

Video Quality Assessment Method: MD-SSIM

Woei-Tan Loh and David B.L. Bong

Faculty of Engineering Universiti Malaysia Sarawak, 94300, Kota Samarahan, Malaysia

Abstract-- In this paper, video quality assessment (VQA) for compression losses is the main focus. A new method, MD-SSIM (Mean Squared Error Difference SSIM) is used for detecting the spatial distortion. For the temporal part, the differences of the SSIM scores of each frame are used to form the quality scores. In addition, this method has higher computational speed and competent performance as compared to other existing algorithms.

I. INTRODUCTION

Video comprises the spatial and temporal effects which can affect the quality of the video as observed by the human visual system (HVS) [1]. Most of the video quality assessment (VQA) methods are direct implementation of image quality assessment (IQA) methods on each video frame such as the structural similarity index (SSIM) [2] and multi-scale SSIM (MSSIM) [3]. Some IQA or VQA emphasizes only on a specific type of distortion in videos or images, for example the IQA method proposed by Bong and Khoo [4] which emphasizes on images which are corrupted with Gaussian white noise and impulse noise. In this paper, an objective VQA, Mean Squared Error Difference SSIM (MD-SSIM) is proposed to measure the quality of the videos which are distorted by compression error.

II. PROPOSED VQA

MD-SSIM consists of two parts, i.e. local and global quality. This is inspired by the work of Junyong et al. [5]. For the local part, it further divided into spatial and temporal parts. Mean Squared Error (MSE) is used to produce the error map in the spatial part which is then used to weight the SSIM map produced by SSIM index. For the temporal part, the difference of global quality is used to weight the spatial quality of each frame. This approach is inspired by the work of Narwaria *et al.* [6] Lastly, the local quality and global quality are averaged to obtain the overall video quality.

SSIM map measures the similarity between the same frame of original and distorted videos. Mathematically, SSIM function is defined as [2]

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (1)$$

The mean of image x and image y is denoted as μ_x and μ_y respectively, whereas σ_x and σ_y are the standard deviations of x and y respectively. Parameter σ_{xy} is the cross-correlation of

This work was supported by Ministry of Higher Education Malaysia for the provision of research grant: RACE/c(2)/1109/2014(17) and Faculty of Engineering, Universiti Malaysia Sarawak for the provision of research facilities.

mean shifted images and C_i for $i=1,2$ are positive constant values. For the error map, MSE is used. In mathematics, MSE can be written as [7]

$$MSE(x, y) = [R(x, y) - D(x, y)]^2, \quad (2)$$

where $R(x, y)$ is the reference image and $D(x, y)$ is the distorted image. For the spatial part, SSIM maps are weighted by the error map. Mathematically this is shown as

$$MDSSIM(spatial) = \frac{\sum_x \sum_y E(x, y) S(x, y)}{\sum_x \sum_y E(x, y)}. \quad (3)$$

$E(x, y)$ represents the error map while $S(x, y)$ represents SSIM map. Instead of taking the averages of the quality scores of all frames, weightage is given to each frame of the video. The local MD-SSIM is represented as

$$MDSSIM(local) = \frac{\sum_f^F W_f MDSSIM(spatial)}{\sum_f^F W_f}, \quad (4)$$

where f is the frame number and F is the total number of frames in the video. For the global part, SSIM index is applied on frame basis and is averaged. This is shown as

$$MDSSIM(global) = \frac{1}{F} \sum_f^F SSIM(x, y). \quad (5)$$

Overall, MD-SSIM is mathematically presented as:

$$MDSSIM = \frac{1}{2} [MDSSIM(local) + MDSSIM(global)]. \quad (6)$$

III. RESULTS AND DISCUSSION

MD-SSIM has been tested with the videos from LIVE video database [8] [9] and CSIQ video database [10]. For LIVE video database, there are 2 categories of compression error which are H.264 compression and Motion Joint Photographic Experts Group (MPEG) 2 compression. For each category, there are 4 levels of distortions for each reference videos. For each compression, 12868 frames are being tested. For CSIQ database, 3 videos are distorted by H.264/AVC (Advanced Video Coding) compression, 3 videos are distorted by MJPEG compression, 3 videos distorted by wavelet compression and 3 videos distorted from HEVC (High Efficiency Video Coding) compression for each reference video. Altogether there are 13710 images are being tested for each compression.

MD-SSIM is compared with MSE/ peak signal to noise ratio (PSNR), SSIM, MSSIM, motion-based video integrity evaluation (MOVIE) [12], and ViS₃ [10]. MOVIE and ViS₃ are applied with their default implementation and settings. The other methods are applied on a frame by frame basis.

In this paper, two correlation coefficients as recommended by VQEG [11] are used, i.e. Spearman rank-order correlation coefficient (SROCC) and Pearson linear correlation coefficient (CC). For CC, a 3 parameter logistic function, which transform the quality scores to raw predicted scores, is