A MATHEMATICAL MODEL FOR VIDEO QUALITY ASSESSMENT

Deshvir Malik*

Abstract—

Image Quality Assessment (IQA) has become an important issue in many fields and applications. Currently used subjective method to analyse the quality of video is time consuming and not accurate. An algorithm has been chosen to analyse the video quality. Through the algorithm chosen I wish to attempt a different strategy and technique in the process of IQA. The general algorithms use a technique of operation on each frame of a video. I wish to consider the complete video as a stream of data rather than a set of frames. The algorithm works by extraction of features from the target video.The algorithm will target the features like Quantization parameter, motion factors and Bit allocation factor which are calculated using the data from compressed bit stream.

The experiment results show that the proposed algorithm has better performance than many conventional methods in terms of both the prediction accuracy and the computation complexity.

Keywords—**video quality assessment, no-reference, full reference, reduced reference, quantization parameter, bit allocation, SSIM, MS-SSIM, F-SSIM and DMOS.**

* E&TC Department, MITCOE (PUNE), PUNE, INDIA

I. INTRODUCTION

In recent years, image quality assessment (IQA) has become an important issue in many fields and applications such as image acquisition, transmission, compression, restoration and enhancement. As the ultimate consumers, humans can easily give subjective scores to measure the qualities of images they observe. However, it is a challenging work to embed such a mechanism into an image processing system whose goal is to maximize visual quality at a given cost. Therefore, an automatic quality measurement method that can give scores to images in a meaningful agreement with subjective judgment of human being is needed.

The field of image and video processing generally deals with signals that are meant for human consumption, such as images or videos over the Internet. An image or video may go through many stages of processing before being presented to a human observer, and each stage of processing may introduce distortions that could reduce the quality of the final display. For example, images and videos are acquired by camera devices that may introduce distortions due to optics, sensor noise, colourcalibration, exposure control, camera motion etc. After acquisition, the image or video may further be Processed by a compression algorithm that reduces the bandwidth requirements for storage or transmission. Such compression algorithms are generally designed to achieve greater savings in bandwidth by letting certain distortions happen to the signal. Similarly, bit errors, which occur while an image is being transmitted over a channel or (rarely) when it is stored, also tend to introduce distortions. Finally, the display device used to render the final output may introduce some of its own distortion, such as low reproduction resolution, bad calibration etc. The amount of distortion that each of these stages could add depends mostly on economics and/or physical limitations of the devices.

One is obviously interested in being able to measure the quality of an image or video, and to gauge the distortion that has been added to it during different stages. One obvious way of determining the quality of an image or video is to solicit opinion from human observers. After all, these signals are meant for human consumption. However, such a method is not feasible not only due to the shear number of images and videos that are "out there" but also because we want to be able to embed quality measurement techniques into the very algorithms that process images and videos, so that their output quality may be maximized for a given set of resources.

The goal of research in objective image quality assessment is to develop quantitative measures that can automatically predict perceived image quality. Generally speaking, an objective image quality metric can play an important role in a broad range of applications, such as image acquisition, compression, communication, displaying, printing, restoration, enhancement, analysis and watermarking [2]. First, it can be used to dynamically monitor and adjust image quality. Second, it can be used to optimize algorithms and parameter settings of image processing systems. Third, it can be used to benchmark image processing systems and algorithms.

In short, objective quality measurement (as opposed to subjective quality assessment by human observers) seeks to determine the quality of images or videos algorithmically. The goal of objective quality assessment (QA) research is to design algorithms whose quality prediction is in good agreement with subjective scores from human observers.

Image and video QA algorithms may be classified into three broad categories:

A. Full-reference

Full-Reference (FR) QAmethods, in which the QA algorithm has access to a 'perfect version' of the image or video against which it can compare a 'distorted version'. The 'perfect version' generally comes from a high-quality acquisition device, before it is distorted by, say, compression artefacts and transmission errors. However, the reference image or video generally requires much more resources than the distorted version, and hence FR QA is generally only used as a tool for designing image and video processing algorithms for in-lab testing, and cannot be deployed as an application.

B. Reduced reference

Reduced-Reference (RR) QA methods, in which partial information regarding the 'perfect version' is available. A side-channel (called an RR channel) exists through which some information regarding the reference can be made available to the QA algorithm. RR QA algorithms use this partial reference information to judge the quality of the distorted signal.

C. No reference

No-Reference (NR) QA methods, in which the QA algorithm has access only to the distorted signal and must estimate the quality of the signal without any knowledge of the 'perfect version'. Since NR methods do not require any reference information, they can be used in any application where a quality measurement is required. However, the price paid for this flexibility is in terms of the ability of the algorithm to make accurate quality predictions, or a limited scope of the NR QA algorithm (such as NR QA for JPEG images only etc.[1].

