Université du Québec
École de technologie supérieure
Department of software and IT engineering

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

Authors:

Soroosh Rezazadeh , Stéphane Coulombe

ÉTS
le génie pour l'industrie





2011 3rd International Conference on Signal Acquisition and Processing

AGENDA

1. **Introduction to error-based visual quality assessment**
   - Full-reference image quality assessment
   - Error-based quality metrics
2. **The Proposed Method**
   - Quality calculation framework
   - Description of computation steps
   - Computational complexity of the algorithm
3. **Experimental Results**
   - Image database
   - Video database
4. **Conclusion**

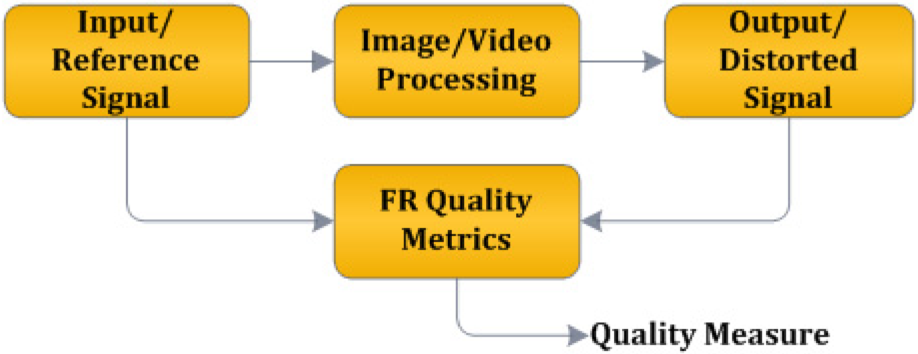
Université du Québec
École de technologie supérieure
Department of software and IT engineering

1. Introduction to Error-Based Visual Quality Assessment

ÉTS
le génie
pour l'industrie

Full-Reference Image Quality Assessment

- **Image quality assessment strategies:**
  - Subjectively by human viewers
  - Objectively by mathematical models



- **Full-reference (FR) quality assessment of image signals:**
  - Bottom-up (error-based) approach:
    - Perceptual quality scores are estimated by quantifying the visibility of errors: PSNR, WSNR, NQM, and VSNR
  - Top-down approach:
    - The whole HVS is considered as a black box and its input/output relationship is of interest: SSIM, and VIF
      - higher prediction accuracy, but high computational complexity

