
Department of Mechatronics Engineering,
International Islamic University Malaysia,
Email: farahdy@iium.edu.my

1. INTRODUCTION

Video quality assessment is crucial in everyday life as it is the key function in visual processing and interaction between humans, machines and systems. The ultimate end-users are humans, hence it is ideal to have video perceptual quality assessment that reflects or matches the requirement and satisfaction of human end-users. However, the subjective assessment is time-consuming, not ideal for real-time applications and very restrictive in nature as it requires a standardized environment. Therefore, an automatic and computable quantitative visual quality measure of visual content is in need to evaluate the quality and assist any further visual processing.

The aim of this paper is to provide an automated quality of experience indices, which can be used for quality control that reflects on human visual perception. The challenge of formulation of an objective video information content measure that is consistent with its subjective video information content measure. In order to be consistent, the assessment needs to be sensitive to underlying substantial intensity dissimilarities in image or video. The usage of human visual system (HVS) knowledge in video processing may be performed by recognizing the characteristics of human perceptions. HVS characteristics may be divided into its physical structure, visual perception and its image processing theories [1]. The two most common theories are the line-edge detection theory, and spatial frequency theory. The line-edge detection

theory is based on the primary visual cortex cells that have areas of excitation and inhibition that respond to a luminance edge, as well as to bright or dark lines, which ultimately detect edge and line, respectively. This finding is used as the basis for the proposed metric that is based on edges as well as the shapes of objects.

In this work, a novel Edge Dissimilarity Reduced-Reference (EDIRR) metric for reduced reference video quality assessment with low overhead bitrate is proposed. The metric is derived from the motivation that human perception understands an image mainly according to its low-level features, specifically the edges, which are a measure of the significance of a local structure. Detecting accurate information of distortion level on received videos is an important technique to measure the visual quality of transmitted video over unreliable wireless channels. During pre- or post-processing, compression, acquisition, transmission and storage, video frames may change due to various artifacts or noise which is viewed as distortion that degrades the visual quality [2]. The artifacts can be divided into two types, compression artifacts and also transmission errors. In this paper, both types of errors are introduced to the tested sequences as it is important to have a diversity of distorted test sequences. In this work, the last frame in each Group-of-Pictures (GOP) of the reference and distorted frames are used as inputs. The system produces an output in a form of a value that quantifies the quality of the distorted visual signal.

In video transmission wireless networks, there is rather limited support for bandwidths capacity. Even more, when a reduced-reference metric is implemented on the video transmission, a side information would need to be transmitted via ancillary channel with some information from the reference video. Therefore, it is pertinent to have low overhead bitrate for the side information side. In [3], a novel EDIRR metric has been proposed, where its assessment has been found to correlate well with DMOS obtained from LIVE. However, since the overhead bitrates from the metric are not sufficiently low (2.8 Mbps), an investigation focuses on compressing and evaluating the ability of the metric after being compressed is proposed to be investigated in this paper. The compression performance is investigated by looking at different combinations of spatial and temporal resolutions. This algorithm allows an effective assessment of reduced-reference video quality with much lower overhead cost.

2. RESEARCH METHOD

The proposed RR metric in this paper is performed by comparing the edge information of the original and distorted sequence in terms of structural distortion. The edge degradation can be detected in this manner as it is highly associated with the structural edge. The development of the edge-based distortion measure is by assessing the decoded grey scale images. The low-level content of visual importance or salient image features such as the edge intensity definitely has image information and this is perceptually important. Therefore, using this observation, the edge information is incorporated into the RR dissimilarity measure to develop video quality indices – EDIRR: Edge-based Dissimilarity Reduced Reference Metric.

The test sequences are obtained from the Laboratory for Image & Video (LIVE) Video Database [26] and the proposed metric would be tested against the subjective quality score, Differential Mean Opinion Score (DMOS), provided from the subjective study carried out by LIVE [27, 28]. The DMOS values range from 0 to 100, where the smaller value expresses the greater quality and the larger value states the worse quality, are collected using the subjective test model specified in ITU-R BT 500.11 [29-31]. The subjective study was conducted using a single stimulus procedure with hidden reference removal and the subjects indicated the quality of the video on a continuous scale. Subjects also viewed each of the reference videos to facilitate computation of difference scores using hidden reference removal. From the database, 80 distorted video sequences were obtained from 10 different high-quality videos with a wide variety of content as reference videos.

