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

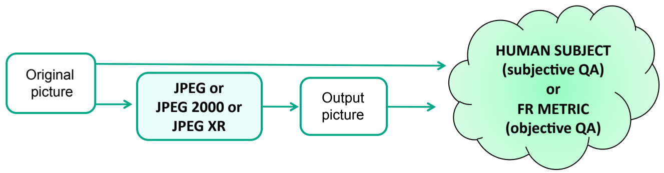
- Introduction
  - Quality Assessment (QA) and codec performance evaluation
  - Status
  - Our previous contributions
- Objective QA
  - Test material
  - Codecs and configuration parameters
  - Quality metrics
  - Selected results
- Subjective QA
  - Proposed methodology
  - Test conditions
  - Preliminary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



Status

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"**, wg1n4404, 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"**, wg1n4701, JPEG meeting Poitiers (July 2008)

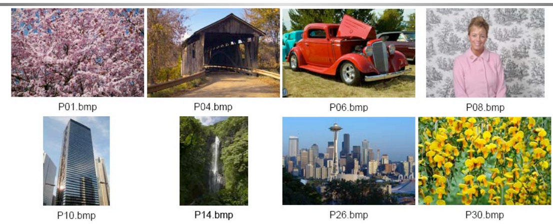
Conference publications:

- F. De Simone et al., **"A comparative study of JPEG 2000, AVC/H.264, and HD Photo"**, SPIE Optics and Photonics, Applications of Digital Image Processing XXX, 6696 (Aug. 2007)
- F. De Simone et al., **"A comparative study of color image compression standards using perceptually driven quality metrics"**, SPIE Optics and Photonics, Applications of Digital Image Processing XXXI (Aug. 2008)

Objective QA

- **Test material**
- **Codecs and configuration parameters**
- **Quality metrics**
- **Selected results**

Test Material – 24 bpp pictures



(sample pictures from Microsoft dataset,
6 different spatial resolutions: 4064x2704, 2268x1512, 2592x1944, 2128x2832, 2704x3499, 4288x2848)

(sample pictures from Thomas Richter dataset,
2 different spatial resolutions: 3888x2592, 2592x3888)

Codecs and configuration parameters

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.

Codecs and configuration parameters

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 Richter'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 Microsoft 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_{inf})

- Considering RGB color space:

$$L_{inf\,R} = \max\left[\mathrm{abs}(Im_{aR}(x,y) - Im_{bR}(x,y))\right]$$
$$L_{inf\,G} = \max\left[\mathrm{abs}(Im_{aG}(x,y) - Im_{bG}(x,y))\right]$$
$$L_{inf\,B} = \max\left[\mathrm{abs}(Im_{aB}(x,y) - Im_{bB}(x,y))\right]$$

$(L_{inf} \in [0,1])$

where:

Im_a, Im_b = pictures to compare

Metric 2: single channel PSNR

$$\boxed{PSNR = 10\log_{10}\frac{(2^B-1)^2}{MSE}}$$

where: $MSE = \frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}[Im_a(x,y) - Im_b(x,y)]^2$

M, N = image dimensions

Im_a, Im_b = pictures to compare

B= 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’, Cb and Cr components (ITU-R Rec. BT. 601):

$$\boxed{WPSNR = w_1 PSNR_1 + w_2 PSNR_2 + w_3 PSNR_3}$$

where:

$w_1 = 1/3,\ w_2 = 1/3,\ w_3 = 1/3$ 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’, Cb and Cr components (ITU-R Rec. BT. 601):

$$\boxed{WPSNR\_MSE = 10\log_{10} \frac{(2^B - 1)^2}{(w_1 MSE_1 + w_2 MSE_2 + w_3 MSE_3)}}$$

where:

$w_1 = 1/3,\ w_2 = 1/3,\ w_3 = 1/3$ 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.

Multimedia Signal Processing Group
Swiss Federal Institute of Technology
VISNET II
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Metric 3: PSNR weighted average (WPSNR_PIX)

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

$$\boxed{WPSNR\_PIX = 10\log_{10} \frac{(2^B - 1)^2}{\frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}\left[\left(w_1 Im_{a1}(x,y) + w_2 Im_{a2}(x,y) + w_3 Im_{a3}(x,y)\right) - \left(w_1 Im_{b1}(x,y) + w_2 Im_{b2}(x,y) + w_3 Im_{b3}(x,y)\right)\right]^2}}$$

where:

M, N = image dimensions

Im_a, Im_b = pictures to compare

B= bit depth

$w_1 = 1/3,\ w_2 = 1/3,\ w_3 = 1/3$ considering R,G, and B components.

