Content-weighted video quality assessment using a three-component image model

Chaofeng Li
Jiangnan University
School of Information Technology
Wuxi, Jiangsu 214122
China
wxlichaofeng@yahoo.com.cn

Alan Conrad Bovik
The University of Texas at Austin
Department of Electrical and Computer Engineering
Austin, Texas 78712-1084

Abstract. *Objective image and video quality measures play important roles in numerous image and video processing applications. In this work, we propose a new content-weighted method for full-reference (FR) video quality assessment using a three-component image model. Using the idea that different image regions have different perceptual significance relative to quality, we deploy a model that classifies image local regions according to their image gradient properties, then apply variable weights to structural similarity image index (SSIM) [and peak signal-to-noise ratio (PSNR)] scores according to region. A frame-based video quality assessment algorithm is thereby derived. Experimental results on the Video Quality Experts Group (VQEG) FR-TV Phase 1 test dataset show that the proposed algorithm outperforms existing video quality assessment methods.* © *2010 SPIE and IS&T.* [DOI: 10.1117/1.3267087]

1 Introduction

Applications of digital video have been steadily increasing, owing to rapid developments in many technologies such as cameras, camcorders, DVD players, set-top boxes, mobile video, video teleconferencing, streaming video over Internet, video chat, etc. Since most applications are directed toward human end-users, the visual quality of a video signal is quite important. Quality monitoring of such videos generally requires automatic methods of assessment. The goal of video quality assessment (VQA) algorithms is to automatically assess the quality of videos in a manner that is consistent with human visual judgment. Such algorithms are widely applied in video acquisition, communication, and display systems, and for evaluating and testing video coders, for online quality monitoring and control, for end-to-end video transmission systems, for perceptual video compression and restoration, and so on.

The simplest and most commonly used (at least until recently) full-reference (FR, meaning that a reference comparison image is available to assess quality against) objective image and video distortion quality metrics are the mean squared error (MSE), computed by averaging the squared intensity differences of distorted and reference image pixels, and the log-reciprocal peak signal-to-noise ratio (PSNR). MSE and PSNR have been widely used because they are simple to calculate, have clear physical meanings, and are mathematically convenient. However, they have been widely criticized for not correlating well with perceived quality.[1–3]

In an effort to address this problem, many VQA metrics have been recently proposed. For conceptual convenience, we divide existing VQA methods into three different but related types. The first type is based on modeling the human visual system (HVS).[4–8] This approach to VQA is intuitive, since the goal of objective VQA systems is to match visual performance in predicting quality. The second type of approach is based on the extraction of image features and/or statistics. These are computed from the reference and test videos, and include such perceptually relevant quantities as local brightness, contrast, edges, textures, color, blockiness, and so on. Several popular algorithms utilize a signal statistic based approach to VQA.[9–13] The third type of approach computes local motion information,[9,10,14–16] for example, video structural similarity image index (SSIM),[9] speed-weighted SSIM,[10] and the MOVIE index.[16]

Several of these are closely related to the still image QA algorithm called the structural similarity image index (SSIM),[2] which operates under the assumption that visual perception is highly adapted for extracting structural information from a scene. The authors of Ref. 10, for example, impose a motion-weighting model[9] onto SSIM to account for the fact that the accuracy of visual perception is significantly reduced when the speed of motion is large. Lu *et al.*[17] proposed a perceptual quality significance map (PQSM) generated by using local perceptual stimuli from color contrast, texture contrast, motion, and cognitive features. The authors showed that the ideas in PQSM can improve the VQA performance of MSE/PSNR and SSIM. Se-

Paper 09071SSR received May 3, 2009; revised manuscript received Jun. 23, 2009; accepted for publication Jul. 21, 2009; published online Jan. 7, 2010.

1017-9909/2010/19(1)/011003/9/$25.00 © 2010 SPIE and IS&T.

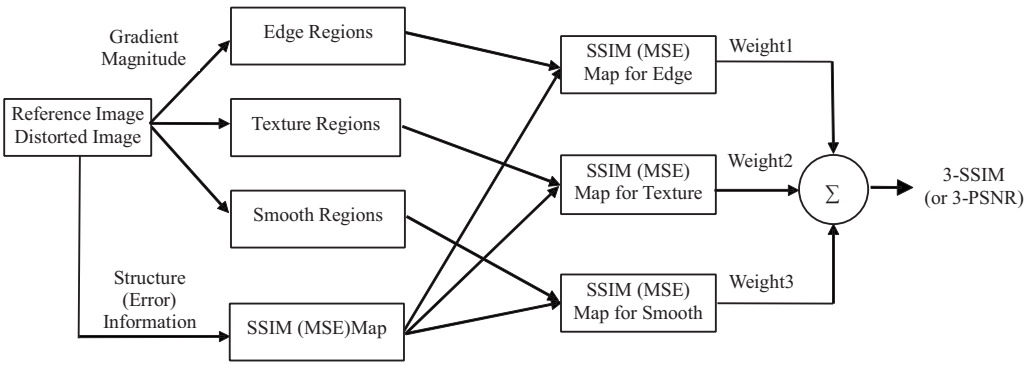


Fig. 1 Diagram for calculating 3-SSIM (or 3-PSNR).

shadrinathan and Bovik[15,16] employed local motion information obtained from optical flow to adaptively guide the orientation of a set of 3-D Gabor filters. The Gabor responses are motion tuned using a model of cortical area MT to compute VQA. Wang and Li[10] incorporated a recent model of human visual speed perception, and used a motion information framework to adapt SSIM for VQA.