Error-Based Quality Metrics

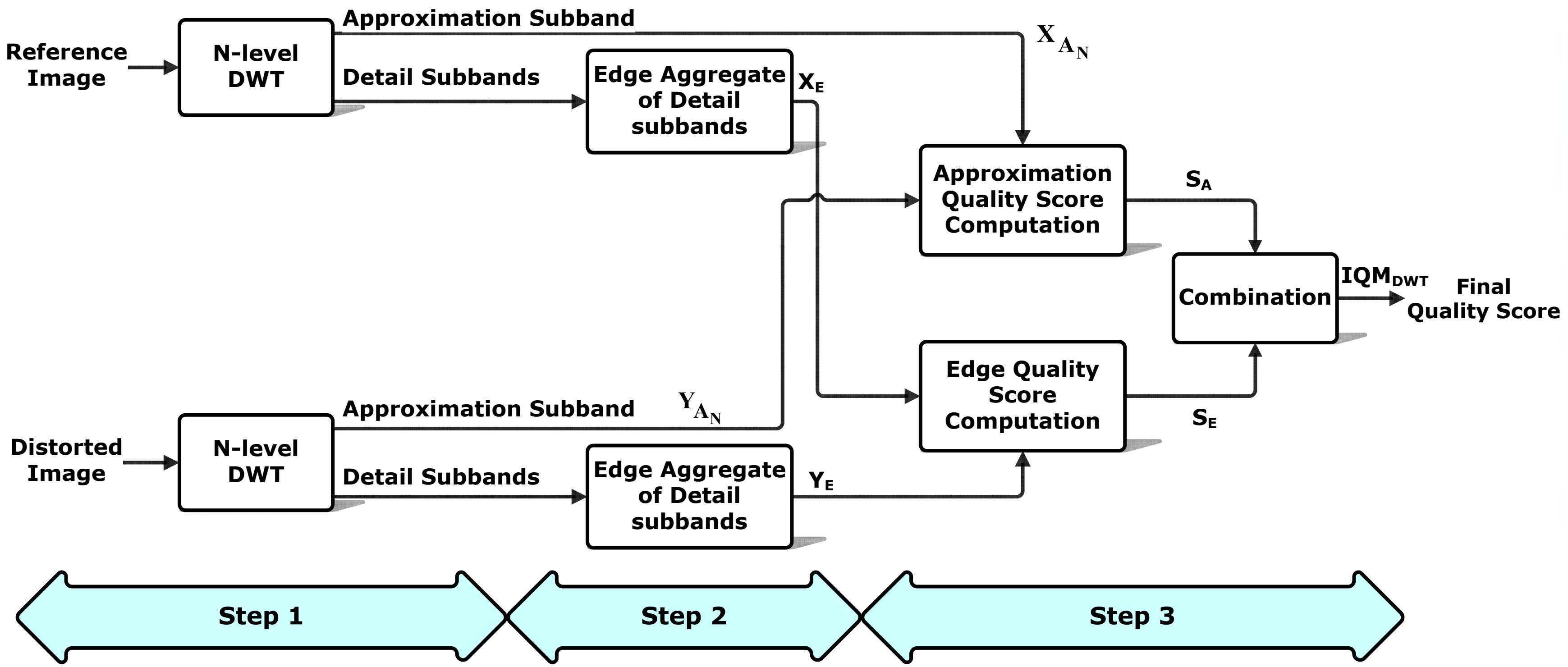
- **Peak Signal-to-Noise Ratio (PSNR):**
  - Most commonly used image and video quality metric
    - PSNR (MSE) is simple, parameter free, and has a clear physical meaning
  - PSNR does not accurately reflect the perceived image/video quality
    - A large gain in PSNR may result in a small improvement in visual quality

- **Disadvantages of Error-based Metrics (such as WSNR, and VSNR):**
  - Accurate quality prediction techniques need to extract HVS parameters or image statistics
    - high computational complexity
  - HVS is a complex system and the sensitivity of the HVS to different scales or subbands is not completely known to us
    - effective combining of various subbands distortions into a final score is difficult
  - **A simpler approach, but a carefully designed one, may achieve accuracy close to that of the complex methods**

2. The Proposed Method

ÉTS le génie pour l'industrie

The Proposed Quality Calculation Framework

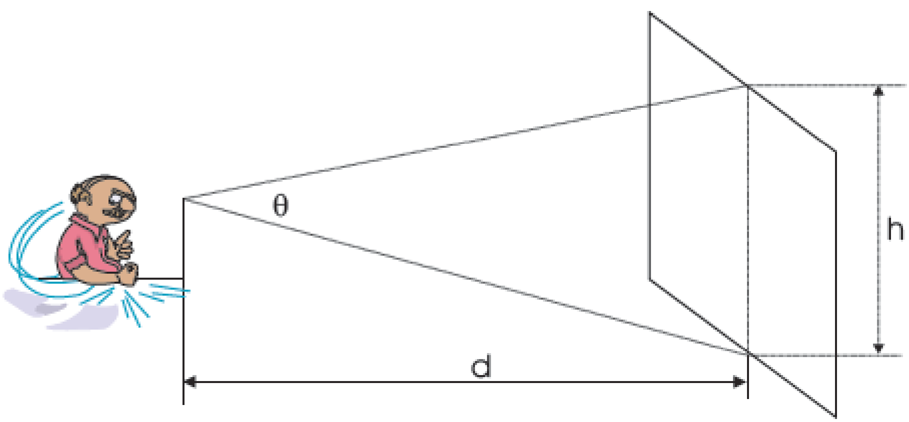


Computation Steps: Wavelet Decomposition

- **Step 1:** N-level wavelet decomposition using Haar wavelet filter
  - Haar basis is used because of simplicity and symmetry
  - Good performance ⇒ the Haar wavelet provides more accurate quality scores than other wavelet bases
- For an image of size H×W:
  - Peak response frequency of HVS is at about 3 cpd

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

N = the required decomposition levels
H = height of the image (pixels)
W= width of the image (pi
h = display height
d = viewing distance = kh



