LOW-COMPLEXITY COMPUTATION OF VISUAL INFORMATION FIDELITY IN THE DISCRETE WAVELET DOMAIN

Soroosh Rezazadeh , Stéphane Coulombe

École de technologie supérieure, Université du Québec, Montréal, Canada
soroosh.rezazadeh.1@ens.etsmtl.ca , stephane.coulombe@etsmtl.ca

ABSTRACT

The visual information fidelity (VIF) index is an objective quality metric that gives very accurate image similarity scores, but at the cost of very high computational complexity. In this paper, a method is presented for calculating VIF in the discrete wavelet domain using the Haar wavelet. The proposed method exploits scalar Gaussian Scale Mixture (GSM) instead of vector GSM for calculating the prediction scores. The complexity of the proposed method is assessed for five different popular image sizes and compared to other methods based on a C/C++ implementation of the algorithms. Experimental results show that the proposed method can compute the visual quality score with less than 30% of the computational complexity of the well-known SSIM index, with greater accuracy than that achieved by the original VIF index method (at about 5% of its computational complexity).

Index Terms— Information fidelity, image quality assessment, discrete wavelet transform

1. INTRODUCTION

Generally speaking, the full-reference (FR) quality assessment of image signals involves two categories of approach: bottom-up and top-down [1]. In the bottom-up approaches, the perceptual quality scores are best estimated by quantifying the visibility of errors. These methods have several important limitations, however, outlined in [1]. In the top-down approaches, the whole human visual system (HVS) is considered as a black box, and the input/output relationship is of interest.

The Structural SIMilarity (SSIM) index was the first method to be developed in the top-down category and was introduced in [2]. The basic form of SSIM is relatively simple, and, because of its simplicity, it has attracted a great deal of attention in recent years and been considered for a wide range of applications. In [3], an Information Fidelity Criterion (IFC) for image quality measurement is presented

which works based on natural scene statistics. In the IFC, the image source is modeled using Gaussian scale mixture (GSM), while the image distortion process is modeled as an error-prone communication channel. Another information-theoretic quality metric is the Visual Information Fidelity (VIF) index [4] (shown in Fig. 1). The VIF index follows the same procedure as the IFC, except that in VIF both the image distortion process and the visual perception process are modeled as error-prone communication channels.
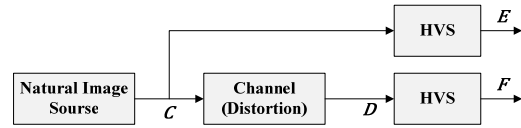

Fig. 1. Block diagram of the VIF index [4]

The VIF index is the most accurate image quality metric according to the performance evaluation of prominent image quality assessment algorithms performed in [5]. In spite of its high level of accuracy, this index has not been given as much consideration as the SSIM index in a variety of applications. This is probably because of its high computational complexity (6.5 times the computation time of the SSIM index according to [4]). Most of the complexity in the VIF index comes from over-complete steerable pyramid decomposition, in which the neighboring coefficients from the same subband are linearly correlated. Consequently, the vector GSM is applied for accurate quality prediction.

In this paper, we propose an approach for calculating VIF in the discrete wavelet domain. The proposed approach is more accurate than the original VIF index, and yet is less complex than the VIF index and even the SSIM index. It applies real Cartesian-separable wavelets and uses scalar GSM instead of vector GSM in modeling the images for VIF computation.

2. VISUAL INFORMATION FIDELITY COMPUTATION

2.1. Scalar GSM-based VIF

Scalar GSM has been described and applied in the computation of IFC [3]. We repeat that procedure here for VIF index calculation using scalar GSM. Considering Fig.1,

let $C^N=(C_1,C_2,...,C_N)$ denote N elements from C, and let $D^N=(D_1,D_2,...,D_N)$ be the corresponding N elements from D. C and D denote the RFs from the reference and distorted signals respectively (as in [3] the models correspond to one subband). C is a product of two stationary random fields (RFs) that are independent of each other:

$$C = \{C_i : i \in I\} = S.U = \{S_i.U_i : i \in I\} \qquad (1)$$

where I denotes the set of spatial indices for the RF, S is an RF of positive scalars, and U is a Gaussian scalar RF with mean zero and variance σ_U^2. The distortion model is a signal attenuation and additive Gaussian noise, defined as follows:

$$D = \{D_i : i \in I\} = GC + V = \{g_i C_i + V_i : i \in I\} \qquad (2)$$

where G is a deterministic scalar attenuation field, and V is a stationary additive zero-mean Gaussian noise RF with variance σ_V^2. The perceived signals in Fig. 1 are defined as follows (see [4]):

$$E = C + N \ , \quad F = D + N' \qquad (3)$$

where N and N' represent stationary white Gaussian noise RFs with variance σ_N^2. If we take the steps outlined in [4] for VIF index calculation considering scalar GSM, we obtain:

$$I(C^N; E^N | S^N = s^N) = I(C^N; E^N | s^N)$$
$$= \frac{1}{2} \sum_{i=1}^{N} log_2 \left(\frac{s_i^2 \sigma_U^2 + \sigma_N^2}{\sigma_N^2} \right) \qquad (4)$$

In the GSM model, the reference image coefficients are assumed to have zero mean. So, for the scalar GSM model, estimates of s_i^2 can be obtained by localized sample variance estimation. The variance σ_U^2 can be assumed to be unity without loss of generality [3]. Thus, eq. (4) is simplified to eq. (5).

$$I(C^N; E^N | s^N) = \frac{1}{2} \sum_{i=1}^{N} log_2 \left(1 + \frac{\sigma_{C_i}^2}{\sigma_N^2} \right) \qquad (5)$$

Similarly, we arrive at eq. (6):

$$I(C^N; F^N | s^N) = \frac{1}{2} \sum_{i=1}^{N} log_2 \left(1 + \frac{g_i^2 \sigma_{C_i}^2}{\sigma_V^2 + \sigma_N^2} \right) \qquad (6)$$

The final VIF index is defined by eq. (7), as in [4] but considering a single subband:

$$VIF = \frac{I(C^N; F^N | s^N)}{I(C^N; E^N | s^N)} \qquad (7)$$

2.2. Description of the proposed approach

Let $\mathbf{X}$ and $\mathbf{Y}$ denote the reference and distorted images respectively. The procedure for calculating the proposed discrete wavelet-based VIF is described, and explained, in the following steps.

Step 1. We perform one-level discrete wavelet decomposition on both the reference and the distorted images using Haar wavelet basis to obtain $\mathbf{X}_A$ and $\mathbf{Y}_A$. With one-level decomposition, the approximation subbands $\mathbf{X}_A$ and $\mathbf{Y}_A$ are still large enough relative to the original images to provide accurate image statistics. Symmetric Haar filters have a generalized linear phase, so the perceptual image structures can be preserved. Moreover, the Haar wavelet lends itself to a fast implementation, which reduces the computational complexity of the algorithm.

Step 2. We calculate the quality score between the approximation subbands $\mathbf{X}_A$ and $\mathbf{Y}_A$ using eq. (7), and call it DWT_VIF_A:

$$DWT\_VIF_A = \frac{\sum_{i=1}^{N} log_2 \left(1 + \frac{g_i^2 \sigma_{\mathbf{x}_{A,i}}^2}{\sigma_{V_i}^2 + \sigma_N^2} \right)}{\sum_{i=1}^{N} log_2 \left(1 + \frac{\sigma_{\mathbf{x}_{A,i}}^2}{\sigma_N^2} \right)} \qquad (8)$$

where N is the number of samples in the approximation subband, $\mathbf{x}_{A,i}$ is the ith image patch within the approximation subband $\mathbf{X}_A$, and $\sigma_{\mathbf{x}_{A,i}}^2$ is the variance of $\mathbf{x}_{A,i}$. The noise variance σ_N^2 is set to 5 in our approach. The parameters g_i and $\sigma_{V_i}^2$ are estimated as described in [3], which results in eq. (9) and eq. (10).