All of these successful algorithms operate solely using low-level video features, while neglecting the important influence of image content. Video quality is highly correlated with video content. There are a number of perceptual factors[18] that influence human perception of visual quality. Successfully incorporating these into objective image quality assessment (IQA) metrics have the potential to achieve improved correlations with visual perception.

In this work, we commence our attack on this problem by taking a low-level image-content approach to the problem by seeking to parse the image to be assessed into regions of different low-level content. Specifically, we use a popular three-component image model to parse the reference and distorted images, then derive region-weighted versions of PSNR and SSIM to achieve video quality assessment. The resulting algorithms are termed three component PSNR (3-PSNR) and three component SSIM (3-SSIM). We find that we are able to improve the performance of PSNR and SSIM relative to human subjectivity on the Video Quality Experts Group (VQEG) Phase 1 test database[19] and Laboratory for Image and Video Engineering (LIVE) Image Quality Database.[20]

The rest of the work is organized as follows. Section 2 introduces our three-component weighting method. The proposed 3-SSIM (and 3-PSNR) indices and some examples are described in Sec. 3. Section 4 gives experimental results. Finally, in Sec. 5, future thoughts are given.

2 Three-Component Weighting Method

IQA/VQA algorithms generally operate without attempting to take into account image/video content. Since algorithms for image/video content identification remain in a nascent state, IQA/VQA algorithms that succeed in assessing quality as a function of content will await developments in that direction. However, low-level content of visual importance, sometimes called salient image features, might be used to improve IQA/VQA algorithms. For example, intensity edges certainly contain considerable image information and are perceptually significant. Using this observation we incorporate a three-component image model into SSIM (or PSNR), and thereby develop three-component weighted SSIM (3-SSIM) and PSNR (3-PSNR) indices.

The development of 3-SSIM and 3-PSNR follows four steps. 1. Calculate the SSIM (or PSNR) map. 2. Independent of the SSIM (or PSNR) results, segment the original (reference and distorted) image into three categories of regions (edges, textures, and smooth regions). Edge regions are found where a gradient magnitude estimate is large, while smooth regions are determined where the gradient magnitude estimate is small. Textured regions are taken to fall between these two thresholds. 3. Apply nonuniform weights to the SSIM (or PSNR) values over the three regions. 4. Pool the weighted SSIM (or PSNR) values, e.g., their weighted average, thus defining a single quality index for the image (3-SSIM or 3-PSNR). A diagram depicting calculation of 3-SSIM (or 3-PSNR) is shown in Fig. 1.

In our approach, an image is partitioned into three parts: edges, textures, and smooth regions as proposed in Ref. 21. We seek to more heavily weight degradations of intensity edges in the IQA process. Textured regions, by contrast, often mask degradations. Artifacts in smooth regions may be quite obvious, especially if they are high frequency or edge-like.

We can partition an image into three components using the computed gradient magnitude; Ref. 22 gives a simple method to obtain a partition into three types of regions. The following steps explain the process.

Step 1. Compute the gradient magnitudes using a Sobel operator on the reference and the distorted images.
Step 2. Determine thresholds $TH_1=(0.12)\ g_{\max}$ and $TH_2=(0.06)\ g_{\max}$, where $g_{\max}$ is the maximum gradient magnitude value computed over the reference image.
Step 3. Assign pixels as belonging to edge, texture, and smooth regions as follows.

Denoting the gradient at coordinate (i,j) on the reference image by $p_o(i,j)$, and the gradient on the distorted image as $p_d(i,j)$, pixel classification is carried out according to the following rules:

R_1: if $p_o(i,j)>TH_1$ or $p_d(i,j)>TH_1$, then the pixel is considered to be an edge pixel.
R_2: if $p_o(i,j)<TH_2$ and $p_d(i,j)\leqslant TH_1$, then the pixel is regarded as part of a smooth region.