Computation Steps: Edge Map Definition

- **Step 2:** the edge map (estimate) of image **X** $\Leftrightarrow$ $\mathbf{X}_{\mathrm{E}}(m,n)=\sum_{\mathrm{L}=1}^{\mathrm{N}}\mathbf{X}_{\mathrm{E,L}}(m,n)$
  - $\mathbf{X}_{\mathrm{E,L}}$ is the image edge map at the decomposition level L

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

$$\mu+\lambda+\psi=1$$

The wavelet subbands for a two-level decomposed image (N=2)

$\mathbf{X}_{\mathrm{A_2}}$	$\mathbf{X}_{\mathrm{H_2}}$	$\mathbf{X}_{\mathrm{H_1,A_1}}$	$\mathbf{X}_{\mathrm{H_1,H_1}}$
$\mathbf{X}_{\mathrm{V_2}}$	$\mathbf{X}_{\mathrm{D_2}}$	$\mathbf{X}_{\mathrm{H_1,V_1}}$	$\mathbf{X}_{\mathrm{H_1,D_1}}$
$\mathbf{X}_{\mathrm{V_1,A_1}}$	$\mathbf{X}_{\mathrm{V_1,H_1}}$	$\mathbf{X}_{\mathrm{D_1,A_1}}$	$\mathbf{X}_{\mathrm{D_1,H_1}}$
$\mathbf{X}_{\mathrm{V_1,V_1}}$	$\mathbf{X}_{\mathrm{V_1,D_1}}$	$\mathbf{X}_{\mathrm{D_1,V_1}}$	$\mathbf{X}_{\mathrm{D_1,D_1}}$

$\mathbf{X}_{\mathrm{H_2}}$, $\mathbf{X}_{\mathrm{V_2}}$, and $\mathbf{X}_{\mathrm{D_2}}$ (blue boxes) $\Rightarrow \mathbf{X}_{\mathrm{E,2}}$

$\mathbf{X}_{\mathrm{H_1,A_1}}$, $\mathbf{X}_{\mathrm{V_1,A_1}}$, and $\mathbf{X}_{\mathrm{D_1,A_1}}$ (green boxes) $\Rightarrow \mathbf{X}_{\mathrm{E,1}}$

Computation Steps: The Final Quality Score

- the HVS is more sensitive to the horizontal and vertical subbands and less sensitive to the diagonal one

$$\mu = \lambda = 4.5\psi \;\Rightarrow\; \begin{cases} \mu = \lambda = 0.45 \\ \psi = 0.10 \end{cases}$$

- **Step 3:** approximation quality score S_A, and edge quality score S_E

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

- IQM_{DWT} : overall quality score between images **X** and **Y**

$$IQM_{DWT}(\mathbf{X}, \mathbf{Y}) = \beta \cdot S_A + (1 - \beta) \cdot S_E$$
$$0 < \beta \leq 1$$

- the approximation subband contains the main content of image
  - β should be close to 1 to give the approximation quality score greater importance
- We will set β to 0.85 in our simulations

Computational Complexity of the Algorithm

- PSNR calculation needs 3 operations per input pixel:
  - 1 addition, 1 subtraction, and 1 multiplication
- Obtaining a desired image subband using the Haar wavelet needs one operation per input pixel
  - the second level approximation subband, we need to perform 15 additions and 1 division for every 4×4=16 neighboring pixels $\Rightarrow$ 1 operation per input pixel
- S_A calculation needs $2+(3/4^N)$ operations per input pixel ($N \geq 1$)
  - S_A is much more accurate than the PSNR in predicting quality scores (while less complex)
  - C++ implementation $\Rightarrow$ S_A is more than 50 times faster than SSIM
- considering the square root as s operations:

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

  - for Intel processor architectures: $s \approx 30$
    - At a typical N=2, the complexity of S_E is about 7.24 times that of the PSNR
  - C++ implementation $\Rightarrow$ SSIM is approximately 115 times more complex than the PSNR
  - S_E is most effective for certain distortions, like fast fading channel distortion