While full-reference (FR) and reduced-reference (RR) IQA algorithms provide a useful and effective way to evaluate quality of distorted images, in many cases however, the reference image or even the partial information of images are unknown, in which case a no-reference (NR) IQA algorithm is desired. NR IQA algorithms can be further classified into distortion-specific and non-distortion-specific, based on the prior knowledge of the distortion type. Most existing NR IQA methods are distortion-specific, assuming that the distortion type is known, such as the white noise and blurring, JPEG/JPEG2000 [1].

This underlying assumption limits the application domain of these algorithms. Non-distortion-specific algorithms do not consider the prior knowledge of distortion type, but instead they give quality scores assuming that image to be evaluated has a same distortion type as those in the training database. These methods usually involve machine learning techniques, where distinct features related to image quality are extracted to train learning models and then these models are used to evaluate the quality of testing images [3].

II. VIDEO COMPRESSION AND ANALYSING PARAMETERS

A. *Need of Video Compression*

High Definition Television (HDTV)

- 1920x1080
- 24 bit for 1 pixel (8 bits for each RGB)
- 6.22 Mb per Frame
- Total 1.5 Gb/sec for 30 frames per second (full motion)

Cable TV: Each cable channel is 6 MHz

- Max data rate of 19.2 Mb/sec
- Reduced to 18 Mb/sec (To support audio, transport)
- Compression ratio required ~ 80:1

Uncompressed video (and audio) data are huge. In HDTV, the bit rate easily exceeds 1 Gbps. It is a big problems for storage and network communications [5]. Hence data compression becomes really essential component of video storage and transmission.

Lossy methods have to be employed since the compression ratio of lossless methods (e.g., Huffman, Arithmetic, and LZW) is not high enough for image and video compression, especially when distribution of pixel values is relatively flat [11].

The following compression types are commonly used in Video compression:

- Spatial Redundancy Removal - Intraframe coding (JPEG).
- Spatial and Temporal Redundancy Removal - Intraframe and Interframe coding (H.261, MPEG) [8], [11].

Network transmission and intelligent terminals represent the current trends in the video industry. Due to the limitation of bandwidth and storage capacity, video compression techniques such as MPEG-2, MPEG-4, H.263, and H.264 are widely used. Although these techniques can decrease the bandwidth in transmission significantly, they introduce distortion in videos. Since human observers at the receiver side are sensitive to the video quality, for many applications, such as video conferences and broadcasting, it is important to have a good estimate of the quality of the material being received. Unfortunately, the most accurate way to determine the quality of a video is by measuring it using psychophysical experiments with human subjects, which is a time-consuming and expensive process. As a result, objective quality assessments aiming to measure the video quality without the presence of human is more desirable.

Conventional objective quality metrics can be classified into the following three categories by its dependence on the information of the original videos: Full-reference (FR) metrics require the original videos as reference , reduced reference (RR) metrics rely on partial information related to the original signal and no-reference (NR) metrics that assess quality score only using the data receive at the terminals . It is quite difficult to achieve the original video information in most video related application terminals, and the transmission of such information will also occupy certain limited bandwidth. Comparatively, NR metrics show better applicability.

There are many algorithms based on Full-Reference, Reduced Reference and No-Reference methods. It quite difficult to achieve the original video information in most video related application terminals and the transmission of such information will also occupy certain limited bandwidth. Comparatively, NR metrics show better applicability.

No reference quality metric is used to judge the networked video, the value of average bit rate per pixel is used to predict the spatial and temporal complexity [7]. But this value varies wildly with the change of the encoder and its algorithm, which could not reflect the spatial and temporal characteristics of video precisely.

The information extracted from the bit stream during the decoding process including quantization parameter, motion vector, and bits consumption which exist in every bit stream, are used to calculate three key factors; namely

- The Quantization parameter factor
- The motion factor
- The bit allocation factor

The above parameters could be utilized to reflect the video quality. Different from the conventional VQA method, multiple video frames are processed together to obtain the objective quality score. C-VQA will be tested on the H.264/AVC bit streams in LIVE Video Quality Database [14][15]. This will help us to compare the results of subjective methods with the proposed method.

B. Quantization parameter factor

The quantization step is where a significant portion of data compression takes place. In H.264, the transform coefficients are quantized using scalar quantization with no widened dead-zone. Fifty-two different quantization step sizes can be chosen on a macro block basis - this being different from prior standards. Moreover, in H.264 the step sizes are increased at a compounding rate of approximately 12.5%, rather than increasing it by a constant increment. The fidelity of chrominance components is improved by using finer quantization step sizes compared to those used for the luminance coefficients, particularly when the luminance coefficients are coarsely quantized.