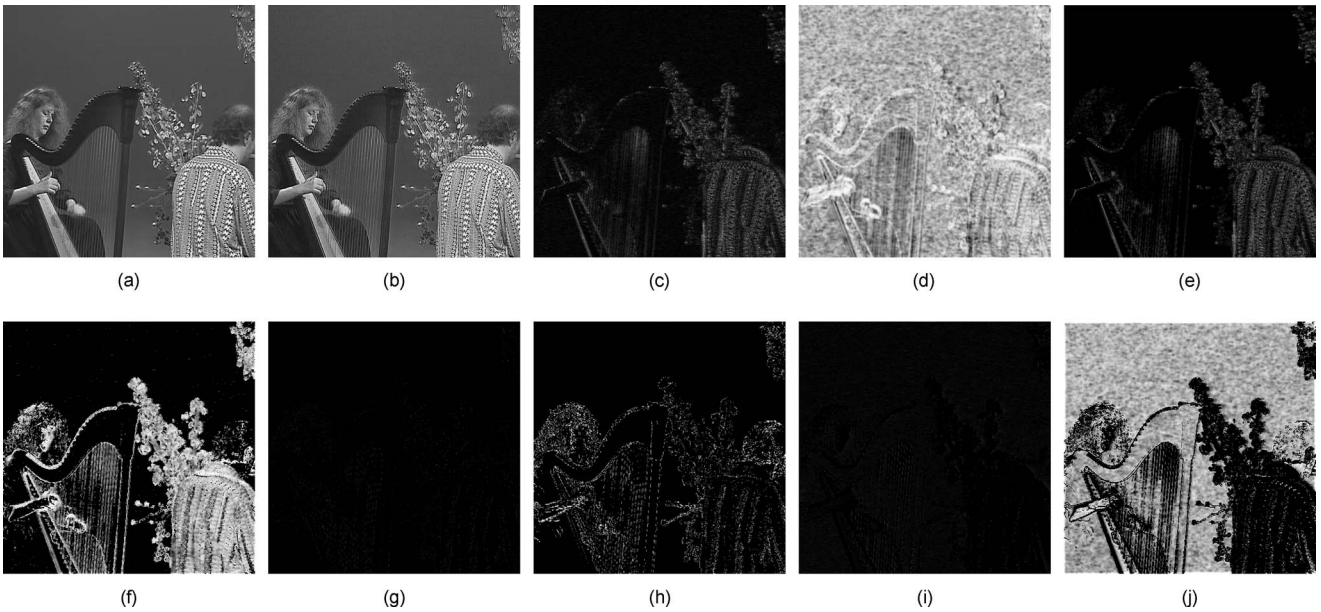


Fig. 2 Illustration of quality maps: (a) original image, (b) distorted image, (c) absolute error map, (d) SSIM index map, (e) absolute error map for edge regions, (f) SSIM index map for edge regions, (g) absolute error map for texture regions, (h) SSIM index map for texture regions, (i) absolute error map for smooth regions, and (j) SSIM index map for smooth regions.

R_3: otherwise, the pixel is regarded as part of a textured region.

Figure 2 shows an original image and a distorted version of it. Also shown are the absolute error map, the SSIM index map, and the edge, smooth, and textured regions.

3 Content-Based Image and Video Quality Assessment

The primary goal of IQA/VQA is to produce automatic image and video ratings that correlate well with the mean opinion scores (MOS) obtained by subjective trials. Current leading algorithms for FR IQA and VQA do not consider image content of the type we have been discussing. We believe that many of these could be improved by incorporating content into them in simple ways. In this study, we consider this possibility using two popular metrics: the PSNR and SSIM.

3.1 *Peak Signal-to-Noise Ratio*

The MSE and the related peak signal-to-noise ratio (PSNR) are popularly used to assess image quality. Given two vectors $x=\{x_i|i=1,\ldots,N\}$ and $y=\{y_i|i=1,\ldots,N\}$, then

$$\mathrm{MSE}(x,y)=\frac{1}{N}\sum_{i=1}^{N}(x_i-y_i)^2. \tag{1}$$

Then

$$PSNR(x,y)=10\log_{10}\left(\frac{L^2}{MSE(x,y)}\right), \tag{2}$$

where L is a constant, representing the image dynamic range (e.g., for 8-bits/pixel grayscale image, $L=2^8-1=255$).

Of course, the PSNR is easy to compute and implement in software and hardware. However, the PSNR is a very poor measure of image quality! A simple illustration is shown in Fig. 3. Figure 3(a) is an original Einstein image. We used the MATLAB function "imadjust" to produce a contrast-enhanced image [adjusted input ranges from 0.1 to 0.811, shown in Fig. 3(b)]; the MATLAB function "imnoise" to produce noise contaminated images (Gaussian noise density is set to 0.0044, salt-pepper noise density is 0.015, and speckle noise density is set to 0.0233), shown in Figs. 3(c)–3(e), respectively; the MATLAB function "imwrite" to produce a JPEG compressed image (quality parameter set to 1, shown in Fig. 3(f)); and the MATLAB function "imfilter" to produce a blurred image [window size set to 7×7, shown in Fig. 3(g)]. A human would likely rate the distorted images in order (best) Figs. 3(b)–3(g) (worst) relative to the reference [Fig. 3(a)]. However, all six images have identical PSNR of 23.58 relative to the reference.

Using the three-component weighted approach described in Sec. 2, it is a simple matter to define the three-component weighted PSNR, termed 3-PSNR. The proposed 3-PSNR is defined here with weights for the three types of regions set as follows: for edges the weighting is 0.5, while for the texture and smooth regions the weighting is 0.25. The resulting 3-PSNR yields results that are more consistent with human subjective perception than PSNR, as shown in Table 1.

3.2 *Structural Similarity Index*

The SSIM index is a recent and very popular IQA/VQA algorithm. The idea behind SSIM[2] is that natural images are highly structured, and that the human visual system is sensitive to structural distortion. It defines the function for the

