OVERVIEW AND BENCHMARKING SUMMARY FOR THE ICIP 2016 COMPRESSION CHALLENGE

Evangelos Alexiou , Irene Viola,
Lukas Krasula, Thomas Richter, Tim Bruylants, Antonio Pinheiro, Karel Fliegel,
Martin Rerabek, Athanassios Skodras, Peter Schelkens and Touradj Ebrahimi

A contribution from Qualinet to JPEG call for information

Background

- Call for information on still image coding
  - *Issued by JPEG committee on February 2015*
  - Broad scope not only limited to compression efficiency
    - New imaging modalities (more than 8-bit, HDR, …)
    - Features (scalability, random access, …)
    - Characteristics (complexity, latency, …)
  - A first response produced during PCS 2015
  - Both lossy and lossless

http://www.jpeg.org

- ICIP 2016 Feature Event
  - Evaluation of current and future Image compression technologies
- This contribution only focuses on compression efficiency of conventional images in lossy and lossless without taking into account other criteria (features, complexity, delay, etc.)
- Objective and subjective evaluations in lossy case carried out by Qualinet
  - VUB/iMinds (Belgium)
  - UBI (Portugal)
  - CTU (Czech Republic)
  - University Stuttgart(Germany)
  - University Patras(Greece)
  - EPFL (Switzerland)
- Lossless evaluations carried out by University of Stuttgart (Germany)

http://www.qualinet.eu

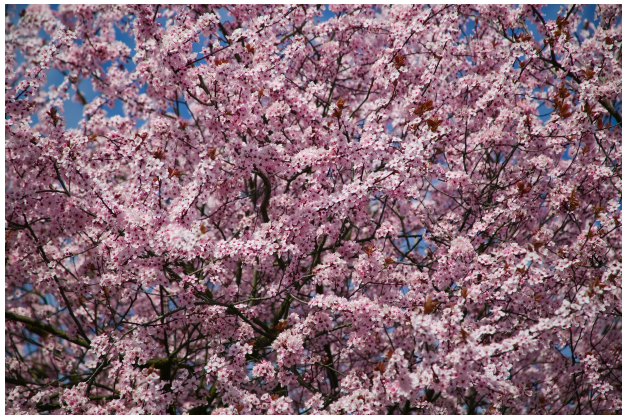
Test material in lossy evaluations

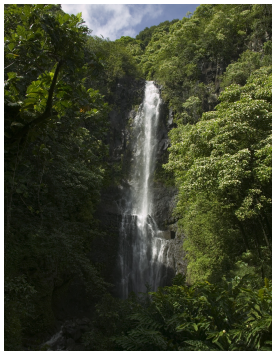
- Contents: 7 (1 training + 6 test):
  - Resolutions - 800x1152 or 800x1280 depending on content
  - Subjective evaluations on cropped versions to fit display
  - Objective metrics performed on the cropped versions

- Stimuli:
  - Original images
  - Compressed/decompressed images with 10 codecs
    - JPEG (default)
    - JPEG (PSNR)
    - JPEG (visual)
    - JPEG 2000 (PSNR)
    - JPEG 2000 (visual)
    - JPEG XR (444)
    - JPEG XR (420)
    - HEVC (SCC ext.)
    - Daala
    - WebP

- 8 bit rates for objective metrics:
  - 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75 and 2 bpp
- 4 bit rates for subjective evaluations:
  - 0.25, 0.5, 0.75 and 1 bpp or 0.75, 1, 1.25 and 1.5 bpp depending on content

Test material in lossless evaluations



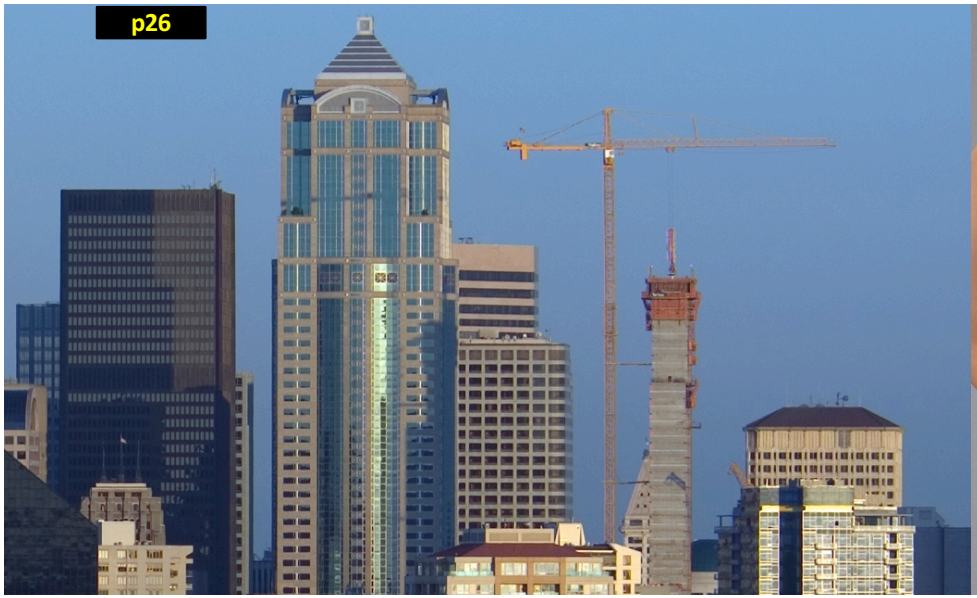






RGB, 444, 24 bpp

Cropped images used in lossy evaluations

Cropped images used in lossy evaluations



Codecs used in lossless evaluations

1. JPEG XR Lossless
2. JPEG 2000 lossless
3. FLIF (submitted to ICIP2016 Grand Challenge)
4. JPEG LS part-1 without color transformation
5. JPEG LS part-2 with lossless color transformation
6. PNG
7. WebP lossless
8. JPEG XT part 8 lossless with residual coding
9. JPEG XT part 8 lossless with arithmetic coding
10. JPEG XT part 8 with lossless DCT
11. JPEG-1 lossless
12. JPEG-1 lossless with arithmetic coding
13. JPEG XT part 8, residual coding with optimized Huffman coding, progressive mode
14. JPEG XT part 8, lossless DCT with optimized Huffman coding, progressive mode.
15. Hierarchical JPEG with the initial pass a DCT coding and the second level of the pyramid lossless coding and optimized Huffman coding
16. Same as 15, but with arithmetic coding instead of Huffman coding
17. Same as 15, except that the first level of the pyramid is a down scaled ¼ and processed by the DCT. compressed with DCT, then up scaled, and the residual is compressed by the predictive mode of JPEG.
18. Same as 17, but with arithmetic coding.

Codecs used in lossy evaluations

- **JPEG (default):** IJG implementation of JPEG with default options (i.e. Annex K quantization settings, 420 color subsampling, baseline configuration) – "cjpeg" without further options.

- **JPEG (PSNR):** JPEG XT reference software using JPEG XT part 1 baseline compressor *compatible with 10918-1*. PNSR optimized version: -oz -h -qt 1 –v
  - Enabling: 444 subsampling, flat quantization matrix, optimized Huffman coding, progressive scan order.

- **JPEG (visual):** JPEG XT reference software using JPEG XT part 1 baseline compressor *compatible with 10918-1*. Visually optimized version: -oz -h -qt 3 -v -s 1x1,2x2,2x2
  - Enabling: 420 subsampling, ImageMagick quantization matrix, optimized Huffman coding, progressive coding.

Codecs used in lossy evaluations

- **JPEG 2000 (PSNR):** JPEG 2000 compressor using Accusoft software with the following command line options: -lo -as -cn 1
  - Enabling: strict rate allocator, 5 decomposition levels, one layer, no precincts, no tiles. lossy 5/3 and not 9/7 wavelet.

- **JPEG 2000 (visual):** JPEG 2000 compressor using Accusoft software. Visually optimized version: -lo -as -cn 1 -w 1000
  - Enabling: features identical to JPEG 2000 (PSNR), but with visual weighting.

- **JPEG XR (444):** JPEG XR reference software with the following options: -f YUV444 -l 1 –d
  - Enabling: 444 chroma subsampling, one level overlap, derived chroma quantization.

- **JPEG XR (420):** JPEG XR reference software with the following options: -f YUV420 -l 2 –d
  - Enabling: 420 chroma subsampling, two level overlap, derived chroma quantization.

Codecs used in lossy evaluations

- **HEVC (SCC ext.):** HEVC with Screen Content Coding extension configured to the main intra-profile for 420 chroma input, with the following command line options: -c cfg/encoder_intra_main_scc.cfg --InputChromaFormat=420 --ProgressiveSource --FrameOnly -cf 420 --FrameRate=30 --FramesToBeEncoded=1 --QuadtreeTULog2MaxSize=5 --GOPSize=1 --IntraPeriod=1 --ConformanceWindowMode=1 --AdaptiveQP=1 --RateControl=0 --TransquantBypassEnable=1 --CrossComponentPrediction=0 --ColourTransform=0

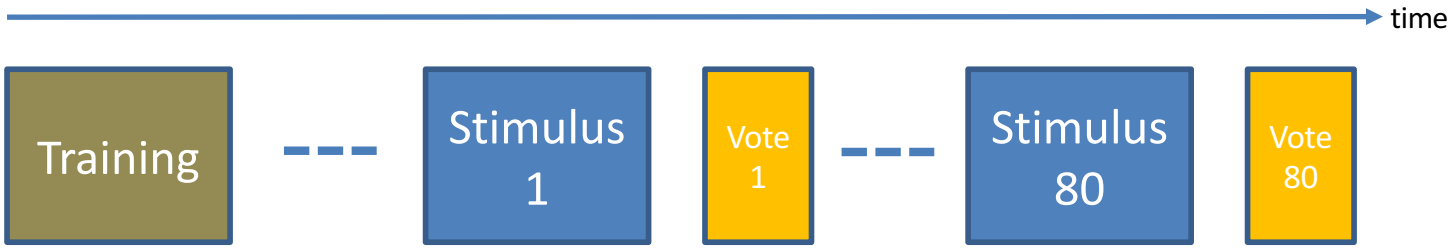
- **Daala:** mozilla Daala with default configuration. No further options.

