A Novel Discrete Wavelet Domain Error-Based Image Quality Metric with Enhanced Perceptual Performance

Soroosh Rezazadeh and Stéphane Coulombe

Abstract—**In this paper, we present an error-based method to compute image quality scores in the discrete wavelet domain using the Haar wavelet. The proposed method does not extract any human visual system parameters, which noticeably reduces its complexity, and relies on computation of the errors (differences) at an appropriate level of wavelet decomposition. We also propose a formula to automatically calculate the required level of wavelet decomposition at a desired viewing distance. To consider the effect of very fine image details in quality assessment, the proposed method defines a multi-level edge map for each image which comprises only the most informative image subbands. We present the mathematical complexity analysis of our method and compare it to the conventional PSNR. We show that when the defined edge map is not taken into account we can calculate the quality score with lower complexity than with the PSNR, and with a prediction accuracy close to that of SSIM. We tested the performance of our method on two different databases: an image database and a video database.**

Index Terms—**Discrete wavelet transform (DWT), image quality assessment, peak signal-to-noise ratio (PSNR).**

I. Introduction

Image quality is best evaluated subjectively by human viewers. However, subjective quality assessment is time-consuming and expensive, and cannot be performed in real-time applications. It is therefore necessary to define an objective criterion that can gauge the difference between an undistorted reference and distorted image signals. Ideally, such an objective measure should correlate well with the perceived difference between two image signals and should vary linearly with subjective quality.

Objective methods are usually classified based on the availability of reference images. If an undistorted reference image is available, the quality metric considered as a full-reference (FR) assessment method. Generally speaking, the FR quality assessment of image signals involves two categories of approach: bottom-up (error-based) and top-down [1]. In the bottom-up approaches, the perceptual quality scores are best estimated by quantifying the visibility of errors, while in the top-down approaches, the whole human visual system (HVS) is considered as a black box and it is the input/output relationship that is of interest.

Manuscript received April 9, 2012; revised May 23, 2012. This work was funded by Vantrix Corporation and by the Natural Sciences and Engineering Research Council of Canada under the Collaborative Research and Development Program (NSERC-CRD 326637-05).

The authors are with the department of Software and IT Engineering, École de technologie supérieure, Université du Québec, 1100 rue Notre-Dame Ouest, Montréal, Québec, H3C 1K3, Canada (e-mail: soroosh.rezazadeh.1@ens.etsmtl.ca, stephane.coulombe@etsmtl.ca)

The peak signal-to-noise ratio (PSNR), and its equivalent, mean square error (MSE), are the oldest FR image quality evaluation measures. The PSNR is the one most commonly used because of its interesting characteristics (it is simple, has clear physical meaning, is parameter free, and performs superbly in the optimization context) [2]. It is defined in eq. (1).

$$\mathrm{PSNR}(\mathbf{X},\mathbf{Y}) = 10 \cdot \log_{10}\left(\frac{\mathbf{X}_{max}^2}{\mathrm{MSE}(\mathbf{X},\mathbf{Y})}\right) \quad (1)$$

$$\mathrm{MSE}(\mathbf{X},\mathbf{Y}) = \frac{1}{\mathrm{N_P}} \cdot \sum_{m,n} \left(\mathbf{X}(m,n) - \mathbf{Y}(m,n)\right)^2 \quad (2)$$

where $\mathbf{X}$ and $\mathbf{Y}$ denote the reference and distorted images respectively, $\mathbf{X}_{max}$ is the maximum possible pixel value of the reference image $\mathbf{X}$ (the minimum pixel value is assumed to be zero), and $\mathrm{N_P}$ is the number of pixels in each of the images.

In spite of the widespread use of the PSNR, it cannot adequately reflect the human perception of image fidelity, that is, a large PSNR gain may result in a small improvement in visual quality. As a result, many other quality measures have been developed.

The PSNR is still popular because of its ability to easily compute quality in decibels (dB). Other error-based techniques, such as WSNR[3], NQM [3], and VSNR [4], are less simple to use, as they follow sophisticated procedures to compute the human visual system (HVS) parameters.