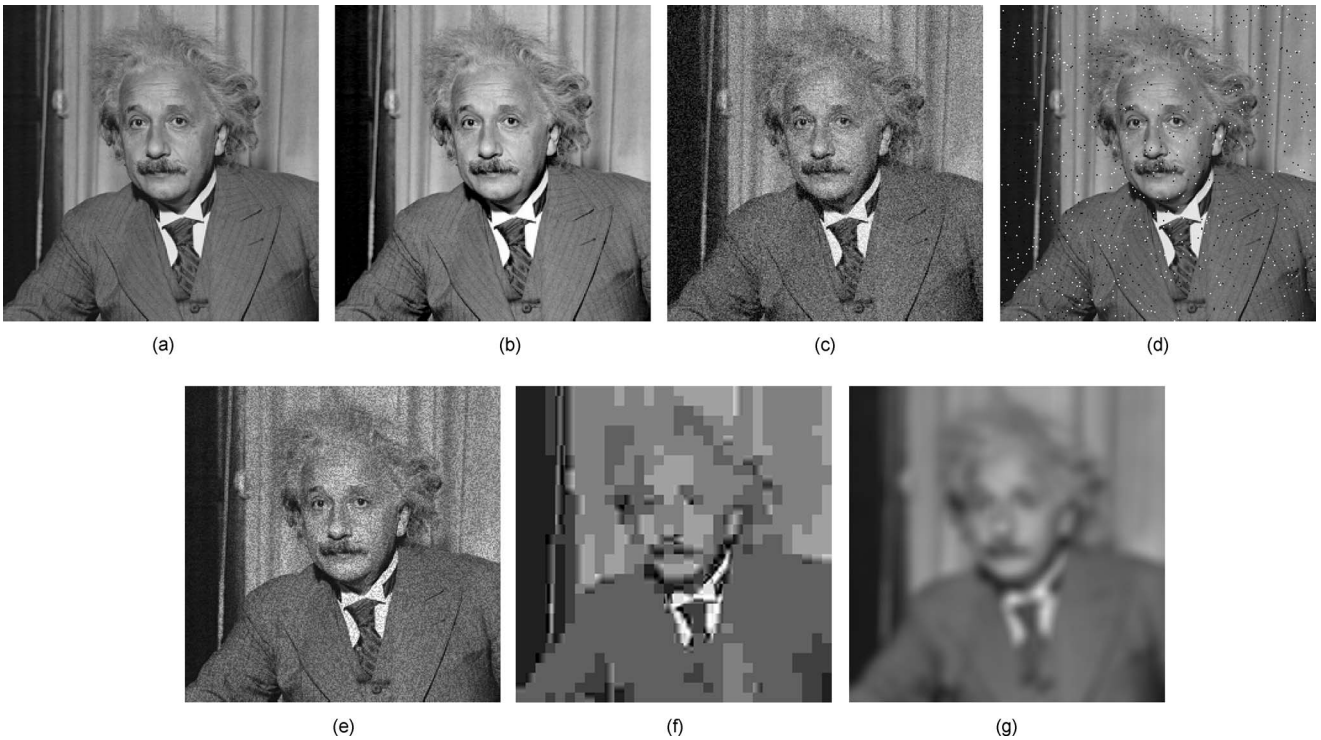


Fig. 3 Distorted images: (a) reference Einstein image, (b) contrast enhanced image, (c) Gaussian noise contaminated image, (d) salt-pepper noise contaminated image, (e) speckle noise contaminated, (f) JPEG compressed image, and (g) blurred image.

luminance comparison of the signals, the contrast comparison of the signals, and the structure comparison of the signals, respectively, as follows:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{3}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{4}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{5}$$

where μ_x and μ_y are (local) sample means of x and y, respectively, σ_x and σ_y are (local) sample standard deviations of x and y, respectively, and σ_{xy} is the (local) sample correlation coefficient between x and y. Generally, these local sample statistics are computed within overlapping windows and weighted within each window, e.g., by a Gaussian-like profile. The small constants C_1, C_2, and C_3

Table 1 Quality assessment results on images having similar PSNR.

	PSNR	3-PSNR	PSNR for edge regions	PSNR for texture regions	PSNR for smooth regions
Fig. 3(b)	23.58	24.21	24.8644	23.8684	23.2391
Fig. 3(c)	23.58	23.61	23.6083	23.6772	23.5430
Fig. 3(d)	23.58	23.57	23.5998	23.4619	23.6118
Fig. 3(e)	23.58	23.37	22.7049	23.8962	24.1569
Fig. 3(f)	23.58	22.61	20.2308	23.8573	26.1282
Fig. 3(g)	23.58	21.52	17.9036	23.6010	26.6760

