Journal of Artificial Intelligence in Electrical Engineering, Vol. 2, No. 7, November 2013

A Machine Learning Approach to No-Reference Objective Video Quality Assessment for High Definition Resources

Nacer Farajzadeh[1], Maliheh Mazloumi[2]
[1]Faculty of Information Technology, Azarbaijan Shahid Madani University, Tabriz, Iran
Email: nfarajzadeh@azaruniv.edu (Corresponding author)
[2]Department of Electrical Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran
Email: m-mazloumi@iau-ahar.ac.ir

ABSTRACT

The video quality assessment must be adapted to the human visual system, which is why researchers have performed subjective viewing experiments in order to obtain the conditions of encoding of video systems to provide the best quality to the user. The objective of this study is to assess the video quality using image features extraction without using reference video. RMSE values and processing time of SVR for BMP and JPEG formats in quality assessment were 0.78×10-2, 0.81×10-2, 6.0s and 4.8s, respectively. In this study, a metric system for no-reference assessing the video quality is presented using wavelet transform and generalized Gaussian distribution parameters. Results of ITU-BT tests for each video were used to train SVR and its performance for video frames is evaluated.

KEYWORDS: Video quality assessments, Generalized Gaussian distribution, Wavelet transform.

1. INTRODUCTION

Video Quality Assessment (VQA) is one of the important aspects that affects video acquisition, compression, processing, transmission and reproduction. Transmission of video over distribution systems, such as broadcast or a digital storage of data requires a process of compression and processing in order to offer a suitable and appropriate quality to the human eyes [1]. VQA is done in two methods: Subjective VQA (SVQA) and Objective VQA (OVQA) [2, 3]. In SVQA, video sequences are usually shown to the group of viewers and then their opinion is recorded and averaged to evaluate the quality of each video sequence [4]. OVQA techniques are mathematical models based on criteria and metrics that can be measured objectively and automatically evaluated by a computer program [5].

There are three types of OVQA depending on the presence and availability of a reference image or any of its features to develop the study: Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR). Old metrics designed for digital imaging systems, such as Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are utilized in FR OVQA [6]. This error sensitivity based approach needs a number of assumptions such as: reference signal is of perfect quality; Light adaptation follows the Weber’s law and so on [7]. NR quality assessment (sometimes called blind quality assessment) is complex due to the fact that many unquantifiable factors play role in human assessment of quality, such as aesthetics, cognitive relevance, learning, visual context, etc., when the reference signal is not available [8].

Only a few methods have been proposed in the literature for NR OVQA [8]. In [9] a new approach is proposed that blindly measures blocking artifacts in images without reference resources. They modeled the blocky image as a non-blocky image interfered with a pure blocky signal. Reference [10] presented a methodology using circular back-propagation artificial neural network to assess objective quality of MPEG video streams. Another study continuously extracted objective features from compressed video streams on a frame-by-frame basis; they feed the network estimating the corresponding perceived quality. Because the frame quality measure is not sufficient for an objective video quality assessment, it is necessary to adjust the frame quality value to the information from contents and motion respectively, such as Spatial Information (SI) and Temporal Information (TI) [11].

Quality metrics commonly model the frequency selective visual stimulus within the constraints of application and computation using wavelet and discrete transforms [8]. A research proposed a metric for quantifying performance of image restoration systems, in which the degradation is modeled as a linear frequency distortion and additive noise injection [12].

The Gaussian distribution is a typical model for signals and noise in many applications in science and engineering. GGD has been proposed for modeling atmospheric noise, sub-band encoding of audio and video signals [13], impulsive noise, and direction of arrival, independent component analysis [14] and blind signal separation [15].

Although Short-Time Fourier Transform (STFT) uses a sliding window to find spectrogram which gives the information of both time and frequency in signal processing, the length of window limits the resolution in frequency. Therefore, wavelet transform seems to be a solution to this problem. Wavelet transforms are based on small wavelets with limited duration. The translated-version wavelets locate whereas the scaled-version wavelets allow us to analyze the signal in different scales [16].