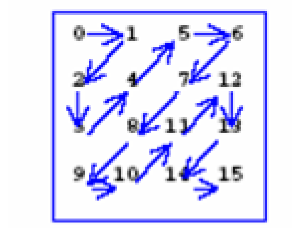

Figure.1Zig-Zag Scanning of Coefficients

C. Motion Factor

Motion plays an important role in human perception of video due to effects such as visual perception of speed and direction of motion, visual tracking of moving objects and motion saliency. Human visual perceptions are very sensitive to the motion information in the video, considerable resources in the HVS are devoted to motion perception. Unfortunately, motion information usually suffer from serious distortion due to the inadequate inter frame prediction technique and coarse quantization. Since motion shows great influence in human visual perception, it is reasonable to take motion into consideration when measuring the objective video quality [6], [9].

It is believed that object motion is associated with visual attention and can be used for predicting visual fixations. This is intuitively sensible because statistically, most of the objects in the visual scene are static (or close to static) relative to the background. As a result, an object with significant motion relative to the background would be a strong surprise to the visual system. If the HVS is an optimal information ex-tractor, as discussed, then it should pay attention to such a surprising event. This intuitive idea can be converted into a mathematical measure if the prior distribution about the speed of motion is known.

Conventional VQM examines mismatches between original and distorted video pixel by pixel, and estimates the artifacts caused by motion distortion such as blockiness, blur, and mosaic. But this method is not feasible in the compressed domain.

D. Bit allocation factor

Bit rate control scheme (RC) has been adopted in most networked related applications. Due to the limitation of the bandwidth, the bitstreams that can be used for the transmission of videos are restricted. Researches on the topic of RC which aims to achieve best video quality with limited bitstream, have been carried out for decades. Many ROI based bit allocation algorithms have also been employed in video compression applications. Thus, it is necessary to take the influence of bit allocation into consideration when applying VQA.

III. ALGORITHMS AND EXPERIMENT DESIGN

A. *Flowchart*

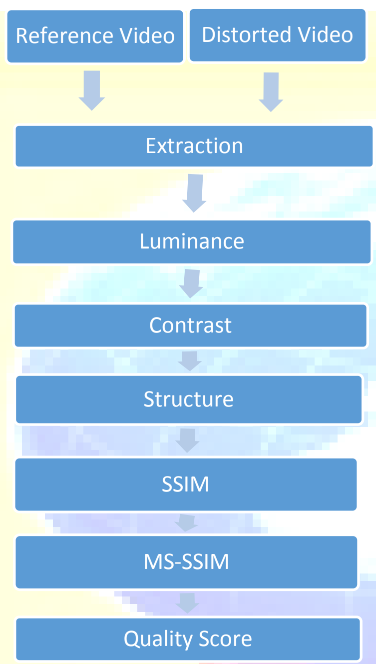

Fig.2 Flow chart for Quality Score

B. *SSIM*

Objective methods for assessing perceptual image quality traditionally attempted to quantify the visibility of errors (differences) between a distorted image and a reference image using a variety of known properties of the human visual system. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, we introduce an alternative complementary framework for quality assessment based on the degradation of structural information.

Natural image signals are highly structured: their pixels exhibit strong dependencies, especially when they are spatially proximate, and these dependencies carry important information about the structure of the objects in the visual scene.

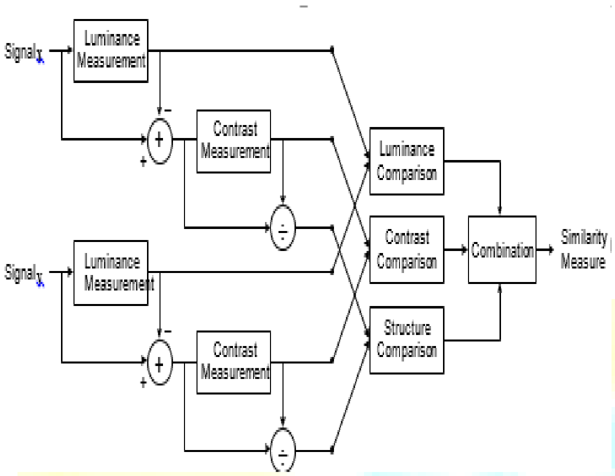

Fig.3 Block diagram of SSIM

This new framework for the design of image quality measures was proposed, based on the assumption that the human visual system is highly adapted to extract structural information from the viewing field. It follows that a measure of structural information change can provide a good approximation to perceived image distortion.

First, the error sensitivity approach estimates perceived errors to quantify image degradations, while the new philosophy considers image degradations as perceived changes in structural information variation. A motivating example is shown in Fig. 3, where the original “Boat” image is altered with different distortions, each adjusted to yield nearly identical MSE relative to the original image. Despite this, the images can be seen to have drastically different perceptual quality [12].