- **WebP:** Google WebP (part of WebM) with the following command line options: -m 6 -partition_limit 50 -af

Subjective evaluation methodology

- Subjective evaluation methodology based on ITU-T P.910
- ACR-HR: Absolute Category Rating with Hidden Reference
- Randomization of presentation order
- 5-level discrete scale: bad, poor, fair, good, excellent
- 10 codecs tested for their subjective quality
  - 10(codecs) x 6(images) x 4(bit rates) + 6(originals) = 246 stimuli
- 21 naïve subjects participated in VUB, UBI and EPFL labs
- Each subject completed 3 sessions of 80 stimuli (circa 15 min per session, 30 min break)
- Short training for bad, fair and excellent quality illustrations
- Display: Apple MacBook Pro Retina 15 inch or equivalent
- Typical office environment

Excellent
Good
Fair
Poor
Bad



Objective evaluation metrics

- PSNR
  - Widely used quality metric in image processing community.
  - Performed for both Y channel and RGB.

- SSIM: Structural Similarity Index
  - Mean of similarity between an image under test and its reference based on structural information.

- MSSIM: Multiscale Structural Similarity Index
  - Multiscale version of SSIM.

- FSIM: Feature Similarity Index
  - Based on SSIM.
  - Adds a comparison of low-level feature sets between the reference and the distorted images.
  - analyzes the high phase congruency extracting highly informative features and the gradient magnitude, to encode the contrast information.
  - This analysis is complementary and reflects different aspects of the HVS in assessing the local quality of an image.
  - Performed for both Y and C channels.

Objective evaluation metrics

- HDR-VDP2.2: High Dynamic Range Visible Difference Predictor
  - Calibrated metric developed for HDR images
  - Considers a light-adaptive contrast sensitivity function, as the ranges of light adaptation can vary substantially.
  - Includes a specific model of the point spread function (PSF) of the eye optics, as human optical lens flare can be very strong in high contrast HDR content.
  - The front-end amplitude non-linearity is based on integration of the Weber-Fechner law.
  - Takes into account the angular resolution.
  - Uses a multi-scale decomposition.
  - A neural noise block is defined to calculate per-pixel probabilities maps of visibility and the predicted quality metric.

- CIEDE2000: Color difference metric
  - Includes weighting factors for lightness, chroma, and hue (like the CIE1976 L*a*b* perceptual space).
  - Also includes factors to handle the relationship between chroma and hue.

- VIF: Visual Information Fidelity
  - Analyses the natural scene statistics.
  - Uses an image degradation model and the HVS model.
  - Based on the quantification of the Shannon information present in both the reference and the distorted images.

Thank you for your attention!

Results of evaluations will be presented at the end of this session!!!

23rd International Conference on Image Processing
September 25 – 28, 2016, Phoenix, Arizona

SUMMARY AND NEXT STEPS FOR THE IMAGE COMPRESSION CHALLENGE

Evangelos Alexiou , Irene Viola,
Lukas Krasula, Thomas Richter, Tim Bruylants, Antonio Pinheiro, Karel Fliegel,
Martin Rerabek, Athanassios Skodras, Peter Schelkens and Touradj Ebrahimi

A contribution from Qualinet to JPEG call for information

Overview

- ICIP 2016 Feature Event
  - Evaluation of current and future Image compression technologies
- 7 Contents used in lossy evaluations (1 training + 6 tests)
  - Resolutions 800x1152 or 800x1280
- 10 codecs tested in lossy evaluations (anchors and proponents)
- 8 bit rates for objective metrics
  - 0.25, 0.5. 0.75, 1, 1.25, 1.5, 1.75, 2 bpp
- 4 bit rates for subjective evaluations
  - 0.25, 0.5, 0.75, 1 bpp or 0.75, 1, 1.25 and 1.5 bpp
- Subjective assessment protocol based on ITU-T P.910 for lossy evaluations
  - ACR-HR: Absolute Category Rating with Hidden Reference
  - 21 naïve subjects in VUB, UBI and EPFL
  - 3 sessions of 80 stimuli
  - 5-level discrete scale
  - Outliers detection based on boxplot algorithm
  - Display: Apple MacBook Pro Retina 15 inch
  - Typical office environment
- Objective evaluation in lossy case using 9 metrics
  - Cross-checked between CTU, University Stuttgart and University Patras