A set of 80 distorted videos are tested using two different distortion types: H.264 compression and simulated transmission of H.264 compressed bitstreams through error-prone wireless networks, as these type of distortions relate the most to the work performed in this chapter. The diversity of distortion types is to test the ability of the proposed objective model to predict visual quality consistently across distortions. The H.264 compression system produces fairly uniform spatial and temporal distortions in the video. Network losses, however, cause transient distortions in the video, both spatially and temporally. The H.264 compressed videos exhibit a visual appearance of typical compression artifacts such as blur, blocking, ringing and motion compensation mismatches around the edges of the main body in the frame. Videos obtained from the wireless transmission error exhibit errors that are restricted to small regions of a frame. Errors sustained by an H.264 compressed video stream in a wireless environment are also spatio-temporally localized distortions, due to the small packet sizes or temporally transient and appear as glitches in the video. A packet transmitted over a wireless channel is susceptible to transmission errors due to various factors such as shadowing, attenuation, fading and multi-user interference in wireless channels.

All of the ten uncompressed high-quality YUV sequences used have the resolution of 768 x 432 pixels. Each sequence was assessed by 29 valid human subjects in a single stimulus study where the scores are based on a continuous quality scale. The DMOS from the subjective evaluations are used to compare with some of the similarity measures. In this work, all of the frames are used to determine the most suitable similarity measure. However, only the last reference frames in each GOP from both the reference and received sequences perform as inputs in the proposed quality assessment system. This is due to the fact that it is crucial in keeping the overhead bit rate as low as possible, as well its practicality and realistic in keeping with real-time wireless transmission over a multicast network scenario. The system outputs a value to quantify the quality of the distorted visual. The LIVE Video Database has been evaluated by many researchers and has been verified with various objective performance metrics.

3. RESULTS AND METHOD

The proposed RR metric in this paper is performed by comparing the edge information of the original and distorted sequence in terms of structural distortion. The edge degradation can be detected in this manner as it is highly associated with the structural edge. The development of the edge-based distortion measure is by assessing the decoded grey scale images. The low-level content of visual importance or salient image features such as the edge intensity definitely has image information and this is perceptually important. Therefore, using this observation, the edge information is incorporated into the RR dissimilarity measure to develop video quality indices – EDIRR: Edge-based Dissimilarity Reduced Reference Metric.

The test sequences are obtained from the Laboratory for Image & Video (LIVE) Video Database [26] and the proposed metric would be tested against the subjective quality score, Differential Mean Opinion Score (DMOS), provided from the subjective study carried out by LIVE [27, 28]. The DMOS values range from 0 to 100, where the smaller value expresses the greater quality and the larger value states the worse quality, are collected using the subjective test model specified in ITU-R BT 500.11 [29-31]. The subjective study was conducted using a single stimulus procedure with hidden reference removal and the subjects indicated the quality of the video on a continuous scale. Subjects also viewed each of the reference videos to facilitate computation of difference scores using hidden reference removal. From the database, 80 distorted video sequences were obtained from 10 different high-quality videos with a wide variety of content as reference videos.

A set of 80 distorted videos are tested using two different distortion types: H.264 compression and simulated transmission of H.264 compressed bitstreams through error-prone wireless networks, as these type of distortions relate the most to the work performed in this chapter. The diversity of distortion types is to test the ability of the proposed objective model to predict visual quality consistently across distortions. The H.264 compression system produces fairly uniform spatial and temporal distortions in the video. Network losses, however, cause transient distortions in the video, both spatially and temporally. The H.264 compressed videos exhibit a visual appearance of typical compression artifacts such as blur, blocking, ringing and motion compensation mismatches around the edges of the main body in the frame. Videos obtained from the wireless transmission error exhibit errors that are restricted to small regions of a frame. Errors sustained by an H.264 compressed video stream in a wireless environment are also spatio-temporally localized distortions, due to the small packet sizes or temporally transient and appear as glitches in the video. A packet transmitted over a wireless channel is susceptible to transmission errors due to various factors such as shadowing, attenuation, fading and multi-user interference in wireless channels.

All of the ten uncompressed high-quality YUV sequences used have the resolution of 768 x 432 pixels. Each sequence was assessed by 29 valid human subjects in a single stimulus study where the scores are based on a continuous quality scale. The DMOS from the subjective evaluations are used to compare with some of the similarity measures. In this work, all of the frames are used to determine the most suitable similarity measure. However, only the last reference frames in each GOP from both the reference and received sequences perform as inputs in the proposed quality assessment system. This is due to the fact that it is crucial in keeping the overhead bit rate as low as possible, as well its practicality and realistic in keeping with real-time wireless transmission over a multicast network scenario. The system outputs a value to quantify the quality of the distorted visual. The LIVE Video Database has been evaluated by many researchers and has been verified with various objective performance metrics.

3.1. Edge Detector Evaluation Methodology

Ideally it is aspiring to have a set of connected curves that indicate the boundaries of objects, while preserving the important structural properties of an image, after applying an edge detector. However, it is not always possible to obtain such ideal edges from real life images of moderate complexity. In reality, edges extracted from significant images are often hindered where the edge curves are not connected; missing edge

segments as well as false or weak edges emerge where they are not relating to appealing phenomena in the image. This complicates the process of interpreting the image data consequently. For standardisation and simplification purposes, Canny and Sobel edge detectors are going to be implemented throughout this paper as opposed to use only Sobel edge detector as in [3].

