Image Quality Assessment Based on Improved Feature Similarity Metric

Zhengyou Wang [†*], Zhenxing Li [*] and Weisi Lin [#], Chenchen Liu [†]
[†] Shijiazhuang Tiedao University, Shijiazhuang
E-mail: zhengyouwang@stdu.edu.cn, lcczzm@126.com
[*] Jiangxi University of Finance & Economics, Nanchang
E-mail: lzx.good@163.com
[#] Nanyang Technological University, 50 Nanyang Avenue, Singapore
E-mail: wslin@ntu.edu.sg

***Abstract*—In this paper, a new full-reference metric for image quality assessment is proposed, which is based on the recent feature similarity (FSIM) index and incorporates proper human visual system (HVS) characteristics. This method improves FSIM by using the CSF (Contrast Sensitivity Function) operator and the contrast masking operator in DCT domain. To test the performance of the proposed metric, we have carried out experiments on LIVE database. Experimental results demonstrate that the improved metric can achieve higher consistency with the subjective evaluation than FSIM and other relevant state-of-the-art image quality assessment metrics.**

I. INTRODUCTION

Image quality is an important indicator for the performance of imaging and image processing (such as image acquisition, transmission, compression, storage, enhancement, etc.), and therefore, image quality assessment has become a research hotspot in visual signal processing.

How to measure the image quality? Because human observers are the ultimate receivers of the visual information contained in an image, quality of users' experience may be the most reliable method for image quality assessment (i.e., subjective methods). A number of different subjective methods are represented by ITU Recommendation BT.500 [1]. However, subjective experiments are time-consuming and expensive; they require many human viewers and cannot be easily and routinely performed for many scenarios, e.g., real time systems.

Objective image quality assessment is proposed to provide a computational model to measure the perceptual quality of an image. There are traditional metrics such as the peak signal to noise ratio (PSNR), mean square error (MSE), etc., and these methods are very popular because of their simple calculation; but they exhibit poor correlation with subjective evaluation. Much work has been done in modeling the human visual system (HVS) to better approximate subjective metrics [2-5]. These objective metrics emphasize the HVS' sensitivity to different visual signals, such as brightness, contrast, masking effect, frequency and other information. The examples are the PSNR-HVS metric (PSNRHVS) [2], the noise quality measure (NQM) [3], the visual signal-to-noise ratio (VSNR) [4] and so on. The performance of these metrics is better than PSNR, but they still calculate in the pixel by pixel manner (without consideration of structure among pixels/regions). The well-known structural similarity (SSIM) index [6] avoids pixel-by-pixel calculation, and considers the loss of structure in the image. So SSIM provides a good approximation of perceived image quality. A number of improved SSIM algorithms are presented, including The multi-scale extension of SSIM (MS-SSIM) [7], three-component weighted SSIM (3-SSIM) [8] and so on. Another state-of-the-art metric is the visual information fidelity (VIF) metric [5] which is an extended version of IFC [9].

In [10-11], the performance of SSIM, MS-SSIM and VIF metrics are demonstrated to be better than other state-of-the-art image quality assessment metrics. In [14], the feature similarity (FSIM) index is proposed as the latest full-reference metric, in which the phase congruency (PC) is extracted as a low-level image feature, because the PC is with a simple but biologically plausible ground on how mammalian visual systems detect and identify features in an image [12-13]. At the same time, because of the visual contrast sensitivity, the image gradient magnitude (GM) is employed as the secondary feature. After obtaining the PC and the GM features, the FSIM index can be computed in two stages. In the first stage, the local similarity map is computed, and then in the second stage, the similarity map is pooled into a single similarity score.

The FSIM metric proposed in [14] based on the fact that the HVS understands an image mainly according to its low-level features. But it does not consider various visual masking effects. Because the HVS cannot detect subtle changes in image content, this work proposes a new full-reference metric for image quality assessment, which improves the FSIM by incorporating the CSF (Contrast Sensitivity Function) operator and the contrast masking operator in DCT domain. The proposed method maintains the simplicity and efficiency as the FSIM. Experiments performed on the LIVE database demonstrate that the method we proposed in this paper is better than the FSIM and other relevant state-of-the-art image quality assessment metrics.

