Assessment of Color Video Quality with Singular Value Decomposition of Complex Matrix

Fuqiang Zhang, Junli Li, Gang Chen
The College of Information Science & Engineering
Ningbo University
Ningbo, China
e-mail:www.kevin.cn@gmail.com

Jiaju Man
Hunan Normal University
Changsha, China

Abstract—Most proposed video quality assessment approaches in the literature are only applied to the luminance layer of frame. Others are applied to luminance and chrominance layers respectively and then pooling by means of grayscale image quality metrics. In this paper, we proposed a full-reference objective video quality assessment called CMSVD, which treats luminance and chrominance parts as a whole and considers multiple factors of the human visual system (HVS) and video features. The algorithm is tested on the video quality experts group (VQEG) phase I FR-TV test data set. Experiments show that it has good correlation with perceived video quality.

Keywords: objective video quality assessment; complex matrix; singular value decomposition (SVD); human visual system (HVS); video quality experts group (VQEG)

I. INTRODUCTION

Up to date, a number of objective video quality assessment algorithms have been proposed to predict human perception of video quality. Currently, the most commonly used full-reference (FR) objective quality metrics are mean squared error (MSE) and peak signal-to-noise ratio (PSNR) due to its simplicity and mathematical convenience [1]. However, they have been widely criticized as well for not correlating well with perceived quality measurement, because they ignore the fact that the end-user of the video sequence is a human observer. Most of the research has focused on methods that attempt to simulate the functionality of the relevant early human visual system (HVS) components. Examples of video quality metrics based on the HVS-based philosophy include the Digital Video Quality (DVQ) metric[2] and the Perceptual Distortion Model (PDM)[3]. Although such HVS-based methods have met with good success, there are many questions that arise in their design such as complexity, suprathreshold problem and inaccurate modeling of the HVS. Some researchers have therefore proposed the Structural Similarity (SSIM) metric[4] and Visual Information Fidelity (VIF)[5] based on overarching principles such as structure or information extraction. These metrics are not affected by assumptions about HVS models, and have also met with good success.

Up to now, most of color video quality assessments are generally applied to its luminance layer and discard important chrominance information. Others assess quality of luminance and chrominance layers respectively and then pooling through the grayscale image quality assessment methods and ignore the correlation between luminance component and color components. It is obvious that the number of color that can be distinguished by the eye is much more than that of the grayscale level, so color information is important to assessment of color video quality. In this paper, we proposed the complex model for color image which treat luminance and chrominance as a whole. The luminance layer of color image is taken as the real part of a complex number, then the chrominance information of the color image are encoded into the imaginary part of the complex number.

The rest of the paper is organized as follows. In section 2, we give a brief introduction about SVD and quaternion model for color image; in section 3, we describe a new color video quality assessment called CMSVD; in section 4, we compare the test results of different quality assessment models tested on the VQEG Phase I FR-TV video data set; finally, Section 5 draws conclusions and provides further discussions.

II. SINGULAR VALUE DECOMPOSITION AND QUATERNIONS

A. Singular Value Decomposition (SVD)

Suppose A is an m-by-n matrix whose entries come from the field K, which is either the field of real numbers or the field of complex numbers. Then there exists a factorization of the form \( A = U S V^T \), where \( U \) is an m-by-m unitary matrix over K, the matrix \( S \) is m-by-n diagonal matrix with nonnegative numbers on the diagonal, and \( V^T \) denotes the conjugate transpose of \( V \), an n-by-n unitary matrix over K. Such a factorization is called singular value decomposition (SVD) of \( A \).