3. Experimental Results

ÉTS
le génie pour l'industrie

Simulation Results for Image Database

- Performance evaluation carried out on LIVE Image Quality Assessment Database Release 2: 779 distorted images and 5 types of distortions
- Performance measures adopted:
  - Pearson correlation coefficients (LCC) ⇒ prediction accuracy
  - Root mean square error (RMSE) ⇒ prediction consistency
  - Spearman rank order correlation coefficient (ROCC) ⇒ prediction monotonicity
  - Kendall rank correlation coefficient (KRCC) ⇒ association or statistical dependence
    - Nonlinear regression between DMOS and output values of models before calculating performance measures
  - A two-tailed *F*-test on the residual differences between the models predictions and the DMOS
    - $F > F_{critical}$ or $F < 1/F_{critical}$: residuals of one quality metric are statistically distinguishable from the residuals of another quality metric
    - significance level $\alpha = 0.05 \Rightarrow F_{critical} = 1.151$ for 779 images
- Comparison with other metrics :
  - PSNR, spatial mean SSIM, autoscale version of SSIM, and WSNR
  - Viewing distance ratio k=3, based on the database experimental setup

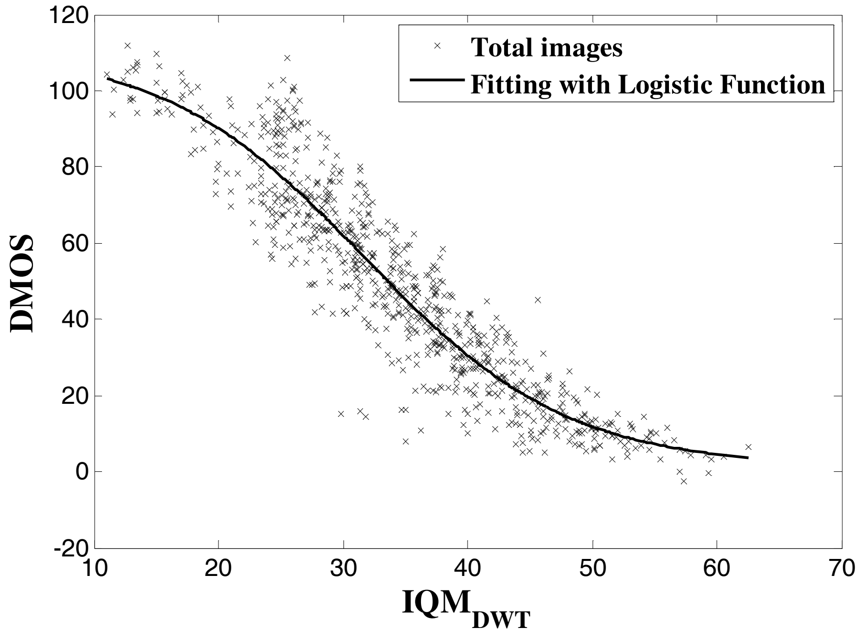
Simulation Results for the Image Database

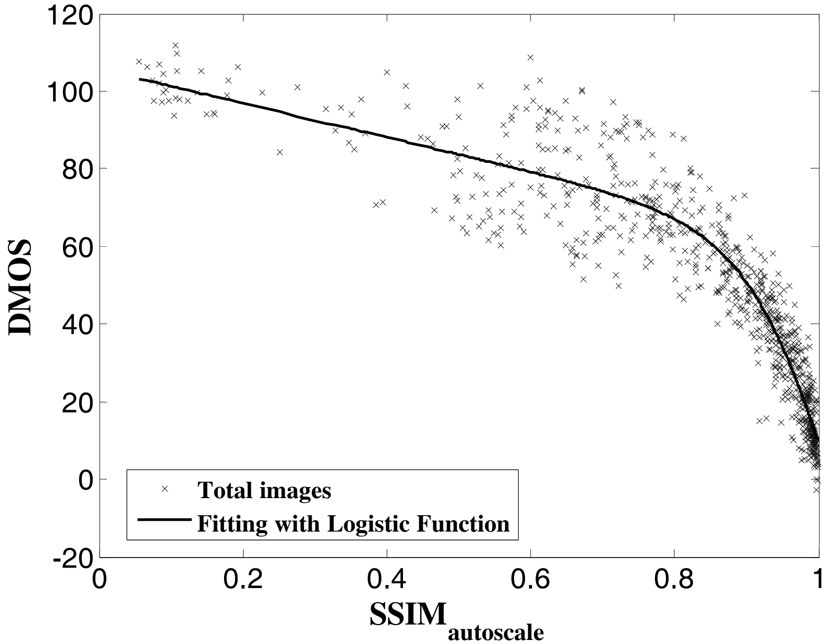
- The IQM_{DWT} generally outperforms the PSNR and $SSIM_{spatial}$, and is slightly better than the WSNR for all types of distortion
- The performance of IQM_{DWT} is close to $SSIM_{autoscale}$, especially if it is considered separately for JPEG, JPEG2000, and Gaussian noise distortions