The process of filtering the edges and its quality metrics that can retain the metric score correlated with subjective quality assessment from LIVE database after undergo a few sub-sampling and image processing is depicted as this process flow as shown in Figure 1.
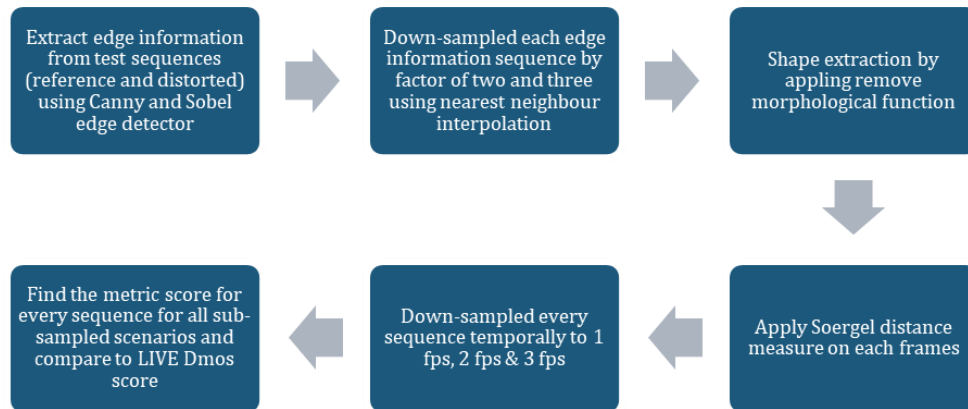

Figure 1. Methodology of Finding Out the Correlation between Metric Score for Spatially and Temporally Down-sampled Test Sequences with LIVE Database Score

3.2. Spatial Resolution Reduction Methodology

There are various ways to reduce spatial and temporal resolutions. It is important to compare the effects of compressing methods on the images to be used as side information in the proposed Soergel distance from [3]. One of the compression processes is down-sampling where the reduced sequence would have a horizontal and a vertical sampling frequency that is less than the maximum value of all components in the original sequence. It is decided to use sampling frequency of half and one third of the maximum value using the method of nearest-neighbour interpolation where it estimates an image value in between image pixels that the point falls within. Sampling reduces the size of an image by extracting pixels and the reduced image takes a value of interpolation for every block of two by two pixels for down-sample by factor of two and a value of interpolation for every block of three by three pixels for down-sample by factor of three. This type of interpolation is low in computation complexity as the number of pixels considered is low as compared to other types of interpolation such bilinear or bicubic interpolations. It also does not use any low-pass filter to prevent aliasing which is an unnecessary computation complexity cost since no high-contrast images are in used in this experiment. The nearest neighbour assignment resampling technique is used because rarely do the centres of the input cells align with the transformed cell centres of the desired resolution. The other spatial resolution reduction process is done through pixel sampling. The reduced sequence takes a sample value of every block of two by two pixels for spatial reduction by factor of two and a sample value of every block of three by three pixels for spatial reduction by factor of three. It reduces the spatial resolution of a component in an image. Both spatial resolution reduction methods are tested in order to decide the better method with low complexity in mind. The results of both spatial reduction methods are presented in Figure 2 which used the first frame from Tractor sequence of original sequence as an example.
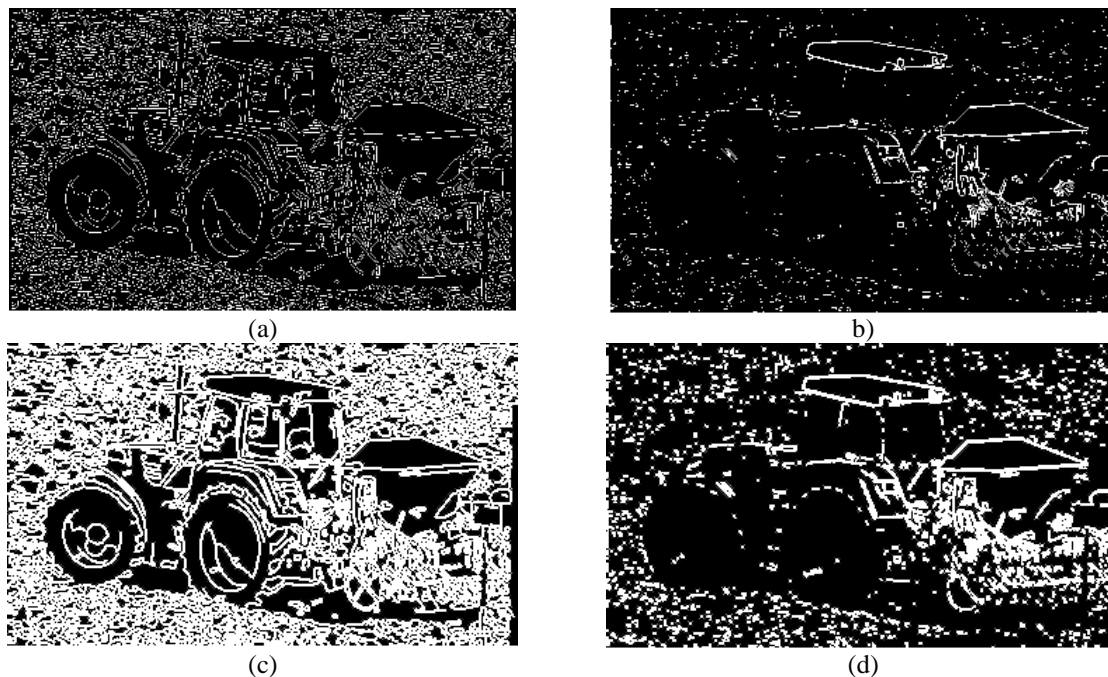
(a)