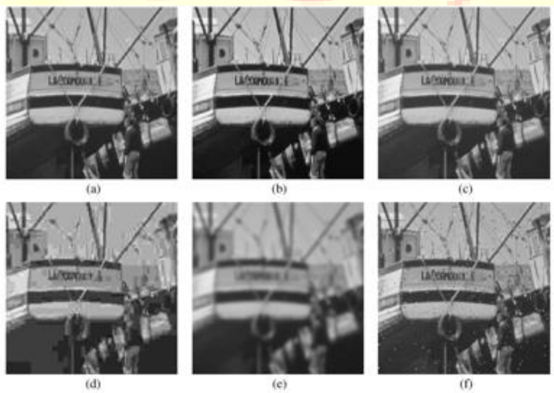

Fig.4Comparison of “Boat” images with different types of distortions, all with MSE = 210. (a) Original image (8 bits/pixel; cropped from 512 _ 512 to 256_ 256 for visibility). (b) Contrast-stretched image, MSSIM = 0:9168. (c) Mean-shifted image, MSSIM = 0:9900. (d) JPEG compressed image, MSSIM =0:6949. (e) Blurred image, MSSIM = 0:7052. (f) Salt-pepper impulsive noise contaminated image, MSSIM = 0:7748.

1) Luminance

The luminance of the surface of an object being observed is the product of the illumination and the reflectance, but the structures of the objects in the scene are independent of the illumination. Consequently, to explore the structural information in an image, we wish to separate the influence of the illumination. We define the structural information in an image as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast. Since luminance and contrast can vary across a scene, we use the local luminance and contrast for our definition.

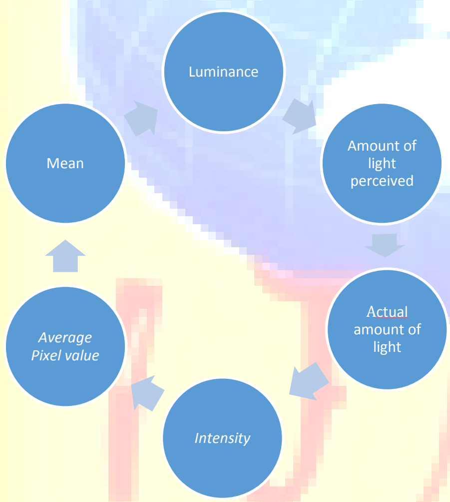

Fig.5 Luminance

2) Contrast Comparison

For comparing contrast we need to remove the effect of luminance which is done by removing the mean intensity from the signal. Indiscrete form, the resulting signal corresponds to the projection of vector onto the hyperplane. The statistical tool used for this purpose is standard deviation.

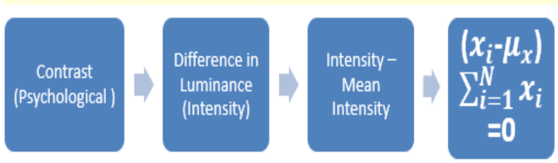

Fig.6 Contrast Comparison

Standard Deviation

$$\boldsymbol{\sigma_x = \left[\frac{1}{N-1}\sum_{i=1}^{N} w_i\,(x_i-\mu_x)^2\right]^{\frac{1}{2}}}$$

Contrast Comparison Term

$$\mathrm{C}(x,y)=\left(\frac{2\sigma_x\sigma_y+c2}{\sigma_x{}^2+\sigma_y{}^2+c2}\right)$$

3) Structure Comparison

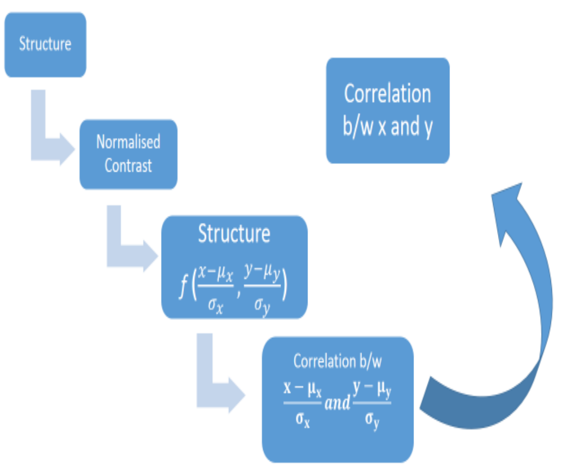

Fig.7 Structure Comparison

For the purpose of structure comparison the signal is normalized (divided) by its own standard deviation, so that the two signals being compared have unit standard deviation. The structure comparison is conducted on these normalized signals.

$$S(x,y)=\frac{\sigma_{xy}+c3}{\sigma_x\sigma_y+c3}$$

Where $c3 = c2/2$

4) Overall SSIM

SSIM (x,y)=*f* {*l* (x,y), *c* (x,y), *s*(x,y) }

$$\mathrm{SSIM}=\left[l\,(\mathrm{x,\,y})^{\alpha}\right]\left[c\,(\mathrm{x,\,y})^{\beta}\right][s\,(\mathrm{x,\,y})^{\gamma}]$$

$$\text{SSIM}=\frac{(2\mu_x\mu_y+c1)(\sigma_{xy}+c2)}{(\mu_x{}^2+\mu_y{}^2+c1)(\sigma_x{}^2+\sigma_y{}^2+c2)}$$

Where $\alpha = \beta = \gamma = 1$ for a normal case.