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

Simulation Results: Scatter Plots

- Scatter plots of DMOS versus model prediction for all 779 distorted images
  - The $SSIM_{autoscale}$ prediction scores are mostly concentrated between 0.8 and 1
  - The IQM_{DWT} quality prediction scores are scattered nearly monotonically along the plot





Simulation Results for the Video Database

- Performance of our algorithm tested on the LIVE Video Quality Database
  - 150 distorted videos and four distortion types:
    - MPEG-2 and H.264 compressions, transmission of H.264 bitstreams through IP networks and through wireless networks
- **The performance of IQM_{DWT} is close to SSIM and much better than the conventional PSNR**

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

4. Conclusion

ÉTS
le génie
pour l'industrie

Conclusion

- A simple DWT-based metric to accurately calculate the image quality (in dB)
  - excellent trade-off between accuracy and computational complexity
  - the proposed algorithm does not use any HVS parameter

- A proposed formula to calculate the required level of wavelet decomposition at a desired viewing distance

- Our proposed method was tested on two well-known databases
  - our method predicts quality scores more accurately than the conventional PSNR and is competitive with SSIM

- An analysis of our method shows that it has the potential for calculating quality with lower complexity than the PSNR

- The proposed method can be used efficiently in real-time applications
  - obtaining accurate quality in decibels with low computational complexity

THANKS FOR YOUR ATTENTION

Error-Based Quality Metrics

- **Peak Signal-to-Noise Ratio (PSNR):**
  - Most commonly used image and video quality metric
    - PSNR (MSE) is simple, parameter free, and has a clear physical meaning
  - PSNR does not accurately reflect the perceived image/video quality
    - A large gain in PSNR may result in a small improvement in visual quality

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

X&Y: reference and distorted images , N_p: # of pixels in each of the images

- **Weighted SNR (WSNR):**
  - Calculates quality in decibels (dB)
  - Works in the Fourier transform domain for HVS modeling
  - Uses the contrast sensitivity function (CSF) as the weighting function to weight the Fourier transforms of the error image
    - CSF is a linear spatially invariant approximation to the HVS
  - ✘ Its computational complexity is high

Error-Based Quality Metrics

- **Visual SNR (VSNR):**
  - After preprocessing, both reference and error images are wavelet decomposed
  - Five levels of wavelet decomposition by using 9/7 biorthogonal filters
  - The contrast detection threshold is computed for each subband to assess the detectability of the distortions
  - VSNR attempts to capture a mid-level property of the HVS known as global precedence

- **Disadvantages of Error-based Metrics:**
  - Accurate quality prediction techniques need to extract HVS parameters or image statistics
    - high computational complexity
  - HVS is a complex system and the sensitivity of the HVS to different scales or subbands is not completely known to us
    - effective combining of various subbands distortions into a final score is difficult
  - **A simpler approach, but a carefully designed one, may achieve accuracy close to that of the complex methods**

The Proposed Algorithm Examples: Lena 256×256

JPEG compressed QF=95

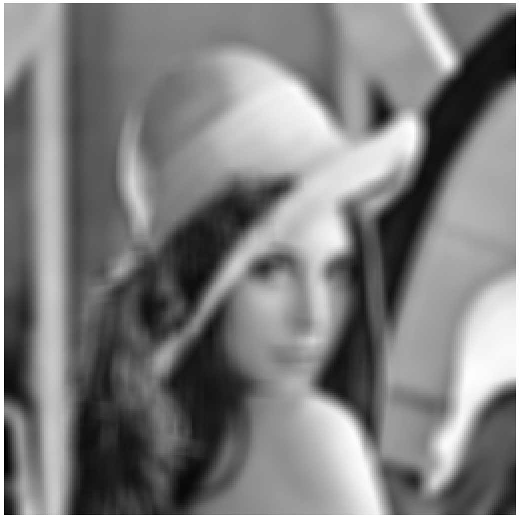
JPEG compressed QF=5

PSNR = 43.1060 $\xleftrightarrow{18.1478\text{ dB}}$ PSNR = 24.9582

$IQM_{DWT} = 51.3717$ $\xleftrightarrow{25.5884\text{ dB}}$ $IQM_{DWT} = 25.7833$