b)

(c)

(d)

Figure 2. Down-sampled by Factor of 2 on Canny (left) and Sobel (right) Edges using Pixel Sampling (top) and Nearest Neighbour Interpolation (bottom)

Figure 2(a) and (b), show the edge extracted using Canny and Sobel edge detector from the first frame of Tractor sequence and down-sampled using pixel sampling method respectively. On the other hand Figure 2(c) and (d) show the edge extracted using Canny and Sobel edge detector from the first frame of Tractor sequence and are down-sampled using nearest neighbour interpolation respectively. Referring to Figure 1, the process after reduce the spatial resolution is to ensure the edge extracted and down-sampled can be process accordingly. Morphology function in MATLAB use set of image processing operations that process images based on shapes. It is decided to use a built-in Matlab function bwmorph with operation 'remove' in order to reduce the unnecessary information which will increase the overhead cost where it applies a morphological remove operation to the binary image. This function removes the interior pixels to leave an outline of the edges where it sets an interior pixel to 0 if all of its 4-connected neighbouring pixels are 1, which will become its boundary pixels and thus leaving only the boundary pixels on which reconnects many disconnected edges. This process is ideal in reducing the complexity and the side-information overhead as it removes the non-edge information and suppresses noisy pixels from the main structure within the frames. The morphological filter is suited to the purpose of extraction of main objects or structures from the scenes.

The whole process is to reduce the overhead while keeping and capturing the main structure of the frame as it is one of the most important aspects in video quality evaluation. It is a fact that human vision focuses on certain areas of interest within the frame and their sensitivity is greatly reduced outside the areas of interest. In this case the boundary of edge pixels comprises as the interest region. Proposed quality metric takes this aspect into account and attempt to model the focus of attention that is the main structure of the frame for overall video quality score computation.
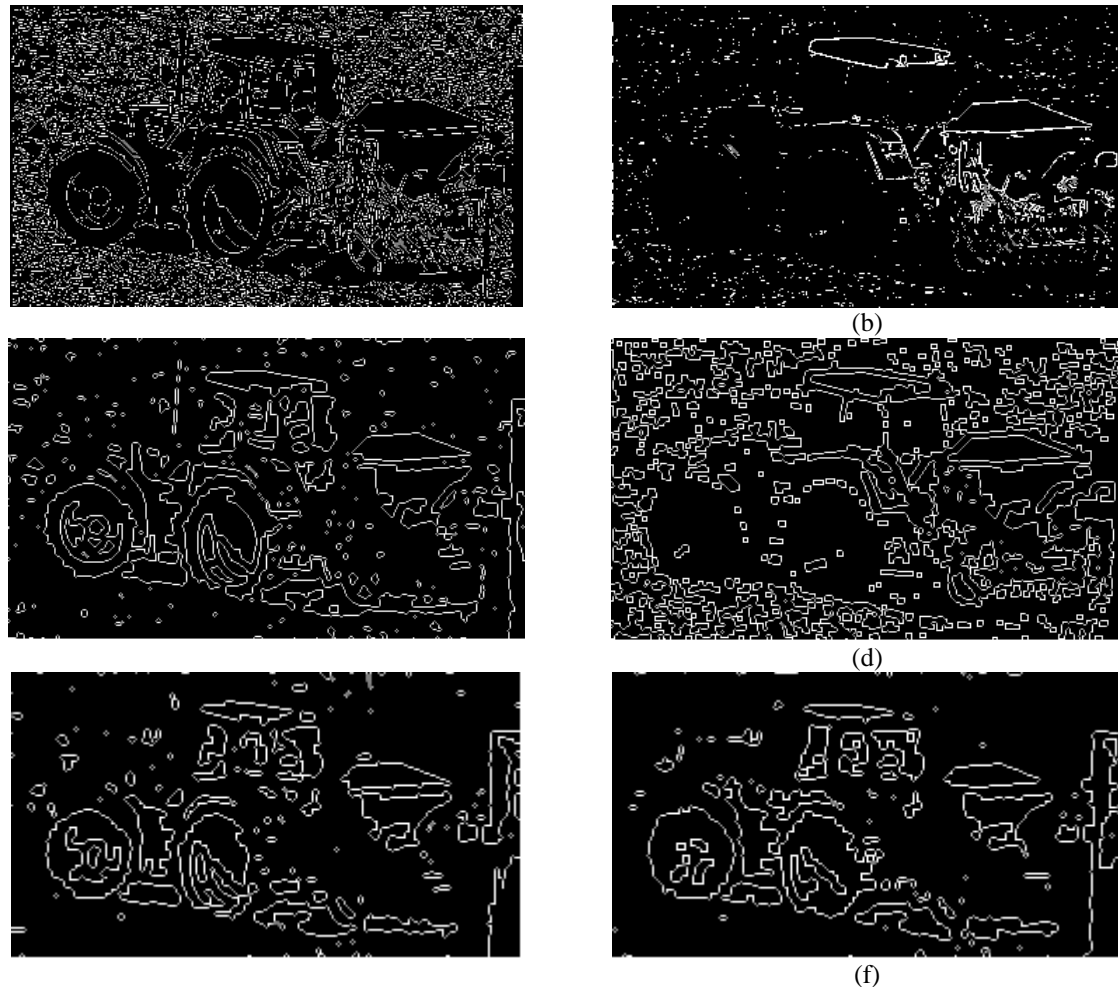
Figure 3. Sequence of Boundary Images Produced by Canny (left) and Sobel (right) detector. The Images are Down-sample by Factor of Two in (a)(b)(c)(d) and by Factor of Three in (e)(f) All with Morphological Post-Processing

