Objective Video Quality Assessment
N.Krishna, M. Sivasubramanyam
Department of Electronics and Communication engineering
Sree Vidyanikethan Engineering College, Tirupati, 517102, A.P. India

Abstract: This paper proposes an objective quality assessment of videos by combining structural fidelity and statistical naturalness of the frames. Initially input video is converted into the number of frames and then calculate the mean of the quality of these frames, the frames are converted into gray scale images. Validations using an independent subject rated image, database shows good correlations between subjective ranking scores and proposed tone mapped quality index (TMQI). The results are carried out by MATLAB/SIMULINK software.

Index terms-high dynamic range image, image fusion, video quality assessment, naturalness, statistical naturalness, structural similarity, tone mapping operator.

I. INTRODUCTION

Our obvious way to implement video quality metric is to apply a still image quality assessment metric on a frame by frame basis. However, a more efficient approach would model the temporal aspects of the human visual system in the design of the metric. Then implemented a video distortion meter by using video quality assessment meter followed by a “cognitive emulator” that temporal effect such as smoothing and temporal masking of the frame quality measure, saturation and symmetric tracking, Van den Braden Lambright eat has extended the HVS modelling into the time dimension by modelling the temporal dimension of the CSF, and generating by two visual streams turned to different temporal aspects of the stimulus from the output of each spatial channel Watson’s digital video quality (DVQ) metric operates in the DCT domain and is therefore more attractive from an implementation point of view since the DCT is efficient to implement and most video coding standards are based on the DCT[7].

Here calculate the Objective Quality assessment of videos is followed by using main referenced paper is “Objective Quality assessment of Tone mapped images”, so used same procedure to calculate the quality of the frames in the input video. An overview and a subjective comparison of 8 TMOs were reported in [2]. HDR capable monitor was employed into compare 6TMOs in a subjective experiment using a paired comparison method. In 14 subjects were asked to rate two architectural interior scenes produced by 7 TMOs based on basic image attributes as well as the naturalness of the LDR images. A more comprehensive subjective evaluation was carried out in where tone mapped images generated by 14 TMOs were shown to 2 groups of 10 human observers to rate LDR images. Our work is inspired by the success of two design principles in IQA literature. The first is the structural similarity (SSIM) approach [11] and its multi-scale derivations [8], [9]. Here we propose a method that combines a multi-scale structural fidelity measure and a statistical naturalness measure, leading to Tone mapped Quality index (TMQI).

Moreover, we demonstrate that TMQI can be employed for optimizing parameters in TMOs and for adaptively fusing multiple tone mapped images.

A. HIGH DYNAMIC RANGE IMAGE (HDR)

High dynamic range image (HDRI or HDR) is a set of techniques used in imaging and photography to reproduce a greater dynamic range of luminosity than possible using standard digital imaging or photographic techniques. HDR images can represent more accurately the range of intensity levels found in real scenes, from direct sunlight to faint starlight, and is often captured by way of a plurality of differently exposed pictures of the same subject matter. Because in the display devices are not capable to display the high dynamic range images, so HDR images are converted into the Low dynamic range images in this paper.

Non HDR cameras means Low Dynamic range cameras take photographs with a limited exposure range, resulting in the loss of detail in bright or dark areas and some blur in that images. Here some loss occurred in LDR cameras so this loss of details are capturing multiple photographs at different exposure levels and adding them to produce a photograph representative of a broader tonal range.

HDR images can also be getting using special image sensors, like oversampled binary image sensor. Tone mapping methods, which decreases total contrast to make possible display of high dynamic range images on particular devices with lower dynamic range, can be applied to produce image with conserved or overstated local contrast for artistic effect.