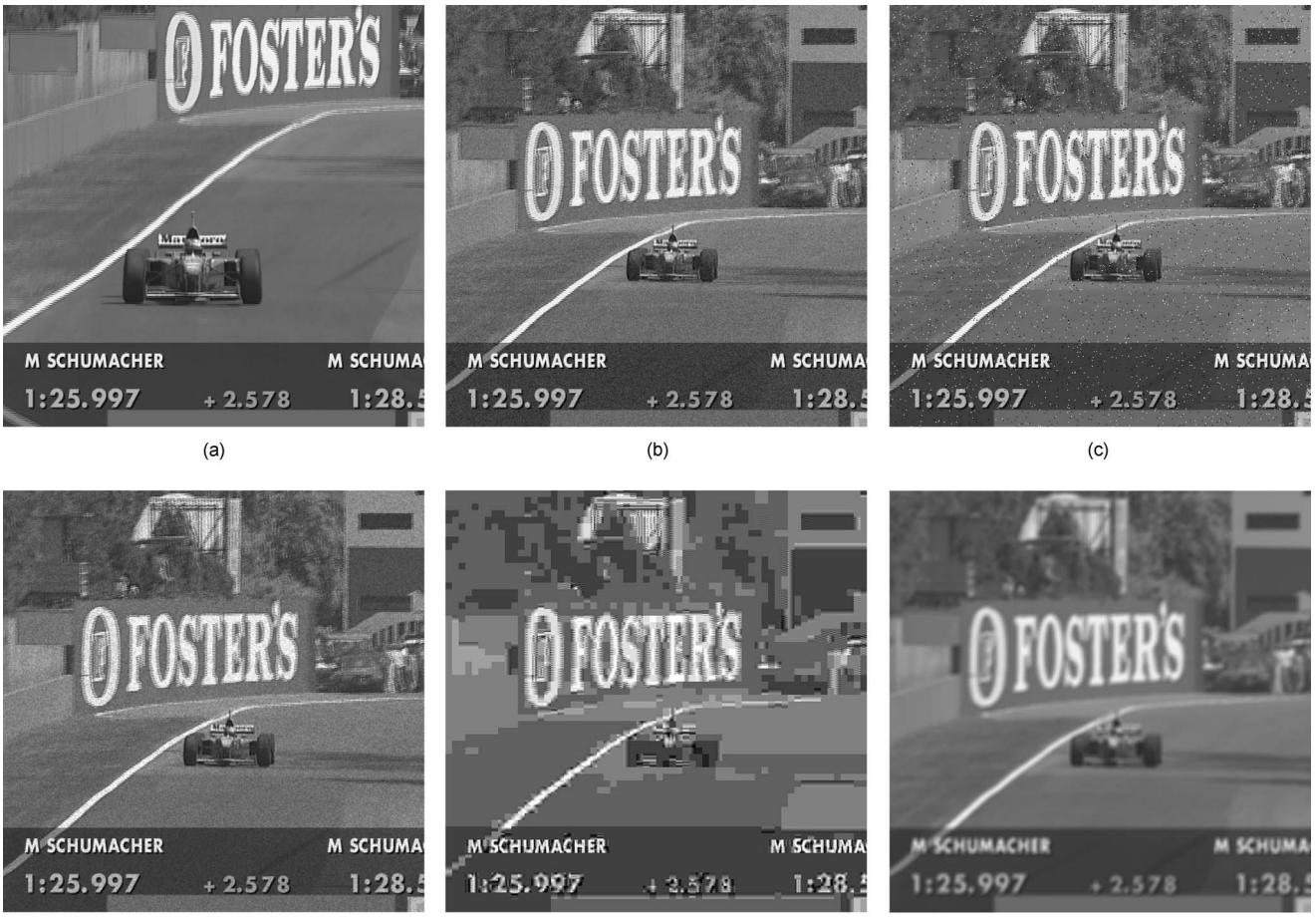


Fig. 4 Illustration of distorted images: (a) reference image, (b) Gaussian noise contaminated image, (c) salt-pepper noise contaminated image, (d) speckle noise contaminated, (e) JPEG compressed image, and (f) blurred image.

stabilize the computations of Eqs. (3)–(5) when the denominator(s) become small.

Combining the three comparison functions of Eqs. (3)–(5) yields a general form of the SSIM index:

$$\mathrm{SSIM}(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}, \tag{6}$$

where α, β, and γ are parameters that mediate the relative importance of the three components. Usually, $\alpha=\beta=\gamma=1$, yielding the now-familiar specific form of the SSIM index:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{7}$$

In Ref. 2, the SSIM index is deployed using an 11 $\times$ 11 sliding window over the entire image space. At each image coordinate, the SSIM index is calculated within the local window; the resulting SSIM index map can be used to visualize the quality map of the distorted images. Finally, the SSIM index values can be spatially pooled, e.g., by taking their sample mean, yielding a single descriptor of the image objective quality.

Results in large human studies have shown that the SSIM index performs quite well relative to human subjectivity. However, the performance of SSIM has been observed to be less competitive when used to assess blurred and noisy images. A simple example is shown in Fig. 4. Figure 4(a) is a reference video frame image. We used the MATLAB function "imnoise" to produce a noise contaminated image (Gaussian noise density set to 0.00166, salt-pepper noise density to 0.02, and speckle noise density set to 0.0111), shown in Figs. 4(b)–4(d), respectively; the MATLAB function "imwrite" to produce a JPEG compressed image (quality parameter set to 1), shown in Fig. 4(e); and the MATLAB function "imfilter" to produce a blurred image (window size set to 7×7), shown in Fig. 4(f). The human subjective impression of the quality of these images is markedly different, yet the simple SSIM index does not distinguish them (see Table 2).

Using the three-component weighted approach describe in Sec. 2, it is a simple matter to define the three-component weighted SSIM, termed 3-SSIM. The 3-SSIM scores shown in Table 2 do distinguish the distorted images, and in a manner that appears coincident with human perception of quality. If this proves to be verifiable, then it provides evidence for an apparent advantage of the idea of using low-level content in IQA/VQA algorithms. Table 2 also shows SSIM for edge, texture, and smooth regions,

Table 2 Quality assessment results on images with similar SSIM values.

	SSIM	3-SSIM	SSIM for edge	SSIM for texture	SSIM for smooth
Fig. 4(b)	0.648	0.811	0.9224	0.8414	0.5574
Fig. 4(c)	0.648	0.793	0.8852	0.8394	0.5624
Fig. 4(d)	0.648	0.676	0.6169	0.8297	0.6412
Fig. 4(e)	0.648	0.565	0.5735	0.3839	0.7281
Fig. 4(f)	0.648	0.48	0.3732	0.4084	0.7655

respectively. As can be seen, the SSIM values on edge regions are more consistent with human subjective judgment.

4 Experimental Results and Discussion

We tested the 3-SSIM and 3-PSNR indices on the VQEG Phase 1 test dataset.[23] Showing demonstrable success on this widely used video quality database would be a powerful indication of the relevance of low-level content to quality assessment, since the two algorithms make no use of computed motion information. The VQEG dataset contains 20 reference video sequences. A large number of test sequences were obtained by distorting each reference video with 16 different distortion types. Subjective scores were recorded for all test sequences.[19] Video quality is recorded on a frame-by-frame basis only. As suggested in Ref. 9, we set the color weighting parameters $W_Y=1$ and $W_{Cb}=W_{Cr}=0$, i.e., we only use the luminance channel Y for video quality assessment.