$w_1 = 0.8,\ w_2 = 0.1,\ w_3 = 0.1$, considering Y’, Cb, and Cr components.

Multimedia Signal Processing Group
Swiss Federal Institute of Technology
VISNET II
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Metric 4: Mean SSIM (MSSIM) (I)

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Measurement to Structural Similarity” (2004).

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

➢ Estimate of luminance = mean intensity: $\mu = \frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N} Im(x,y)$

➢ Estimate of contrast = standard deviation: $\sigma = \left(\frac{1}{MN-1}\sum_{y=1}^{M}\sum_{x=1}^{N}(Im(x,y)-\mu)^2\right)^{1/2}$

➢ Estimate of picture structure: $\frac{(Im - \mu)}{\sigma}$

Multimedia Signal Processing Group
Swiss Federal Institute of Technology
VISNET II
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Metric 4: Mean SSIM (MSSIM) (II)

$$\mathrm{SSIM}(\mathrm{Im}_1,\mathrm{Im}_2) = \left[l(\mathrm{Im}_1,\mathrm{Im}_2)\right]^{\alpha}\left[c(\mathrm{Im}_1,\mathrm{Im}_2)\right]^{\beta}\left[s(\mathrm{Im}_1,\mathrm{Im}_2)\right]^{\gamma} \quad (\alpha>0,\beta>0,\gamma>0)$$

- **Luminance comparison function:** $l(\mathrm{Im}_1,\mathrm{Im}_2) = \dfrac{2\mu_1\mu_2 + C_1}{\mu_1^2+\mu_2^2+C_1}$ (C_1=constant)
- **Contrast comparison function:** $c(\mathrm{Im}_1,\mathrm{Im}_2) = \dfrac{2\sigma_1\sigma_2 + C_2}{\sigma_1^2+\sigma_2^2+C_2}$ (C_2=constant)
- Measure of structural similarity = correlation between $\dfrac{(\mathrm{Im}_1-\mu_1)}{\sigma_1}$ and $\dfrac{(\mathrm{Im}_2-\mu_2)}{\sigma_2}$

Structure comparison function: $s(\mathrm{Im}_1,\mathrm{Im}_2) = \dfrac{\sigma_{1,2} + C_3}{\sigma_1\sigma_2+C_3}$ (C_3=constant)

where $\sigma_{1,2} = \dfrac{1}{MN-1}\sum_{y=1}^{M}\sum_{x=1}^{N}(\mathrm{Im}_1(x,y)-\mu_1)(\mathrm{Im}_2(x,y)-\mu_2)$

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):

$$\boxed{\mathrm{MSSIM} = w_y \mathrm{MSSIM}_Y + w_{Cb}\mathrm{MSSIM}_{Cb} + w_{Cr}\mathrm{MSSIM}_{Cr}} \quad (\mathrm{MSSIM} \in [0,1])$$

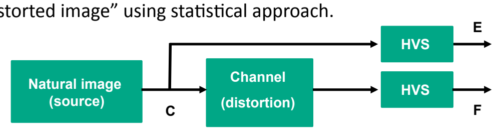
where: $w_Y = 0.8,\ w_{Cb} = 0.1,\ w_{Cr} = 0.1.$

Multimedia Signal Processing Group
Swiss Federal Institute of Technology
VISNET II
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

[2] H. R. Sheikh, A. C. Bovik "Image Information And Visual Quality" (2004).

- "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

Multimedia Signal Processing Group
Swiss Federal Institute of Technology
VISNET II
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

- Natural image modeling in wavelet domain using Gaussian scale mixtures (GSMs)
- **Information that the brain could ideally extract from reference image** = mutual information between C and E: $I(C;E|z)$
- **Corresponding information that could be extracted from test image** = mutual information between C and F: $I(C;F|z)$

$$\mathrm{VIF} = \frac{I(C,F|z)}{I(C,E|z)}$$

(VIF $\in [0,1]$ and VIF>1 if the test image is enhanced version of the original)

where:

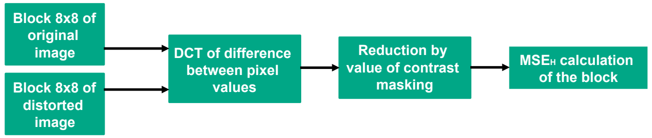
z= source model parameters.