B. Tone Mapping

Tone mapping is a technique used in image processing and computer graphics to map one set of colours to another in order to approximate the appearance of high dynamic range images in a medium that has a more limited dynamic range. More dynamic range means high contrast and brightness of the images. Some of the computer print outs and CRT, LCD monitors are the limited dynamic range values only doesn’t exceed the limited range of the dynamic range. Different methods of tone mapping operators have been developed in the previous years. They all can be divided in two main types Global (or spatiality uniform) operators and Local (or spatial varying) operators. The effect of the algorithm changes in each pixel according to the local features of the image. These algorithms are high difficulty than the global operators. They can show halo effect and ringing and the output can look unrealistic, but they can provide the best performance, since human vision is mainly sensitive to local contrast.
\[ V_{out} = \frac{V_{in}}{V_{in} + 1} \]  

A simple example of global tone mapping filter is where \( V_{in} \) is the luminance of the original pixel and \( V_{out} \) is the luminance of the filtered pixel. This function will map the luminance \( V_{in} \) in the domain (0, \( \infty \)) to a displayable output range of (0,1) while this filter provides a decent contrast for parts of the image with low luminance (particularly when \( V_{in} < 1 \), parts of the image with higher luminance will get increasingly lower contrast as the luminance of the filtered image goes to ‘1’. Those tone mapping methods are frequently produce very sharp image, which protect very well small contrast details; however, this is occurred at the cost of obliteration on total image contrast, and may be a side effect produce halo like glows a surroundings dark objects. Examples of these tone mapping methods are include: gradient domain HDR image compression and a perceptual frame work for contrast processing of high dynamic range images [3]. This is the one of application of frame work for tone mapping.

Another approach to tone mapping of HDR image is inspired by the anchoring theory of lightness perception. This theory explains many kind of the human visual system such as lightness constancy and its failures which are important in the perception of images. The reference of this tone mapping method is a decay of an HDR image into areas (frameworks) of consistent illumination and the local calculation of the lightness values. The total lightness of an image is calculated by merging the frameworks are proportionally to their strength. Mainly important is the anchoring relating of the luminance to a known luminance, namely estimating which luminance value is perceived as white in the scene. This method is to tone mapping does not affect the local contrast and preserves the natural colours of an HDR image due to the linear conduct of luminance. One example form of tone mapping takes a standard

Examples of the tone mapping images by gamma exposur

![Image 1](http://www.ijettjournal.org)  

Fig.1. tone mapped HDR image of Dundas square; tone mapping was done as post processing technique, using photo metric software.

![Image 2](http://www.ijettjournal.org)  

Fig.2. the six individual exposures used to create the previous image, in the low exposure images, the room is dark and undecided, but the details of the windows are visible, in the high exposure images, the windows bright and unclear, but the details of the room are revealed.

II. A PERCEPTUAL IMAGE QUALITY ASSESSMENT METRIC STRUCTURAL SIMILARITY (SSIM)

This paper is nearly reference taken from structural similarity index (SSIM) this represents perceptual image quality based on the structural information. SSIM is an objective image quality method and is better to simple understanding quantitative measure such as MSE and PSNR values this image quality assessment and illustrates its validity in terms of human visual perception. It will be very useful to understand the SSIM and its applications by reviewing this paper and videos extension.

A general form of SSIM is

\[
SSIM(x, y) = \left[ I(x, y) \right]^\alpha [C(x, y)]^\beta [S(x, y)]^\gamma 
\]

[Note that \( \alpha > 0, \beta > 0 \) \& \( \gamma > 0 \) these parameters are used to adjust the relative importance of the three components, Where \( x, y \) are image patches].

\[
I(x, y) = \frac{2 \mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{3}
\]

\[
s(x, y) = \frac{\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{4}
\]

\[
c(x, y) = \frac{2 \sigma_{xy} + C_3}{\sigma_x^2 + \sigma_y^2 + C_3}
\]

\( I(x, y) \) is luminance comparison (eq.6, \( C(x, y) \) is contrast comparison of (eq. 9), and \( s(x, y) \) is structural comparison (eq. 10) \( C_1, C_2, C_3 \) are constants, \( \mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy} \) are defined in above in the reference paper. Gaussian weight function has the form

\[
\exp \left( -\frac{n_1^2 + n_2^2}{2\sigma^2} \right)
\]

\( n_1, n_2 = 1, 2, \cdots, 11 \)

\[ \frac{1}{\sqrt{N}}\sum_{x=1}^{N} x, \frac{1}{\sqrt{N}}\sum_{y=1}^{N} y \]

\[ \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2 \]

\[ \frac{1}{N-1}\sum_{i=1}^{N} (y_i - \bar{y})^2 \]

\[ \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \]

(a) Set \( \alpha, \beta, \alpha, \beta, \gamma, C_1, C_2, C_3 \) As in the paper and apply your function to the different distorted Lena images (512*512) with the same mean square error (MSE).