The remainder of this paper is organized as follows. Section II gives a simple description of feature similarity (FSIM) metric. Section III presents in detail the new image quality assessment method to improve FSIM by incorporating the related HVS characteristics. Section IV reports the experimental results and analysis. Finally, Section V concludes the paper.

II. FEATURE SIMILARITY (FSIM) METRIC

The computation of the FSIM index consists of two stages. In the first stage, the phase congruency (PC) and gradient magnitude (GM) features will be extracted. At each point x in the 2D image, the PC [15-16] is defined as

$$PC_{2D}(x) = \frac{\sum_j E_{\theta_j}(x)}{\varepsilon + \sum_n \sum_j A_{n,\theta_j}(x)} \tag{1}$$

where $A_{n,\theta_j}(x)$ is the local amplitude of point x at scale n and orientation θ_j, ε is a small positive constant. and $A_{n,\theta_j}(x) = \sqrt{e_{n,\theta_j}(x)^2 + o_{n,\theta_j}(x)^2}$; $E_{\theta_j}(x)$ is the local energy of point x along orientation θ_j, $E_{\theta_j}(x) = \sqrt{F_{\theta_j}(x)^2 + H_{\theta_j}(x)^2}$, $F_{\theta_j}(x) = \sum_n e_{n,\theta_j}(x)$, and $H_{\theta_j}(x) = \sum_n o_{n,\theta_j}(x)$.

Gradient operators can be expressed by convolution masks. Three commonly used gradient operators are the Sobel operator [17], the Prewitt operator [18] and the Scharr operator [19]. The GM of the image $f(x)$ is defined as

$$G = \sqrt{G_x^2 + G_y^2} \tag{2}$$

where $G_x(x)$ and $G_y(x)$ is the partial derivatives of $f(x)$ along horizontal and vertical directions using one of the three gradient operators.

And then to separate the feature similarity measurement between the original image $f_1(x)$ and the distorted image $f_2(x)$ into two components, each for PC and GM. The similarity measure for PC values and GM values are defined as

$$S_{PC}(x) = \frac{2PC_1(x) \cdot PC_2(x) + T_1}{PC_1^2(x) + PC_2^2(x) + T_1} \tag{3}$$

$$S_G(x) = \frac{2G_1(x) \cdot G_2(x) + T_2}{G_1^2(x) + G_2^2(x) + T_2} \tag{4}$$

Where PC_1, PC_2 is the PC value of $f_1(x)$ and $f_2(x)$, G_1, G_2 is the GM value of $f_1(x)$ and $f_2(x)$, and $T_1 = 0.85, T_2 = 160$.

After obtaining $S_{PC}(x)$ and $S_G(x)$, we combine them to get the similarity $S_L(x)$ of $f_1(x)$ and $f_2(x)$, and $S_L(x)$ is defined as

$$S_L(x) = [S_{PC}(x)]^\alpha + [S_G(x)]^\beta \tag{5}$$

The second stage is to compute the FSIM index between $f_1(x)$ and $f_2(x)$ is given as

$$FSIM = \frac{\sum_{x\in\Omega} S_L(x) \cdot PC_m(x)}{\sum_{x\in\Omega} PC_m(x)} \tag{6}$$

where $PC_m(x) = \max(PC_1(x), PC_2(x))$. For RGB color images, the images will be converted into YIQ color space firstly, defined $S_I(x)$, $S_Q(x)$ Similar to the definitions of $S_{PC}(x)$ and $S_G(x)$, the chrominance similarity measures are then obtained: $S_C(x) = S_I(x) \cdot S_Q(x)$. The feature similarity metric of color images $FSIM_c$ is define as follow. More detail about FSIM and $FSIM_c$ the reader can refer to [15].

$$FSIM_c = \frac{\sum_{x\in\Omega} S_L(x) \cdot [S_C(x)]^\lambda PC_m(x)}{\sum_{x\in\Omega} PC_m(x)} \tag{7}$$

III. IMPROVED QUALITY METRIC BASED ON THE HVS TO IMPROVE FSIM

As aforementioned, the PC and the GM are low-level features, and the FSIM considers these two features for the improved consistency between subjective and objective evaluation. Many of the quality metrics proposed were also based on other properties of the HVS, such as CSF and luminance masking [20-21], and contrast sensitivity masking [22-24] .