Support Vector Machine (SVM) is a supervised learning model with associated learning algorithms that analyzes data and recognizes patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier [17]. A version of (SVM) for regression was proposed in 1996 called Support Vector Regression (SVR) [17]. The model produced by SVR depends only on a subset of the training data, because the cost function for building the model ignores any training data close to the model prediction (within a threshold ε).

The objective of this study is to assess the video quality using image features extraction without using reference video. In this research, a metric system for NR assessing the high Definition (HD) video quality is presented using wavelet transform and GGD parameters. Results of ITU-BT tests for each video were used to train SVM and its performance for BMP and JPEG formats of video frames is evaluated. The rest of this paper is organized as follows: in the next section, the proposed system is introduced. Section 3 provides experimental results and section 4 presents conclusion of the study.

2. METHODOLOGY

The aim of an objective quality metric is to show a high correlation with the mean subjective scores. The proposed method for NR OVQA is shown in Figure 1. Since the method of assessing the video quality is NR in this study, several noises add to the videos and noisy videos are created from each reference video. The method commences by decoding the video stream. Following decoding, noise is added to the video stream by adding noise to each pixel in an amount correlated to the additive noise of pixels in a prior picture. Thus, in accordance with the present principles, temporal noise correlation aids in determining the additive noise to reduce large frame-to-frame differences, a disadvantage of prior noise additive techniques.

Subjective tests are run in order to establish the rank order of the noisy sequences using a quality scale from 1-10, later normalized to 0.1-1. This study follows the protocol which is basically described in ITU-R Recommendation BT-500 about SVQA. The selection of video contents and the duration of sequences should be determined carefully to do a proper job and to be able to compare with similar tests. Although the SVQA studies offer real results as the response of the observers is collected, this kind of studies are expensive in time and money and they are not always so efficient, because sometimes it depends on the place to elaborate the study and its conditions of lighting or comfort of the user, being able to change a valid result because of an external conditions.

After gathering SVQA test, all frames of considered noisy videos are extracted and stores as BMP and JPEG encoded images with dimension 768×432 pixels to evaluate the performance of proposed method.

To extract the obtained image features, a two-dimensional scaling function, $\varphi(x,y)$ and three two dimensional wavelets, $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$ are usually considered. Each is the product of a non-dimensional scaling function φ and corresponding wavelet ψ. Excluding products that produce one-dimensional results, like $\varphi(x)\psi(x)$, the four remaining products produce the separable scaling function shown in equation (1) [18],

$$\varphi(x,y)=\varphi(x)\varphi(y) \tag{1}$$

and separable directionally sensitive wavelets as equations (2) to (4).

$$\psi^H(x,y)=\psi(x)\varphi(y) \tag{2}$$

$$\psi^V(x,y)=\varphi(x)\psi(y) \tag{3}$$

$$\psi^D(x,y)=\psi(x)\psi(y) \tag{4}$$

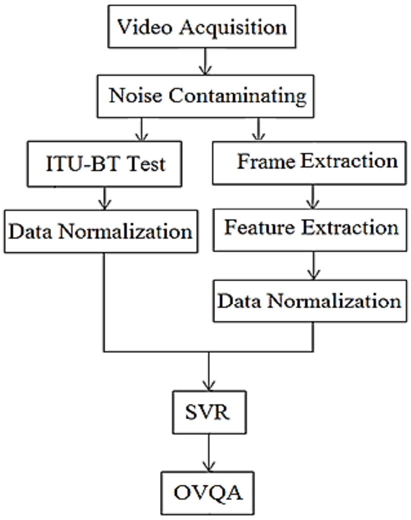

Fig.1. Block diagram of the proposed method

The wavelet transform of function $f(x,y)$ of size $M\times N$ is then can be calculated by equations (5) and (6).

$$W_{\varphi}(j_0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\varphi_{j_0,m,n}(x,y) \tag{5}$$

$$W_{\psi}^{i}(j,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y){\psi^{i}}_{j,m,n}(x,y)$$
$$i=\{H,V,D\} \tag{6}$$

After wavelet transform of each images at three levels, the parameters of GGD probability density function (μ and σ^2) from two-dimensional wavelet transform is calculated using equation (7). A random variable X is distributed as generalized Gaussian if its probability density function is given by equation (7).