(b) Fix \( C_1, C_2, C_3 \) as in (b) and set \( \alpha \) as 1. Choose any 5 pairs of \( \beta, \gamma \) for example \( \beta = 1, \gamma = 2 \), \( \beta = 1, \gamma = 3 \), \( \beta = 1, \gamma = 4 \), \( \beta = 2, \gamma = 1 \), \( \beta = 3, \gamma = 1 \) and apply the function to the distorted.
III. QUALITY ASSESSMENT METHOD

Here first extract the input video into the frames and then work out individual quality of these frames. And the frames are taken here like an image, that’s why the quality of video is calculated by using quality of an image, and then calculate the mean of these frames gives the quality of the video. So the basic concepts are same as the quality assessment of an image, here structural fidelity only does not suffice to give an overall quality evaluation of the image. So another parameter is statistical naturalness of the same image. When a good quality tone mapped image has to get a good arbitration are getting these factors like structural fidelity and statistical naturalness, these are sometimes competing factors of an LDR image [6].

a) STRUCTURAL FIDELITY

The structural similarity method (SSIM) gives a useful design viewpoint as well as a practical method for measuring structural fidelities between two images [12]. This method is the simplest to calculate the local map of the tone mapping HDR images. Here define our local structural fidelity measured by using below formula

\[ S'_{\text{local}}(x, y) = \frac{2\sigma_x'\sigma_y' + c_1}{\sigma_x'^2 + \sigma_y'^2 + c_1} \]  

(5)

Where \(\sigma_x,\sigma_y\) and \(\sigma_{x'}\), \(\sigma_{y'}\) are the two corresponding patches in HDR and LDR images of local standard deviations and cross correlation between these two images correspondingly, and \(C_1\) and \(C_2\) are helpful stabilizing constants. Here Compared with the SSIM [13], psychometric helpfulness is known as Galton’s method gives [13], which takes the form of an increasing normal distribution function given by

\[ p(s) = \frac{1}{\sqrt{2 \pi \theta}} \int_{-\infty}^{\sigma} \exp \left[ -\frac{(x - \tau_s)^2}{2 \sigma_s^2} \right] dx \]  

(6)

Where \(\psi\), \(s\) are the detection probability density and amplitude of the sinusoidal stimulus respectively, \(\tau_s\) is the modulation threshold, and \(\theta_s\) is the standard deviation of the normal distribution, it Controls variation the slope of detection probability density function. It was established that these ratio given by

\[ k = \frac{\tau_s}{\theta_s} \]  

(7)

This is an approximately a stable constant, it is known as Crosier’s law, Typical values of \(k\) ranges between 2.3 and 4, and \(k = 3\) makes the possibility of false alarm significantly small value [21].

\[ A(f) \approx 2.6[0.0912 + 0.114f] \exp \left[ -0.114f \right] \]  

(8)

Where \(f\) denotes spatial frequency of the SSIM, this function is normalize to have maximum value is 1, and therefore only provides relative understanding across the frequency spectrum. Adding this with above equation then obtain threshold value

\[ \tau_s(f) = \frac{1}{\lambda A(f)} \]  

(9)

The above threshold value is calculated by using the contrast sensitivity dimension assuming accurate sinusoidal incentive. The threshold value is defined as signal standard deviation can be computed as \(1\)

\[ \tau_s(f) = \frac{\mu}{\sqrt{2\lambda A(f)}} \]  

(10)

Where \(\mu\) is based on “Crosier’s law” the mean intensity value [7], [9], we have mean value is

\[ \theta_s(f) = \frac{\tau_s(f)}{k} \]  

(11)