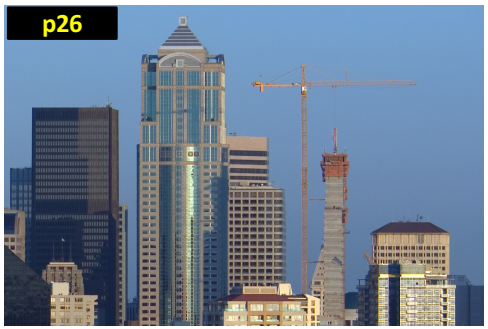
JPEG AIC

http://www.jpeg.org

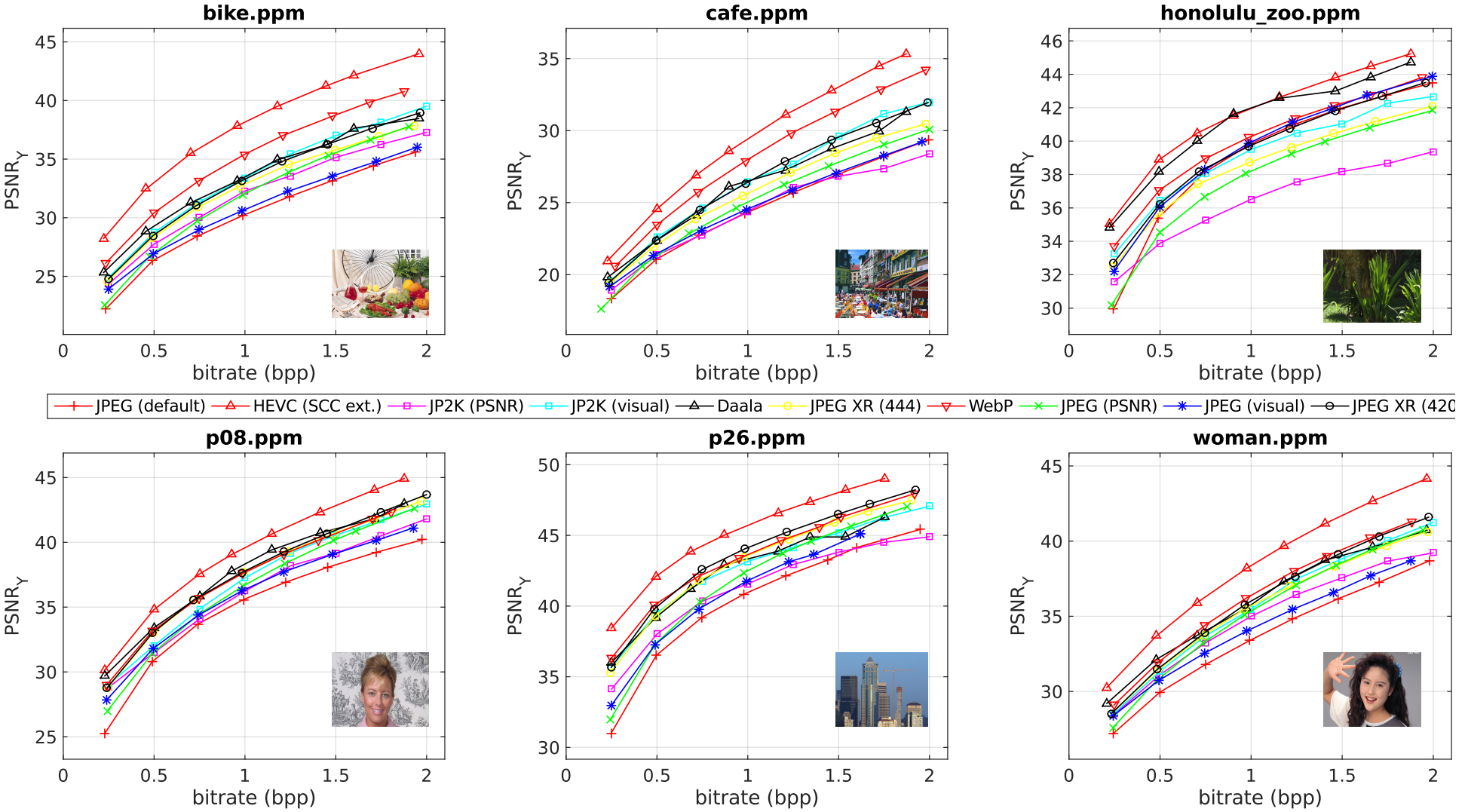
Qcost IC1003

http://www.qualinet.eu

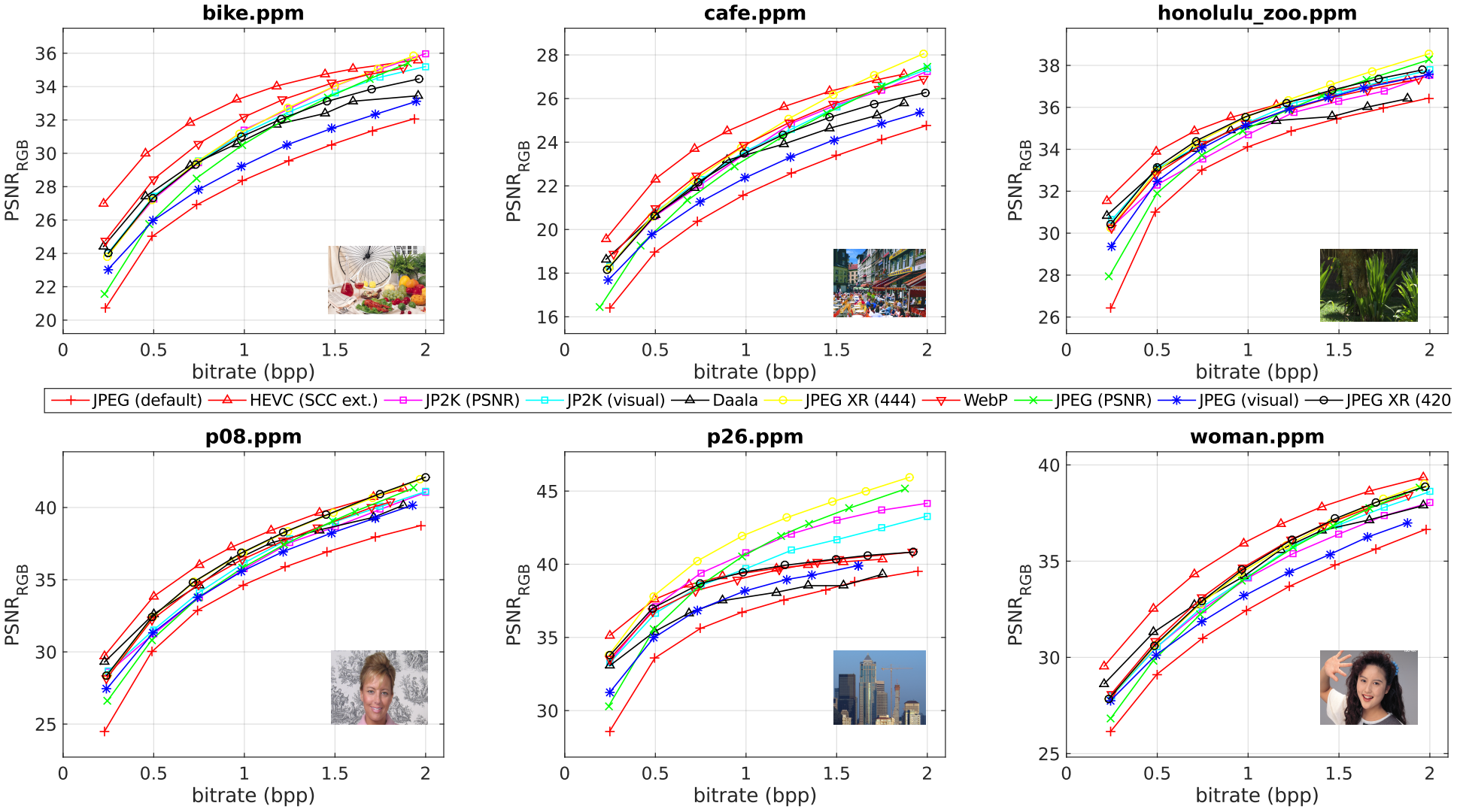
Test images in lossy evaluations



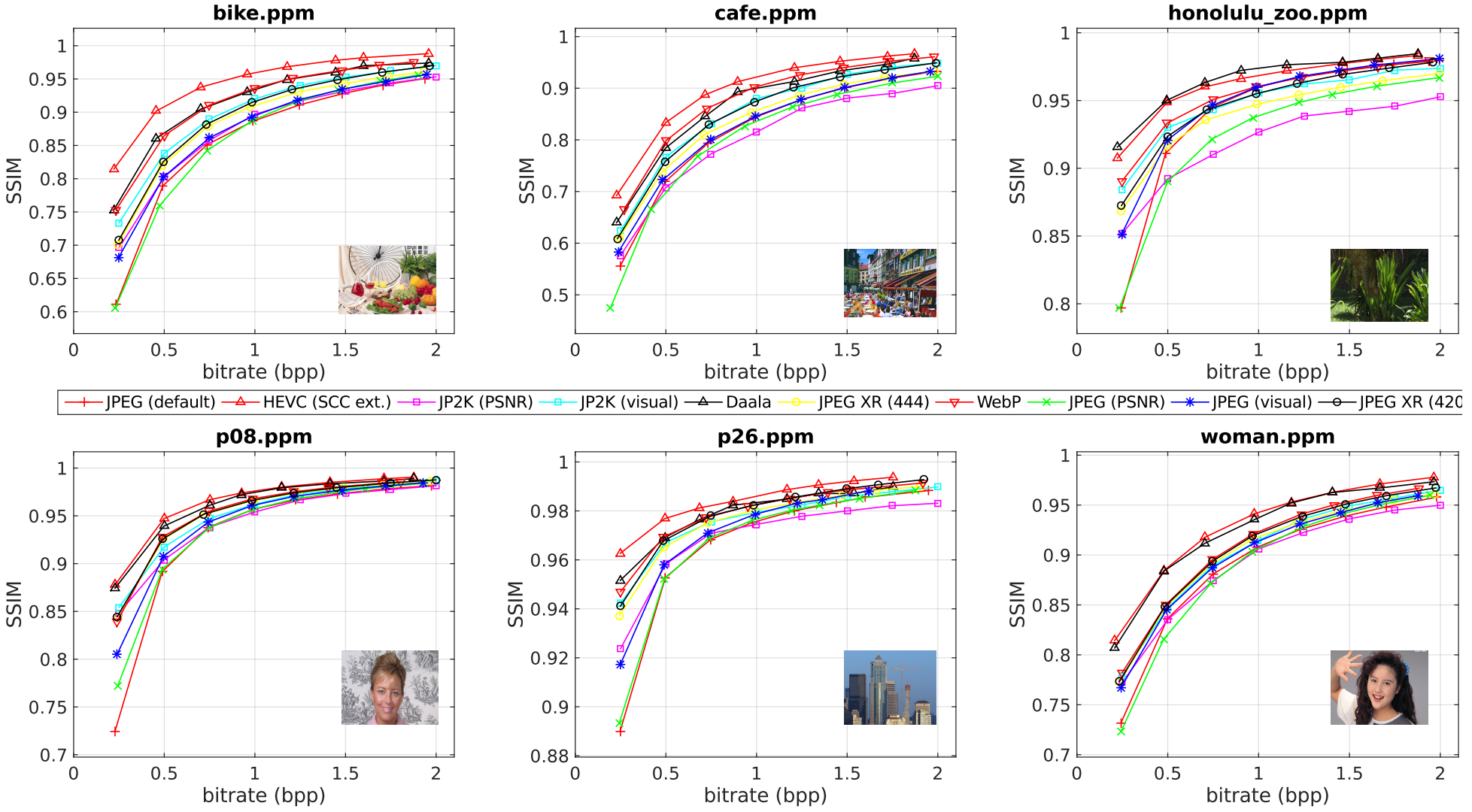
Objective evaluation: $PSNR_Y$ results



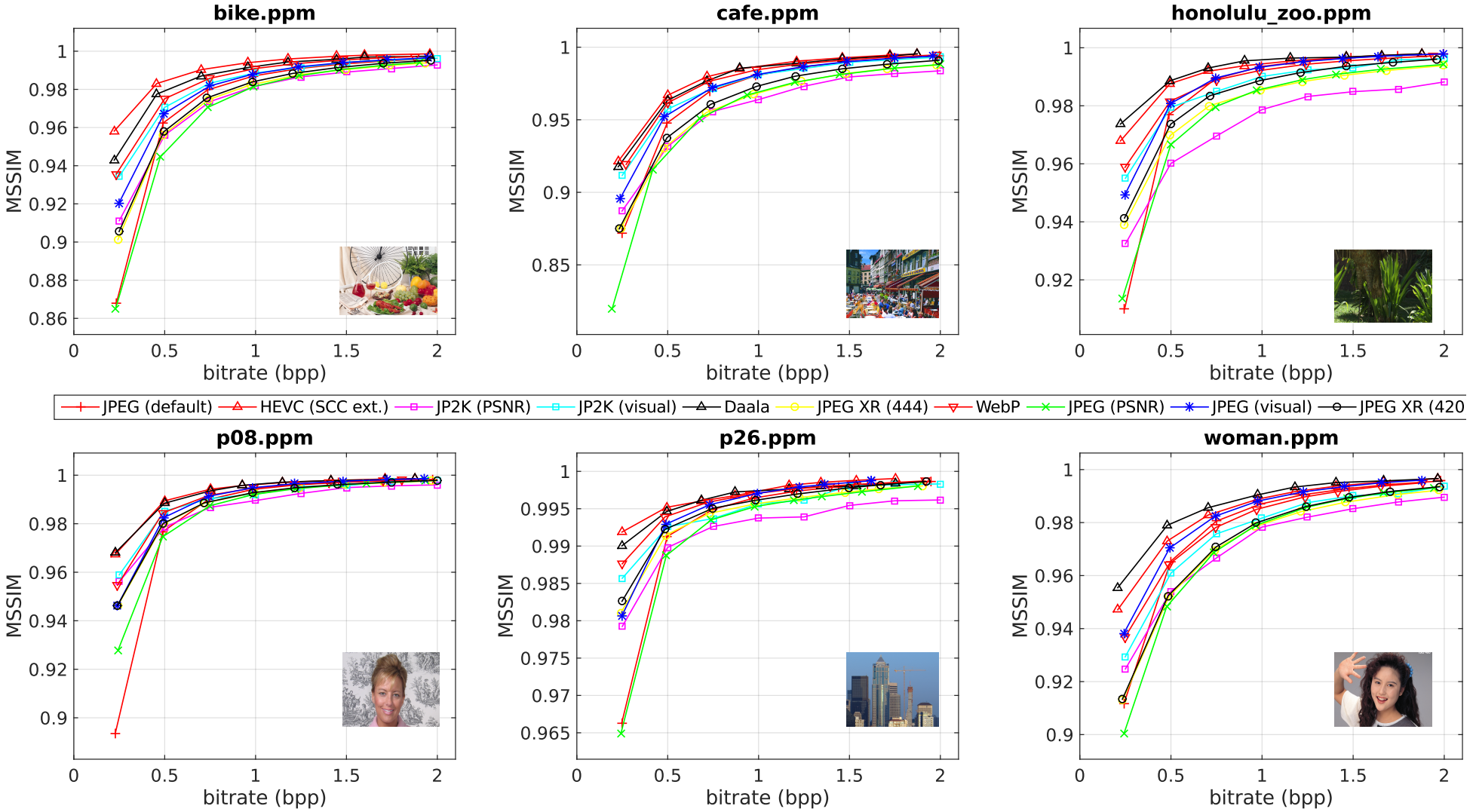
Objective evaluation: $PSNR_{RGB}$ results



