TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 18/19 pp112-116
Volume 13, Number 1, February 2008

Image and Video Quality Assessment Using Neural Network and SVM

DING Wenrui (丁文锐)**, TONG Yubing (佟雨兵), ZHANG Qishan (张其善), YANG Dongkai (杨东凯)

School of Electronic and Information Engineering, Beihang University, Beijing 100083, China

Abstract: An image and video quality assessment method was developed using neural network and support vector machines (SVM) with the peak signal to noise ratio (PSNR) and the structure similarity indexes used to describe image quality. The neural network was used to obtain the mapping functions between the objective quality assessment indexes and subjective quality assessment. The SVM was used to classify the images into different types which were accessed using different mapping functions. Video quality was assessed based on the quality of each frame in the video sequence with various weights to describe motion and scene changes in the video. The number of isolated points in the correlations of the image and video subjective and objective quality assessments was reduced by this method. Simulation results show that the method accurately accesses image quality. The monotonicity of the method for images is 6.94% higher than with the PSNR method, and the root mean square error is at least 35.90% higher than with the PSNR.

Key words: image quality assessment; support vector machines (SVM); neural network; isolated points

Introduction

More and more digital images and videos are being captured, transmitted, and used in computer systems worldwide. Most images and video sequences are compressed with losses before transmission. Therefore, methods are needed to evaluate image quality with the reconstructed images quality directly reflecting the performance of the compression method, much attention has been placed on objective quality assessments which reflect the subjective image and video subjective quality with the main task being to reduce the offset between subjective and objective assessments. Images and video are assessed based on the human visual system (HVS), but the HVS is more complicated than the peak signal to noise ratio (PSNR) and some physiological and psychophysical characteristics of the HVS are not yet well understood. PSNR does not involve much characteristic information concerning the image, so distinct differences occur between the PSNR and the subjective image quality. The structure similarity (SSIM) method accesses the structure similarity between the original signal and the reconstructed signal[1-3]. Here, the images are classified based on neural network (NN) and support vector machines (SVM) first. The PSNR and SSIM are used as indexes describing the image and video sequence quality. Motion and scene changes are described with the mean absolute difference (MAD). Results using the image database of the University of Texas and common video sequences show that the method can accurately assess subjective quality[4].

1 PSNR and SSIM Analysis

PSNR is defined as

$$\mathrm{PSNR} = 10\ \lg\left(255^2 MN / \sum_{i=1}^{M}\sum_{j=1}^{N}[\boldsymbol{a}(i,j) - \hat{\boldsymbol{a}}(i,j)]^2\right) \quad (1)$$

where M represents the number of rows in the image, N represents the number of columns in the image,

Received: 2006-09-05; revised: 2007-09-07
** To whom correspondence should be addressed.
E-mail: ding@buaa.edu.cn; Tel: 86-10-82317392

$\boldsymbol{a}(i,j)$ is the pixel value of the original image, and $\hat{\boldsymbol{a}}(i,j)$ is the pixel value of the distorted image. PSNR is the most common index for assessing signal quality and is valid for most images, but has some weaknesses. For the simplest case, if the following is true,

$$\left|\boldsymbol{a}(i,j)-\boldsymbol{a}_1(\hat{i},j)\right|=\left|\boldsymbol{a}(i,j)-\boldsymbol{a}_2(\hat{i},j)\right| \tag{2}$$

Then

$$\hat{\boldsymbol{a}}_1(i,j)=\hat{\boldsymbol{a}}_2(i,j) \text{ or} \tag{3}$$

$$\hat{\boldsymbol{a}}_1(i,j)+\hat{\boldsymbol{a}}_2(i,j)=2\boldsymbol{a}(i,j) \tag{4}$$

where $\hat{\boldsymbol{a}}_1(i,j)$ and $\hat{\boldsymbol{a}}_2(i,j)$ are two distorted images with the same PSNR in which $\hat{\boldsymbol{a}}_1(i,j)$ and $\hat{\boldsymbol{a}}_2(i,j)$ may not be equal and may even look quite different.