In this paper, we propose a low-complexity error-based approach to calculate quality very accurately in the discrete wavelet domain. The method does not require computation of any HVS parameters or image statistics, yet it predicts an accurate quality score at a desired viewing distance. Our development of the proposed approach was based on two major observations. First, the computational complexity of the assessment techniques capable of accurately predicting quality scores is very high, because of the need to extract HVS parameters or gather image statistics. Second, the techniques that apply a multi-resolution transform (like VSNR) decompose the input image into a large number of resolutions (five or more). Since the HVS is a complex system that is not completely known to us, properly combining the various bands perfectly into the final metric is difficult. Therefore, a simpler approach, but a carefully designed one, may achieve accuracy close to that of the complex methods.

The rest of paper is organized as follows. In section II, we describe the proposed approach for quality calculation. In section III, the experimental results are presented. Finally, our concluding remarks are given in section IV.

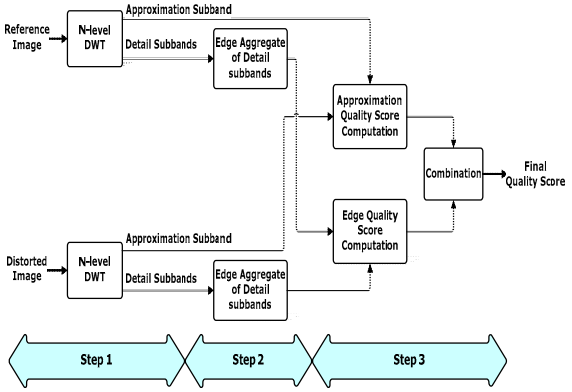

Fig. 1. The proposed quality calculation system in the discrete wavelet domain.

II. Description of The Proposed Method

A. Computation Steps

Let **X** and **Y** denote the reference and distorted images respectively. The proposed procedure for computing the PSNR is depicted in the diagram in Fig. 1, and is described and explained in the following steps:

Step 1. We apply an N-level discrete wavelet transform (DWT) on both the reference and distorted images based on the Haar wavelet filter. With N-level decomposition, the approximation subbands $\mathbf{X}_{A_N}$, and $\mathbf{Y}_{A_N}$, as well as a number of detail subbands, are obtained. We chose the Haar filter for its simplicity and good performance. That simplicity imposes negligible computational burden on the algorithm. Based on our simulations, the Haar wavelet also provides more accurate quality scores than other wavelet bases.

We can formulate the required decomposition levels N as follows. When an image is viewed at distance d from a display of height h, we have [5]:

$$f_\theta = \frac{\pi}{180}\frac{\mathrm{d}}{\mathrm{h}} f_s \quad \text{(cycles/degree)} \tag{3}$$

where f_θ is the angular frequency that has a cycle/degree (cpd) unit; and f_s denotes the spatial frequency. For an image of height H pixels, the Nyquist theorem results in eq. (4):

$$\left(f_s\right)_{max} = \frac{\mathrm{H}}{2} \quad \text{(cycles / picture height)} \tag{4}$$

It is known that the HVS has a peak response for frequencies at about 2-4 cpd. We chose $f_\theta = 3$. If the image is assessed at a viewing distance of d = kh, using eq. (3) and eq. (4), we deduce eq. (5):

$$\mathrm{H} \geq \frac{360 f_\theta}{\pi(\mathrm{d/h})} = \frac{360\times 3}{3.14\times k} \approx \frac{344}{k} \tag{5}$$

So, the effective size of an image for human eye assessment should be around (344/k). Accordingly, the minimum size of the approximation subband after N-level decomposition should be approximately equal to (344/k). For an image of size H×W, N is calculated as follows (considering that N must be non negative):

$$\frac{\min(\mathrm{H},\mathrm{W})}{2^{\mathrm{N}}} \approx \frac{344}{k} \Rightarrow \mathrm{N} = \mathrm{round}\left(\log_2\left(\frac{\min(\mathrm{H},\mathrm{W})}{(344/k)}\right)\right) \tag{6}$$

$$\mathrm{N} \geq 0 \Rightarrow \mathrm{N} = \max\left(0, \mathrm{round}\left(\log_2\left(\frac{\min(\mathrm{H},\mathrm{W})}{(344/k)}\right)\right)\right) \tag{7}$$

Step 2. An estimate of the image edges is formed for each image using an aggregate of detail subbands. If we apply the N-level DWT to the images, the edge map (estimate) of image **X** is defined as in eq. (8).