$$gg(x,\mu,\sigma,p)=\frac{1}{2\Gamma(1+1/p)A(p,\sigma)}e^{-\left|\frac{x-\mu}{A(p,\sigma)}\right|^{p}},x\in R \tag{7}$$

where:

$$A(p,\sigma)=\left[\frac{\sigma^2\Gamma(1/p)}{\Gamma(3/p)}\right]^{1/2} \tag{8}$$

The parameter μ is the mean, the function $A(p,\sigma)$ is an scaling factor which allows that $Var(X)=\sigma^2$, and p is the shape parameter. By calculating μ and σ^2 for each obtained wavelet, 18 features will be obtained for each image. To predict the SVQA values from input features, SVR is used to evaluate the performance of machine learning techniques in VQA. Similarly to classification problems, a non-linear model is usually required to adequately model data. In the same manner as the non-linear Support Vector Classification (SVC) approach, a non-linear mapping can be used to map the data into a high dimensional feature space where linear regression is performed. The kernel approach is again employed to address the curse of dimensionality. The input dataset of SVM is extracted noisy image features for each image (18 features) and its output dataset is normalized averaged ranks for each noisy video that considered for constitutive images of that video in SVM training and testing.

3. EXPERIMENTAL RESULTS

3.1 Dataset

To carry out this study, 10 HD video sequences including 218 or 501 frames with varied contents and characteristics were considered as resources. First frame of each video is shown figure 2. Fifteen noisy videos were created from each reference video. Thirty persons with ages between 20 to 30 years old were selected to rank all the noisy videos and obtained results were averaged for each video.

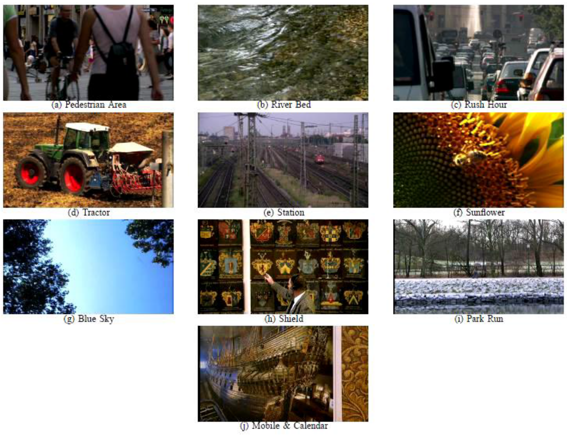

Fig.2. First frame of each HD video resource

It was considered that the obtained images distributed as GGD and then, the parameters of GGD probability density function at three directions (vertical, horizontal and diagonal) and 2D wavelet transform were calculated and normalized to 0.1-1. Table 1 shows normalized extracted image features for two noisy resources shown in Figure 3. As it is shown in Fig. 3, the quality of the images presented in Fig. 3(b) is better than the quality of the images presented in Fig. 3(c).

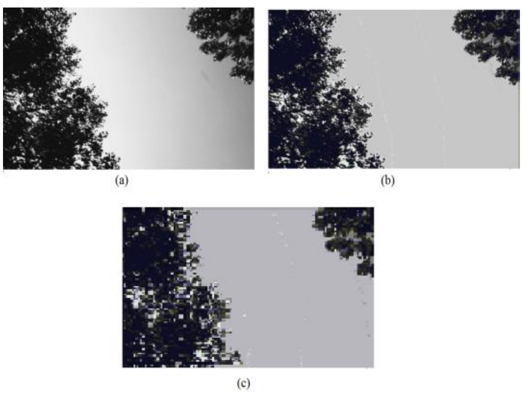

Fig.3. (a) An example of reference BMP image (b) Noisy image from reference image considered as resource no. 1 (c) Noisy image from reference image considered as resource no. 2.

Table 1. Normalized image feature extraction for resources shown in Figure 2.