Aleksandr proposed an SVD-Based Gray-Scale Image Quality Measure called M_SVD[6] which divide grayscale image into \( n \times n \) blocks, then each block is considered to be a matrix with an integer number corresponding to each pixel.
B. Quaternions

A quaternion \( q \) is made of one real and three imaginary parts:
\[
q = a + bi + cj + dk
\]
where \( i, j, k \) and \( k \) obey the rules as below:
\[
i^2 = j^2 = k^2 = -1 \quad i \cdot j = k \quad j \cdot k = i \quad k \cdot i = j \quad j \cdot i = -k \quad k \cdot j = -i \quad i \cdot k = -j.
\]

According to quaternion model for color image, the Red, Green and Blue values of each pixel of a color image are represented as a single pure quaternion valued pixel. A \( N \times M \) color image is then represented as a pure quaternion image:
\[
S(x, y) = r(x, y)i + g(x, y)j + b(x, y)k
\]
where \( r(x, y), g(x, y) \) and \( b(x, y) \) are respectively the red, green and blue components of the pixel at position \((x, y)\) in the image \( S(x, y) \).

F. Zhang [7] proved the existence of SVD of any quaternion matrix. So the quaternion singular value decomposition (QSVD) can be extended to color image processing without separating the color image into three channel images.

III. ASSESSMENT OF COLOR VIDEO QUALITY

Video is composed of a sequence of still images representing scenes in motion. We must consider the differences between image and video in format and the characteristics of the HVS, when image quality metrics are extended to assess the video quality, especially color video quality. Our method consists of four steps as follows.

A. Complex Matrix Composition

Different from color image, The YCbCr color space and its variations (sometimes referred to as YUV) is a popular way of efficiently representing color images in video. In terms of quaternion model, a color image also can be represented as below:
\[
S(x, y) = Y(x, y)i + Cb(x, y)j + Cr(x, y)k
\]

Because the human visual system (HVS) is less sensitive to color than to luminance (brightness), the luminance component is more important than the chrominance components in video quality assessment field. Besides, MPEG-4 Visual and H.264 support three sampling patterns for Y, Cb and Cr, among which 4:2:0 and 4:2:2 are widely-used formats[8]. In these formats, the luminance(Y) component have higher resolution than the chrominance components(Cb and Cr), so we adopt a complex model to represent a color image. Different from quaternion model, the Y, Cb (or Cr) values of each pixel of a color image are represented as a single complex valued pixel in complex model. A \( N \times M \) color image is then represented as a complex image:
\[
S(x, y) = Y(x, y) + C(x, y)j
\]

where \( Y(x, y) \) and \( C(x, y) \) are respectively the luminance(Y) and chrominance(Cb or Cr) components of the pixel at position \((x, y)\) in the image \( S(x, y) \).

For simplicity, the original and distorted images are divided into 8 x 8 non-overlapping blocks each of which is considered to be a matrix with a complex number corresponding to each pixel.

B. Block Measure and Weighted

The M_SVD algorithm introduced in [6] compute distortion in every block using equation:
\[
D_i = \sqrt{\sum_{j=1}^{n}(s_j - \hat{s}_j)^2}
\]
where \( s_j \) and \( \hat{s}_j \) are the singular values of the original block \( i \) and the distorted block \( i \) respectively, and \( n \) is the block size, which gives equal importance to distortion in every block.

However, HVS researchers find that human visual perception is characterized by a variable resolution across the field of vision because of the uneven distribution of photoreceptors on the retina, which implies that every block plays different role in quality assessment. Therefore, we compute entropy in each block, which denote the degree of visual interest of regions within the distorted image using equation:
\[
\hat{e}_i = -\sum_{j=0}^{L-1} p(z_j) \log_2 p(z_j)
\]
where \( z_j \) is a random variable indicating intensity, \( p(z_j) \) is the histogram of the intensity levels in block \( i \), \( L \) is the number of possible intensity levels.

Equation(5) must be substituted with equation as below:
\[
D_i = \hat{e}_i \cdot \sqrt{\sum_{j=1}^{n}(s_j - \hat{s}_j)^2}
\]
where the block \( i \) is weighted with the corresponding degree of interest of the region.

C. Frame Measure

As mentioned in [6], the distortion of the image in video can be measured using equation as below:
\[
CMSVD_f = \frac{\sum_{i=1}^{H/n}(W/n)|D_i - D_{mid}|}{(H/n)\times(W/n)}
\]
where \( D_{mid} \) represents the midpoint of the sorted \( D_i \) s, \( H \) and \( W \) are the height and width of the image respectively, and \( n \) is the block size.