$$\mathbf{X}_{\mathrm{E}}(m,n) = \sum_{\mathrm{L}=1}^{\mathrm{N}} \mathbf{X}_{\mathrm{E,L}}(m,n) \tag{8}$$

where $\mathbf{X}_{\mathrm{E}}$ is the edge map of **X**, and $\mathbf{X}_{\mathrm{E,L}}$ is the image edge map at the decomposition level L, computed as defined in eq. (9). In eq. (9), $\mathbf{X}_{\mathrm{H_L}}$, $\mathbf{X}_{\mathrm{V_L}}$, and $\mathbf{X}_{\mathrm{D_L}}$ denote the horizontal, vertical, and diagonal detail subbands obtained at the decomposition level L for image **X**. $\mathbf{X}_{\mathrm{H_L,A_{N\text{-}L}}}$, $\mathbf{X}_{\mathrm{V_L,A_{N\text{-}L}}}$, and $\mathbf{X}_{\mathrm{D_L,A_{N\text{-}L}}}$ are the wavelet packet approximation subbands obtained by applying an (N–L)-level DWT on $\mathbf{X}_{\mathrm{H_L}}$, $\mathbf{X}_{\mathrm{V_L}}$, and $\mathbf{X}_{\mathrm{D_L}}$ respectively. The parameters μ, λ, and ψ are constant. As the HVS is more sensitive to the horizontal and vertical subbands and less sensitive to the diagonal one, greater weight is given to the horizontal and vertical subbands and a smaller weight to the diagonal subband. We arbitrarily propose $\mu = \lambda = 4.5\psi$ in this paper, which results in $\mu = \lambda = 0.45$ and $\psi = 0.10$ to satisfy eq. (10).

The edge map of **Y** is defined in a similar way to **X**. As an example, Fig. 2 depicts the subbands of image **X** for N=2. The subbands involved in computing the edge map are shown in color in this figure. It is notable that the edge map is intended to be an estimate of image edges. Thus, the most informative subbands are used in forming the edge map, rather than considering all of them. According to our simulations, considering all image subbands in calculating the edge map does not have a significant impact on increasing prediction accuracy.

$$\mathbf{X}_{\mathrm{E,L}}(m,n) = \begin{cases} \sqrt{\mu\cdot\left(\mathbf{X}_{\mathrm{H_L}}(m,n)\right)^2 + \lambda\left(\mathbf{X}_{\mathrm{V_L}}(m,n)\right)^2 + \psi\left(\mathbf{X}_{\mathrm{D_L}}(m,n)\right)^2} & \text{if} \quad \mathrm{L} = \mathrm{N} \\ \sqrt{\mu\cdot\left(\mathbf{X}_{\mathrm{H_L,A_{N\text{-}L}}}(m,n)\right)^2 + \lambda\left(\mathbf{X}_{\mathrm{V_L,A_{N\text{-}L}}}(m,n)\right)^2 + \psi\left(\mathbf{X}_{\mathrm{D_L,A_{N\text{-}L}}}(m,n)\right)^2} & \text{if} \quad \mathrm{L} < \mathrm{N} \end{cases} \tag{9}$$

$$\mu + \lambda + \psi = 1 \tag{10}$$

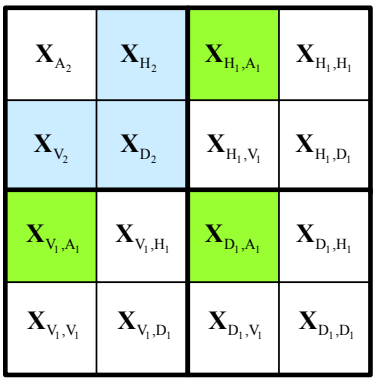

Fig. 2. The wavelet subbands for a two-level decomposed image

Step 3. We determine the final quality score, by calculating the PSNR scores between the approximation subbands and the edge maps using eq. (1) and calling them S_A and S_E respectively, as defined in eq. (11) and eq. (12).