We follow the performance evaluation procedures employed in the VQEG Phase 1 FRTV test[19] to provide quantitative measures on the performance of the objective quality assessment models. Two metrics are employed. The first is the Spearman rank order correlation coefficient (SROCC), which is an indicator of the prediction monotonicity of the quality index. The second is the linear correlation coefficient (LCC) between the difference mean opinion scores (DMOS) and the algorithm scores following nonlinear regression. The nonlinearity chosen for regression was a five-parameter logistic function.[24]

Edges play an important role in the perception of video frame images, and edges that are distorted, e.g., by blur, can greatly impact the perceived quality of a video frame image. We therefore modify the SSIM and PSNR indices by allocating greater weight to the scores on edge regions than on smooth and texture regions. The distortion of textures can be perceptually significant also, and some distortions can be obscured or masked by the presence of textures. Smooth regions are also important. The eye is sensitive to artifacts such as false contouring, blocking, and high-frequency noise in these areas.

Training quality assessment algorithms on any single QA database to achieve the most competitive performance on that database is, of course, to be deplored. However, it can be quite instructive to examine the performance of a VQA algorithm on the VQEG database while varying the parameters of the algorithm to gain insights into those aspects of the algorithm that contribute to its performance relative to human opinion. Reporting the performance over all parameter values tested in such studies is the only fair method of comparison. Another way of stating this is that the selection of parameters in any VQA algorithm needs careful explanation.

In any case, to explore the relevance of the various parameters of 3-PSNR and 3-SSIM, we varied the weight for edge regions from 0.5 to 1.0, in steps of 0.1. In each case equal weights were allocated to the texture and smooth regions, so that the total weights summed to unity. Using this protocol, the plot of SROCC against the value of the QA weight on the edge regions for 3-PSNR and 3-SSIM is shown in Figs. 5 and 6, respectively. As can be seen, when the edge weight takes the value 0.7, the SROCC for 3-PSNR reaches the highest (discrete) score, suggesting that the 3-PSNR QA index should weight edge regions much more heavily than other regions, but not eliminate textured and smooth regions. This is intuitive, since perceptually significant errors can occur in any of the regions.

Surprisingly though, the plot of SROCC monotonically increases with the edge weight, reaching a highest value at unity (the other weights being zero). This suggests that with respect to QA by 3-SSIM, edge regions play a dominant role in video quality perception. While this conclusion is

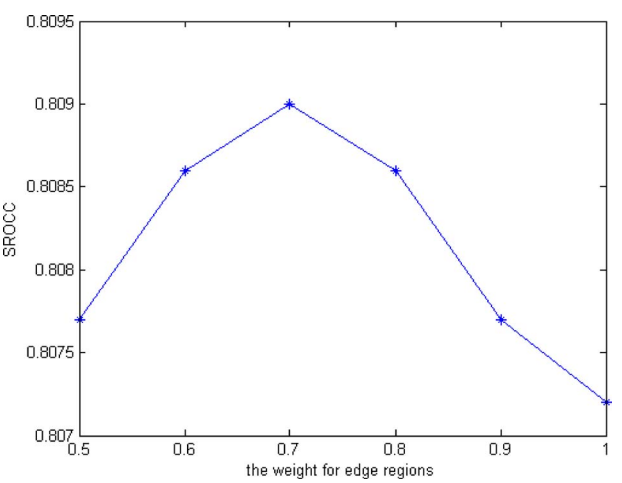


Fig. 5 Plot of SROCC of 3-PSNR against edge region weight.

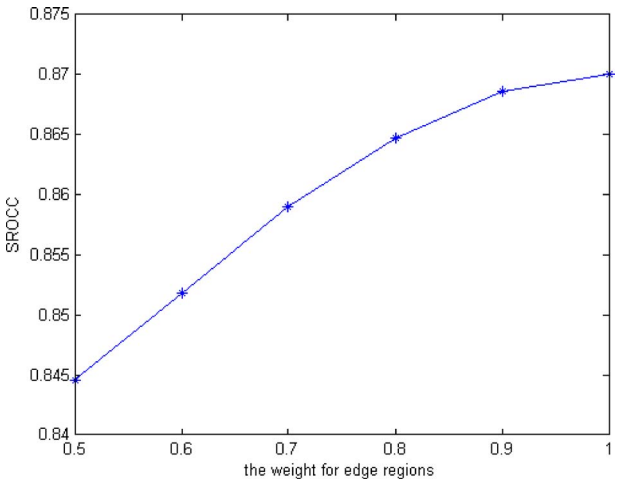


Fig. 6 Plot of SROCC of 3-SSIM against edge region weight.

probably too much to accept, 3-SSIM with edge weight unity achieves an SROCC score higher than any other prior reported method by a considerable margin.

Figures 7(a)–7(d) show the scatter plots of the subjective/objective comparisons on all test video sequences for PSNR, SSIM, 3-PSNR (with edge, texture, and smooth region weights set to 0.7, 0.15, and 0.15, i.e., the best weights), and 3-SSIM (with edge region weight of 1). Table 3 shows comparisons of the two metrics. Another recently proposed algorithm,[10] which we label speed-weighted SSIM (SW-SSIM) is listed as well. Here we compute quality at every pixel location in the frame for SW-SSIM by using dense sampling, rather than the sparse sampling strategy used in Ref. 10. Despite its simplicity, the proposed 3-PSNR method outperforms PSNR, Proponent P8 (the best performing index among the ten different proponent models tested by the VQEG[23]) and a motion-weighted PSNR.[10] Likewise, 3-SSIM greatly improves the performance of SSIM, yielding performance that is better than all other tested methods.