$SSIM_{autoscale} = 0.9831$ $SSIM_{autoscale} = 0.6766$

The Proposed Algorithm Examples: Lena 256×256

3×3 Averaging	9×9 Averaging
	

$\text{PSNR} = 28.6065 \xleftrightarrow{5.5938\ \text{dB}} \text{PSNR} = 23.0127$

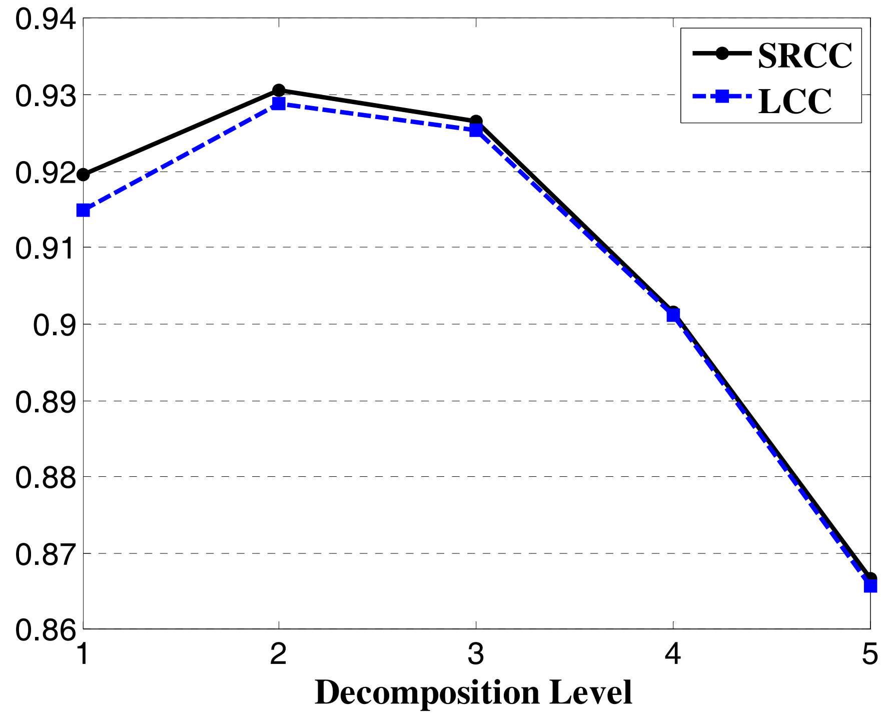
$\text{IQM}_{\text{DWT}} = 31.7243 \xleftrightarrow{8.5202\ \text{dB}} \text{IQM}_{\text{DWT}} = 23.2041$

$\text{SSIM}_{\text{autoscale}} = 0.8582$ $\quad$ $\text{SSIM}_{\text{autoscale}} = 0.6216$

of Decomposition Levels (N) Verification

- 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

LCC and SRCC between the DMOS and S_A prediction values for various decomposition levels



- The theoretical value of N according to the proposed formula is N=2

of Decomposition Levels (N) Verification

SRCC values for individual types of image distortion in the LIVE image database

Distortion	PSNR	IQM_{DWT} N=2, β=0.84	IQM_{DWT} N=3, β=0.64
JPEG	0.8812	0.9647	0.9742
JPEG2000	0.8951	0.9493	0.9585
GWN	0.9853	0.9820	0.9795
GBlur	0.7812	**0.9228**	0.9097
FF	0.8904	**0.9012**	0.8841
All Data	0.8754	0.9324	0.9334

- The theoretical value of N according to the proposed formula is N=2
- β = 0.84 for N=2, and β = 0.64 for N=3 globally minimize the RMSE for IQM_{DWT}
  - Since N=2 is the appropriate decomposition level, we choose β=0.85 for simplicity in the final step
- Small variations of β do not alter the performance notably