Figure 3 shows the example of outputs after applying the function onto all the down-sampled factor of two test sequence Tractor and the down-sampled factor of three test sequence Tractor. Figure 3 proved that it is pertinent to use the nearest neighbour interpolation instead of pixel sampling as the later method has no constructive effect and none of the edges seem to be connected when remove morphological function is applied. Instead the outputs from using pixel sampling gave either too much information, from edges extracted by Canny edge detector, or too little information, from edges extracted by Sobel edge detector, but both outputs did not manage to capture the main structures or essence of the image. Therefore, an algorithm consists of reducing the spatial resolution employing nearest neighbour interpolation, using remove morphological function on both Canny and Sobel edge extracted frames, as well as applying Soergel distance measure is proposed. The down-sampling means a lot of information within the sequences are thrown away and might affects the correlation of Soergel distance measure with subjective quality as found in [3]. Therefore, it is necessary to employ the Soergel distance measure on the sub-sampled and processed sequences and observe the apprised metric performance in quantifying the video quality and its correlation with DMOS.

It is decided to only apply nearest neighbour interpolation edge detector in order to standardise the results and the process continued by applying remove morphological operation before implementing the Soergel distance measure on the processed sequences. The reason being is to reduce any weak edges as well as unwanted noises. Elimination of the noisiest bit levels from the original image gives better results for the edge image analysis or edge detection. Mathematical morphology is interesting because it involves simple logical operations, thus making real-time application possible.

3.3. Temporal Resolution Variations

All sub-sampled sequences are compressed using CALIC and the bitrates for each sequences are calculated. The compressor called CALIC which stands for Context Adaptive Lossless Image Compression uses both context to obtain the distribution of the symbol being encoded, and prediction of the pixel values by using previous values of the sequence to obtain prediction of the value of the symbol being encoded. Bitrate of the side information is calculated using Equation (1):

$$Bitrate = Ave.\, no.\, bits\, per\, pixel\, (bpp) \times video\, resolution \times frame\, rate \qquad (1)$$

After the side information is compressed using CALIC with the original sequences bitrate (25 fps or 50 fps), the side information overhead cost is found to be still high. Video streaming multicast would generally deteriorate as the frames transmitted accumulate, therefore usually the highest distortions could be observed in the last frame of each GOP. For practicality, the side information is transmitted at the end of every GOP, where the frames would have the worst signal if compared to the signal in previous frames. The size of GOP can be decided by the video transmitter and provider depending on the bandwidth availability of the user's channel. Therefore, it is determined to reduce the temporal frequency. The effect of reduction in temporal frequency is investigated on the test sequences using three different scenarios. These scenarios are transmitting side information every one frame per second (fps), two fps and three fps. The test sequences in the LIVE database have either 25 fps or 50 fps bitrate and the scenarios for each temporally down-sampled are described as follow in Table 1. Temporal down-sample is to cater the need to reduce the overhead cost. The original frame rate may be 25 fps or 50 fps and the frame rates are reduced by performing frame drop in temporal down-sampling.