5) Properties of SSIM

Symmetry : S(x,y) = S(y,x)

Boundedness : S(x,y) ≤ 1

Unique maximum : S(x,y) = 1 if and only if x= y

(In discrete representations x_i= y_i, for all i= 1, 2……N).

C. Multiscale SSIM

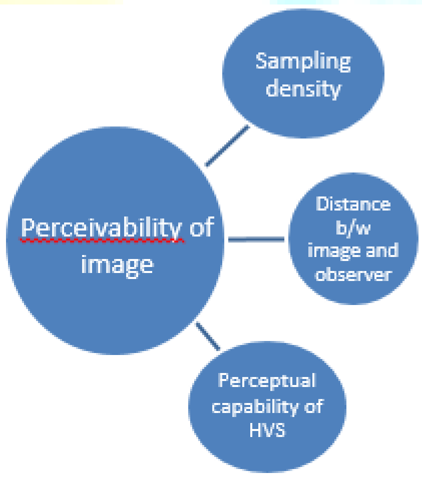

Fig.8 MS-SSIM

There is a drawback of single scale SSIM [13] because the right scale depends on viewing conditions (e.g., display resolution and viewing distance). The perceivability of image details depends the sampling density of the image signal, the distance from the image plane to the observer and the perceptual capability of the observer's visual system. Multi-scale method is a convenient way to incorporate image details at different resolutions.

A multi-scale SSIM[14] method for image quality assessment whose system diagram is illustrated in Fig. 9. Taking the reference and distorted image signals as the input, the system iteratively applies a low-pass filter and downsamples the filtered image by a factor of 2. We index the original image as Scale 1, and the highest scale as Scale M, which is obtained after M - 1 iterations. At the j-th scale, the contrast comparison and the structure comparison are calculated and denoted as $c_j(x; y)$ and $s_j(x; y)$, respectively. The luminance comparison is computed only at Scale M and is denoted as $l_M(x,y)$. The overall SSIM evaluation is obtained by combining the measurement at different scales.

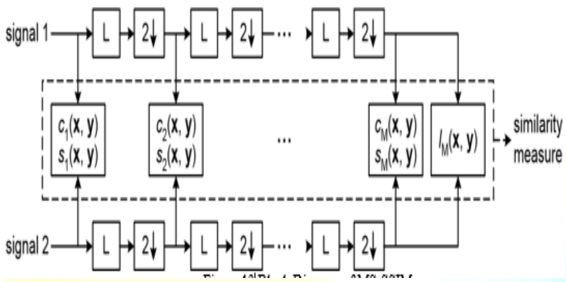

Fig.9 Block Diagram of MS-SSIM

$$MS - SSIM = [l_M(x,y)]^{\alpha_M} . \prod_{j=1}^{M} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}$$

Where scaling ranges from 1 to M by M-1 iterations

Original image is at Scale 1

$[c_j(x,y)]$-Contrast comparison at j^{th} scale

$[s_j(x,y)]$-Structure comparison at j^{th} scale

$[l_M(x,y)]$-Luminance comparison at scale M

The exponents α_M, β_j and γ_j are used to adjust the relative importance of different components. It also includes the single-scale method as a special case. In particular, a single-scale implementation for Scale M applies the iterative filtering and downsampling procedure up to Scale M and only the exponents α_M, β_j and γ_j are given nonzero values.

D. F-SSIM

Fast Structural Similarity Index is an advance version of SSIM. Algorithms used in F-SSIM are different and much efficient than the one used in SSIM. These algorithms make computational complexity quite easy and give result in lesser time as compared to SSIM.

F-SSIM focuses on these three parameters:

- Luminance Term
- Contrast

- Structure

1) Luminance

The luminance term of the SSIM index often plays a less significant perceptual role in predicting visual quality that the other terms. They propose eliminating it to reduce complexity. We choose to preserve the luminance term since images may suffer from a luminance bias, even if image quality databases do not explicitly include such distortions. Nevertheless, we have sought to expend as little computation as possible on the luminance term.