Such error needs to be reduced so that PSNR can be more effective for assessing quality. This analysis focuses on how to reduce the worst case scenarios. SSIM provides a good approximation of the image distortion by measuring structural information changes with the assumption that the HVS is highly adapted to extract structural information from the viewing field. If x and y represent images, then the SSIM is defined as

$$\mathrm{SSIM}(x,y)=[\boldsymbol{l}(x,y)]^{\alpha}[\boldsymbol{c}(x,y)]^{\beta}[\boldsymbol{s}(x,y)]^{\gamma} \tag{5}$$

where $\boldsymbol{l}(x,y)$, $\boldsymbol{c}(x,y)$, and $\boldsymbol{s}(x,y)$ are the luminance, contrast, and structure comparison functions with $\alpha, \beta, \gamma > 0$. $\mathrm{SSIM}(x,y)$ is then a quality assessment index. In reality, the mapping functions between $\mathrm{SSIM}(x,y)$ and $\boldsymbol{l}(x,y)$, $\boldsymbol{c}(x,y)$, and $\boldsymbol{s}(x,y)$ are nonlinear and probably quite different from Eq. (5). As a result, there are points whose objective and subjective assessments differ, which limits the performance of this image quality assessment method. If the assessments of these points were modified and there were less such points, the objective results would be more closely related to the subjective assessments. For this reason SSIM and PSNR are used together here to get a better quality assessment[5,6].

2 Quality Assessments Using Neural Network and SVM

The neural network and SVM methods were combined to set up a new method with two image quality assessment indexes, PSNR and SSIM. The flow chart is shown in Fig. 1.

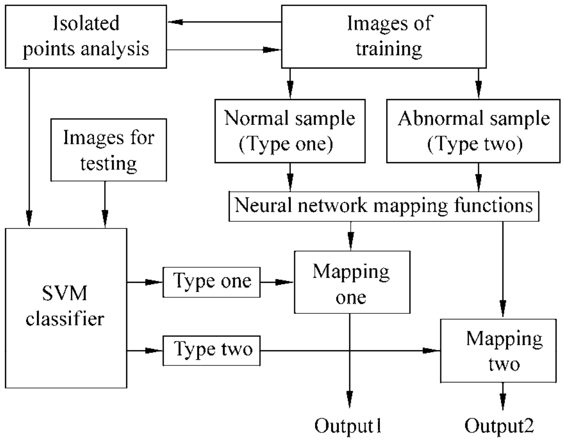


Fig. 1 Image quality assessment method

The image quality assessment is divided into training and testing parts[7]. In the training part, the isolated points were analyzed with the mapping functions between the objective and subjective quality assessments obtained using the neural network. In the testing part, the isolated points were identified and the testing images was then assessed by using the new method.

2.1 Isolated points analysis and mapping functions for assessing image quality

Definition 1 In the correlation of the subjective and objective assessments, if the offset between the assessments is larger than a threshold, V_{T}, then the image corresponding to that point in the curve is defined as an isolated point.

1st evaluation (space distance):

$$\left|\mathrm{MOS}_{\text{subjective}}-\mathrm{EVA}_{\text{objective}}\right| \geqslant V_{\mathrm{T}} \tag{6}$$

2nd evaluation (slope ratio):

$$\begin{cases}(\mathrm{MOS}_{\text{subjective}}/\mathrm{EVA}_{\text{objective}}) \geqslant V_{\mathrm{T}}, \\ \text{or} \\ (\mathrm{MOS}_{\text{subjective}}/\mathrm{EVA}_{\text{objective}}) \leqslant 1/V_{\mathrm{T}}\end{cases} \tag{7}$$

where V_{T} is the threshold which may be derived from tests, $\mathrm{MOS}_{\text{subjective}}$ is the subjective quality assessment, and $\mathrm{EVA}_{\text{objective}}$ is the objective quality assessment. In this section, the 2nd evaluation method is used with the correlation between the subjective and objective assessment shown in Fig. 2.