Table 1. Temporal Down-sampled Scenarios.

Temporal Down-sampled	25 fps sequences	50 fps sequences
1 fps	Transmit side information every 25 frames	Transmit side information every 50 frames
2 fps	Transmit side information every 13 frames	Transmit side information every 26 frames
3 fps	Transmit side information every 8 frames	Transmit side information every 16 frames

4.   Results and Analysis

Experimental tests are conducted based on the LIVE Video Database in order to justify the proposed metric and its correlation with the HVS. Correlations between the subjective score and the objective metric have to be computed in order to verify the usefulness of the objective metrics using various performance metrics.The first analysis performed is the correlation between the DMOS and the quality indices from the variety of similarity measures. The correlation coefficients are acquired between the two parameters in order to compare and justify the performances relatively. The correlations of every sequence are averaged and the findings are reported in Table 2. Table 2 shows the correlations between spatially down-sampled, with remove morphological function applied and original resolution from [3], with LIVE DMOS score. The Soergel distance measure calculated on every sequences were averaged for each sequence over the range of wireless and compression distortions.

Table 2. Pearson Correlation between Spatially-down-sampled then Morphologically Removed and Original Resolution Sequences with LIVE DMOS score.

Sequences	Sobel 768 x 432	Canny 384 x 216	Sobel 384 x 216	Canny 256 x 144	Sobel 256 x 144
Blue Sky	0.9506	0.9493	0.8934	0.9449	0.8813
Pedestrian Area	0.7278	0.5190	0.6535	0.4869	0.6419
Park Run	0.8856	0.8249	0.7365	0.7981	0.7166
River Bed	0.7661	0.3352	0.7028	0.3014	0.6743
Rush Hour	0.9526	0.6182	0.9251	0.5224	0.9114
Shields	0.9749	0.9771	0.96845	0.9752	0.9614
Station	0.8674	0.9022	0.9031	0.8949	0.8988
Sunflower	0.6307	0.6349	0.6177	0.6201	0.6110
Tractor	0.9625	0.6423	0.9620	0.5463	0.9163

Table 2 shows that Soergel distance measures on spatially down-sampled and morphologically removed sequences that based on edges extracted using Sobel edge detector produce consistent and high

correlation to DMOS score. However, the correlations are not consistent when the Soergel distance measures applied on edges extracted using Canny edge detector, as there are quite high correlations and there are also quite low correlations with DMOS score on certain sequences such as River Bed and Tractor. Temporal down-sample is to cater to the need to reduce the overhead cost. The original frame rate may be 25 fps or 50 fps and the frame rates are reduced by performing frame drop in temporal down-sampling. An example of average bitrate of side information from Sobel edge detected Tractor sequence compressed using CALIC specifically for different spatial and temporal down-sampled factors are as shown in Table 3 to 6.

Table 3. Spatial Down-sampled Bitrates for 25fps

Spatial downsampled	Pixel resolution	Overhead bitrate
None	768 x 432	2.8 Mbps
2	384 x 216	769.3 kbps
3	256 x 144	411.0 kbps

Table 4. Temporal down-sampled Bitrates for Original Sequence

Temporal downsampled	Pixel resolution	Overhead bitrate
25fps	768 x 432	2.8 Mbps
3fps	768 x 432	333.6 kbps
2fps	768 x 432	222.4 kbps
1fps	768 x 432	111.2 kbps

Table 5. Temporal down-sampled bitrates for spatial down-sampled factor of two

Temporal downsampled	Pixel resolution	Overhead bitrate
25fps	384 x 216	769.3 kbps
3fps	384 x 216	92.3 kbps
2fps	384 x 216	61.5 kbps
1fps	384 x 216	30.7 kbps

Table 6. Temporal Down-sampled Bitrates for Spatial Down-sampled Factor of Three

Temporal downsampled	Pixel resolution	Overhead bitrate
25fps	256 x 144	411.0 kbps
3fps	256 x 144	49.3 kbps
2fps	256 x 144	32.9 kbps
1fps	256 x 144	16.4 kbps