D. Video Measure

The overall quality of the entire video sequence is commonly given by:
\[
CMSVD_{AVG} = \frac{\sum_{f=1}^{F} CMSVD_f}{F}
\]
where \( F \) is the number of frames.
Some experiments show that when most of the frames in a video sequence have high quality, but only a few are damaged and have extremely low quality, which is called burst-of-error[4]. The observers tend to give a lower quality score than averaging all the frames, so we use the max of the CMSVDs as the overall quality of the distorted video.

\[
CMSVD_{\text{MAX}} = \max(CMSVD_1, CMSVD_2...CMSVD_j)
\]  

IV. EXPERIMENTAL RESULTS

The VQEG Phase I test dataset for FR-TV video quality assessment [9] is public and commonly be used to test the objective video quality assessment system. We follow the performance evaluation procedures employed in the VQEG Phase I FR-TV test[10] to provide quantitative measures on the performance of the objective quality assessment models. Four metrics are employed. Many detailed information about these metrics can be found in [10].

Fig.1 (a) ~ (i) show the scatter plots of the subjective/objective comparisons on all test video sequences(625@50Hz) given by PSNR, SSIM, M_SVD((5),(8),(9)), QSVD_RGB((2),(5),(8),(9)), QSVD_YCbCr ((3),(5),(8),(9)), CMSVD(No weighted)_AVG((4),(5),(8),(9)), CMSVD(No weighted)_MAX((4),(5),(8),(10)),CMSVD(Weighted)_AVG ((4),(6),(7),(8),(9)),CMSVD(Weighted)_MAX((4),(6),(7),(8),(10)), respectively. It also needs to be mentioned that PSNR, SSIM and M_SVD are separate the luminance information from the color information, and use the luminance layer only. As a visual illustration of the relationship between data and model predictions, It is clear that the proposed CMSVD consistent with the subjective scores much better than others.

In Table I, we give the comparison results of the three metrics with the 4-parameter cubic polynomial regression, when all the 625@50Hz test video sequences are included. First, the 4-parameter cubic polynomial functions are used in a fitting procedure to provide a non-linear mapping between the objective/subjective scores. The Pearson linear correlation coefficient after non-linear regression means the correlation degree between each model and DMOS, they provide the prediction accuracy evaluation, and the large OR value means the better performance. We can see that M_SVD performs worse than SSIM, because it ignores important information and the HVS properties. It is obvious that complex model(4) , weighting method with texture ((6)) and approach for the overall quality of the entire video sequence((10)) can all efficiently improve the prediction performance of M_SVD, because these method is based on some researchers’ studying results, e.g. the HVS is more sensitive to luminance than chrominance, and so on.

V. CONCLUSIONS AND DISCUSSIONS

In this paper, we have presented a novel color video quality assessment based on SVD considering some video features, characteristics of the HVS and burst-of-error issue. We validate the performance of our algorithm using The VQEG Phase I test dataset with 170 video sequences, and showed that the proposed method is competitive with the state-of-the-art metrics and outperform them in our simulations. Currently we are continuing to make efforts towards improving CMSVD by developing quaternion which contain structural information.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China under Grant No.60672072 and No.60832003, Zhejiang Provincial Natural Science Foundation of China under Grant No.Y106505, Hunan Provincial Natural Science Foundation of China under Grant No.08JJ3130 and K.C. Wong Magna Fund in NingBo University.

REFERENCES


Figure 1. Scatter plot comparison of different video quality assessment models on VQEG Phase I test dataset (170 625@50Hz test video sequences included). Vertical and horizontal axes are for subjective and objective measurements, respectively. Each sample point represents one test video. (a) PSNR; (b) SSIM; (c) M_SVD; (d) QSVD_RGB; (e) QSVD_YCbCr; (f) CMSVD(No weighted)_AVG; (g) CMSVD(No weighted)_MAX; (h) CMSVD(Weighted)_AVG; (i) CMSVD(Weighted)_MAX.