The isolated points were denoted by “+” in Fig. 2. From the above definition, the training data set was divided into two subsets, a normal set and an abnormal set.

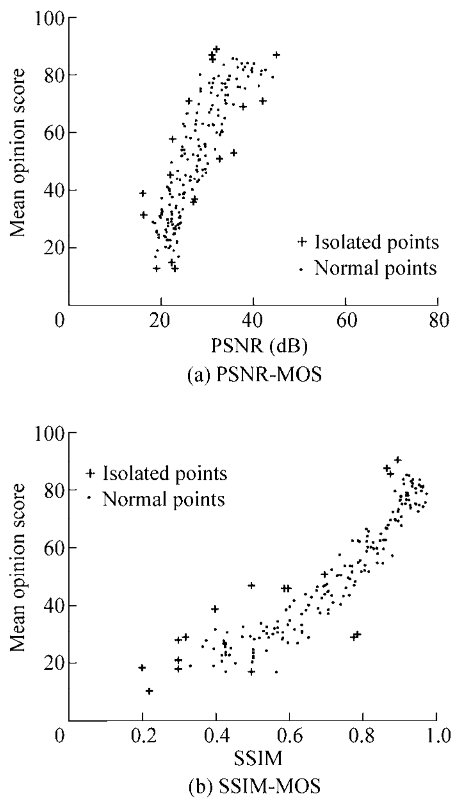


Fig. 2 Correlations of subjective and objective assessments

Then the training subset is used to train the back-propagation (BP) neural to obtain the mappings for the PSNR-MOS count and the SSIM-MOS count. The number of isolated points can also be controlled by adjusting the threshold in definition Eq. (7). Different kinds of images can be assessed by using different mappings according to the structure in Fig. 3.

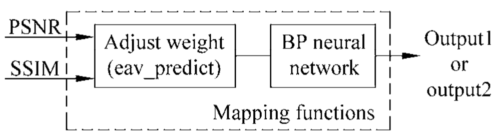


Fig. 3 Mappings for various kinds of image quality assessments

PSNR and SSIM were combined as a weighted sum before using the BP neural network,

$$\text{eav\_predict} = p\text{PSNR} + s\text{SSIM} \tag{8}$$

where p and s are weights derived from experiments.

For normal samples, p=0.4 and s=0.6 . For abnormal samples, p=0.4 and s=0.6. The BP-NN then uses the training data to obtain the final mapping functions by adjusting the weights among the network nodes.

2.2 Isolated points predictions with the SVM classifier

The isolated points were defined using the subjective quality assessment, $\text{MOS}_{\text{subjective}}$. In reality, there are often no subjective assessments for assessing image quality. Therefore, isolated points were identified using an SVM classifier, which classifies samples into two kinds without using $\text{MOS}_{\text{subjective}}$. The SVM uses a kernel function to map the data in the input space to a high-dimensional feature space where the data becomes linearly separable. The SVM determines a generalized optimal classifying plane with the following form,

$$y = f(\boldsymbol{x};a) = \sum_{i=1}^{N}(a_i^* - a_i)(\boldsymbol{x}_i \cdot \boldsymbol{x}) + b = \sum_{i=1}^{N}(a_i^* - a_i)K(\boldsymbol{x},\boldsymbol{x}_i) + b \tag{9}$$

where $K(\cdot,\cdot) = (\boldsymbol{x}_i \cdot \boldsymbol{x})$ is a kernel function satisfying the Mercey theory, b is a constant, and a and a^* are the optimal results of a quadratic programming (QP) problem[8,9]. Here, the SVM classifier used the results of an isolated points analysis. Then the image to be assessed was classified into one of various kinds of images with different classification values. A result of "Positive one" means that image quality is assessed in the normal way. "Negative one" means that abnormal assessment of the image quality is made and isolated points are generated. The distance between these two kinds is the largest spread predicted by the SVM classifier.