- **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.

Multimedia Signal Processing Group
Swiss Federal Institute of Technology
VISNET II
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Metric 6: PSNR-HVS-M (I)

[3] N. Ponomarenko, F. Silvestri, K. Egiazarian, M.Carli, J. Astola, and V. Lukin, “On between-coefficient contrast masking of DCT basis functions” (2007).



- DCT coefficients of 8x8 pixel blocks X and Y are visually undistinguished if:

  $E_w(X-Y) < \max(E_m(X), E_m(Y))$

  where E_w(block) is the energy of DCT coefficients of the block weighted according to CSF and E_m(block) is the masking effect of DCT coefficients of the block which depends upon E_w(block) and upon the local variances.

Metric 6: PSNR-HVS-M (II)

$$\boxed{PSNR-HVS-M = 10\log_{10}\frac{(2^B-1)^2}{MSE_H}}$$

where: $$MSE_H = K\sum_{i=1}^{M-7}\sum_{j=1}^{N-7}\sum_{m=1}^{8}\sum_{n=1}^{8}\left[X(m,n)_{\Delta ij}\,T_c(m,n)\right]^2$$

M, N = image dimensions

K= constant

$X(m,n)_{\Delta ij}$ = visible difference between DCT coefficient of the original image and distorted image 8x8 blocks, depending upon contrast masking

T_c = matrix of correcting factors based on standard visually optimized JPEG quantization tables

B= bit depth

Metric 7: DC Tune

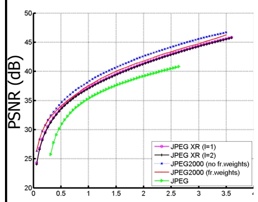
[4] A. B. Watson, A. P. Gale, J. A. Solomon, and A. J. Ahumada JR., “DCTune: A Techinque For Visual Optimization Of DCT Quantization Matrices For Individual Images” (1994).

- 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

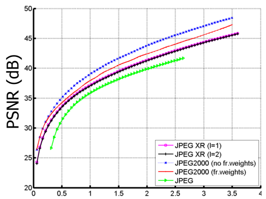
Average over image dataset of PSNR values

on R component:



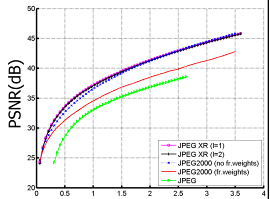
bpp (bits/pixel)

on G component:



bpp (bits/pixel)

on B component:

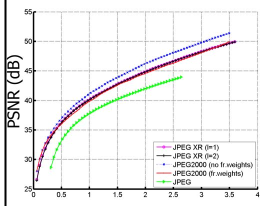


bpp (bits/pixel)

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

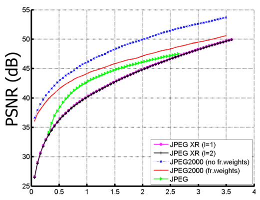
Average over image dataset of PSNR values

on Y’ component:



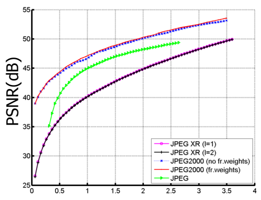
bpp (bits/pixel)

on Cb component:



bpp (bits/pixel)

on Cr component:

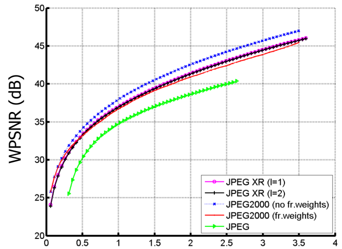


bpp (bits/pixel)

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

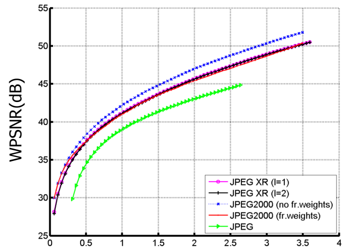
Average over image dataset of WPSNR values

on RGB components:



bpp (bits/pixel)

on Y’CbCr components:

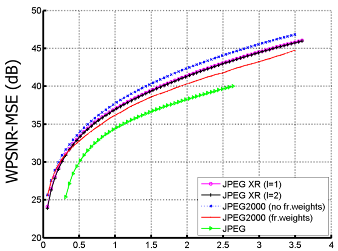


bpp (bits/pixel)

