Rate-distortion performance evaluation of JPEG-XR

Objective results and proposed methodology for subjective quality assessment

Francesca De Simone, Frederic Dufaux, Touradj Ebrahimi

JPEG XR Camera Raw coding AHG meeting
18-19 February 2009, Tokyo, Japan

Outline of the presentation

- Introduction
  - Quality Assessment (QA) and codec performance evaluation
  - Status
  - Our previous contributions
- Objective QA
  - Test material
  - Codecs and configuration parameters
  - Quality metrics
  - Selected results
- Subjective QA
  - Proposed methodology
  - Test conditions
  - Preliminary results

Introduction

- Quality Assessment (QA) and codec performance evaluation
- Status
- Our previous contributions

QA and codec performance evaluation

- Codec performance evaluation in terms of:
  - Compression efficiency.
  - Computational requirements.
  - Additional functionalities.
- Rate-Distortion (RD) curves = quality measure vs bit per pixel
There are not yet reliable and standard objective methods for image quality assessment.

- Image and video systems complexity
- Human Visual System (HVS) complexity
- Lack of standardization

- **Objective QA** can be performed to provide a first comparison of a wide range of conditions.

- **Subjective QA** needs to be performed as benchmark to validate the results of the objective metrics.

Our previous contributions:

- **JPEG contributions:**
  - F. De Simone et al., "Comparison of PSNR performance of HD Photo and JPEG2000", wg1n3404, JPEG meeting Kobe (Nov. 2007)
  - F. De Simone et al., "Objective evaluation of the rate-distortion performance of JPEG-XR", wg1n4552, JPEG Interim meeting Poitiers (Feb. 2008)
  - F. De Simone et al., "Still image coding algorithms performance comparison: objective quality metrics", wg1n4497, JPEG meeting San Francisco (Apr. 2008)
  - F. De Simone et al., "Objective rate-distortion performance of different JPEG-XR implementations", wg1n4703, JPEG meeting Poitiers (July 2008)

Conference publications:


Test Material – 24 bpp pictures

- 24 bpp pictures (sample pictures from Thomas Richter dataset, 2 different spatial resolutions: 3888x2592, 2592x3888)
- 24 bpp pictures (sample pictures from Microsoft dataset, 6 different spatial resolutions: 4064x2704, 2268x1512, 2592x1944, 2128x2832, 2704x3499, 4288x2848)
JPEG XR vs JPEG2000 vs JPEG:

- JPEG XR (DPK version 1.0):
  - one level overlapping and two level overlapping.
  - 4:4:4 and 4:2:0 chroma subsampling.

- JPEG 2000 (Kakadu version 6.0):
  - default settings (64x64 code-block size, 1 quality layer, no precincts, 1 tile, 9x7 wavelet, 5 decomposition levels).
  - rate control.
  - no visual frequency weighting and visual frequency weighting.
  - 4:4:4 and 4:2:0 chroma subsampling.

- JPEG (IJG version 6b):
  - default settings (Huffman coding).
  - default visually optimized quantization tables.
  - 4:4:4 and 4:2:0 chroma subsampling.

Different JPEG XR implementations:

- JPEG XR DPK version 1.0:
  - different quantization steps for different color channels (default).
  - same quantization steps for different frequency bands (default).

- JPEG XR Reference Software version 1.0:
  - same quantization steps for different color channels (default).
  - same quantization steps for different frequency bands (default).

- JPEG XR Reference Software version 1.2 - i.e. Thomas Richer's version:
  - different quantization steps for different color channels (same as DPK).
  - different quantization steps for different frequency bands (default).
  - new POT (leakage fix described in wg1n4660) (default).

- JPEG XR Microsob implementation described in HDPn21 / wg1n4549:
  - different quantization steps for different color channels (enhanced encoding techniques described in HDPn21 / wg1n4549) (default).
  - different quantization steps for different frequency bands (enhanced encoding techniques of HDPn21 / wg1n4549) (default).
  - new POT (leakage fix described in wg1n4660) (default).