3 Video Quality Assessment

Video quality can be assessed using a single image quality assessment with motion information included. Since the video display speed was fixed, the video motion information mainly refers to the changes. The mean absolute difference (MAD) between two continuous video images was used to describe the motion degree[10].

Since the luminance and chroma play different roles in the HVS, in the YUV space luminance and chroma contribute differently to the final value of MAD given by

$$\text{MAD}_i = 0.7\text{MAD}_Y(i) + 0.15\text{MAD}_U(i) + 0.15\text{MAD}_V(i) \tag{10}$$

where $\mathrm{MAD}_Y(i)$ is the MAD value of component Y, $\mathrm{MAD}_U(i)$ is the MAD value of component U and $\mathrm{MAD}_V(i)$ is the MAD value of component V.

Then

$$\beta_i = \mathrm{MAD}_{i+1} / \mathrm{MAD}_i \tag{11}$$

where $i = 1, \ldots, n$. The degree of motion or scene changes, β_i, is used to describe the motion in the video. For example, for Quickchange.yuv and Slowchange.yuv with 4:2:0 video format, the scene changes and motion in Quickchange.yuv were more severe than in Slowchange.yuv. The video motion degree is shown in Fig. 4, where the horizontal axis is the number of frames and the vertical axis is the change degree.

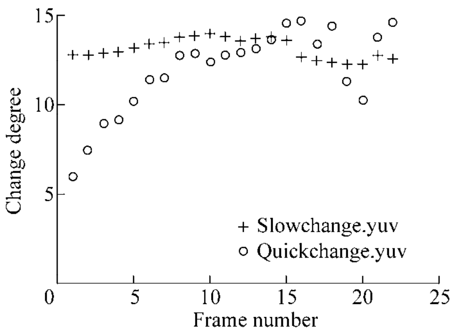


Fig. 4 Degree of motion for two sample videos

Then the video sequence quality is assessed using

$$\mathrm{output}_{\mathrm{video}} = \frac{(\mathrm{output}_1 + \mathrm{output}_2)}{\left(2 + \sum_{i=1}^{n-2} \beta_i\right)} + \frac{\beta_i}{\left(2 + \sum_{i=1}^{n-2} \beta_i\right)} \times \mathrm{output}_{i+2},$$

where i is the sequence number of frame.

4 Results

The correlation between the subjective and objective quality assessments using the present method is shown in Fig. 5.

The results in Fig. 5 show that the number of isolated points is lower in the output-MOS curve using the present method than that in the PSNR-MOS or SSIM-MOS curves. Table 1 lists the statistical data from the Video Quality Experts Group (VQEG) final report[5] used to evaluate the performance of image quality assessment methods with two methods to describe the isolated points. r_{p} is the linear correlation coefficient between the objective assessment value and MOS. Then r_{p} provides an evaluation of the prediction monotonicity. The outlier ratio evaluates the prediction consistency of a method. The root mean square error (RMSE) evaluates the prediction accuracy.

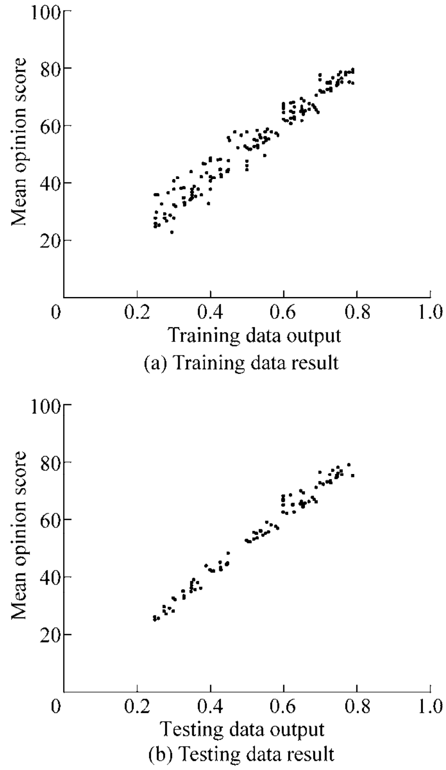


Fig. 5 Correlation of subjective and objective quality assessment

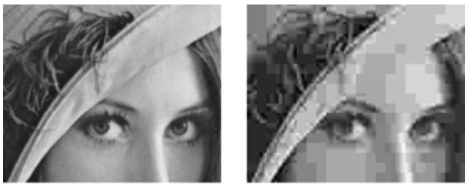
Table 1 Statistical data for image quality assessment