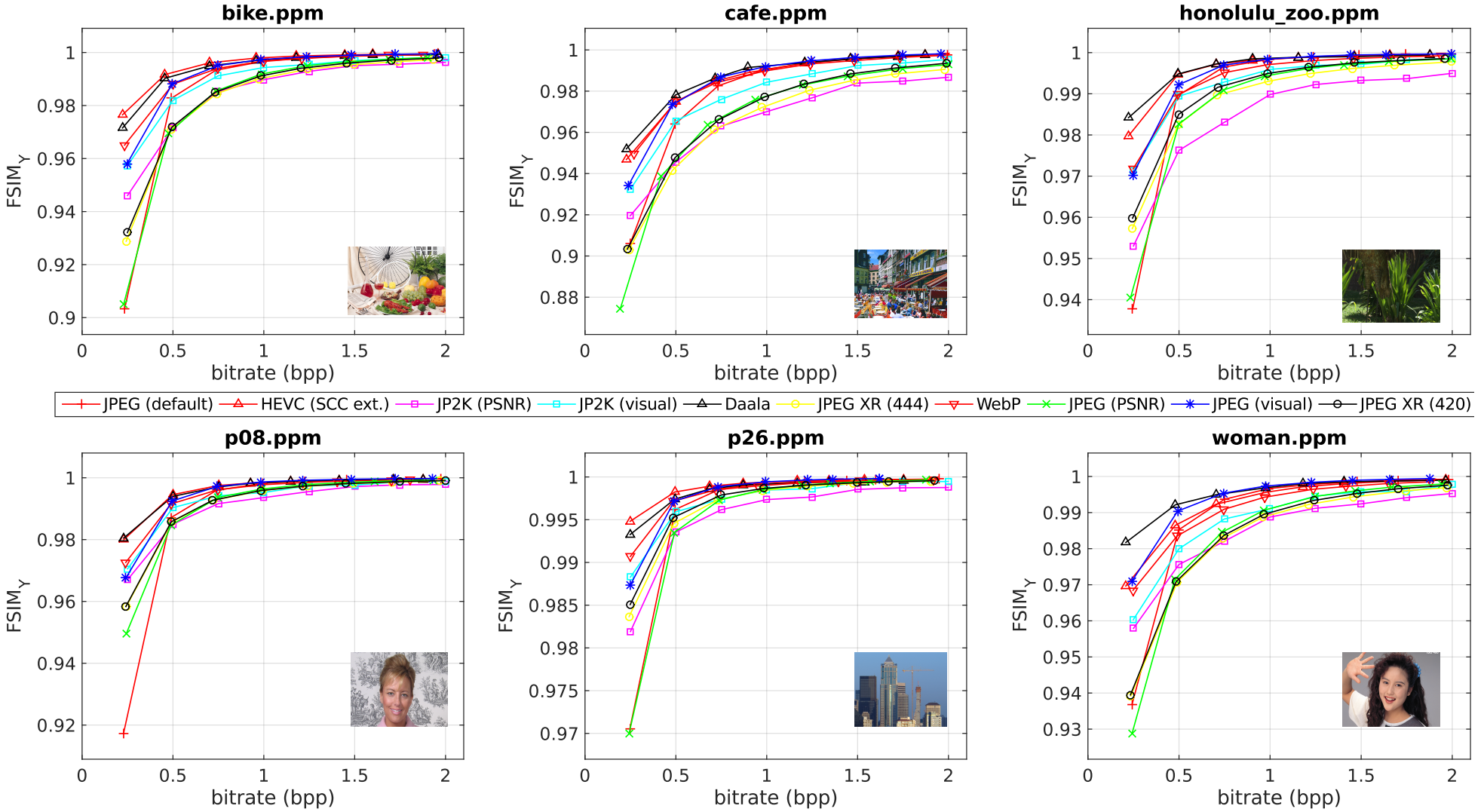
Objective evaluation: SSIM results



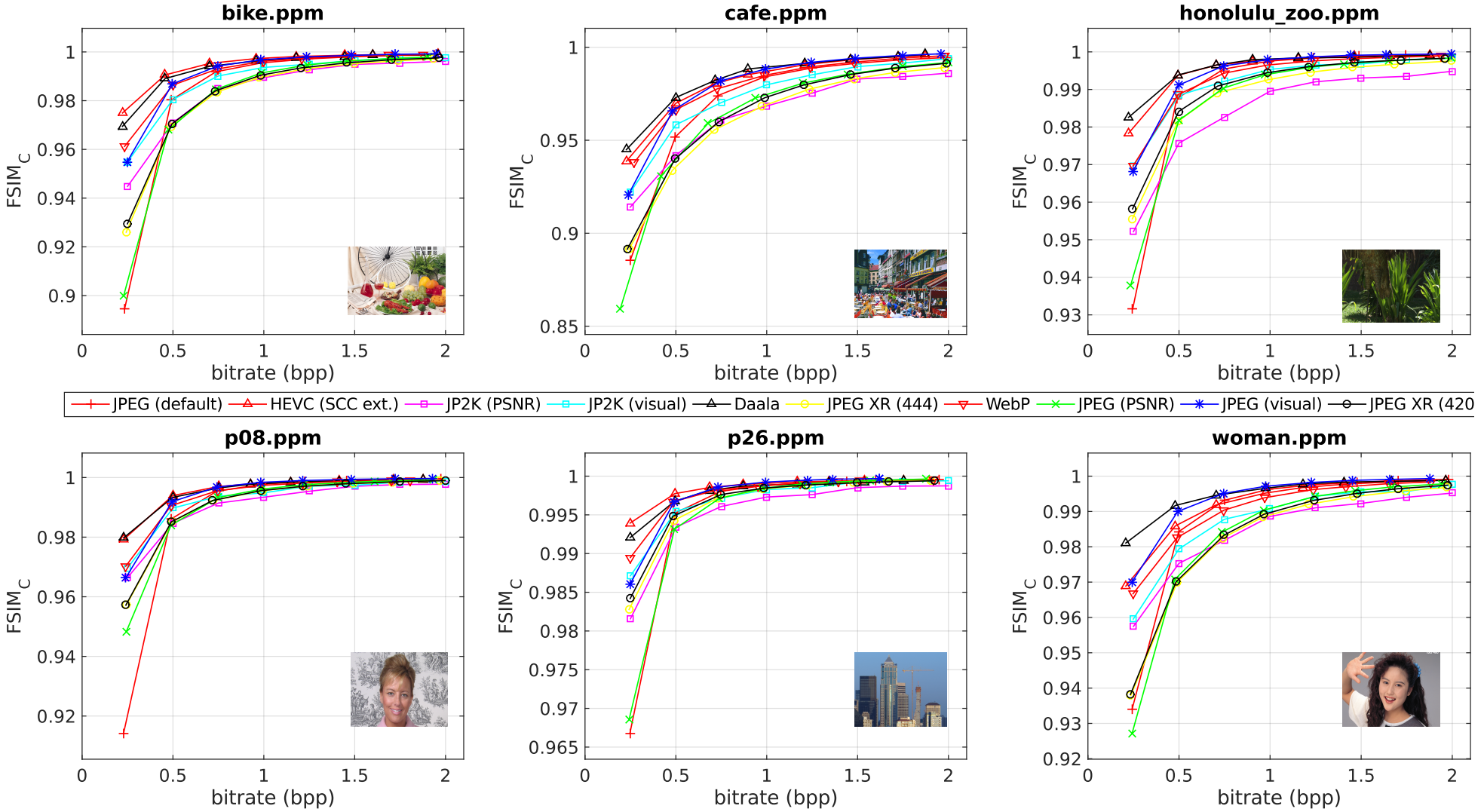
Objective evaluation: MSSIM results



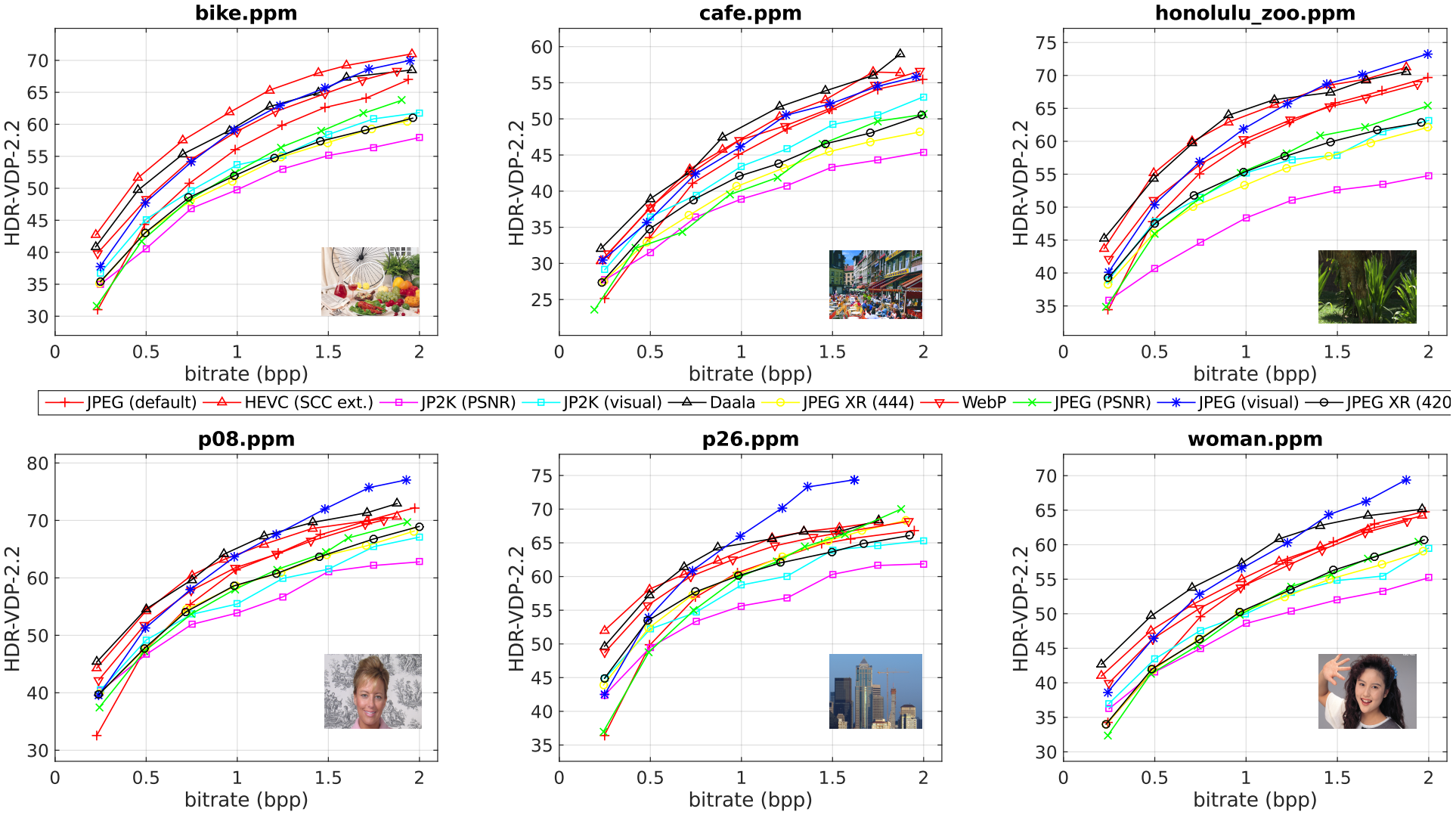