Metric 1: Maximum Pixel Deviation (L_\text{inf})

- Considering RGB color space:
  \[
  L_{\text{inf} R} = \max \{ \text{abs}(\text{Im}_{aR}(x,y) - \text{Im}_{bR}(x,y)) \} \\
  L_{\text{inf} G} = \max \{ \text{abs}(\text{Im}_{aG}(x,y) - \text{Im}_{bG}(x,y)) \} \\
  L_{\text{inf} B} = \max \{ \text{abs}(\text{Im}_{aB}(x,y) - \text{Im}_{bB}(x,y)) \} 
  \]
  \((L_{\text{inf}} \in [0,1])

where:

- \text{Im}_a, \text{Im}_b = \text{pictures to compare}

Metric 2: single channel PSNR

\[
\text{PSNR} = 10 \log_{10} \left( \frac{2^8 - 1}{\text{MSE}} \right)
\]

where:

- \text{MSE} = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} [\text{Im}_a(x,y) - \text{Im}_b(x,y)]^2
- M, N = \text{image dimensions}
- \text{Im}_a, \text{Im}_b = \text{pictures to compare}
- B = \text{bit depth}

- PSNR evaluation considering:
  - R, G and B components
  - Y', C_b and C_r components (ITU-R Rec. BT.601)
### Metric 3: PSNR weighted average (WPSNR)

- PSNR considering weighted summation of the PSNRs evaluated on R, G and B components or Y', C_b and C_r components (ITU-R Rec. BT. 601):

\[
WPSNR = w_1 \cdot PSNR_1 + w_2 \cdot PSNR_2 + w_3 \cdot PSNR_3
\]

where:
- \( w_1 = \frac{1}{3}, \ w_2 = \frac{1}{3}, \ w_3 \) considering R, G, and B components.
- \( w_1 = 0.8, \ w_2 = 0.1, \ w_3 = 0.1 \) considering Y', C_b, and C_r components.

### Metric 3: PSNR weighted average (WPSNR_MSE)

- PSNR considering weighted summation of the MSEs evaluated on R, G and B components or Y', C_b and C_r components (ITU-R Rec. BT. 601):

\[
WPSNR_{MSE} = 10 \log_{10} \left( \frac{(2^B - 1)^2}{w_1 \cdot MSE_1 + w_2 \cdot MSE_2 + w_3 \cdot MSE_3} \right)
\]

where:
- \( w_1 = \frac{1}{3}, \ w_2 = \frac{1}{3}, \ w_3 \) considering R, G, and B components.
- \( w_1 = 0.8, \ w_2 = 0.1, \ w_3 = 0.1 \) considering Y', C_b, and C_r components.

### Metric 3: PSNR weighted average (WPSNR_PIX)

- PSNR considering MSE evaluated on weighted summation of the image R, G and B components:

\[
WPSNR_{PIX} = 10 \log_{10} \left( \frac{1}{M \cdot N} \sum_{x,y} \left[ (\mu_2 - \mu_1)^2 + \frac{1}{\sigma^2} (\mu_1 - \mu_2)^2 \right] \right)
\]

where:
- \( M, N \) = image dimensions
- \( I_m, I_b \) = pictures to compare
- \( B \) = bit depth
- \( w_1 = \frac{1}{3}, \ w_2 = \frac{1}{3}, \ w_3 \) considering R, G, and B components.
- \( w_1 = 0.8, \ w_2 = 0.1, \ w_3 = 0.1 \) considering Y', C_b, and C_r components.

### Metric 4: Mean SSIM (MSSIM) (I)

- Estimate of luminance = “attributes that represent the structure of objects in the scene, independent of the average luminance and contrast”.

\[
\mu = \frac{1}{MN} \sum_{x,y} [I(x,y)]
\]

- Estimate of contrast = standard deviation:

\[
\sigma = \left( \frac{1}{MN - 1} \sum_{x,y} [I(x,y) - \mu]^2 \right)^{1/2}
\]

- Estimate of picture structure:

\[
\left( \frac{I(x,y) - \mu}{\sigma} \right)
\]
Metric 4: Mean SSIM (MSSIM) (II)

\[
\text{SSIM}(I_m, I_m') = \frac{[2\mu_m\mu_m' + C_1][2\sigma_m\sigma_m' + C_2]}{[\mu_m^2 + \mu_m'^2 + C_1][\sigma_m^2 + \sigma_m'^2 + C_2]}
\]

\[\alpha > 0, \beta > 0, \gamma > 0\]

- Luminance comparison function:
  \[
  l(I_m, I_m') = \frac{2\mu_m\mu_m' + C_1}{\mu_m^2 + \mu_m'^2 + C_1}
  \]

- Contrast comparison function:
  \[
  c(I_m, I_m') = \frac{2\sigma_m\sigma_m' + C_2}{\sigma_m^2 + \sigma_m'^2 + C_2}
  \]

- Measure of structural similarity = correlation between \( \frac{m_m - \mu_m}{\sigma_m} \) and \( \frac{m_m' - \mu_m'}{\sigma_m'} \)

Structure comparison function:

\[
\sigma_{m2} = \frac{1}{MN-1} \sum_{x=1}^{M} \sum_{y=1}^{N} (m_m(x,y) - \mu_m)(m_m'(x,y) - \mu_m')
\]

Metric 4: Mean SSIM (MSSIM) (III)

- The SSIM indexing algorithm is applied using a sliding window approach which results in a SSIM index quality map of the image.