$$g_i = \frac{\sigma_{\mathbf{x}_{A,i}\mathbf{y}_{A,i}}}{\sigma_{\mathbf{x}_{A,i}}^2 + \varepsilon} \qquad (9)$$

where $\sigma_{\mathbf{x}_{A,i}\mathbf{y}_{A,i}}$ is the covariance between image patches $\mathbf{x}_{A,i}$ and $\mathbf{y}_{A,i}$, and ε is a very small constant to avoid instability when $\sigma_{\mathbf{x}_{A,i}}^2$ is zero. In our approach, $\varepsilon = 10^{-20}$.

$$\sigma_{V_i}^2 = \sigma_{\mathbf{y}_{A,i}}^2 - g_i \cdot \sigma_{\mathbf{x}_{A,i}\mathbf{y}_{A,i}} \qquad (10)$$

All the statistics (the variance and covariance of image patches) are computed within a local Gaussian square window, which moves (pixel-by-pixel) over the entire approximation subbands $\mathbf{X}_A$ and $\mathbf{Y}_A$. As with the SSIM index [2], a Gaussian sliding window is defined, in this case $\mathbf{W} = \{w_k | k = 1,2,\cdots,M\}$, with a standard deviation of 1.5 samples, normalized to unit sum. Because of the smaller resolution of the subbands in the wavelet domain, we can extract accurate local statistics with a small sliding window. We show in the next section that the DWT_VIF_A can provide accurate scores with a window of 3×3 (M=9). A small window size reduces the number of computations required to obtain local statistics and consequently reduces the complexity of the algorithm.

Step 3. In this step, an estimate of the image edge is formed for each image using an aggregate of detail subbands (eqs 11 and 12).

$$\mathbf{X}_E(m,n) =$$
$$\sqrt{0.45 \cdot \mathbf{X}_H^2(m,n) + 0.45 \cdot \mathbf{X}_V^2(m,n) + 0.1 \cdot \mathbf{X}_D^2(m,n)} \qquad (11)$$

$$\mathbf{Y}_E(m,n) =$$
$$\sqrt{0.45 \cdot \mathbf{Y}_H^2(m,n) + 0.45 \cdot \mathbf{Y}_V^2(m,n) + 0.1 \cdot \mathbf{Y}_D^2(m,n)} \qquad (12)$$

where $\mathbf{X}_E$ and $\mathbf{Y}_E$ represent the edge maps of $\mathbf{X}$ and $\mathbf{Y}$ respectively; (m,n) shows the sample position within the wavelet subbands; $\mathbf{X}_H$, $\mathbf{X}_V$, and $\mathbf{X}_D$ denote horizontal, vertical, and diagonal detail subbands of image $\mathbf{X}$; $\mathbf{Y}_H$, $\mathbf{Y}_V$, and $\mathbf{Y}_D$ are detail subbands of image $\mathbf{Y}$.

Step 4. The similarity between image edge aggregates $\mathbf{X}_E$ and $\mathbf{Y}_E$ is calculated as in step 2, and called DWT_VIF_E.

Step 5. Finally, the overall quality measure between images $\mathbf{X}$ and $\mathbf{Y}$ is obtained by linear combination of the approximation and edge similarity scores in steps 2 and 4:

$$DWT\_VIF = \alpha \cdot DWT\_VIF_A + (1 - \alpha) \cdot DWT\_VIF_E \quad (13)$$
$$0 < \alpha \le 1$$

where DWT_VIF gives the final similarity score of images in the range [0,1], and α is a constant. As the approximation subband contains main image contents, α should be close to one. We set $\alpha = 0.93$ for our experiments.

3. SIMULATION RESULTS AND ANALYSIS

3.1. Verification of accuracy

Evaluation of the performance of the proposed method is carried out on *LIVE Image Quality Assessment Database Release 2* [6]. This database consists of 779 distorted images derived from 29 original color images using five types of distortion. Distortion types are JPEG compression, JPEG2000 compression, Gaussian white noise, Gaussian blurring, and the Rayleigh fast fading channel model. The realigned subjective quality data for the database are used in all experiments [6].