	Wavelet	Direction	μ	σ^2	SVQA Result (Based on ITU-BT)
Resource No.1	Level 1	Horizontal	0.53	0.66	0.53
		Vertical	0.61	0.35	
		Diagonal	0.32	0.34	
	Level 2	Horizontal	0.39	0.69	
		Vertical	0.50	0.45	
		Diagonal	0.33	0.32	
	Level 3	Horizontal	0.91	0.75	
		Vertical	0.52	0.35	
		Diagonal	0.57	0.74	
Resource No.2	Level 1	Horizontal	0.33	0.45	0.26
		Vertical	0.45	0.67	
		Diagonal	0.83	0.29	
	Level 2	Horizontal	0.37	0.53	
		Vertical	0.42	0.79	
		Diagonal	0.55	0.74	
	Level 3	Horizontal	0.73	0.23	
		Vertical	0.38	0.56	
		Diagonal	0.54	0.34	

3.2 SVR RESULTS

As it is mentioned before, assessing the video quality using image features extraction and SVR is the goal of this research. Since 15 noisy videos were created from each reference video and these videos included 218 or 501 frames, 46850 datasets were prepared to train and test SVR for BMP and JPEG images, separately. SVR program was written using MATLAB and ran on MD101 MacBook Pro (Apple Inc., California, United States). Extracted images from 11 of each 15 noisy video obtained from reference videos were used to train SVR and rest of data was used to test the SVR for BMP and JPEG images, separately. Effects of different types and number of kernels on RMSE values and processing speed of SVR for video quality determination of normalized dataset of BMP images are shown in table 2 considering epsilon equals to 0.01. Table 3 shows the effects of different types and number of kernels on RMSE values and processing speed of SVR for video quality determination of normalized dataset of JPEG images considering epsilon equals to 0.01.

Table 2. Effects of different types and number of kernels on RMSE values and processing speed of SVR for video quality determination of normalized dataset of BMP images (epsilon=0.01)

Kernel type	C=10		C=100		C=1000		C=10000	
	RMSE ($\times10^{-2}$)	Speed (s)	RMSE ($\times10^{-2}$)	Speed (s)	RMSE ($\times10^{-2}$)	Speed (s)	RMSE ($\times10^{-2}$)	Speed (s)
Linear	0.88	3.4	0.80	4.7	0.78	5.9	0.79	6.8
Polynomial	0.85	4.2	0.73	4.8	0.81	5.7	0.82	6.9
Gaussian	0.92	3.9	0.84	5.2	0.72	6.2	0.91	6.3
Sigmoid	0.82	4.9	0.82	5.6	0.79	6.1	0.85	6.1
Average	0.87	4.1	0.80	5.1	0.78	6.0	0.84	6.5

Table 3. Effects of different types and number of kernels on RMSE values and processing speed of SVR for video quality determination of normalized dataset of JPEG images (epsilon=0.01)

Kernel type	C=10		C=100		C=1000		C=10000	
	RMSE ($\times10^{-2}$)	Speed (s)	RMSE ($\times10^{-2}$)	Speed (s)	RMSE ($\times10^{-2}$)	Speed (s)	RMSE ($\times10^{-2}$)	Speed (s)
Linear	0.93	2.9	0.86	3.2	0.87	4.2	0.81	5.1
Polynomial	0.88	3.1	0.82	4.4	0.78	4.9	0.92	4.9
Gaussian	0.83	3.0	0.79	4.3	0.76	4.8	0.99	5.8
Sigmoid	0.94	3.6	0.80	4.9	0.83	5.3	0.83	5.9
Average	0.89	3.2	0.82	4.2	0.81	4.8	0.89	5.4

As seen in Tables 2 and 3, SVR with Gaussian kernel with C and epsilon parameters equal to 1000 and 0.01 had the best performance for OVQA. RMSE values and processing time of SVR for BMP and JPEG formats in quality assessment were 0.78×10^{-2}, 0.81×10^{-2}, 6.0s and 4.8s, respectively. Since the dataset was normalized, these obtained values of RMSE were acceptable. Therefore, proposed method is an appropriate method to assess video quality. Proposed method is attractive not only because of its promising results, but also because of its simplicity. However, more theoretical analysis and subjective experimental work is needed to provide direct evidence on how it is connected with visual perception and natural image statistics. Many other issues may also be considered, such as multi-scale analysis, adaptive windowing and space-variant pooling using a statistical fixation model.