$$S_A = \mathrm{PSNR}(\mathbf{X}_{A_N}, \mathbf{Y}_{A_N}) \tag{11}$$

$$S_E = \mathrm{PSNR}(\mathbf{X}_E, \mathbf{Y}_E) \tag{12}$$

Finally, the approximation and edge quality indices are combined linearly, as defined in eq. (13), to obtain the overall quality score $\mathrm{IQM}_{\mathrm{DWT}}$ between images **X** and **Y**.

$$\mathrm{IQM}_{\mathrm{DWT}}(\mathbf{X},\mathbf{Y}) = \beta \cdot S_A + (1-\beta)\cdot S_E \qquad 0 < \beta \le 1 \tag{13}$$

where $\mathrm{IQM}_{\mathrm{DWT}}$ gives the final quality score of the images in dB, and β is a constant. As the approximation subband contains the content of the main image, β should be close to 1 to give the approximation quality score (S_A) much greater importance. We will set β to 0.85 in our simulations, which means the approximation quality score constitutes 85% of the final quality score and only 15% is made up of the edge quality score. This choice will be justified in section III.

B. Computational Complexity of the Method

In this section, we investigate the computational complexity of our method and compare it to that of the PSNR. To calculate the PSNR between two images, we need 1 subtraction, 1 multiplication (square), and 1 addition for every input pixel. Therefore, this calculation requires 3 operations per input pixel.

The requirement is to perform one operation per input pixel to obtain a desired image subband using the Haar wavelet. That is because the Haar wavelet is actually a simple averaging. For example, to obtain the second level approximation subband, we need to perform 15 additions and 1 division for every 4x4=16 neighboring pixels, which results in 1 operation per input pixel. As we apply the DWT to both the reference and the distorted images, we need $(2+3/(4^N))$ operations per input pixel to calculate S_A (with N-level wavelet decomposition). Since N is greater than or equal to unity ($N \ge 1$), the computational complexity of S_A is less than that of the PSNR. In the next section, we show that S_A is much more accurate than the PSNR in predicting quality scores.

In order to calculate S_E, we first need to compute edge maps for the reference and distorted images. By analyzing eq. (8), eq. (9), and eq. (12), we find that the number of operations per input pixel for calculating S_E is found as in eq. (14) (considering the square root as s operations).

$$\#\,of\ operations\ per\ input\ pixel\ (S_E) = 2N\cdot\left(3+\frac{8+s}{4^N}\right)+\frac{3}{4^N} = 3\cdot\left(2N+\frac{1+\frac{1}{3}N(16+2s)}{4^N}\right) \tag{14}$$

The value of s is about 30 for Intel processor architectures [6]. Thus, at a typical N of 2, the complexity of S_E is about 7.24 times that of the PSNR. A comparison based on a C++ implementation of SSIM shows that, for an image of size 640×480, it is approximately 115 times more complex than the PSNR. Thus, calculation of $\mathrm{IQM}_{\mathrm{DWT}}$ is not computationally expensive relative to other metrics. However, we should mention that S_A alone gives very accurate scores. S_E is most effective when taking into account certain distortions, like fast fading channel distortion. Otherwise, it only marginally improves S_A.

III. Simulation Results and Analysis

The performance of the proposed algorithm for the quality calculation is firstly evaluated in the LIVE Image Quality Assessment Database, Release 2 [7]. This database consists of 779 distorted images derived from 29 original color images using five types of distortion: JPEG compression, JPEG2000 compression, Gaussian white noise (GWN), Gaussian blurring (GBlur), and the Rayleigh fast fading (FF) channel model. The realigned subjective quality data for the database are used in all experiments [7].

In this paper, four performance metrics, along with the statistical F-test, are applied to measure the performance of the objective models. The first metric is the Pearson correlation coefficient (LCC) between the Difference Mean Opinion Score (DMOS) and the objective model outputs after nonlinear regression. The correlation coefficient gives an evaluation of prediction accuracy. We use the five-parameter logistical function defined in [8] for nonlinear regression. The second metric is the root mean square error (RMSE) between the DMOS and the objective model outputs after nonlinear regression. The RMSE is considered as a measure of prediction consistency. The third metric is the Spearman rank correlation coefficient (SRCC), which provides a measure of prediction monotonicity. The fourth metric is the Kendall rank correlation coefficient (KRCC), which is used to measure the association or statistical dependence between two measured quantities in cases where the population distribution of either or both variables is unknown, that is, it is a measure of the degree of correspondence between sets of rankings.