Method	r_{p}	Outlier	RMSE
PSNR	0.8979	0	0.3053
SSIM	0.9601	0.3947	0.4860
Present method	0.9640	0	0.1952

With the same level of prediction consistency, r_{p} (monotonicity) of the present method is 7.42% higher than that of the PSNR method, and the RMSE is 36.06% higher than that of the PSNR.

The objective assessment of the distorted image in Fig. 6 derived from the present method was 0.63 with the PSNR being 28.7 dB and SSIM being 0.98. The highest score was 1 without any loss between the two images. The quality of Fig. 6b is much worse than that of Fig. 6a, but Fig. 6b is still distinguishable. Therefore, the results of the present method are closer to the subjective quality assessment.

Totally 70 quarter common intermediate format (QCIF) video sequences and 25 common intermediate format (CIF) video sequences were used for the video



(a) Original (b) Distorted

Fig. 6 Original and distorted images

quality assessment. Table 2 lists r_p, the outlier ratio, and RMSE for the three methods. The results show that r_p of the present method is 10.47% higher than with the PSNR and the RMSE is 10.48% higher than with the PSNR with the consistency being far better than with the PSNR method.

Table 2 Statistical data for the video quality assessment

Method	r_p	Outlier	RMSE
PSNR	0.6559	0.6000	0.4912
SSIM	0.7075	0.2421	0.3238
Present method	0.7246	0.0316	0.3864

5 Conclusions

The concept of isolated points was developed to better define image quality. The isolated points analysis was used to develop a neural network and SVM based image quality assessment method using PSNR and SSIM as two image quality indexes. Tests show that the method more accurately reflects image quality and reduces the number of isolated points in the performance curve. To get more accurate assessment, more HVS characteristics should be analyzed to develop high-performance quality assessment methods[11,12].

References

[1] Fang Z, Bovik A C. A universal image quality index. *IEEE Signal Processing Letters,* 2002, **9**(3): 81-84.

[2] Tong Yubing, Yang Dongkai, Zhang Qishan. Video quality assessment methods overview. *Journal of CAD & Graphics*, 2006, **18**(5): 735-741. (in Chinese)

[3] ITU-R Recommendation BT. 500-10 (2000). Methodology for the subjective assessment of the quality of the television pictures. 2000.

[4] UTexas's image database and common video sequences. http://live.ece.utexas.edu/index.htm. University of Texas, USA, 2005.

[5] Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment. http://www.vqeg.org. 2004.

[6] Zhou Wang, Bovik A C. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **4**(13): 600-612.

[7] Zhou Wang, Bovik A C. Structure approached to image quality accenmet. In: Handbook of Image and Video Proceesing, 2nd Ed. New York: Academic Press, 2005.

[8] Zhang Li, Zhou Weida, Jiao Licheng. Walvelet support vector machine. *IEEE Transactions on System, Man, and Cybernerics—Part B*: *Cybernetics*, 2004, **34**(1): 34-39.

[9] Vapnik V N. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 1999, **10**(5): 988-999.

[10] Li Guiling, Wang Nannan, Zhang Qiang. Study of moving image quality evaluation base on MPEG-2 system. *Journal of Tianjin University*, 2001, **34**(5): 573-576. (in Chinese)

[11] Sanghoon L P, Bovik M S, Foveated A C. Video quality assessment. *IEEE Transactions on Multimedia*, 2002, **4**(1): 129-132.

[12] Wang Nannan. Objective quality evaluation of digital video. In: The 2000 IEEE APCCS. Tianjin, China, 2000: 791-794.