The work uses the CSF operator and contrast sensitivity masking operator in the DCT domain to improve FSIM metric. The 8x8 CSF operator [25] and 8x8 contrast sensitivity masking operator [26] are shown in Table I and TableII. Herinafter, the proposed image quality assessment metric is denoted as $FSIM_{HVS}$ for gray images and $FSIM_{HVS}^C$ for color images. The calculation steps for the proposed metric are as follows.

TABLE I
8X8 CONTRAST MASKING OPERATOR, $MASK(i,j)$, i,j=1,2,...,8.

0.3906	0.8264	1.0000	0.3906	0.1736	0.0625	0.0384	0.0269
0.6944	0.6944	0.5102	0.2770	0.1479	0.0297	0.0278	0.0331
0.5102	0.5917	0.3906	0.1736	0.0625	0.0308	0.0210	0.0319
0.5102	0.3460	0.2066	0.1189	0.0384	0.0132	0.0156	0.0260
0.3086	0.2066	0.0730	0.0319	0.0216	0.0084	0.0094	0.0169
0.1736	0.0816	0.0331	0.0244	0.0152	0.0092	0.0078	0.0118
0.0416	0.0244	0.0164	0.0132	0.0094	0.0068	0.0069	0.0098
0.0193	0.0118	0.0111	0.0104	0.0080	0.0100	0.0094	0.0102

TABLE II
8X8 CSF OPERATOR, $CSF(i,j)$, i,j=1,2,...,8.

1.6084	2.3396	2.5735	1.6084	1.0723	0.6434	0.5046	0.4219
2.1446	2.1446	1.8382	1.3545	0.9898	0.4437	0.4289	0.4679
1.8382	1.9796	1.6084	1.0723	0.6434	0.4515	0.3730	0.4596
1.8382	1.5138	1.1698	0.8874	0.5046	0.2958	0.3217	0.4151
1.4297	1.1698	0.6955	0.4596	0.3785	0.2361	0.2499	0.3342
1.0723	0.7353	0.4679	0.4021	0.3177	0.2475	0.2277	0.2797
0.5252	0.4021	0.3299	0.2958	0.2499	0.2127	0.2145	0.2548
0.3574	0.2797	0.2709	0.2626	0.2298	0.2574	0.2499	0.2600

Step 1: Divide the original image $f_1(x)$ and the distorted image $f_2(x)$ into $N \times N$ blocks; in this paper, $N = 8$. Compute the 2D DCT of block A in original image $f_1(x)$ and block B in the distorted image $f_2(x)$, and A and B are with the same location in the two images, to obtain $A_{DCT} = DCT2(A)$ and $B_{DCT} = DCT2(B)$.

Step 2: As stated in [27], each DCT coefficient $DCT(i, j)$ of an image block masks any other coefficients to some extent，so we combine the energy of DCT coefficients with the mask operator to account for the masking effect of the original image (M_A) and the distorted image (M_B) in the DCT domain, with the formulae as follow:

$$M_A = \sqrt{\left(\left(\sum_{i=1}^{8}\sum_{j=1}^{8}[A_{DCT}(i,j)]^2 \cdot MASK(i,j)\right) \cdot pop_A\right)/(N \times N)} \tag{8}$$

$$M_B = \sqrt{\left(\left(\sum_{i=1}^{8}\sum_{j=1}^{8}[B_{DCT}(i,j)]^2 \cdot MASK(i,j)\right) \cdot pop_B\right)/(N \times N)} \tag{9}$$