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

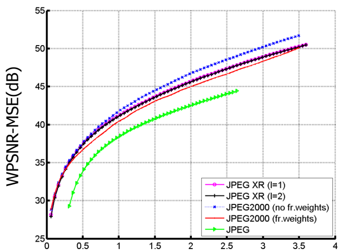
Average over image dataset of WPSNR-MSE values

on RGB components:



bpp (bits/pixel)

on Y’CbCr components:

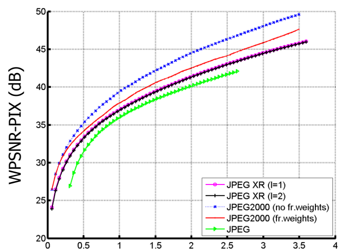


bpp (bits/pixel)

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

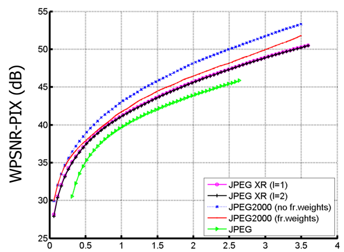
Average over image dataset of WPSNR-PIX values

on RGB components:



bpp (bits/pixel)

on Y’CbCr components:

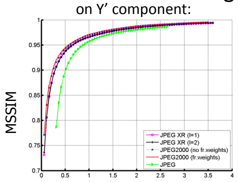


bpp (bits/pixel)

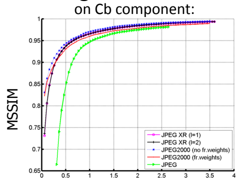
Selected results 4:4:4 – JPEG XR vs JPEG2000 vs JPEG

Average over image dataset of MSSIM values

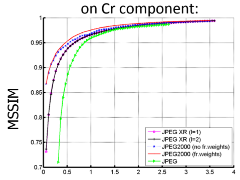
on Y’ component:

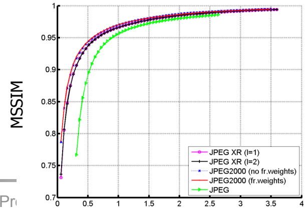


on Cb component:



on Cr component:



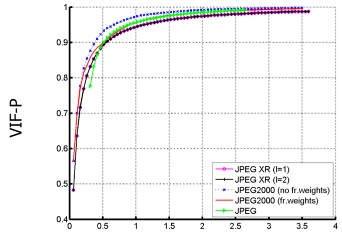


bpp (bits/pixel)

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

Average over image dataset of VIF-P values

on Y’ component only:



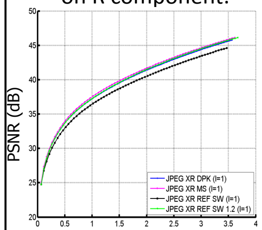
bpp (bits/pixel)

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

Average over image dataset of PSNR values

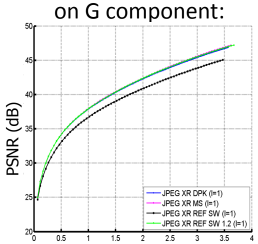
(one level POT)

on R component:



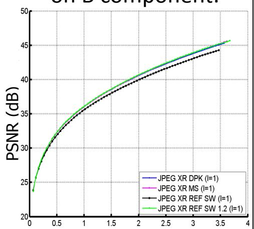
bpp (bits/pixel)

on G component:



bpp (bits/pixel)

on B component:



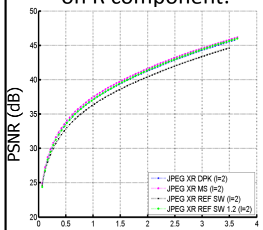
bpp (bits/pixel)

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

Average over image dataset of PSNR values

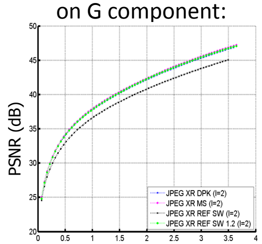
(two levels POT)

on R component:



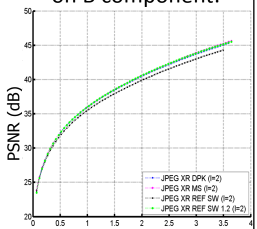
bpp (bits/pixel)

on G component:



bpp (bits/pixel)

on B component:



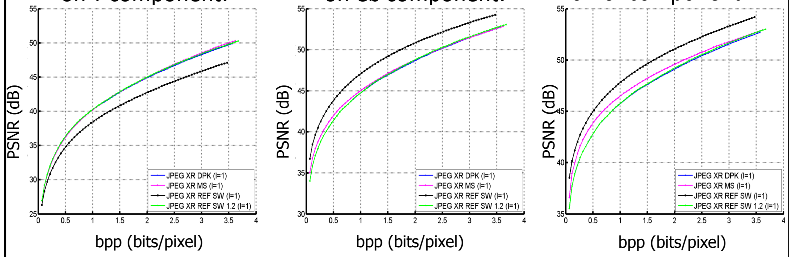
bpp (bits/pixel)

Selected results 4:4:4 – different JPEG XR implem. 33

Average over image dataset of PSNR values

(one level POT)

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

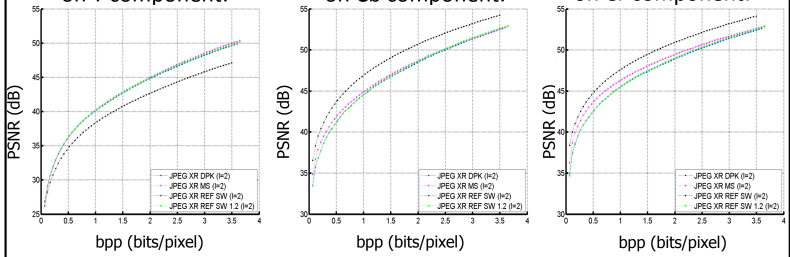


Selected results 4:4:4 – different JPEG XR implem. 34

Average over image dataset of PSNR values

(two levels POT)

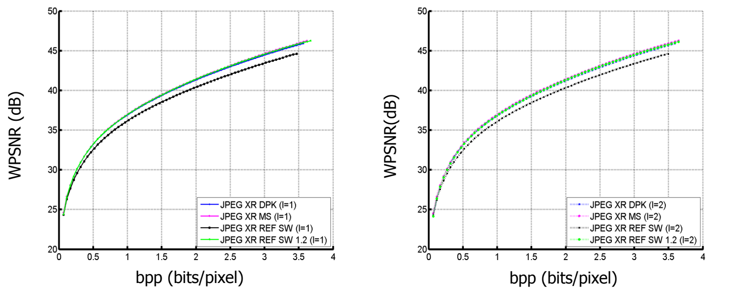
on Y component: on Cb component: on Cr component:



Selected results 4:4:4 – different JPEG XR implem. 35

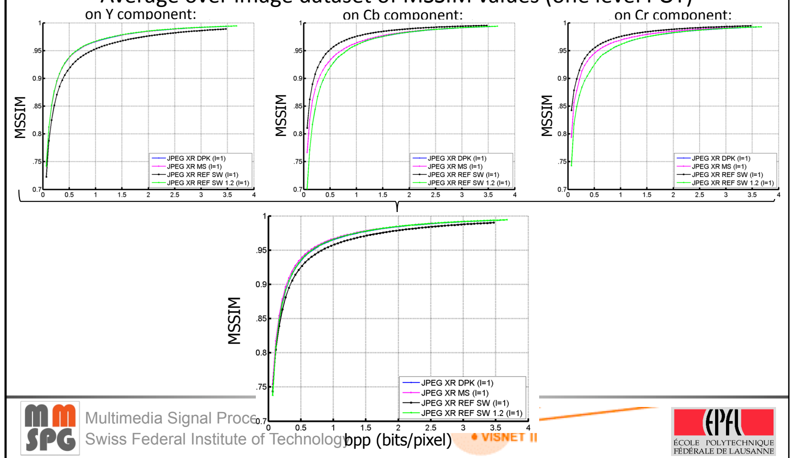
Average over image dataset of WPSNR_MSE values

one level POT: two levels POT:



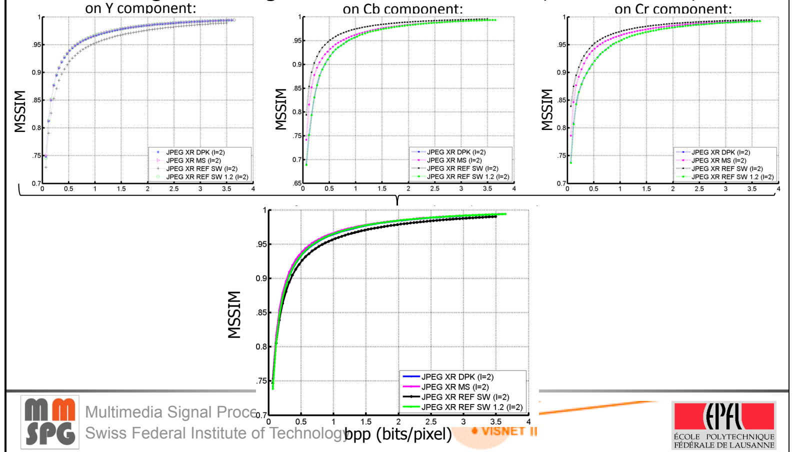
Selected results 4:4:4 – different JPEG XR implem. 36