Table 3 to 6 show the cost of side information in terms of bitrate after implementing spatio-temporal down-sampling. The overhead costs are substantially reduced from 2.8 Mbps to 16.4 kbps if the frame is down-sampled by a factor of three spatially and transmitting one side information for every 1 fps. This is the comparison between original resolution and frame rate side information and down-sampled spatially and temporally side information. These values could be smaller but it could also compromise the quality predictor reliability. The average bitrate across different distortions for down-sampled side information sequences based on Canny edge detector was also calculated and compared to average bitrate from side information based on Sobel edge detector. The results found have shown not much improvement when using Canny as the edge extracted operator. Even though 16.2 kbps is quite a small bitrate as compared to initial 2.8 Mbps, it is still open for improvement in reducing the overhead. The down-sampling process instigates a lot of information lost within the sequences and might affects the correlation of Soergel distance measure with subjective quality as found in [3]. 'Soergel' in Table 7 represents Soergel distance measure, s1 represents original spatial resolution using Sobel edge detector [3], s2 represents using Sobel edge detector and sub-sampled by factor of 2, and s3 represents using Sobel edge detector and sub-sampled by factor of 3. All sequences have been down-sampled using nearest neighbour interpolation, remove morphological function and temporally reduced to 1 fps.

Table 7. Pearson Correlation between Soergel Distance and DMOS Scores on Every Processed Sequences According to Figure 1

	Soergel_s1 _25/50fps	Soergel_s2 _1fps	Soergel_s3 _1fps	Soergel_s2 _2fps	Soergel_s3 _2fps	Soergel_s2 _3fps	Soergel_s3 _3fps
Blue sky	0.9506	0.9095	0.9061	0.8959	0.8767	0.8943	0.8851
Pedest. Area	0.7278	0.6372	0.6194	0.6525	0.6350	0.5878	0.5773
Park Run	0.8856	0.7510	0.7489	0.7423	0.7323	0.6783	0.6569
River Bed	0.7661	0.7016	0.6737	0.7018	0.6719	0.6681	0.6300
Rush Hour	0.9526	0.8583	0.8550	0.9325	0.9094	0.8521	0.8067
Shield	0.9749	0.9640	0.9695	0.9744	0.9701	0.7370	0.7075
Station	0.8674	0.8951	0.8882	0.9092	0.9052	0.9027	0.8888
Sunflower	0.6307	0.5996	0.6194	0.5653	0.5454	0.5421	0.5281
Tractor	0.9625	0.9634	0.9124	0.9593	0.9010	0.9348	0.8638

'Soergel' in Table 13 represents Soergel distance measure, s1 represents original spatial resolution using Sobel edge detector (from Chapter 4), c2 represents using Canny edge detector and sub-sampled by factor of 2, s2 represents using Sobel edge detector and sub-sampled by factor of 2, c3 represents using Canny edge detector and sub-sampled by factor of 3 and s3 represents using Sobel edge detector and sub-sampled by factor of 3. All sequences have been down-sampled using nearest neighbour interpolation, remove morphological function and temporally reduced to 1 fps.

5.   CONCLUSION

The main aim in this paper is to reduce significantly the overhead bitrate required in order to transmit the side information without influencing the reduced reference video quality metric performance from [3]. This consequently evolves to become a modified EDIRR, a reduced reference based on edge dissimilarity with low overhead bitrate. Video quality metrics are essential in reflecting the perceived video quality accurately and they should correlate well with subjective quality assessment of the test sequences. The correlation coefficients between the resulting Soergel distance measure sequences temporally reduced to 1 fps with LIVE DMOS score are presented in Table 7. It shows that the side information extracted using Sobel edge detector maintained consistency throughout the reduction in both in spatial and temporal down-sample considerably with the amount of information available. The correlations using Canny edge detector are almost similar to the previous Soergel measure without any down-sampling but they are not thoroughly consistent. The inconsistency only occurred on test sequences with fast motion and a lot of textures. This is influenced by the Canny edge detector which detects more details than Sobel edge detector and increases the probability to capture the weak edges. The problem may be solved by increasing the threshold but a higher threshold causes the inability to capture the main structure after down-sampling and remove morphological function applied according to prior initial test. Results also show the possibility to transmit the side information whether it is down-sampled to factor of three and even factor of two if the potential bandwidth allowance is convenient as the bitrate of the overhead is very small.

REFERENCES

[1]   M. Čadík. Human perception and computer graphics. Postgraduate Study Report, Czech Technical University. 2004.

[2]   S. Winkler, M. Kunt, and C. J. van den Branden Lambrecht. Vision and video: models and applications. *Vision Models and Applications to Image and Video Processing. Springer.* 2001: 201-229.

[3]   F. D. Abdul Rahman, et al. .*Reduced-Reference Video Quality Metric based on Edge Information.* Proceeding of 4th IEEE International Conference on Smart Instrumentations, Measurement, and Aplications (ICSIMA) 2017, Putrajaya. 2017: 1-5.

[4]   T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2002; 24(7): 971-987.

[5]   W. Jinjian, L. Weisi, S. Guangming, and X. Long. *Reduced-reference image quality assessment with local binary structural pattern.* 2014 IEEE International Symposium on Circuits and Systems (ISCAS). 2014: 898-901.