where $\sum_{i=1}^{8}\sum_{j=1}^{8}[A_{DCT}(i,j)]^2$ denotes the energy of DCT coefficients of a 8x8 image block, $MASK(i, j)$ is the (i,j) element in table I, which Note that the values in quantization table JPEG have been normalized and then squared. and *pop* (to be defined in (11) below and explained next) is used to adjust the value of M_A and M_B. The variance of a $N \times N$ pixel block is calculated as:

$$D = \left[A(i,j) - \sum_{i=1}^{N}\sum_{j=1}^{N} A(i,j)/(N \cdot N)\right]^2 \cdot \frac{N \cdot N}{N \cdot N - 1} \tag{10}$$

then *pop* is computed as follows:

$$pop = \begin{cases} 0 & if \ D = 0 \\ (D_1 + D_2 + D_3 + D_4)/D & else \end{cases} \tag{11}$$

The value of masking coefficient in Table I can be too high if an image block belongs to edges; in such a case we divide block into four parts uniformly, where $D_i (i = 1,2,3,4)$ is the local varianceof each part (4x4) in the block, do this can reduce masking effect.

Step 3: When $M_A < M_B$, by adjusting the masking effect value M_A, take $M_A = M_B$; the difference between A_{DCT} and B_{DCT} is defined as u, and $u = abs(A_{DCT}(i,j) - B_{DCT}(i,j))$. The DCT coefficients A_{DCT} and B_{DCT} are visually undistinguished if $u < M_A / MASK(i,j)$, where u' is the adjusted masking effect M_A of DCT coefficients A_{DCT}. u is adjusted as follow

$$u' = \begin{cases} 0 & u < M_A / MASK(i,j) \\ u - M_A / MASK(i,j) & u \geq M_A / MASK(i,j) \end{cases} \tag{12}$$

Step 4: Because the Contrast Sensitivity of every position of block in image is different, we combine with the CSF operator and define the cover factor of block k as S_k.

$$S_k = \sum_{i=1}^{8}\sum_{j=1}^{8}[u \cdot CSF(i,j)]^2 \tag{13}$$

Step 5: Assuming the image is divided into K blocks, define the cover factor of the whole image as S:

$$S = \sum_{k=1}^{K} S_k \tag{14}$$

Finally, we use the cover factor S to obtain the new feature similarity metrics of gray image and color image.

$$FSIM_{HVS} = 10 \cdot FSIM \cdot \log_{10}(255 \cdot 255 / S) \tag{15}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to evaluate the accuracy of $FSIM_{HVS}$ and $FSIM_{HVS}^{C}$ metrics, we carried out the experiment on LIVE database (LIVE database have 29 original images and 779 distorted images, including 6 distortion types) [28]. With the objective score for each image, we use a nonlinear fitting function to fit the objective scores with DMOS (Difference Mean Opinion Score) [29]. We take three different methods to evaluat the performance [29]:

Method1: The correlation coefficient between DMOS and model predictions used non-linear regression analysis.

Method2: The Spearman rank order correlations test for agreement between the rank orders of DMOS and model predictions.

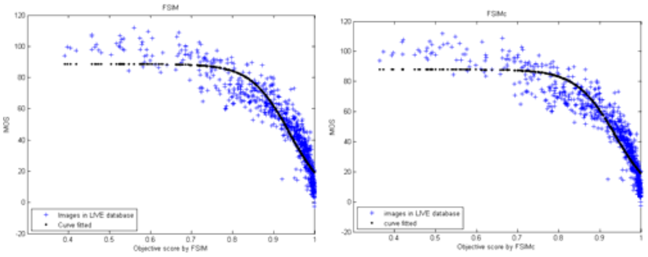
Method 3: The outlier ratio is a percentage of the number of predictions outside the range of times of the standard deviations.

Method1 is a measure of prediction accuracy, method 2 is a measure of prediction monotonicity, and method 3 is a measure of predictions consistency. The detailed definitions of these three methods are given in [30] .