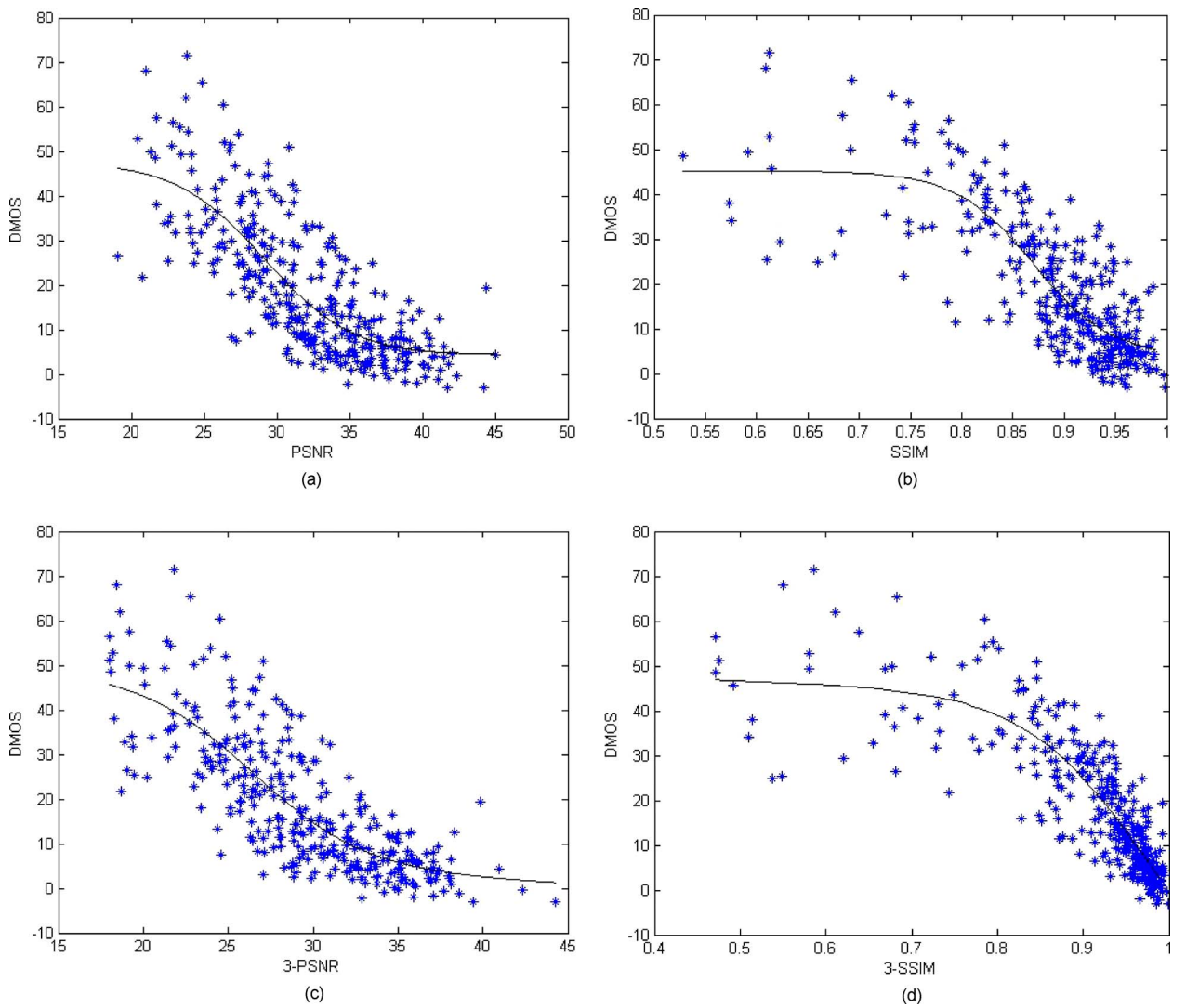


Fig. 7 Scatter plot comparison of different video quality assessment models on the VQEG Phase 1 test dataset. The vertical and horizontal axes plot subjective and objective measurements, respectively. Each sample point represents one test video: (a) PSNR, (b) SSIM, (c) 3-PSNR, and (d) 3-SSIM.

Table 3 Performance comparison of video quality assessment indices on VQEG Phase 1 test dataset.

Model	SROCC	CC
PSNR (Y only)	0.780	0.781
P8[23]	0.803	0.827
Weighted PSNR[10]	0.805	N/A
3-PSNR (Y only)	0.809	0.796
SSIM (Y only)	0.773	0.812
SW-SSIM (dense-Y only)	0.837	0.810
MOVIE (Y only)[14]	0.833	0.821
3-SSIM (Y only)	0.870	0.865

5 Conclusions and Discussions

In this work, we propose a new three-component weighting method for objective video quality assessment. The key feature of the proposed method is the idea of parsing an image into regions of different content types, then weighting QA scores according to content type. This can be viewed as an exploration into the question of the degree to which low-level image content measures affect the perception of quality. Experiments on the LIVE Image Quality Database and the VQEG FR-TV Phase 1 test dataset show that accounting for low-level image content can significantly affect the relevance of QA algorithms to perceived video quality.

One of the most attractive features of the proposed method is its simplicity. Only a simple three-component partition of the image (or frame) is used, without recourse to other more complex measurements, such as motion estimation, spatial, and/or temporal filtering, linear transformations, or evaluation of specific distortions such as blur or blockiness. Naturally, incorporating such aspects into the 3-PSNR or 3-SSIM algorithms may produce yet further improvements.