TABLE I: SRCC VALUES FOR DIFFERENT TYPES OF IMAGE DISTORTION IN THE LIVE IMAGE DATABASE

	PSNR	IQM_{DWT} (N=2, β=0.84)	IQM_{DWT} (N=3, β=0.64)	S_A (N=2)	S_A (N=3)
JPEG	0.8812	0.9647	0.9742	0.9793	0.9847
JPEG2000	0.8951	0.9493	0.9585	0.9542	0.9623
GWN	0.9853	0.9820	0.9795	0.9806	0.9719
GBlur	0.7812	0.9228	0.9097	0.9241	0.9010
FF	0.8904	**0.9012**	0.8841	0.8868	0.8505
All Data	0.8754	0.9324	0.9334	0.9288	0.9254

TABLE II: PERFORMANCE COMPARISON OF QUALITY ASSESSMENT MODELS FOR LIVE IMAGE DATABASE (ALL DISTORTED IMAGES INCLUDED)

MODEL	LCC	RMSE	SRCC	KRCC	Residual Variance	*F* statistic
$SSIM_{spatial}$	0.9038	11.6907	0.9104	0.7311	136.8492	**1.3556**
$SSIM_{autoscale}$	0.9446	8.9673	0.9479	0.7963	80.4888	**0.7973**
WSNR	0.9211	10.6353	0.9240	0.7613	113.2543	1.1219
PSNR	0.8700	13.4717	0.8754	0.6861	181.7198	**1.8001**
S_A	0.9288	10.1224	0.9307	0.7723	102.5939	1.0163
IQM_{DWT} (β=0.85)	0.9300	10.0409	0.9325	0.7731	100.9495	1

TABLE III: PERFORMANCE COMPARISON OF QUALITY ASSESSMENT MODELS FOR LIVE VIDEO DATABASE (ALL DISTORTED SEQUENCES INCLUDED)

MODEL	LCC	RMSE	SRCC	KRCC	Residual Variance	*F* statistic
$SSIM_{spatial}$	0.5429	9.2186	0.5251	0.3605	85.5527	1.3918
$SSIM_{autoscale}$	0.7052	7.7827	0.6947	0.5110	60.9765	0.9920
WSNR	0.6706	8.1431	0.6373	0.4553	66.7555	1.0860
PSNR	0.5486	9.1776	0.5233	0.3665	84.7942	1.3795
S_A	0.6875	7.9710	0.6574	0.4689	63.9640	1.0406
IQM_{DWT} (β=0.85)	0.7024	7.8139	0.6793	0.4863	61.4669	1

In order to put the performance evaluation of our proposed scheme into the proper context, we compared our quality assessment algorithm to other quality metrics, including the conventional PSNR, the spatial domain mean SSIM [9], an autoscale version of SSIM that performs downsampling on images to find the right scale for quality assessment [10], and a weighted SNR (WSNR) [3], in which the images are filtered by the CSF specified in [11]. For the LIVE database, we set k at eq.(7) equal to 3, based on the experimental setup and the decomposition level calculated for each image using eq. (7).

First, we verify the number of decomposition levels (N) obtained by eq. (7). Since the image approximation subband plays the major role in our algorithm, the decomposition level N should maximize the prediction accuracy of the approximation quality index S_A. The plots in Fig. 3 show the LCC and SRCC between S_A and the DMOS values for different N. All 779 distorted images of the LIVE database were included in computing the LCC and SRCC. As can be seen from Fig. 3, S_A attains its best performance for N=2.

Table I lists the SRCC values for individual types of distortion. It is observed that the performance of IQM_{DWT} is superior with N=2 when all data (distorted images) is considered, hence N=2 is the appropriate decomposition level for the proposed algorithm. Best performance was obtained by setting β = 0.84 for N=2, and β = 0.64 for N=3 for simulations in Table I. These values of β globally minimize the RMSE for IQM_{DWT}. Since N=2 is the optimal decomposition level and β=0.84 provides the best performance, we suggested β=0.85 for simplicity in the final step (eq. (13)) (small variations of β do not alter the performance notably).