In this paper, three performance metrics are adopted to measure the performance of objective models. The first is the correlation coefficient (CC) between the Difference Mean Opinion Score (DMOS) and the objective model outputs after nonlinear regression. For the latter, we use the five-parameter logistical function defined in [5]. The second is the root mean square error (RMSE) between DMOS and the objective model outputs after nonlinear regression. The third is the Spearman rank order correlation coefficient (ROCC).

In order to put the performance evaluation of our method in the proper context, we compared the proposed method with other quality metrics, including PSNR, the SSIM index [2], and the VIF index [4]. In the DWT_VIF simulation, we used the Haar wavelet and a Gaussian window size of 3×3. The DWT_VIF parameters were selected to minimize the complexity and RMSE as much as possible. The other metrics were implemented and simulated with the default parameters described in their reference papers.

To better understand the effect of the discrete wavelet transform in quality assessment, we considered DWT_VIF_E and DWT_VIF_A as separate objective quality assessment models. Table 1 lists performance metrics values for each of these models. The CC value for DWT_VIF_A is 0.9649 which is higher than the CC value of the VIF index (0.9593)

defined in [4]. As is clear from the results, DWT_VIF_A is quite efficient. The reason is that the bulk of the useful image information is concentrated in the first-level approximation subband. It can be seen that the performance of DWT_VIF_E is relatively good and nearly the same as that of the SSIM index. This observation confirms the validity of the edge maps defined in eq. (11) and eq. (12). The values of the performance metrics of DWT_VIF are slightly better than those of DWT_VIF_A, which means that DWT_VIF_E is dominated by DWT_VIF_A when the two are combined. This shows that we can calculate the similarity of images with very good precision just by considering their first-level approximation subband. It should be mentioned that, in our experiments, computing the final metric by combining the visual information of the subbands as defined in [4] instead of using eq. (13) does not improve the performance values of DWT_VIF_A. It actually reduces them to CC=0.9518, ROCC=0.9559, and RMSE=8.3830. Fig. 1 shows the scatter plots of DMOS versus DWT_VIF predictions for all the distorted images. It is evident that the DWT_VIF prediction is consistent with the subjective scores.

Table 1. Performance comparison of image quality assessment models (all 779 distorted images included)

Model	CC	ROCC	RMSE
PSNR	0.8701	0.8756	13.4685
Mean SSIM	0.9041	0.9104	11.6736
VIF	0.9593	0.9635	7.7122
DWT_VIF$_E$	0.9039	0.9161	11.6883
DWT_VIF$_A$	0.9649	0.9665	7.1763
DWT_VIF	0.9651	0.9671	7.1561

We also tested DWT_VIF with the previously defined parameters for other wavelet filters. We observed that the other wavelet bases were not as good as the Haar basis and reduced the performance of the metric. For example, Daubechies 9/7 wavelet results in CC=0.9470, ROCC=0.9455, and RMSE=8.7806 for DWT_VIF_A. Although these values are lower than those for the VIF index, they are still quite acceptable, and higher than the SSIM performance metrics values.

3.2. Verification of complexity

In order to verify the computational complexity of the proposed method, the execution time of the algorithm based on the elapsed CPU time is measured. As mentioned in [4], the computation time of the VIF index proposed in [4] is much higher (by about 6.5x) than that of the SSIM index in [2]. Thus, we just selected the SSIM index as a benchmark for evaluating timing (complexity) performance. To assess the method's complexity, only the computation time of DWT_VIF_A is measured. The reason for this is that DWT_VIF_A performs very well as a separate metric (as can be seen from Table 1) and its complexity is half that of

DWT_VIF (since most of the complexity is due to VIF computation on the aggregates of $\mathbf{X_E}$ and $\mathbf{Y_E}$).