[6]   M. Zhang, J. Xie, X. Zhou, and H. Fujita. No reference image quality assessment based on local binary pattern statistics. *Visual Communications and Image Processing (VCIP)*. 2013: 1-6.

[7]   X. Cui, Z. Shi, J. Lin, and L. Huang. The Research of Image Quality Assessment Methods. *Physics Procedia*. 2012; 25(0): 485-491.

[8]   S. Rakshit, and A. Mishra. Estimation of Structural Information Content in Images. *Computer Vision (ACCV)*. 2006: 265-275.

[9]   X. Mou, M. Zhang, W. Xue, and L. Zhang. Image quality assessment based on edge. *IS&T/SPIE Electronic Imaging*. 2011: 78760N-78760N-9.

[10]   X. Fei, L. Xiao, Y. Sun, and Z. Wei. Perceptual image quality assessment based on structural similarity and visual masking. *Signal Processing: Image Communication*. 2012; 27(7): 772-783.

[11]   L. Zhichao, T. Jinxu, and Z. Zhufeng. *Reduced-reference image quality assessment based on average directional information*. 2012 IEEE 11th International Conference on Signal Processing (ICSP). 2012: 787-791.

[12]   R. Anderson, N. Kingsbury, and J. Fauqueur. Determining multiscale image feature angles from complex wavelet phases. *Image Analysis and Recognition*. 2005: 490-498.

[13]   P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. Perceptual blur and ringing metrics: application to JPEG2000. *Signal Processing: Image Communication*. 2004; 19(2): 163-172.

[14]   E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti. *A no-reference quality metric for measuring image blurs*. 7th International Symposium on Signal Processing and Its Applications. 2003: 469-472.

[15]   X. Ran, and N. Farvardin. A perceptually motivated three-component image model-Part I: description of the model. *IEEE Transactions on Image Processing*. 1995; 4(4): 401-415.

[16]   M. S. Prieto, and A. R. Allen. A similarity metric for edge images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2003; 25(10): 1265-1273.

[17]   M. A. Rodríguez-Díaz, and H. Sánchez-Cruz. Refined fixed double pass binary object classification for document image compression. *Digital Signal Processing*. 2014; 30(0): 114-130.

[18]   Y. Yaegashi, K. Tateoka, K. Fujimoto, T. Nakazawa, A. Nakata, Y. Saito, T. Abe, M. Yano, and K. Sakata. Assessment of Similarity Measures for Accurate Deformable Image Registration. *Journal of Nuclear Medicine & Radiation Therapy*. 2012.

[19]   P. Willett, J. M. Barnard, and G. M. Downs. Chemical similarity searching. *Journal of chemical information and computer sciences*. 1998; 38(6): 983-996.

[20]   M. Kowalczyk, P. Koza, P. Kupidura, and J. Marciniak. Application of mathematical morphology operations for simplification and improvement of correlation of images in close-range photogrammetry. *The international archives of the photogrammetry, remote sensing and spatial information sciences*. 2008; 37: 153-8.

[21]   R. Krishnapuram, and S. Gupta. Morphological methods for detection and classification of edges in range images. *Journal of Mathematical Imaging and Vision*. 1992; 2(4): 351-375.

[22]   R. M. Haralick, S. R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1987; 4: 532-550.

[23]   B. Kaur, and A. Garg. Comparative study of different edge detection techniques. *International Journal of Engineering Science and Technology (IJEST)*. 2011: 3.

[24]   R. Kaur, and S. Kaur. Object Extraction and Boundary Tracing Algorithms for Digital Image Processing: Comparative Analysis: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2013; 3(5): 6.

[25]   P. Perona, T. Shiota, and J. Malik. Anisotropic diffusion. *Geometry-driven diffusion in computer vision. Springer*. 1994: 73-92.

[26]   C. Tsiotsios, and M. Petrou. On the choice of the parameters for anisotropic diffusion in image processing. *Pattern recognition*. 2013; 46(5): 1369-1381.

[27]   K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack. LIVE video quality database. Laboratory for Image and Video Engineering. 2009.

[28]   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*. 2010; 19(6): 1427-1441.

[29]   K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack. A subjective study to evaluate video quality assessment algorithms. IS&T/SPIE Electronic Imaging. 2010: 75270H-75270H-100.

[30]   I. Recommendation. Methodology for the Subjective Assessment of the Quality of Television Pictures. Recommendation ITU-R BT. 500-11. ITU Telecom. Standardization Sector of ITU. 2002.

[31]   S. Ch and V. Rao. Low bit Rate Video Quality Analysis Using NRDPF-VQA Algorithm. *International Journal of Electrical and Computer Engineering (IJECE),* 5(1), 71, 2015.

.