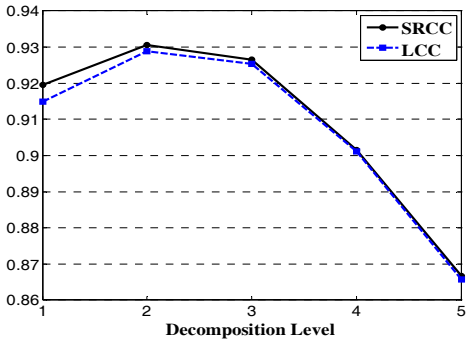

Fig. 3. LCC and SRCC between the DMOS and SA prediction values for various decomposition levels

If eq. (7) is used to compute the appropriate number of decomposition levels for the LIVE database, we obtain N=2. Thus the theoretical value of N obtained in eq. (7) exactly matches the experimental value explained previously.

Table II lists values of performance metrics for each objective model. To assess the statistical significance of each metric's performance compared with that of other metrics, a two-tailed *F*-test was performed on the residual differences between the image quality metric predictions and the DMOS. The *F*-test is used to determine if one metric has significantly larger residuals (greater prediction error) than another [12]. The *F* statistic is defined by a ratio of variances of prediction errors (residuals) from two image quality metrics. The more this ratio deviates from 1, the stronger the evidence for unequal population variances. Values of $F > F_{\text{critical}}$ or $F < 1/F_{\text{critical}}$ indicate that residuals resulting from one quality metric are statistically distinguishable (significantly larger or smaller) from the residuals of another quality metric. $F_{critical}$ is computed based on the number of residuals and a

significance level of α. In this paper, we used $\alpha = 0.05$ which results in $F_{\text{critical}} = 1.151$. Table II also shows the F statistic (F ratio) obtained based on the residuals from each model against the residuals from IQM_{DWT}.

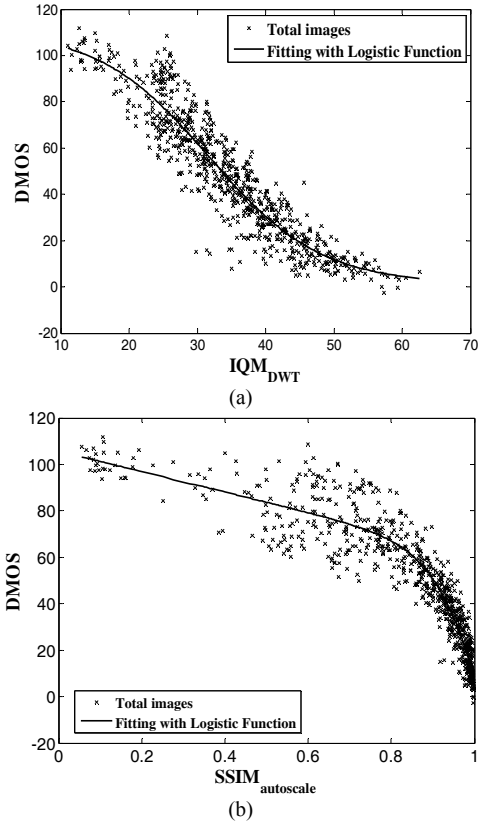

Fig. 4. Scatter plots of the DMOS versus model prediction for all 779 distorted images: (a) IQM_{DWT} model; (b) $SSIM_{autoscale}$ model

As can be seen from Table II, PSNR and $SSIM_{spatial}$ have significantly larger residuals than IQM_{DWT}, and $SSIM_{autoscale}$ has significantly smaller residuals than IQM_{DWT}. Therefore, the proposed IQM_{DWT} generally outperforms the PSNR and $SSIM_{spatial}$, and is slightly better than the WSNR for all types of distortion. $SSIM_{autoscale}$ has the best performance of all the metrics presented, however the performance of IQM_{DWT} is close to it, especially if it is considered separately for JPEG, JPEG2000, and Gaussian noise distortions. Fig. 4 displays the scatter plots of the DMOS versus the IQM_{DWT} and $SSIM_{autoscale}$ predictions for all the distorted images. Fig. 4 shows that, in contrast to the S$SIM_{autoscale}$ predictions, where most of the scores are concentrated between 0.8 and 1, the image quality scores are scattered nearly monotonically along the plot for the IQM_{DWT} model predictions.