Although MATLAB is a useful tool for validating the quality performance of a metric, it does not always provide an accurate indication of computing complexity. Therefore, we used C/C++ implementations for timing measurement. For developing quality metrics in C/C++, the Open Computer Vision (OpenCV) library is exploited [7]. The OpenCV consists of a collection of more than 500 algorithms written in C functions and a few C++ classes for real-time computer vision and image processing. We measured the time required for the luminance component of images with five popular sizes. Table 2 shows the execution time of quality assessment models in C/C++ implementations.

As can be observed from the table, the execution time for calculating *DWT_VIF$_A$* is less than 30% of the execution time for SSIM calculation for different image sizes, while its accuracy is greater than that of the VIF index (at about 5% of its complexity). The effect of the discrete wavelet transform in the less complex *DWT_VIF$_A$* can be considered from three points of view. First, the resolution of the approximation subband is a quarter that of the original image. Second, due to the smaller resolution of the subbands in the wavelet domain, we can extract accurate local statistics with a small sliding window size of 3×3. Third, the employment of orthogonal wavelet decomposition in the proposed method makes it possible to calculate VIF using the scalar GSM model instead of vector GSM, which leads to simpler calculation of VIF.

Finally, it should be mentioned that all simulations were run on a desktop PC with a 2.66-GHz Intel Core 2 Duo CPU and 4 GB of RAM. For C/C++ simulations, we used Microsoft Visual Studio 2005 Professional Edition.

4. CONCLUSION

In this paper, we proposed a low-complexity method for calculating visual information fidelity (VIF) in the discrete wavelet domain using the Haar wavelet. This discrete wavelet allowed us to calculate VIF by applying the scalar GSM model of images instead of vector GSM. We also defined an edge map for the images and investigated the effect of wavelet subbands in the visual quality assessment. Our results show that the first-level approximation subband of decomposed images plays an important role in improving quality assessment performance and also in complexity reduction. The C/C++ implementation of our proposed method shows that the VIF computed based only on the approximation subbands results in higher accuracy than the original VIF proposed in [4], while its complexity is less than 30% of the well-known SSIM index.

Finally, it is worth noting that, since the effect of detail subbands is low in the linear combination as investigated, this proposed method can be improved by devising a technique to determine α adaptively in eq. (13).

5. REFERENCES

[1] Z. Wang, A.C. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool, United States, 2006.

[2] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600-612, April 2004.

[3] H.R. Sheikh, A.C. Bovik, and G. de Veciana, "An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117-2128, December 2005.

[4] H.R. Sheikh, A.C. Bovik, "Image Information and Visual Quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430-444, February 2006.

[5] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing*, vol. 15, pp. 3441-3452, November 2006.

[6] 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.

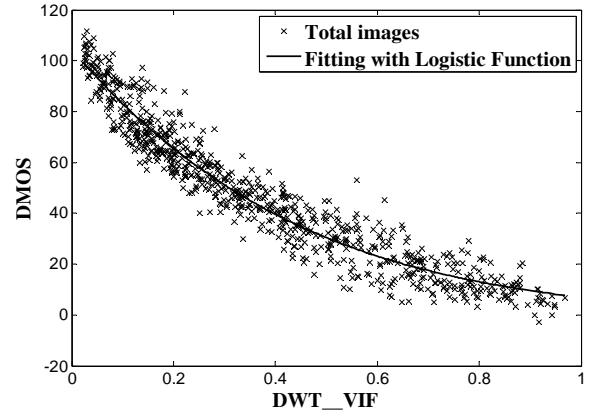[7] OpenCV: Open Source Computer Vision, available at: http://opencv.willowgarage.com/wiki/.

Fig. 1. Scatter plots of DMOS versus model prediction for all 779 distorted images

Table 2. Execution time of quality assessment models measured in seconds

	SSIM	DWT_VIF$_A$	$\dfrac{\text{DWT\_VIF}_A}{\text{SSIM}}$
176×144	0.003465	0.000943	0.2722
320×240	0.010397	0.002698	0.2595
640×480	0.045733	0.011648	0.2547
1280×720	0.149188	0.039575	0.2653
1920×1080	0.336674	0.092802	0.2756