Here define the mapping between \(\sigma\) and \(\overline{\sigma}\) as

\[ \sigma' = \frac{1}{\sqrt{2 \pi \theta}} \int_{-\infty}^{\sigma} \exp \left[ -\frac{(x - \tau_s)^2}{2 \sigma_s^2} \right] dx \]  

(12)

In (1), \(\sigma\) and \(\sigma'\) are the intensity and mean values mapped versions of \(\sigma\) and \(\sigma_s\), respectively. Here bounded between 0 and 1, where 0 and \(1\) is representing completely irrelevant and completely important signal strengths, respectively. At each scale of the mapping is mutual by averaging to provide a single score of the two images are HDR and LDR respectively.

\[ S_i = \frac{1}{N} \sum_{i=1}^{N} S_{\text{local}}(x_i, y_i) \]  

(13)

Where \(x_i\) and \(y_i\) are the \(i^{th}\) patches in the HDR and LDR images are individual compared respectively, where \(N\) is the number of patches in the \(i^{th}\) scale. In this philosophy, highly developed pooling strategy such as information content based pooling of the HDR image [11] have been shown to improve the performance of IQA algorithms. This is only method get the accurate sampling of the HDR and LDR image.

![Fig 3: Framework of multi-scale structural fidelity assessment.](image)

However, in our present experiment, these advanced pooling methods did not give result in notable performance gain in the proposed structural fidelity measure. The overall structural fidelity is calculated by adding the scale level structural fidelity scores using the above procedure in [1]

\[ S = \prod_{i=1}^{L} S_{\beta_i} \]  

(14)

Where \(\beta_i\) is the weight assigned to the \(i^{th}\) scale and \(L\) is the total number of scales. Here the performances of our structural fidelity model have several parameters in

b) Statistical Naturalness

A high quality tone mapped LDR image should not only faithfully preserve the structural fidelity of the HDR image, but
also look natural. So we need the calculation of the particular LDR image naturalness; nevertheless, but the naturalness is a subjective quantity that is difficult to define quantitatively, because the human perception of ranking of the particular image is different to each others. To calculate the mean and standard deviations which are useful measures that reflects the global intensity and contrast of an image. We found that these histograms can be well fitted using a Gaussian and a Beta probability density functions given by

\[ p_m(m) = \frac{1}{\sqrt{2\pi\sigma_m}} \exp \left[ -\frac{m - \mu_m}{2\sigma_m^2} \right] \]  

And

\[ p_d(d) = \frac{(1 - d)^{\beta_d - 1} d^{\alpha_d - 1}}{B(\alpha_d, \beta_d)} \]  

Where \( \cdot \) defines the Beta function. Here the model parameters are calculated by regression, and the best values we found are \( \mu_m = 115.94 \) and \( \sigma_m = 27.99 \) in (11), and \( \alpha_d = 4.4 \) and \( \beta_d = 10.1 \) in (12), respectively. These are getting the gamma exposure of the adobe photo shop. And we may get histogram levels of the dynamic range of an input image; in present research propose that contrast and brightness are more independent patches in terms of both biological computation and natural images [7]. Here the result of their joint probability function should be the product of the two brightness and contrast as follows. Then our statistical naturalness is define as

\[ N = \frac{1}{K} p_m p_d \]  

where \( K \) is the normalization factor is given by \( K = \max \{ p_m, p_d \} \).

This normalization factor gives the naturalness measure to be bounded between ‘0’ and ‘1’. And where ‘m’ and ‘d’ are the mean (fitted by Gaussian PDF) and standard deviation (fitted by Beta PDF).

c) Quality Assessment Model

Given a tone mapped LDR image or frames now we have two available measurements, structural fidelity ‘S’ and naturalness ‘N’, these are given by previous equations respectively. These two factors equations can be used individually or jointly as a 2D vector that characterizes different aspects of the quality of the LDR image, the total tone mapped image (frame) Quality index TMQI is measured as following formula,

\[ Q = \alpha S^\alpha + (1 - \alpha) N^\beta \]  

Where \( 0 < \alpha < 1 \) defines the relative weights assigned to the two aspects, and \( \alpha \) and \( \beta \) defines the sensitivities of the two parameters respectively. From both structural fidelity and statistical naturalness are upper-bounded by ‘1’, this overall quality measure is also Upper-bounded by ‘1’. The parameters in (14) are left to determined, ‘\( \alpha \)’ and ‘\( \beta \)’, in our project implementation; they are mentioned to accurate results of subjective validations by utilizing machine understanding techniques described next. In the end, where \( \alpha = 0.8038 \), \( \beta = 0.3968 \) and \( \beta = 0.8094 \) as our final model parameters respectively.

