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

INTRODUCTION

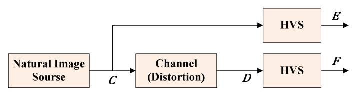
❑ Visual Information Fidelity (VIF)

- ➢ VIF index is the most accurate image quality metric
- ➢ The VIF index applies over-complete steerable pyramid decomposition to images, and exploits vector Gaussian Scale Mixture (GSM) model of images for quality prediction
- ✘ The computational complexity of VIF is very high (about 6.5 times the computation time of the SSIM index)

❑ Proposed Method

- ➢ Calculates VIF in the discrete wavelet domain using Haar wavelet
- ➢ Applies the scalar GSM model of images
- ✓ **Calculates quality with greater accuracy than that achieved by the original VIF index method (at about 5% of its computational complexity)**

SCALAR GSM-BASED VIF



❑ Let C and D denote the RFs from the reference and distorted signals respectively:

$$C = \{C_i : i \in I\} = S \cdot U = \{S_i \cdot U_i : i \in I\}$$

$$D = \{D_i : i \in I\} = GC + V = \{g_i C_i \cdot + V_i : i \in I\}$$

I: set of spatial indices for the RF; S: RF of positive scalars; U: a Gaussian scalar RF with mean zero and variance σ_U^2

G: a deterministic scalar attenuation field; V is a stationary additive zero-mean Gaussian noise RF with variance σ_V^2

SCALAR GSM-BASED VIF (cont.)

❑ C^N : N elements from C and D^N : the corresponding N elements from D

$$C^N = (C_1, C_2, \ldots, C_N) \quad , \quad D^N = (D_1, D_2, \ldots, D_N)$$

❑ Let N and N' represent stationary white Gaussian noise RFs with variance σ_N^2

$$E = C + N \quad , \quad F = D + N'$$

❑ Mutual information between C^N and E^N:

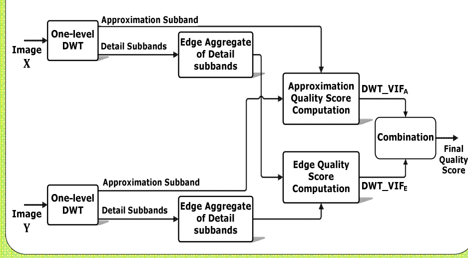
$$I(C^N; E^N | S^N = s^N) = I(C^N; E^N | s^N) = \frac{1}{2}\sum_{i=1}^{N} log_2\left(\frac{s_i^2\sigma_U^2 + \sigma_N^2}{\sigma_N^2}\right)$$

❑ The reference image coefficients are assumed to have zero mean

- ➢ s_i^2 can be estimated by localized sample variance
- ➢ σ_U^2 can be assumed to be unity

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

THE PROPOSED APPROACH



DESCRIPTION OF THE APPROACH

❑ Step 1- Apply one-level discrete wavelet transform, using Haar wavelet filter, on both reference image $\mathbf{X}$ and distorted image $\mathbf{Y}$

❑ Step 2- Form edge aggregate of detail subbands

- ➢ for image $\mathbf{X}$: $\mathbf{X}_E$ and for image $\mathbf{Y}$: $\mathbf{Y}_E$

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

❑ Step 3- Calculate quality score between approximation subbands $\mathbf{X}_A$ and $\mathbf{Y}_A$: DWT_VIF_A

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

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

In our approach: $\sigma_N^2 = 5 \quad , \quad \varepsilon = 10^{-20}$

- ➢ $\sigma_{\mathbf{x}_{A,i},\mathbf{y}_{A,i}}$ is the covariance between image patches $\mathbf{x}_{A,i}$ and $\mathbf{y}_{A,i}$;
- ➢ statistics are computed within a 3 3 Gaussian sliding window with a standard deviation of 1.5 samples, normalized to unit sum

❑ Step 4- Calculate quality score between image edge aggregates $\mathbf{X}_E$ and $\mathbf{Y}_E$: DWT_VIF_E

$$DWT\_VIF = \alpha \cdot DWT\_VIF_A + (1-\alpha) \cdot DWT\_VIF_E$$

$$0 < \alpha \le 1$$

SIMULATIONS

❑ Performance of the proposed method is carried out on *LIVE Image Quality Assessment Database Release* 2

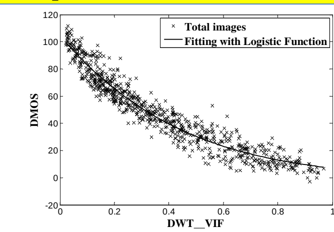
- ➢ 779 distorted images including 5 types of distortion

❑ To verify the models prediction accuracy:

- ➢ **Three performance metrics** computed between DMOS and the objective model outputs after non-linear regression

RESULTS

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



Scatter plot of DMOS versus model prediction

❑ Daubechies 9/7 wavelet results in (for DWT_VIF_A):

CC = 0.9470 , ROCC = 0.9455 , RMSE = 8.7806

❑ To verify the complexity of the proposed method:

- ➢ The elapsed CPU time is measured in seconds
- ➢ The SSIM index selected as benchmark
- ➢ Quality metrics implemented in C/C++ language using OpenCV library
- ➢ Five popular image sizes used for timing measurement

Image Size	SSIM	DWT_VIF_A	$\frac{DWT\_VIF_A}{SSIM}$
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