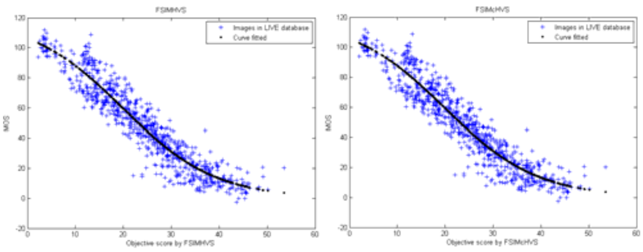
Table 3 shows performance comparison of $FSIM_{HVS}$, $FSIM_{HVS}^{C}$ and $FSIM$, $FSIM_C$ on LIVE database. According to Table 3 , the performance of proposed method $FSIM_{HVS}$, $FSIM_{HVS}^{C}$ is better than $FSIM$, $FSIM_C$; one can see from Fig.1that the objective scores predicted by $FSIM_{HVS}$ correlate much more consistently with the subjective evaluations than the FSIM methods. Becuse the performance of FSIM (as the most recent and the best performing metric) is better than MS-SSIM, SSIM, VIF, VSNR, IFC, NQM, PSNR and so on, and thus the metric this paper proposed demonstrates better performance than state-of-the-art image quality assessment metrics in terms of LIVE database.

TABLE III
PERFORMANCE COMPARISON USED $FSIM_{HVS}$ AND FSIM ON LIVE DATABASE

Model	Metric1:Non-linear regression	Metric2:Spearman rank	Metric3: outline ratio
FSIM	0.9387	0.9348	0.6380
FSIMc	0.9384	0.9347	0.6290
$FSIM_{HVS}$	0.9491	0.9447	0.5623
$FSIM_{HVS}^{C}$	0.9485	0.9440	0.5610



(a) FSIM (b) FSIMc



(c) $FSIM_{HVS}$ (d) $FSIM_{HVS}^{C}$

Fig.1 Scatter plots of subjective MOS versus scores obtained by model prediction on the LIVE database

V. CONCLUSIONS

In this paper, we proposed a new method of objective image quality assessment which based on HVS and FSIM. It fused human perceived performance into feature similarity algorithm effectively, this method not only reserves the FSIM algorithm executed simply and high efficiency, but also reveals the improvement of people visual characteristic and Psychological preferences. The experimental result proved that it can reflect human subjective perception well, and is superior to other traditional image quality assessment methods.

Due to the complexity of HVS which involved in multidisciplinary areas such as biology, anatomy, physiology, psychology, the research maintained to develop. And how to extract low-level features exactly also need to study. And no-reference image quality assessment is applied in most situation, how to apply the characteristic of HVS in it become a tendency of image quality assessment, we should make more effort on it.

VI. ACKNOWLEDGMENT

This work was partially supported by the National Natural Science Foundation of China (60963011, 60963012, 61063034, 61063035), the Sino Social and Science Program of Ministry of Education (No: 09YJA630055), the Jiangxi Natural Science Foundation of China (No: 2009GZS0022), Jiangxi Provincial Innovation Fund of Graduate Student (YC10A087).

REFERENCES

[1] Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU Rec.BT. 500-11.

[2] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, "New full-reference quality metrics based on HVS", CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2006

[3] 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 Process.*, vol. 9, no. 4, pp. 636-650, Apr. 2000.

[4] D.M. Chandler and S.S. Hemami, “VSNR: a wavelet-based visual signal-to-noise ratio for natural images”, *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284-2298, Sep. 2007.

[5] H.R. Sheikh.and A.C. Bovik, "Image information and visual quality," IEEE Transactions on Image Processing, Vol.15, no.2, 2006, pp. 430-444.

[6] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: from error visibility to structural similarity”, *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[7] Z. Wang, E.P. Simoncelli, and A.C. Bovik, “Multi-scale structural similarity for image quality assessment”, presented at the IEEE Asilomar Conf. Signals, Systems and Computers, Nov. 2003.