The luminance term in Fast SSIM[13] utilizes an 8x8 square window, and an integral image technique [9] to compute the luminance similarity between the reference and test images. By utilizing the so-called integral image, extracting the mean value of the pixels within a square window can be made quite efficient. As shown in Fig. 10, the value of the integral image at (x, y) is the sum of the pixels values above and to the left of (x, y), and including the value at (x, y).

Computing the sum over any rectangular area can be achieved with only two additions and one subtraction. As shown in Fig. 10, the sum of the pixel values within the rectangle *D* can be computed using four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle *A*. The value at location 2 is *A*+*B*, at location 3 is *A*+*C*, and at location 4 is *A*+*B*+*C*+*D*. The sum over region *D* can be computed as '4'+'1'-('2'+'3') where '*i*' is the value of the integral image at location *i*.

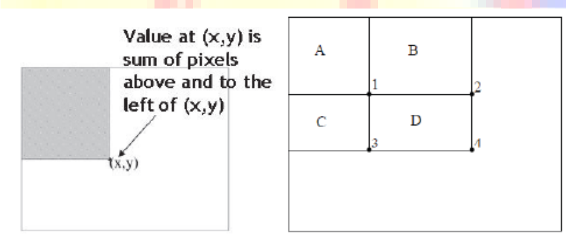

Fig.10Left: Integral image. Right: How to compute sum value over region D in integral image domain.

Using the integral image [9] and a square window, the complexity of computing the luminance term is reduced considerably. Assuming the window size is *n*x*n*, the standard SSIM index algorithm (using a Gaussian weighted window) requires *n*2 multiplies and (*n*2-1) additions to calculate the mean value, while the proposed Fast SSIM algorithm only requires 3 additions and 1 subtraction.

2) Contrast and structure

The computation of the variance term is the most time consuming part of the SSIM algorithm. In order to lower the complexity, we substitute a gradient value in Fast SSIM. Following Field, while images of real-world scenes vary greatly in their absolute luma and chroma distributions, the gradient magnitudes of natural images generally obey heavy tailed distribution laws. Indeed, some no-reference image quality assessment algorithms, use the gradient image to assess blur severity. Similarly, the performance of the Gradient-based SSIM index suggests that applying SSIM on the gradient magnitude may yield higher performance. The gradient is certainly responsive to image variation. Moreover, the gradient magnitude has low complexity and is amenable to integer-only implementation.

We generate the gradient image using the Roberts gradient templates depicted in Fig. 11.

$$\begin{array}{|c|c|}\hline -1 & 0 \\ \hline 0 & 1 \\ \hline \end{array} \qquad \begin{array}{|c|c|}\hline 0 & -1 \\ \hline 1 & 0 \\ \hline \end{array}$$

Fig.11Roberts gradient templates

The gradient magnitude is approximated by1 $|\nabla I| = max\{|\nabla i|, |\nabla j|\} + \left(\frac{1}{4}\right) min\{|\nabla i|, |\nabla j|\}$ where $\nabla i$and $\nabla j$are the Roberts template responses in the two orthogonal directions. This approximation is based upon a simple expansion of the gradient. The contrast $c(x, y)$ and structure $s(x, y)$ terms of the Fast SSIM index algorithm are then defined:

$$C(x,y)=\frac{2\mu_{Gx}\mu_{Gy}+C_2}{\mu_{Gx}^2+\mu_{Gy}^2+C_2}$$

$$S(x,y)=\frac{\mu_{GxGy}+C2}{\mu_{Gx}\mu_{Gy}+C3}$$

Where$C3$ = $C2/2$, and

$$\mu_{Gx} = \frac{1}{N}\sum_{i=1}^{N}|\nabla x_i|$$

$$\mu_{GxGy} = \frac{1}{N}\sum_{i=1}^{N} |\nabla x_i|\,|\nabla y_i|$$

Where $|\nabla x_i|$ and $|\nabla y_i|$ are the gradient magnitude values of the images x and y at spatial coordinate i, estimated using the approximation (1).

The Fast SSIM index between signals x and y is then:

$$\text{Fast-SSIM}(x, y) = \frac{(2\mu_x\mu_y + c1)(\mu_{GxGy} + c_2)}{(\mu_x{}^2 + \mu_y{}^2 + c1)(\mu_{Gx}^2 + \mu_{Gy}^2 + c_2)}$$

IV. RESULT ANALYSES

Analysis was done on LIVE Video Quality Database. A short description of these videos is provided below.

Pedestrian Area Riverbed

Rush hour Tractor

Fig.12Video sequences for training

Four video sequences with different characteristics are chosen in the training process as shown in Fig. 12. They are sampled to a resolution of 768x432 which is consistent with the videos in LIVE Video Quality Database.

MS-SSIM, SSIM, PSNR, MSE and DMOS scores were calculated for these videos and tabulated.