Objective evaluation: $FSIM_Y$ results



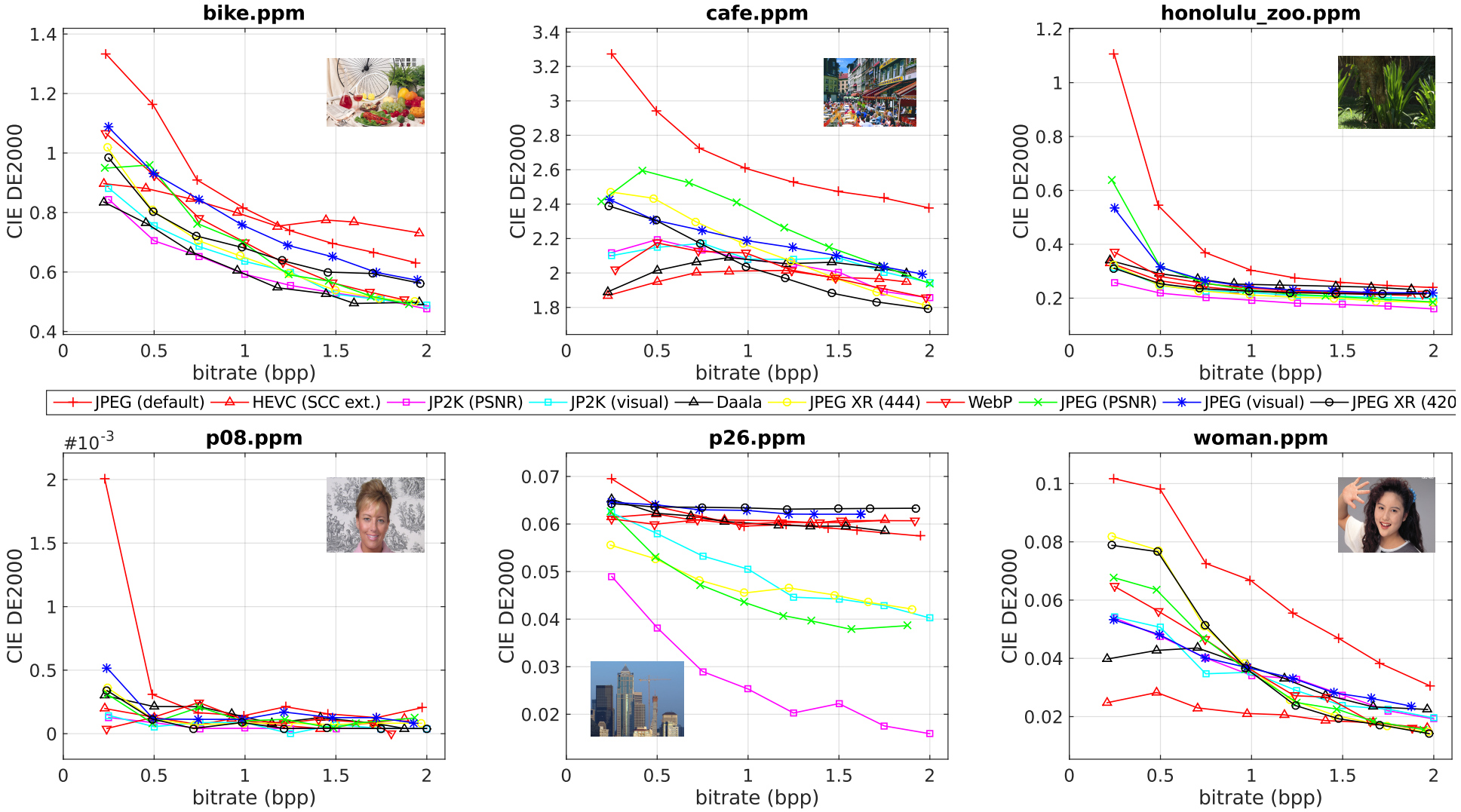
Objective evaluation: $FSIM_C$ results



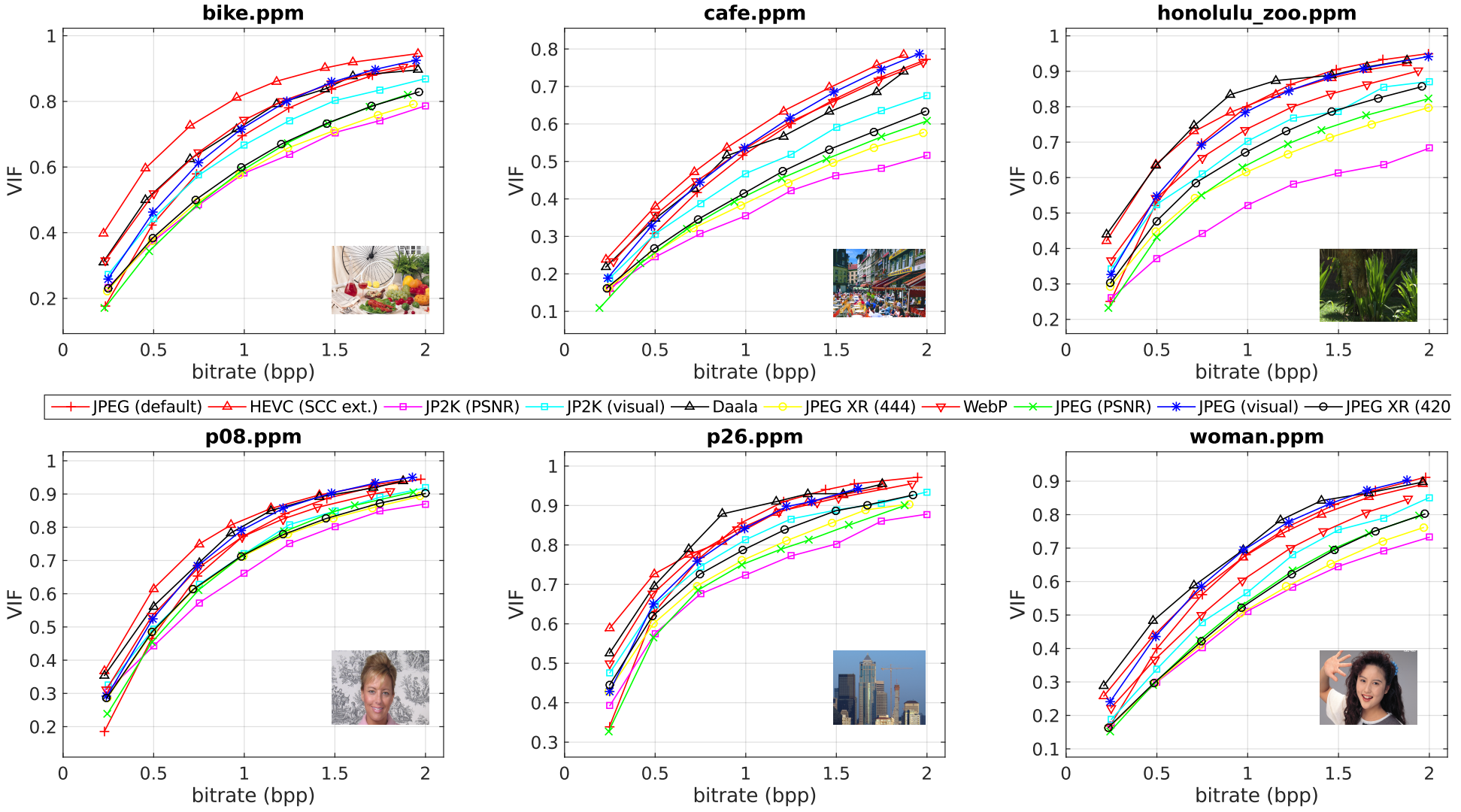
Objective evaluation: HDR-VDP-2.2 results