Quality of the video is calculated by calculate the each frame quality of the video, and then find the mean of the total number of frames in the particular video. Here the run times increased by the number of frames are increase in input of the device, the quality of the frame is same as the calculation of the tone map image in the above quality assessment method, so here follows the same procedure to calculate the quality of the frames for the particular input video.

Directly calculation of the video quality is complicated, so follow this procedure to calculate the objective quality of the videos. Here mention the number of frames in the input so the quality of the video is only at that particular frames only not for reaming frames.

Quality score of the video = mean (Q) (19)

Where Q is the quality of the individual frames of the video,

Here below example clearly mention the quality of the 10 frames of the input video, so the quality of the 10 frames video have been Quality is same as the mean of the 10 frames of the video. This method is the one of the simplest procedure to calculate objective quality assessment of the video, because here only extract the frames from the input video and then calculate the quality of the video. Now we calculate the quality of video by using the “objective quality assessment of tone mapped images” method, here we use same tone mapping and structural fidelity and statistical naturalness of frames. This is the simplest procedure to calculate the objective quality of the video.

IV. RESULT ANALYSIS

![Fig 4: Above frames are extracted from the input video](image)
“Quality of the individual frames are respectively Q= 0.8353 0.8351 0.8344 0.8350 0.8363 0.8382 0.8390 0.8405 0.8411, with default values, are a=0.8012, a=0.3046, β=0.7088. So quality score of the above frames is Q5=0.8372.

V. VALIDATION OF QUALITY ASSESSMENT METRICS

Here we mention two validation methods for validating the quality of the video is accurate or not, so we need to say the project is superior or not that’s why the following methods are given the evaluation of the assessment values. That validation process is done by comparing our objective quality assessment results with the subjective data values. Here two validation methods are given below.

1. Spearman’s rank order correlation coefficient (SROCC) is given as

\[ \text{SROCC} = \frac{\sum_{i=1}^{n} d_i^2}{N(N^2-1)} \]  

(20)

Here \(d_i\) is the represents the difference between the \(i\)th image’s ranks in subjective and objective evaluations respectively. SROCC is a correlation metric based on a non parametric rank order, independent of non linear mapping between subjective and objective scores.

2. Kendall’s rank order correlation coefficient (KROCC) is another rank correlation based non parametric method, which can be calculated using

\[ \text{KROCC} = \frac{N_c - N_d}{0.5N(N-1)} \]  

(21)

Here ‘\(N_c\)’ and ‘\(N_d\)’ denotes the number of concordant and discordant pairs in the data set respectively.

VI. CONCLUSION

We develop the Objective video quality assessment by using mean of the objective quality assessment of images (or frames) of input video, here quality score of the frames are nothing but the quality of the video. Our experiment show that reasonably correlated by TMQI with evaluation subject values of frames, the current procedure is applicable and tested using natural videos only. The application scope is HDR videos and colour videos.

REFERENCES


Authors’ profile

N. Krishna, has received B.Tech degree in Electronics and Communication Engineering, and present pursuing Master’s degree in Digital Electronics and Communication Systems under Sree Vidyanikethan Engineering College, Tirupati. His current research interests are Image processing, Objective Quality Assessment of mapped images, and Objective video Quality assessment.

M. Sivasubramaniam has received the Bachelor’s degree, B.Tech (ECE) from JNTUH and Master degree M.Tech (DECS) from the University of JNTUA. He is currently working as Assistant Professor, Department of ECE, Sree Vidyanikethan Engineering college, Tirupati, India. He has teaching experience in the domain of Electromagnetics, Antennas and Microwaves around 8 years. The work was carried under his supervision out of his interest in image processing domain.