Sr. No	Distortion / IQA	DMOS	MS-SSIM	SSIM	PSNR	MSE
1.	Wireless distortions	44.5104	0.797978	0.998889	6.252178	16.28709
2.	Wireless distortions	70.1054	0.574352	0.996848	10.39341	13.96664
3.	Wireless distortions	66.428	0.33078	0.989768	11.86711	13.75331
4.	Wireless distortions	75.1225	0.176835	0.982677	19.83012	11.78982
5.	IP distortions	73.8803	0.340103	0.994883	12.7244	13.285
6.	IP distortions	63.2564	0.597315	0.996216	8.336473	16.59171
7.	IP distortions	61.2726	0.512388	0.995265	9.912946	15.52273
8.	H.264 compression	40.5551	0.796976	0.999048	6.913062	15.81791
9.	H.264 compression	52.6111	0.669677	0.99829	10.33947	14.05714
10.	H.264 compression	60.2534	0.532117	0.997368	13.6004	12.89104
11.	H.264 compression	68.7186	0.40137	0.996342	16.27181	12.13322
12.	MPEG-2 compression	42.9784	0.894602	0.999571	5.823224	16.58137
13.	MPEG-2 compression	51.053	0.858183	0.999375	7.290719	15.61924
14.	MPEG-2 compression	55.702	0.8184	0.999144	8.686877	14.8663
15.	MPEG-2 compression	65.6457	0.722791	0.998515	11.46535	13.70714

Fig.13 Experiment results on LIVE VIDEO QUALITY DATABASE for pedestrian Area

Sr. No	Distortion / IQA	DMOS	MS-SSIM	SSIM	PSNR	MSE
1.	Wireless distortions	64.9369	0.357316	0.993687	27.54094	9.690943
2.	Wireless distortions	46.2446	0.219974	0.996904	24.44269	10.20044
3.	Wireless distortions	54.3732	0.302617	0.997225	22.01568	10.65758
4.	Wireless distortions	46.4907	0.490326	0.997756	20.0903	11.17547
5.	IP distortions	68.1064	0.33919	0.990321	30.30481	9.741432
6.	IP distortions	54.8101	0.197748	0.996582	22.02622	10.72334
7.	IP distortions	54.6555	0.158978	0.994094	20.23671	11.60694
8.	H.264 compression	39.1978	0.557808	0.99858	21.1269	10.83985
9.	H.264 compression	43.6833	0.35307	0.997795	26.18134	9.915574
10.	H.264 compression	55.8563	0.046302	0.996205	33.44916	8.841012
11.	H.264 compression	63.5809	0.336824	0.99501	37.48975	8.381902
12.	MPEG-2 compression	38.8828	0.722791	0.999191	19.33589	11.31533
13.	MPEG-2 compression	45.6069	0.54328	0.998551	25.44657	10.1116
14.	MPEG-2 compression	48.0089	0.47192	0.998283	27.40144	9.751106
15.	MPEG-2 compression	47.527	0.47192	0.998283	27.40144	9.751106