Objective evaluation: CIE DE2000 results



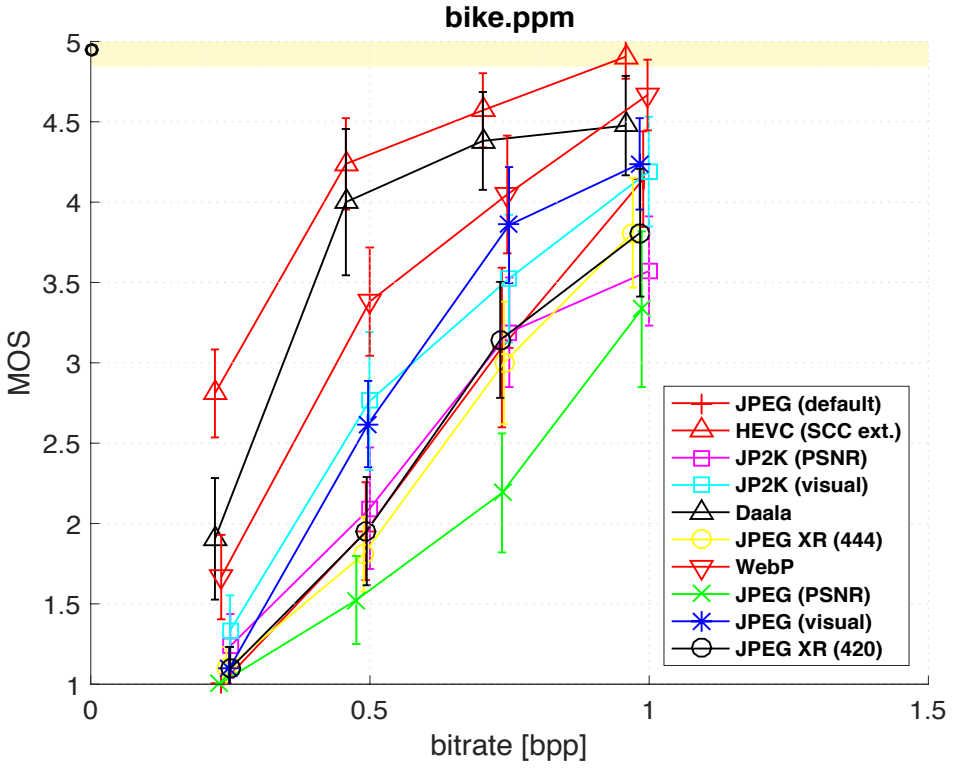
Objective evaluation: VIF results

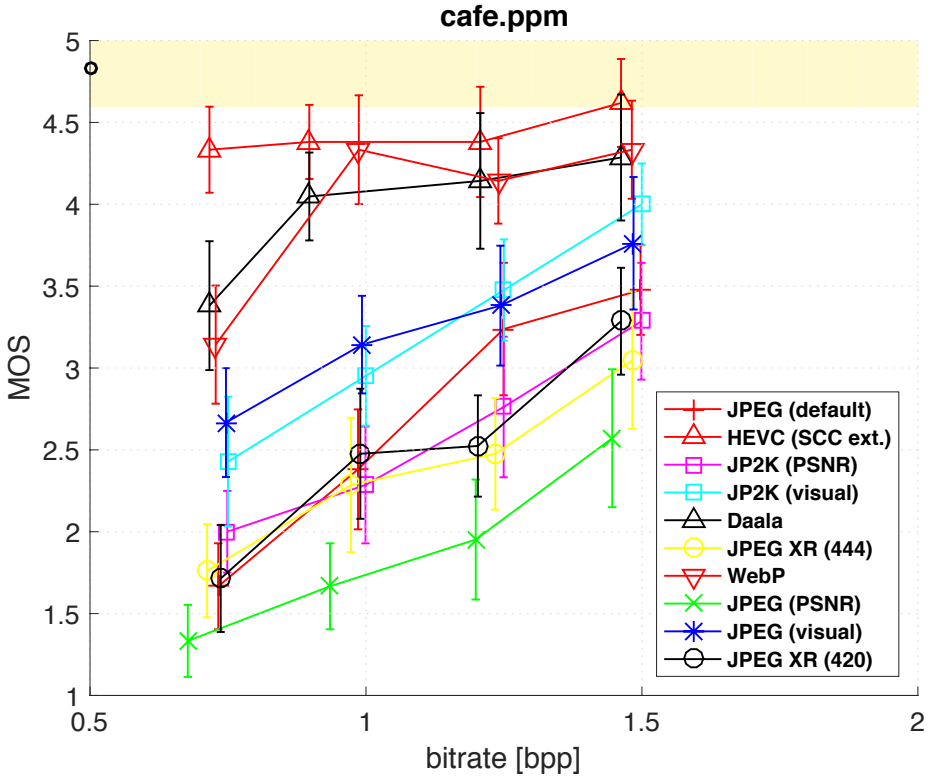


Outlier detection

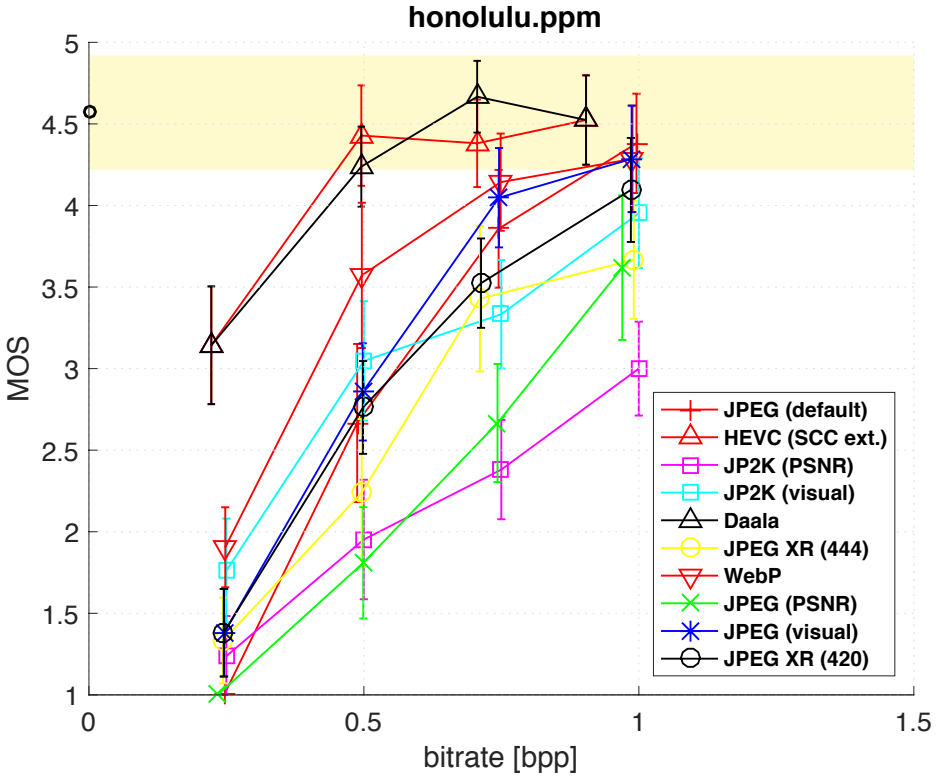
- Statistical analysis using Boxplot method.
- For each test condition, an outlier is defined as a data point that is located outside the interquartile range i.e. above the upper quartile or below the lower quartile of the distribution of the scores multiplied by 1.5.
- If the same subject is identified as outlier in more than 20% of the test conditions, the corresponding scores are discarded.
- No outliers were found.