- The average of the quality map is called Mean SSIM index (MSSIM).

- Weighted summation of MSSIM indexes evaluated on \( Y', \) \( Cb \) and \( Cr \) components (\( Y'CbCr \) color space - Rec. ITU-R BT.601):

\[
\text{MSSIM} = w_Y \text{MSSIM}_Y + w_{Cb} \text{MSSIM}_{Cb} + w_{Cr} \text{MSSIM}_{Cr}
\]

where:

\[
w_Y = 0.8, \ w_{Cb} = 0.1, \ w_{Cr} = 0.1.
\]

Metric 5: Visual Information Fidelity – Pixel (VIF-P) (I)


- “Image information measure that quantifies the information that is present in the reference image and how much this reference information can be extracted from the distorted image” using statistical approach.

- Reference image \( E \) = output of a stochastic natural source that passes through HVS channel and is processed by the brain

- Test image \( F \) = output of an image distortion channel that distorts the output of the natural source before it passes through the HVS channel

\[
\text{VIF} = \frac{I(C,E)}{I(C,F)} \quad \text{VIF} \in [0,1]
\]

where:

- \( z \) = source model parameters.

- \( \text{VIF-P} \) is a new implementation in a multi-scale pixel domain:
  - computationally simpler than Wavelet domain version.
  - performance slightly worse than Wavelet domain version.
Metric 6: PSNR-HVS-M (I)  

- DCT coefficients of 8x8 pixel blocks X and Y are visually undistinguishable if:
  \[ E_{w}(X-Y) < \max(E_{w}(X), E_{w}(Y)) \]

- Block 8x8 of original image
- Block 8x8 of distorted image
- DCT of difference between pixel values
- Reduction by value of contrast masking
- MSE: calculation of the block


Metric 6: PSNR-HVS-M (II)  

\[ \text{PSNR} = \frac{10\log_{10} \left( \frac{2^{2n} - 1}{\text{MSE}_{H}} \right) \text{dB}} \]

where:
- \( \text{MSE}_{H} = \sum_{m=1}^{M} \sum_{n=1}^{N} [x(m,n) - y(m,n)]^2 \]
- \( M, N = \text{image dimensions} \)
- \( K = \text{constant} \)
- \( x(m,n), y(m,n) = \text{visible difference between DCT coefficient of the original image and distorted image 8x8 blocks, depending upon contrast masking} \)
- \( T_{c} = \text{matrix of correcting factors based on standard visually optimized JPEG quantization tables} \)
- \( B = \text{bit depth} \)

Metric 7: DC Tune  


- developed as a method for optimizing JPEG image compression by computing the JPEG quantization matrices which yields a designated perceptual error
- model of perceptual error based upon DCT coefficients analysis, taking into account:
  - luminance masking.
  - contrast masking.
  - spatial error pooling.
  - frequency error pooling.

Selected results 4:4:4 – JPEG XR vs JPEG2000 vs JPEG  

Average over image dataset of PSNR values

- on R component:
- on G component:
- on B component:
Selected results 4:4:4 – JPEG XR vs JPEG2000 vs JPEG

Average over image dataset of PSNR values

- on Y' component:
- on Cb component:
- on Cr component:

Average over image dataset of WPSNR values

- on RGB components:
- on Y'CbCr components:

Average over image dataset of WPSNR-MSE values

- on RGB components:
- on Y'CbCr components:

Average over image dataset of WPSNR-PIX values

- on RGB components:
- on Y'CbCr components:
### Selected results 4:4:4 – JPEG XR vs JPEG2000 vs JPEG

**Average over image dataset of MSSIM values**

- **on Y component:**
- **on Cb component:**
- **on Cr component:**

**MSSIM**

**bpp (bits/pixel)**

**Selected results**

4:4:4 – JPEG XR vs JPEG2000 vs JPEG

---

**Average over image dataset of VIF-P values**

- **on Y’ component only:**

**VIF-P**

**bpp (bits/pixel)**

---

### Selected results 4:4:4 – different JPEG XR implem.

**Average over image dataset of PSNR values**

- **one level POT**
- **on R component:**
- **on G component:**
- **on B component:**

**PSNR (dB)**

**bpp (bits/pixel)**

---

### Selected results 4:4:4 – different JPEG XR implem.

**Average over image dataset of PSNR values**

- **two levels POT**
- **on R component:**
- **on G component:**
- **on B component:**

**PSNR (dB)**

**bpp (bits/pixel)**
Selected results 4:4:4 – different JPEG XR impl.  

Average over image dataset of PSNR values (one level POT)

- on Y component:
- on Cb component:
- on Cr component:

Average over image dataset of PSNR values (two levels POT)

- on Y component:
- on Cb component:
- on Cr component:

Selected results 4:4:4 – different JPEG XR impl.