Fig.14 Experiment results on LIVE VIDEO QUALITY DATABASE for Riverbed

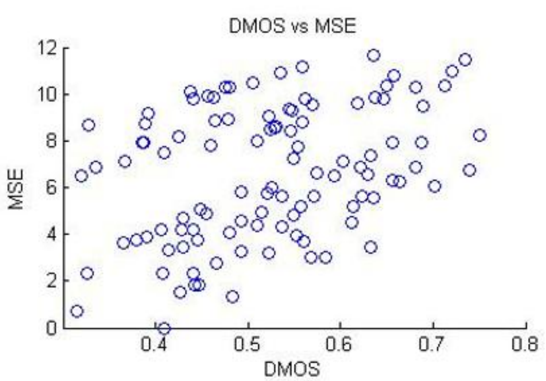

Fig.15 Scatter plots of MSE vs DMOS

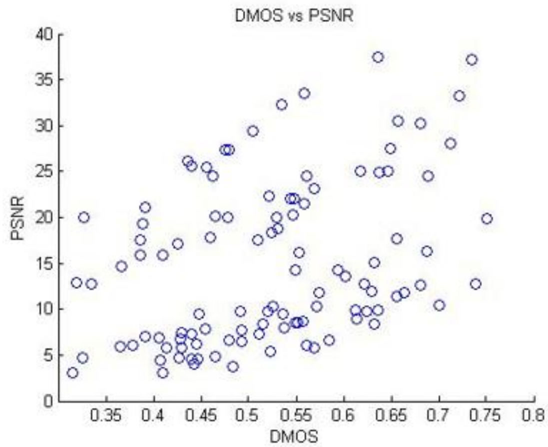

Fig.16 Scatter plots of PSNR vs DMOS

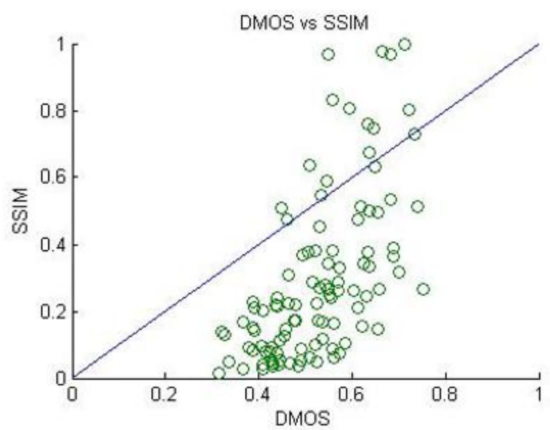

Fig.17 Scatter plots of SSIM vs DMOS

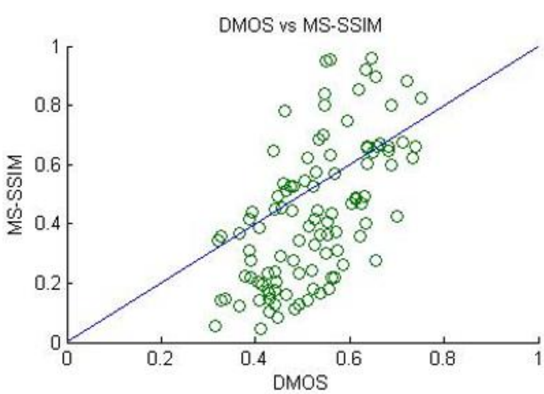

Fig.18 Scatter plots of MS-SSIM vs DMOS

V. CONCLUSION

On comparison of DMOS vs various algorithms like MSE, PSNR, SSIM and MS-SSIM, it is confirmed that MS-SSIM is the best mathematical model in terms of similarity to DMOS. The performance of the algorithm is evaluated by employing both the public database and private encoded bitstreams. The experimental results demonstrate that, in comparing with other leading VQA algorithms, the proposed method i.e. MS-SSIM shows not only higher accuracy, but also lower computation complexity. Besides, the success achieved on bitstreams encoded with various configurations also reflects a good generality.The experiment results also show that the proposed algorithm has better performance than many conventional methods in terms of both the prediction accuracy and the computation complexity**.** All of the advantages suggest that the algorithm can be well utilized in practical applications.

VI. REFERENCES

[1] Xiangyu Lin, Hanjie Ma, Lei Luo and Yaowu Chen "No-Reference Video Quality Assessment in the Compressed Domain,"*IEEE Tran. On Consumer Electronics*, vol. 58, no.2, pp. 505-512, 2012.

[2] M.C.Q. Farias, M. Carli, and S. K. Mitra, "Objective Video QualityMetric based on Data Hiding," *IEEE Trans. Consumer Electron.*, vol.51, no. 3, pp. 983-992, Aug. 2005

[3] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[4] K. Seshadrinathan and A. C. Bovik, "Motion-based Perceptual Quality Assessment of Video," *Proc. SPIE - Human Vision and Electronic Imaging*, 2009.

[5] A. R. Reibman, V. A. Vaishampayan, and Y. Sermadevi, "Quality Monitoring of Video Over a Packet Network," *IEEE Trans. Multimedia*,vol. 6, no. 2, pp. 327–334, Apr. 2004.

[6] D. Culibrk, M. Mirkovic, V. Zloklica, M. Pokric, V. Crnojevic, and D.Kukolk, "Salient Motion Features for Video Quality Assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp.948-958, Apr.2011.

[7] K. Yang, G. Dane, and K. El-Maleh, "Temporal Quality Evaluation for Enhancing compressed Video," *Proc. International Conference on Computer Communications and Networks*, pp. 1160-1165, 2007.

[8] Eden, "No-Reference Estimation of the Coding PSNR for H.264-Coded Sequences," *IEEE Trans. Consumer Electron.*, vol. 53, no. 2, pp.667-674, Aug. 2007

[9] Xiangyu Lin, Hanjie Ma, Lei Luo and Yaowu Chen "No-Reference Video Quality Assessment in the Compressed Domain," *IEEE Transactions on Consumer Electronics*, Vol. 58, No. 2, May 2012.

[10] Alan C. Bovik, *Handbook of The Essential Guide to Video Processing.* Academic Press, 2000.

[11] David Salomon, G. Motta, D. Bryant,*Data Compression: The complete Reference.* Springer,2006.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli,"Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Processing*, vol. 13, Jan. 2004

[13] Zhou Wang, Eero P. Simoncelli and Alan C. Bovik," Multi-scale structural similarity for image quality assessment," *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA*, Nov. 9-12, 2003. ©IEEE

[14] 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.

[15] 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.