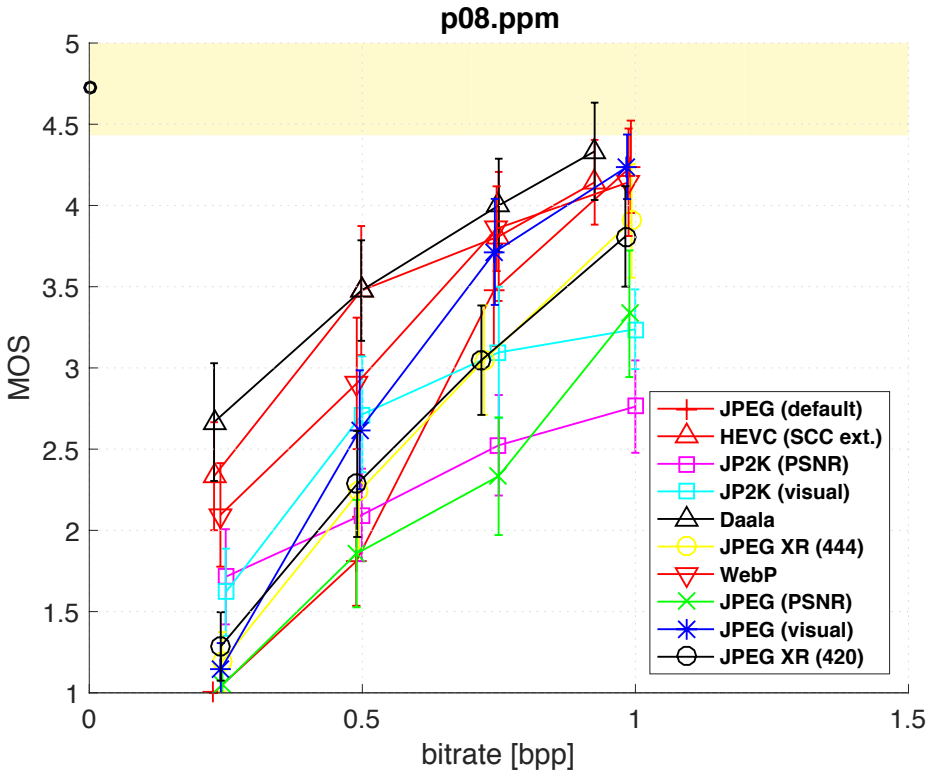
Subjective evaluation results



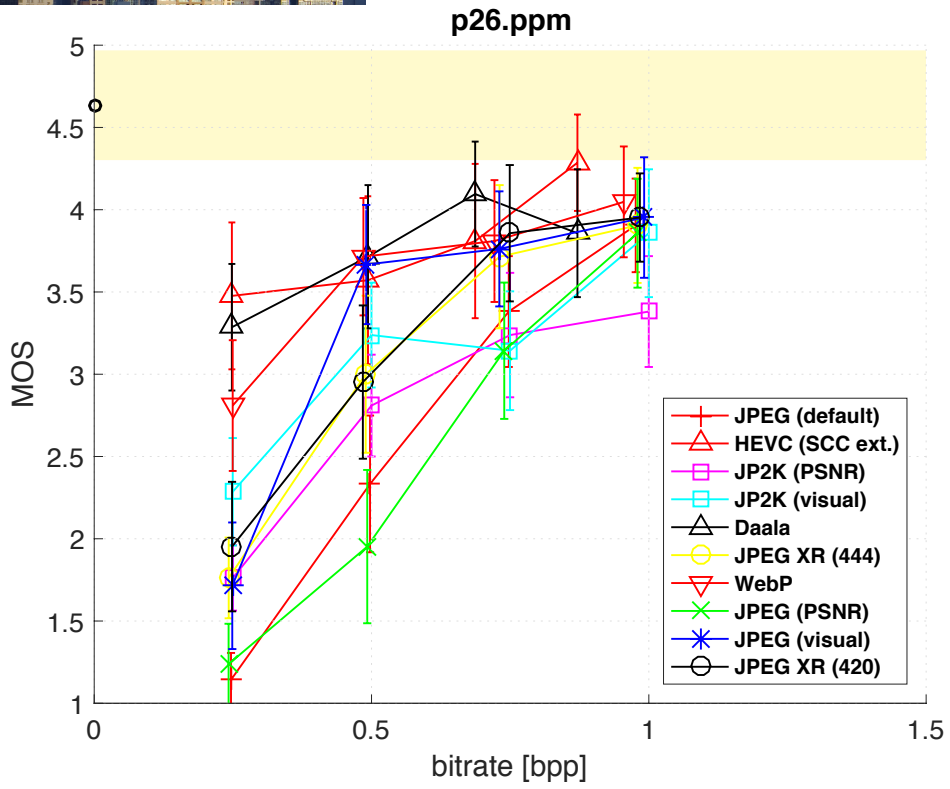


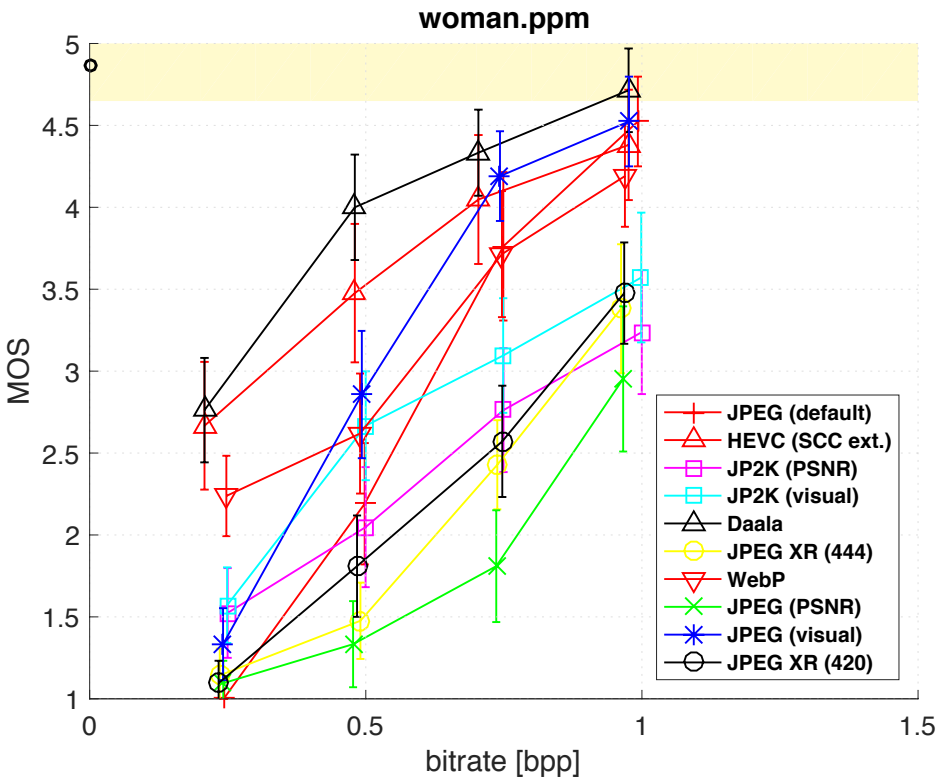
Subjective evaluation results





Subjective evaluation results

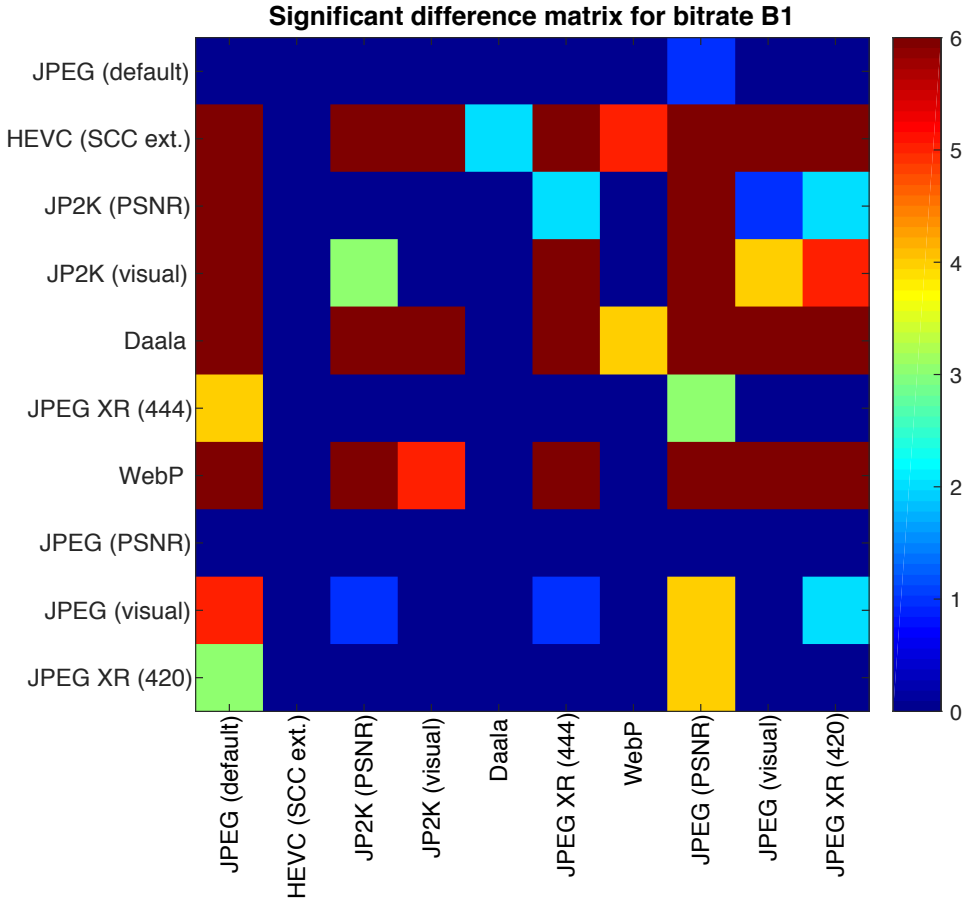


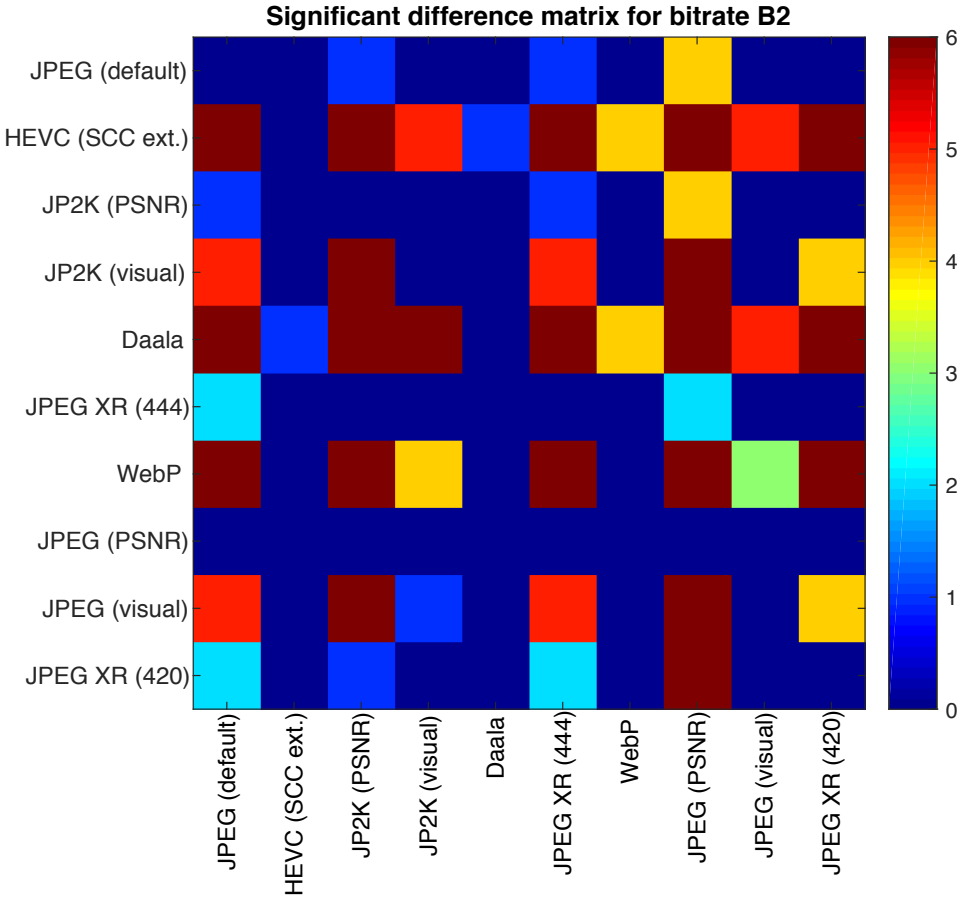


Statistical significance test

- One-tailed Welch's t-test with null hypothesis $H_0 : m_1 \leq m_2$ at 5% significance level.
- m_1 and m_2 are Mean Opinion Scores for a specific content compressed at a specific bitrate with codecs c_1 and c_2
- When the null hypothesis is rejected, the alternative hypothesis indicates that, according to MOS, the first codec is significantly better at the 5% level.
- These tests are performed considering all combinations of codecs: 10x10 matrix.
- For each pair of codecs, we sum up the results for every content at a given quality level: scale 0-6.
- One matrix for each quality level: 4 figures.