[8] C. Li and A.C. Bovik, “Three-component weighted structural similarity index”, in *Proc. SPIE*, vol. 7242, 2009.

[9] H.R. Sheikh, A.C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics”, *IEEE Trans. Image Process.*, vol. 14, no. 12, pp.2117-2128, Dec. 2005.

[10] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms”, *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.

[11] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, “TID2008 - A database for evaluation of full-reference visual quality assessment metrics”, *Advances of Modern Radioelectronics*, vol. 10, pp. 30-45, 2009.

[12] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment”, *IEEE Trans. Image Process, vol.20, no.15, may 2011*

[13] E.C. Larson and D.M. Chandler, “Unveiling relationships between regions of interest and image fidelity metrics”, in *Proc. SPIE Visual Comm. and Image Process.*, vol. 6822, pp. 6822A1-16, Jan. 2008.

[14] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "FSIM: a feature similarity index for image quality assessment", *IEEE Trans. on Image Processing*, vol. 20, no. 8, pp. 2378-2386, 2011

[15] M.C. Morrone, J. Ross, D.C. Burr, and R. Owens, “Mach bands are phase dependent”, *Nature*, vol. 324, no. 6049, pp. 250-253, Nov.1986.

[16] P. Kovesi, “Image features from phase congruency”, *Videre: J. Comp. Vis. Res.*, vol. 1, no. 3, pp. 1-26, 1999.

[17] R. Jain, R. Kasturi, and B.G. Schunck, *Machine Vision*. McGraw-Hill, Inc, 1995.

[18] B. Jähne, H. Haubecker, and P. Geibler, *Handbook of Computer Vision and Applications*. Academic Press, 1999.

[19] C. Yang and S.H. Kwok, “Efficient gamut clipping for color image processing using LHS and YIQ”, *Optical Engineering*, vol. 42, no. 3, pp.701-711, Mar. 2003.

[20] A.B. Watson and J.A. Solomon, “Model of visual contrast gain control and pattern masking”, *J. Opt. Soc. Am. A*, 14(9), pp. 2379–2391, 1997.

[21] Foley, J.M. “Human Luminance Pattern-Vision Mechanisms-Masking Experiments Require a New Model”, *Journal of the Optical Society of America a-Optics Image Science and Vision*, 11 (6), 1710-1719, 1994.

[22] Z. Wang, R.S. Hamid, and A.C. Bovik. *The handbook of video databases: Design and Applications*, CRC Press, 2003.

[23] A. B. Watson, “Perceptual optimization of DCT colour quantization matrices”, *Image Processing ICIP-94*, 1, pp. 100–104, 1994.

[24] Solomon J. A., Watson A. B., and Ahumada A. “Visibility of DCT basis functions: Effects of contrast masking”, *Proceedings*

of Data Compression Conference, Snowbird, Utah: IEEE Computer Society Press, 361-370, 1994.

[25] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on HVS, CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2006, 4 p.

[26] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazarian, Marco Carli, Jaakko Astola, Vladimir Lukin, "On between-coefficient contrast masking of DCT basis functions", CD-ROM Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07, Scottsdale, Arizona, USA, 25-26 January, 2007, 4 p.

[27] Solomon J. A., Watson A. B., and Ahumada A. "Visibility of DCT basis functions: Effects of contrast masking". Proceedings, Data Compression Conference, Snowbird, Utah: IEEE Computer. Society Press, 361-370, 1994

[28] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2", http://live.ece.utexas.edu/research/quality.

[29] Z. Lu, W. Lin, X. Yang, E. Ong and S. Yao, "Modeling Visual Attention's Modulatory Aftereffects on Visual Sensitivity and Quality Evaluation", *IEEE Trans. Image Processing*, Vol.14(11), pp.1928 – 1942, Nov. 2005.

[30] VQEG. Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, PHASE II [EB/OL]. [2006- 03- 01]. ftp://ftp.its.bldrdoc.gov/dist/ituvidq/frtv2_final_report/.