4. CONCLUSION

In this study, an automatic system is proposed for NR OVQA. Wavelet transform and GGD parameters extracted from video frames and results of ITU-BT tests for each video were used to train SVM and its performance for BMP and JPEG formats is evaluated. Experimental results showed that proposed method is capable to determine the quality of video without presence of reference video with an acceptable performance. RMSE values of SVR for BMP and JPEG formats were 0.78×10^{-2} and 0.81×10^{-2}, respectively. However, the JPEG format was less accurate than the BMP format, it was about 80% faster than it. Proposed method can be used for evaluating the video characteristics to provide the observer the best viewing that could expect according the increasing of resolution from standard television to HD or the creation of advanced production of contents systems such as 3-dimensional videos.

REFERENCES

[1] Juan Pedro López Velasco (2012). Video Quality Assessment, Video Compression, Dr. Amal Punchihewa (Ed.), ISBN: 978-953-51-0422-3, InTech.

[2] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," Signal Processing, vol. 70, pp. 177-200, Nov. 1998.

[3] A. M. Eskicioglu and P. S. Fisher, “Image quality measures and their performance,” IEEE Trans Communications, vol. 43, pp. 2959-2965, Dec. 1995.

[4] ITU-R Recommendation BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” Geneva, 2002 (available at www.itu.org).

[5] Y. K. Lai and C.-C. J. Kuo, "A Haar wavelet approach to compressed image quality measurement," Journal of Visual Communication and Image Understanding, vol. 11, pp. 17-40, Mar. 2000.

[6] Z. Wang, A. C. Bovik and L. Lu, “Why is image quality assessment so difficult?” Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc., vol. 4, pp. 3313-3316, May 2002.

[7] J. Lubin, “A visual discrimination model for image system design and evaluation,” in Visual Models for Target Detection and Recognition, E. Peli, ed., pp. 207-220, Singapore: World Scientific Publisher, 1995.

[8] Zhou Wang, Hamid R. Sheikh and Alan C. Bovik, OBJECTIVE VIDEO QUALITY ASSESSMENT, Chapter 41 in The Handbook of Video Databases: Design and Applications, B. Furht and O. Marqure, ed., CRC Press, pp. 1041-1078, September 2003.

[9] Z. Wang, A. C. Bovik and B. L. Evans, “Blind measurement of blocking artifacts in images,” Proc. IEEE Int. Conf. Image Proc., vol. 3, pp. 981-984, Sept. 2000.

[10] P. Gastaldo, S. Rovetta and R. Zunino, “Objective assessment of MPEG-video quality: a neural-network approach,” in Proc. IJCNN, vol. 2, pp. 1432-1437, 2001.

[11] Kawayokeita, Y. ; Horita, Y., NR objective continuous video quality assessment model based on frame quality measure, Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, Year: 2008, Page(s): 385 – 388.

[12] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," IEEE Trans. Image Processing, vol. 4, pp. 636-650, Apr. 2000.

[13] Sharifi, K. and Leon-Garcia, A. (1995). Estimation of shape parameter for generalized Gaussian distribution in subband decomposition of video. IEEE Trans. on Circuits and Systems for Video Technology, Vol 5, No. 1 Feb. 1995 pp. 52-56.

[14] Choi, S. Cichocki, A. and Amari, S. (2000) Local stability analysis of flexible independent component analysis algorithm. Proceedings of 2000 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP2000, Istanbul, Turkey, June 5-9, 2000, pp. 3426-3429.

[15] Wu, H.-C. y Principe, J. (1998). Minimum entropy algorithm for source separation. Proceedings of the 1998 Mindwest symposium on Systems and Circuits.

[16] J. M. Shapiro, “Embedded image coding using zerotrees of wavelet co-efficient,” vol. 41, pp. 3445-3462, Dec. 1993.

[17] Drucker, Harris; Burges, Christopher J. C.; Kaufman, Linda; Smola, Alexander J.; and Vapnik, Vladimir N. (1997); "Support Vector Regression Machines", in Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, MIT Press.

[18] R. C. Gonzalez and R. E. Woods, Digital Image Processing (2nd Edition). Prentice Hall, 2002.