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



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

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



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:

Reference Image

Test Image

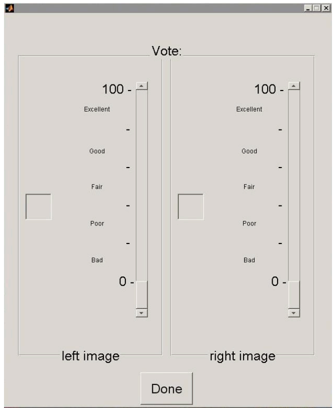
- test picture and its reference are shown at the same time.
- the assessor is not told about the presence of a reference picture.
- positions of reference and test pictures are systematically switched.
- test pairs related to different original contents are always alternated.

Proposed methodology (II)

- when the subject clicks into the active area of the screen a rating window is shown:

Proposed methodology (III)

Rating window
(Continuous Quality Scale)



- 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)

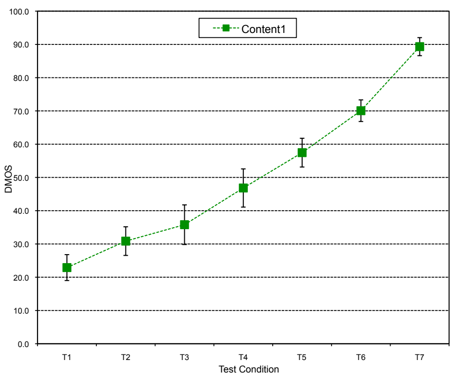
- JPEG XR Microsoft implementation described in HDPn21:
  - different quantization steps for different color channels (enhanced encoding techniques described in HDPn21 / wg1n4549) (default)
  - different quantization steps for different frequency bands (enhanced encoding techniques of HDPn21 / wg1n4549) (default)
  - new POT (leakage fix described in wg1n4660) (default)
- 4:4:4 coding, one level POT
- 4 contents, 7 selected samples corresponding to the following bpp values:

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

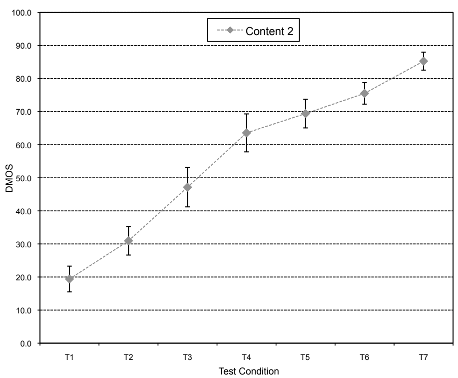
Preliminary results (II)

- 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

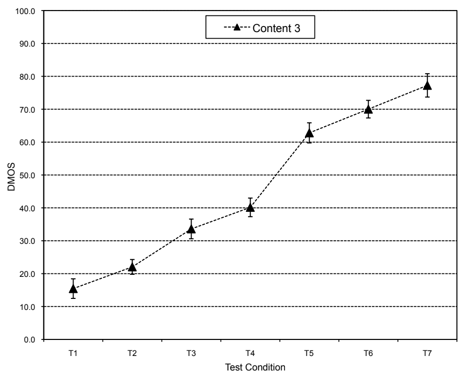
Preliminary results (III)



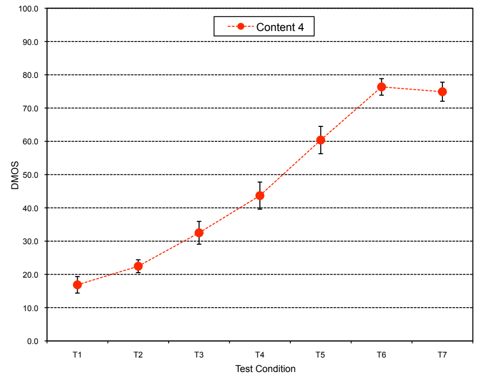
Preliminary results (IV)



Preliminary results (V)



Preliminary results (VI)



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?