Temporal Video Quality Assessment Method Involving Structural Similarity Index

Woei-Tan Loh and David B.L. Bong

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

Abstract—**In this paper, a video quality assessment (VQA) that focuses on temporal part is being proposed. It is based on Structural Similarity Index (SSIM) and the work of previous researchers where differences of the frames are being measured. The proposed VQA possesses lower computational complexity and acceptable performance compared to state-of-the-art VQA methods.**

I. INTRODUCTION

Image quality assessment (IQA) involves spatial part of a frame. Meanwhile, video quality assessment (VQA) further involves temporal aspect of a video. This is because a video consists of three dimensions. Video quality can be measured through direct implementing of IQA over the frames. The video quality as measured by this method ignores the temporal effect in the video. Some existing popular IQA method are mean squared error (MSE) [1] [2], structural similarity (SSIM) index [3] and multi-scale structural similarity (MSSIM) [4]. There are also IQA or VQA methods that focus on certain types of distortions such as blur distortion [5]. This method measures the quality of images that are distorted by Gaussian blur noise. In this paper, a new temporal VQA method is proposed which involved the use of SSIM. It could measure video quality well with lower computational complexity than other VQA methods.

II. PROPOSED VQA

The spatial part of the proposed VQA involves SSIM. SSIM measures the similarity between reference and distorted frames. It is defined as [3]

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

The parameter μ is the mean of a frame while parameter σ with one subscript denotes the standard deviation of particular frame. The parameter σ with two subscripts is the cross correlation of two frames. Parameters C_i where $i = 1,2$ are constants to prevent instability. The spatial aspect of the proposed VQA method is the result of applying SSIM over all frames.

$$spatial = \frac{1}{N} \cdot \sum_{i=1}^{N} SSIM(f_{ri}, f_{di}), \quad (2)$$

where N represents the sum of frames in a video. Parameter f_{ri} is a particular frame in the reference video while f_{di} denotes the same frame from the distorted video.

The temporal aspect of the proposed method involves modification of the methods as found in [6]. The authors had suggested that two differences can be computed to measure video quality from the temporal aspect. The first difference is the difference of the current frame from distorted video (d_n), and the previous frame from reference video (r_{n-1}). The next difference is the variance between current and previous frames of reference video, (r_n and r_{n-1}). Temporal video quality is the result of applying MSSIM to the two differences. This is mathematically represented as:

$$tmp = \frac{1}{N-1} \cdot \sum_{i=1}^{N} MSSIM(d_n - r_{n-1}, r_n - r_{n-1}). \quad (3)$$

In this paper, the method for Equation (3) is improved. First, SSIM is utilized to replace MSSIM. Secondly, instead of measuring the differences of current and previous frames, the differences of current and previous two frames are also measured. This is mathematically shown as:

$$tmp1 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} SSIM(d_i - r_{i-1}, r_i - r_{i-1}), \quad (4)$$

$$tmp2 = \frac{1}{N-2} \cdot \sum_{i=2}^{N} SSIM(d_i - r_{i-2}, r_i - r_{i-2}), \quad (5)$$

$$Temporal = \frac{1}{2} \cdot (tmp1 + tmp2). \quad (6)$$

Parameter N is the overall amount of frames in a video while i denotes a particular frame. The overall video quality can be measured by arithmetic mean of the spatial and temporal video quality. This is expressed as:

$$Video\ quality = \frac{1}{2} \cdot (spatial + temporal). \quad (5)$$

III. RESULTS AND DISCUSSION

The proposed VQA method are verified with a benchmark database, LIVE video database [7] [8]. In this database, there are four categories of video distortions. Two of them are compression errors; H.264 compression error and Motion Joint Photographic Experts Group (MPEG)-2 compression error. The other two types of distortions are wireless distortion and IP distortion. Each distortion is further divided into four levels. A total of 12,868 frames are being tested for each error except for the IP distortion which has only 9,651 frames. Overall, there are 48,255frames that have been tested.

The proposed VQA method is compared with other well-known IQA and VQA methods. Existing IQA methods are MSE / Peak Signal to Noise Ratio (PSNR), SSIM, and MSSIM. Existing VQA methods are MOtion-based Video Integrity Evaluation (MOVIE) [9] and ViS$_3$ [10]. All VQA methods are evaluated with the proposed implementation whereas IQA methods are applied on a frame basis.