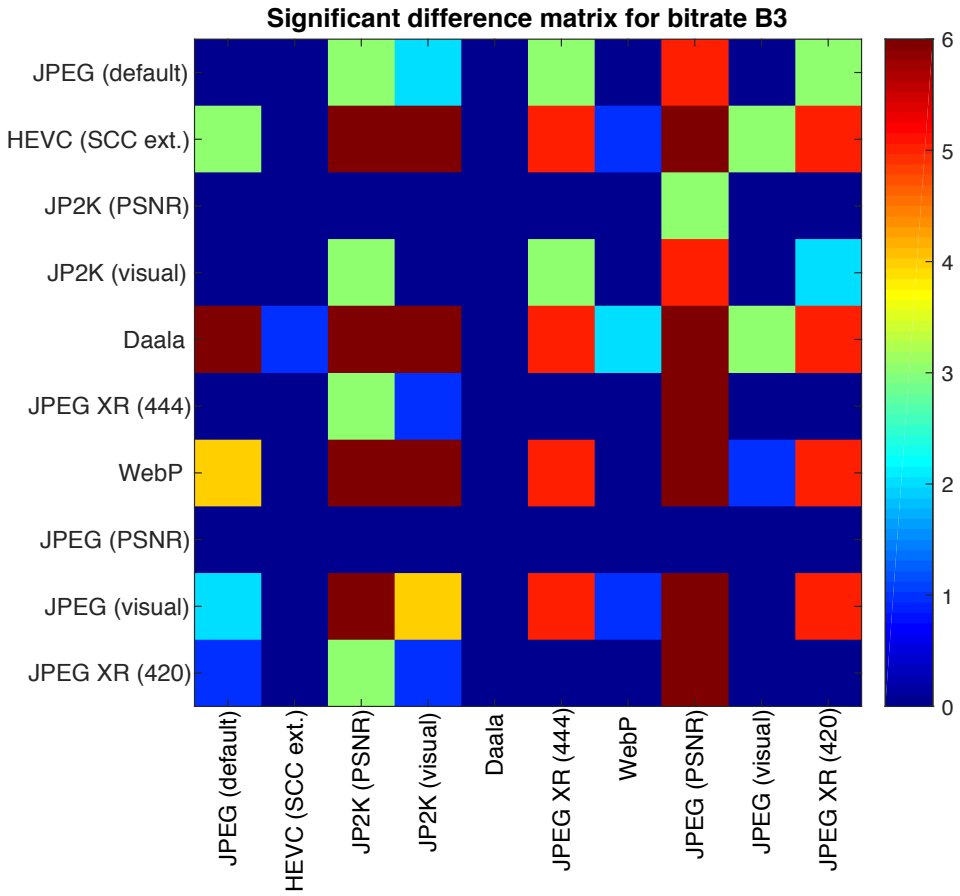
Codec assessment based on subjective tests

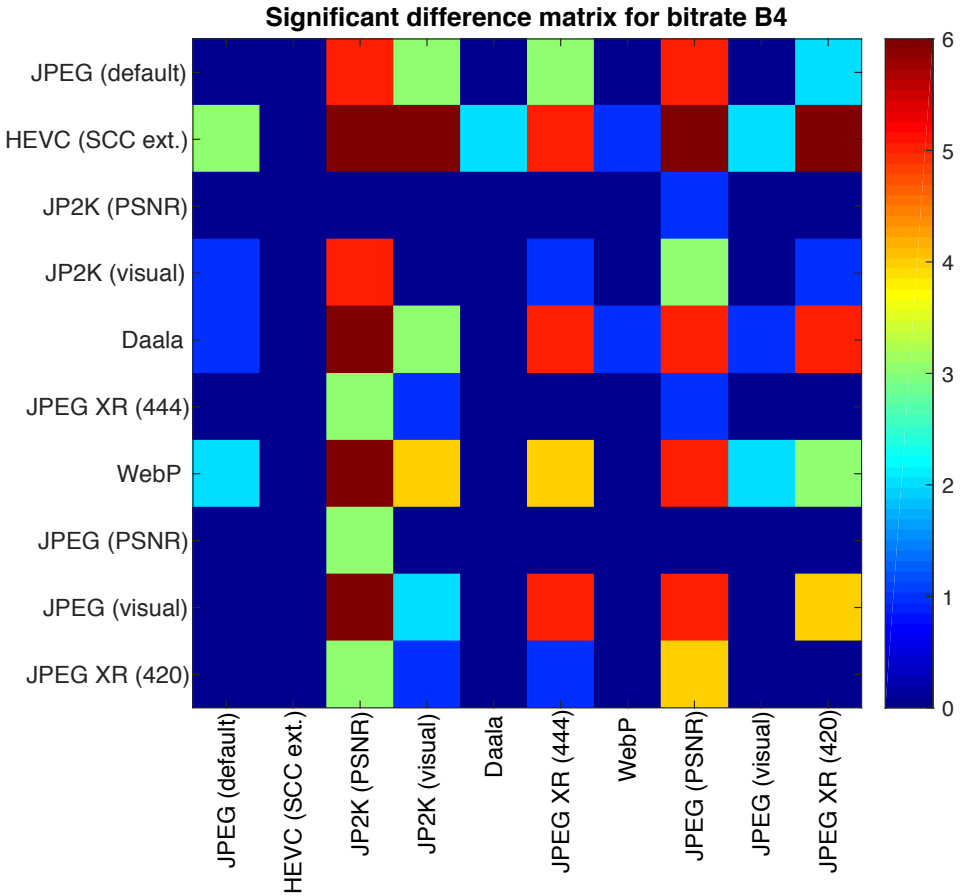




- Bitrate B1 corresponds to the lowest quality
- Bitrate B4 corresponds to the highest quality

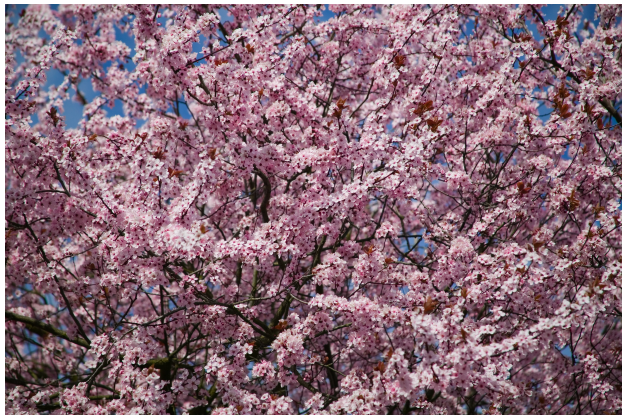
Codec assessment based on subjective tests

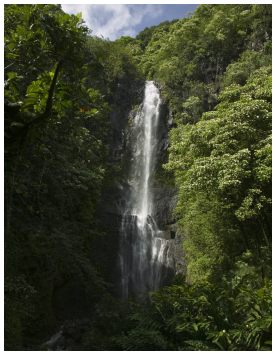




- Bitrate B1 corresponds to the lowest quality
- Bitrate B4 corresponds to the highest quality

Test material in lossless evaluations

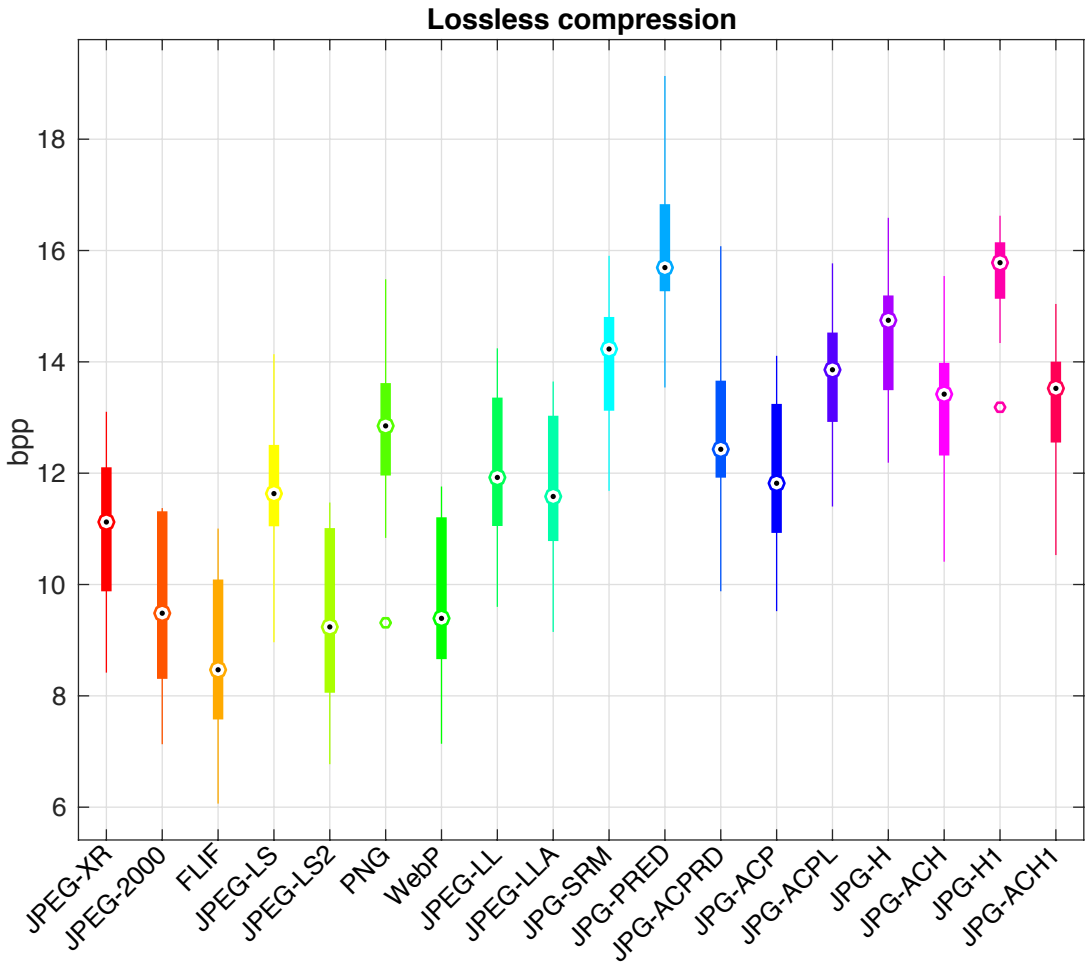





RGB, 444, 24 bpp

Lossless evaluation results



Conclusions

- HEVC and Daala often outperform other codecs in both objective metrics and subjective evaluations in lossy case
  - Daala performs best in images containing faces
- JPEG, JPEG 2000 and JPEG XR perform well in higher bit rates based on $PSNR_{RGB}$ metric in some tested images in lossy case
- JPEG 2000 (PSNR) exhibits good color rendition based on CIE DE2000 metric in several tested images in lossy case
- JPEG (visual) perform well in higher bit rates in lossy case
- FLIF in average produces best lossless compression performance when compared to all alternatives tested for the images tested

Next steps

- There is evidence that significant improvements in compression efficiency can be obtained using latest state of the art in lossy and lossless cases
- Further evaluations are needed to better quantify the cost of such higher efficiency in compression in terms of required resources and other features (delay, etc.)
- Conclusions regarding compression efficiency need to be verified using a larger dataset and through more extensive evaluations campaigns such as crowdsourcing
- Keep in mind these results compare encoders and not coding algorithms (in particular decoders!)
- Many of the algorithms under test are in development and their performance can still improve

Thank you for your attention!