Finally, to more effectively verify the accuracy of the proposed algorithm for quality calculation and also the correctness of the formula defined for deciding on N (eq. (7)), we tested the performance of our algorithm on the LIVE VIDEO Quality Database [13],[14]. To simulate our algorithm, the parameter N was computed using eq. (7) and β was used as before, i.e. $\beta = 0.85$. Table III lists the results of the performance evaluation on the LIVE video database. The results in Table III confirm that the proposed IQM_{DWT} achieves a performance close to SSIM and much better than the conventional PSNR, and also that eq. (7) gives a reasonable N for the algorithm presented.

Although space did not allow presenting the detailed tables for each type of video distortion, the performance of $SSIM_{autoscale}$ is better than IQM_{DWT} for video compression distortions, but IQM_{DWT} performs better quality prediction for video distortions caused by transmission over wireless networks.

Note that S_A is parameterless and therefore was not trained on either databases. The same β we had selected for the image database, performed well on the video database, making IQM_{DWT} more accurate than S_A on both databases. Also, the difference in accuracy between these two metrics is higher for the video database, supporting the validity of the proposed parameter value.

A C++ implementation of various quality metrics showed that S_A was more than 50 times faster than $SSIM_{spatial,}$ making it a very low complexity higher accuracy alternative.

IV. Conclusion

In this paper, we proposed a simple metric to calculate the image quality (in dB) accurately in the discrete wavelet domain using the Haar wavelet. In contrast to the error-based quality assessment methods that take into account HVS characteristics to predict the quality scores, the proposed algorithm does not use any HVS parameters. The main idea behind the proposed method is to compute the errors between images after the appropriate level of wavelet decomposition. We also proposed a formula to calculate the required level of wavelet decomposition at a desired viewing distance (where the DMOS is gathered) for the algorithm.

The proposed method was tested on two different well-known databases. The results showed that our method predicts quality scores more accurately than the conventional PSNR and is competitive with SSIM. Furthermore, an analysis of our method shows that it has the potential for calculating quality with lower complexity than the PSNR. Thus, it can be used efficiently in real-time applications, where we are interested in obtaining accurate quality in decibels and with low computational complexity.

References

[1] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, 1st ed., USA: Morgan and Claypool, 2006.

[2] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[3] N. D. 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, 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, 2007.

[5] Y. Wang, J. Ostermann, and Y. Q. Zhang, *Video Processing and Communications*, New Jersey: Prentice-Hall, 2002.

[6] Intel® 64 and IA32 Architectures Optimization Reference Manual, Intel Corporation, November 2009.

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

[8] 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, 2006.

[9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[10] Z. Wang's SSIM Research Homepage [Online]. Available at: http://www.ece.uwaterloo.ca/~z70wang/research/ssim/.
[11] T. Mitsa and K. L. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 301-304, 1993.
[12] Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II VQEG, Aug. 2003 [Online]. Available at: http://www.vqeg.org.
[13] K. Seshadrinathan, R. Soundararajan, A. C. Bovik , and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video," *IEEE* Trans*. Image Process.*, vol. 19, no. 6, pp. 1427-1441, 2010.
[14] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "A Subjective Study to Evaluate Video Quality Assessment Algorithms," *SPIE Proceedings Human Vision and Electronic Imaging*, Jan. 2010.

Soroosh Rezazadeh received BSc and MSc in Electrical Engineering (Telecommunication Eng.) from Shiraz University, Iran in 2003 and 2006 respectively. Currently, He is doing PhD at the Software and IT Engineering Department, École de technologie supérieure (ÉTS).

Stéphane Coulombe received a B.Eng. in Electrical Engineering from the École Polytechnique de Montréal, Canada, in 1991, and a Ph.D. from INRS-Telecommunications, Montréal, in 1996. He is a Professor at the Software and IT Engineering Department, École de technologie supérieure (ÉTS is a part of the Université du Québec network). From 1997 to 1999, he was with Nortel Wireless Network Group in, Montreal, and from 1999 to 2004, he worked with the Nokia Research Center, Dallas, TX, as Senior Engineer and as Program Manager in the Audiovisual Systems Laboratory. He joined ÉTS in 2004, where he currently carries out research and development on video processing and systems, media adaptation, and transcoding. Since 2009, he has held the Vantrix Industrial Research Chair in Video Optimization

.