Generally, our studies suggest that image content analysis should be a primary direction of inquiry for improving QA algorithms. Of course, this could include creating a finer classification of spatial regions; creating motion-based weightings similar to SW-SSIM;[10] using color more effectively; and ultimately, measuring higher level measures of image content, like located objects like faces. Of course, content-based scene analysis remains in a nascent state. We anticipate future advances in these directions.

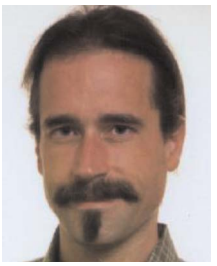
Acknowledgement

This research is supported by the Self-determined Research Program of Jiangnan University (No. JUSRP20915).

References

1. B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision*, A. B. Watson, Ed., pp. 207–220, MIT Press, Cambridge, MA (1993).
2. 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.* **13**(4), 600–612 (2004).
3. K. Seshadrinathan, T. N. Pappas, R. J. Safranek, J. Chen, Z. Wang, H. R. Sheikh, and A. C. Bovik, "Image quality assessment," in *The Essential Guide to Image Processing*, A. C. Bovik, Ed., Academic Press, San Diego, CA (2009).
4. J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed., pp. 163–178, MIT Press, Boston, MA (1993).
5. C. J. van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatiotemporal model of the human visual system," *Proc. SPIE* **2668**, 450–461 (1996).
6. S. Winkler, "Perceptual distortion metric for digital color video," *Proc. SPIE* **3644**, 175–184 (1999).
7. A. B. Watson, J. Hu, and J. F. McGowan III, "Digital video quality metric based on human vision," *J. Electron. Imaging* **10**, 20–29 (2001).
8. M. Masry, S. S. Hemami, and Y. Sermadevi, "A scalable wavelet-based video distortion metric and applications," *IEEE Trans. Circuits Syst. Video Technol.* **16**(2), 260–273, (2006).
9. Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process. Image Commun.* **19**, 121–132 (2004).
10. Z. Wang, and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," *J. Opt. Soc. Am. A* **24**, B61–B69 (2007).
11. H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessment," presented at First Intl. Conf. Video Process Quality Metrics for Consumer Electron., January 2005, pp. 23–25, Scottsdale, AZ.
12. M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.* **50**, 312–322 (2004).
13. A. P. Hekstra, J. G. Beerends, D. Ledermann, F. E. de Caluwe, S. Kohler, R. H. Koenen, S. Rihs, M. Ehrsam, and D. Schlauss, "PVQM - a perceptual video quality measure," *Signal Process. Image Commun.* **17**, 781–798 (2002).
14. K. Seshadrinathan and A. C. Bovik, "Motion based perceptual quality assessment of video," *Proc. SPIE*, **7240**, 72400X (2009).
15. K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," presented at IEEE Intl. Conf. Acoustics, Speech Signal Process., Vol. 1, pp. I-869–I-872, Honolulu, HI, 15–20 April, 2007.
16. K. Seshadrinathan and A. C. Bovik, "An information-theoretic video quality metric based on motion models," presented at Third Intl. Workshop Video Process. Quality Metrics for Consumer Electron., January 25–26, 2007, pp. 25–26, Scottsdale, AZ.
17. Z. K. Lu, W. Lin, X. K. Yang, E. P. Ong, and S. S. Yao, "Modeling visual attention's modulatory after effects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.* **14**, 1928–1942 (2005).
18. B. A. Wandell, *Foundations of Vision*, Sinauer Associates Inc., New York (1995).
19. The Video Quality Experts Group, "Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment," (2000).
20. H. R. Sheikh, Z. Wang, L. K. Cormack, and A. C. Bovik, LIVE Image Quality Database, Sep. 2005, http://live.ece.utexas.edu/research/quality/subjective.htm.
21. X. Ran and N. Farvardin "A perceptually-motivated three-component image model - part I: description of the model," *IEEE Trans. Image Process.* **4**(4), 401–415 (1995).
22. J. L. Li, G. Chen, and Z. R. Chi, "Image coding quality assessment using fuzzy integrals with a three-component image model," *IEEE Trans. Fuzzy Syst.* **12**(1), 99–106 (2004).
23. Video Quality Experts Group, "Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment," see http://www.its.bldrdoc.gov/vqeg/projects/frtv phaseI.
24. H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "An evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.* **15**(11), 3440–3451 (2006).

Chaofeng Li received BS and MSc degrees in mathematical geology, and the PhD degree in remote sensing image processing from the Chinese University of Mining and Technology, Xuzhou, China, in 1995, 1998, and 2001, respectively. From July 2001 to June 2003, he has been finishing postdoctoral research work in pattern recognition and intelligent systems at Nanjing University of Science and Technology. He was a research scholar at the University of Texas at Austin from May 2008 to May 2009. He is now with the School of Information Technology in Jiangnan University, where he developed image processing and pattern recognition algorithms. In particular, he is concerned with the development of image and video quality assessment algorithms. He has published more than 30 technical articles in these areas.



Alan Conrad Bovik is the Curry/Cullen Trust Endowed Chair Professor at The University of Texas at Austin, where he is the director of the Laboratory for Image and Video Engineering (LIVE). He is a faculty member in the Department of Electrical and Computer Engineering, the Department of Biomedical Engineering, and the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published more than 500 technical articles in these areas and holds two U.S. patents. He is the author of *The Handbook of Image and Video Processing* (Academic Press, 2005), *Modern Image Quality Assessment* (Morgan and Claypool, 2006), and two new books, *The Essential Guide to Image Processing* and *The Essential Guide to Image Processing* (Academic Press). He is a Fellow of the IEEE, the Optical Society of America, and SPIE.