Average over image dataset of WPSNR_MSE values

- one level POT:
- two levels POT:

Average over image dataset of MSSIM values (one level POT)

- on Y component:
- on Cb component:
- on Cr component:
Selected results 4:4:4 – different JPEG XR implem.

Average over image dataset of MSSIM values (two levels POT)
on Y component:
on Cb component:
on Cr component:

Subjective QA
- Proposed methodology
- Test conditions
- Preliminary results

Proposed methodology (I)
- **Double Stimulus Continuous Quality Scale (DSCQS) method** [ITU-R Rec. BT.500-11] adapted to deal with the evaluation of still pictures:
  - test picture and its reference are shown at the same time.
  - the assessor is not told about the presence of a reference picture.
  - positions of reference and test pictures are systematically switched.
  - test pairs related to different original contents are always alternated.

Proposed methodology (II)
- when the subject clicks into the active area of the screen a rating window is shown:
Proposed methodology (III)

- the subject has to rate the quality of the two pictures choosing for each a value in between 0 (worse quality possible) to 100 (best quality possible).

Proposed methodology (IV)

- Subjects are checked for visual acuity and color blindness
- Before each session, instructions are provided to subjects and a training session is performed to explain how to use the rating scale
  - contents shown for training are not used for testing
  - data gathered during the training are not included in the final test results
- Some dummy presentations are inserted at the beginning of the test to stabilize subject’s behaviour
  - data gathered from the dummies are not included in the final test results
  - the dummy presentations cover all the quality levels included in the test material
- The test session lasts no more than 20 minutes (including training)

Proposed methodology (V)

- At least 15 subjects
- Subjective data processing:
  - computation of Differential Score (DS):
    - DS = Score for the reference picture – Score for the test picture
  - ANalysis Of Variance (ANOVA) to detect eventual systematic errors and scores normalization to remove them
  - screening to detect outliers (ITU-R Rec. BT.500-11)
  - computation of the Differential Mean Opinion Score (DMOS)

Test conditions

- Eizo CG301W LCD monitor (2560x1600 pixels)
- monitor calibration using color calibration device (EyeOne Display2)
  - Gamut sRGB, white point D65, brightness 120cd/m2, minimum black level.
- controlled lighting system: neon lamps with 6500 K color temperature
- ambient light measurement by EyeOne Display2 tool
Preliminary results (I)

- JPEG XR Microsoft implementation described in HDPn21:
  - different quantization steps for different color channels (enhanced encoding techniques described in HDPn21 / wg4549i (default))
  - different quantization steps for different frequency bands (enhanced encoding techniques of HDPn21 / wg4549 (default))
  - new POT (leakage fix described in wg4660) (default)
- 4:4:4 coding, one level POT
- 4 contents, 7 selected samples corresponding to the following bpp values:

<table>
<thead>
<tr>
<th>Content</th>
<th>q=40 (T1)</th>
<th>q=50 (T2)</th>
<th>q=58 (T3)</th>
<th>q=66 (T4)</th>
<th>q=76 (T5)</th>
<th>q=82 (T6)</th>
<th>q=90 (T7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cont. 1</td>
<td>0.9</td>
<td>0.64</td>
<td>0.46</td>
<td>0.34</td>
<td>0.22</td>
<td>0.18</td>
<td>0.13</td>
</tr>
<tr>
<td>Cont. 2</td>
<td>0.15</td>
<td>0.1</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Cont. 3</td>
<td>0.9</td>
<td>0.61</td>
<td>0.43</td>
<td>0.31</td>
<td>0.19</td>
<td>0.15</td>
<td>0.1</td>
</tr>
<tr>
<td>Cont. 4</td>
<td>0.65</td>
<td>0.44</td>
<td>0.31</td>
<td>0.22</td>
<td>0.13</td>
<td>0.09</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Preliminary results (II)

- 2 contents, other than those used in the test session, have been used for the training session
- 17 subjects have taken part to the experiment:
  - 3 females, 14 males
  - average subject’s age 29
- Statistical analysis of the data:
  - inter-subjects ANOVA
  - offset and gain score normalization
  - outliers’ screening:
    - 4 outliers for content 1
    - 2 outliers for content 2
    - 2 outliers for content 3
    - 5 outliers for content 4

Preliminary results (III)

Preliminary results (IV)
Acknowledgement

• Part of the work reported here has been possible thanks to:
  – European Commission funded Network of Excellence on ‘Networked Audiovisual Media Technologies VISNET II’

Thank you for your attention!
Questions?