# SCALABILITY EVALUATION OF STILL IMAGE CODERS


*Nikola Sprljan, Divna Djordjevic, Ebroul Izquierdo*

Queen Mary, University of London, Mile End Road, E1 4NS, London, UK
{nikola.sprljan, divna.djordjevic, ebroul.izquierdo}@elec.qmul.ac.uk



## ABSTRACT

Transmission of multimedia content over heterogeneous networks requires highly adaptive compression systems. Therefore, its fully scalable performance is an especially attractive feature, since it enables partial decoding that is adaptive to the given requirements. This work is aiming for establishing of a framework for comparison of scalable still image coders. We give a comparison of scalability features for several popular image compression methods and we propose the methodology for testing.


## 1. INTRODUCTION

The popularity of scalability functionality in image coding is increasing rapidly, mainly due to the recent expansion of multimedia applications using heterogeneous networks. The scalability of a compressed bit-stream refers to methods that allow its partial decoding or transmission in order to adapt to different requirements of client devices. Diverse range of devices implies different capacity regarding CPU power and display capability, thus requiring the scalability of the content resolution, fidelity and decoding complexity. Furthermore, for networks with considerable variation of bandwidth due to congestion or transmission errors, fine precision on-the-fly rate adaptation is sought. All these requirements impose a challenging task, if a universal solution that provides multiple dimension scalability is pursued. Apart from the recently introduced JPEG2000 standard, no other standard offers scalability as an inherent part of the employed coding method (although JPEG standard supports scalability functionalities, it does this only in a limited way since enabling of multiple-dimension scalability imposes a serious penalty on coding efficiency). The key difference between the feasibility of scalability in these two methods lies in the transformation technique used. Subband decomposition based compression techniques, as the wavelet transform in JPEG2000, enable natural resolution scalability, as contrast to DCT in JPEG. Although there are many studies of scalable functionalities for different image compression methods, e.g. [1], there is a general lack of comprehensive comparison between them. This paper presents an effort towards devising a standard methodology for assessing the scalability performances of compression methods.

## 2. SCALABILITY IN IMAGE COMPRESSION

Scalability is an ability of a compression system to adapt to the given conditions and to generate a bit-stream that contains the information ordered by importance. If a compressed bit-stream of an image is embedded, then by truncating the bit-stream at any point an efficient representation of that image at the desired level of compression can be obtained. As more and more bits are received the quality of the reconstructed image increases. The effect of embedding is that information content from any spatial part of an image is dispersed throughout the compressed representation ("dispersion principle", [2]). If an opposite case is considered and an image is processed locally, the rate-distortion curve corresponding to a non-embedded bit-stream is linear. If a fully embedded bit-stream is truncated the distortion introduced in this way will be much smaller than if a non-embedded bit-stream was truncated Fig. 1. (dashed line). This is because coefficients in the bit-stream that have the greatest contribution to the reduction of the distortion and correspond to various spatial parts of the image are transmitted first.

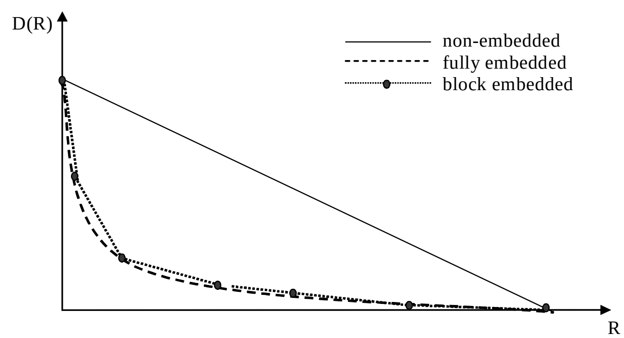



**Fig. 1.** Distortion introduced by truncating a bit-stream

The scalablility of an image is related to the concept of progressive ordering of the bit-stream. If an algorithm produces a bit-stream that is scalable for different image features (e.g. resolution, quality), various progression orders are possible. Several types of progression orders are:

- Quality (or SNR) progression, where the quality of decoded image is improved as the amount of received data is increased;
- Resolution progression, where the first part of the bit-stream is organised to represent a reduced resolution representation of the image and as the portion of transmitted bit-stream increases so does the resolution of the image;
- Spatial progression, where the bit-stream is arranged in such way that elements of the bit-stream that describe the same spatial regions are placed together;
- Component progression, where each component of a multiple component image, e.g. RGB or YUV, can be received and decoded separately.

According to the scalability feature of a given method, compressed bit-streams fall into several categories, ordered in decreasing complexity as follows:

- Transcoding in spatial domain - the complete process of decompression / compression has to be performed;

- Transcoding in compressed/transformation domain - inverse / forward transformations are not needed;
- Parsing - only the relevant portion of the compressed bit-stream is selected and extracted, no compression / decompression functions are needed;
- Truncation - special case of parsing method, when only a prefix of the compressed bit-stream is extracted.

In the literature, only the bit-streams from the two last mentioned categories are regarded as truly scalable bit-streams, since in these cases the resource adaptation engine can operate in generic, and content independent manner.

## 3. EVALUATED IMAGE CODING ALGORITHMS

In this section a brief overview of the evaluated coding techniques, with an emphasis on their main scalable functionalities, is given.

*EBCOT algorithm within the JPEG2000 Standard* - A wavelet based technique named EBCOT (Embedded Block Coding with Optimal Truncation) is the key algorithm in the JPEG2000 standard. The most distinct characteristics of this technique result from the partition of the wavelet subbands into basic coding units, so called code blocks, of typical size of 64×64 or 32×32 coefficients. Each code block is coded independently and as a result, an elementary embedded bit-stream for each code block is created. In order to find optimal truncation points for each elementary bit-stream, post-compression rate-distortion optimisation (PCRD-opt) is employed, followed by combining of bit-streams in a predetermined order. Further details on the EBCOT algorithm can be found in [2].

The bit-stream obtained using JPEG2000 is both resolution and spatialy scalable since code blocks corresponding to a particular subband or spatial part are independent from other code blocks. To achieve the SNR scalability, quality layers are formed by collecting portions of bit-streams from various code blocks. JPEG2000 supports all four scalability functionalities defines five progression schemes with the ability of a bit-stream reorganisation. Lower quality and resolution layers can be obtained by a simple parsing of the bit-stream. Depending on the progression order even simple bit-stream truncation is capable of giving optimal performances since the decoder is capable of reconstructing the image from any prefix of the original bit-stream. The set of truncation points obtained by the PCRD-opt algorithm corresponds to bit rates for which the performance is optimal and the number of these points is a significant parameter for scalable performances. A compromise must be made since selecting too many of points results in excessive additional information which causes impairment of the overall performance, and selecting too few of them results in large suboptimal bit rate segments. In our tests, the number of 10 logarithmically spaced truncation points was found to be optimal, and it was used in all comparisons. Since the bit-stream possesses moderately high amount of header data (178 bytes in our tests), in comparisons the header was neglected in bit rate calculations. The header of the bit-stream for the other tested algorithms was neglected as well.

*SPIHT* - As opposed to JPEG2000, SPIHT (Set Partitioning In Hierarchical Trees, [3]) relies strongly on the correlation that exists among wavelet coefficients across subbands of different resolution levels. SPIHT exploits this property of the wavelet transform by representing clusters of coefficients within zerotree structures and encoding them jointly. The internal state information of the process is efficiently encoded using lists that contain positions of insignificant coefficients, insignificant sets of coefficients and already coded (significant) coefficients. One major drawback of this approach is that it imposes a dependency between bits that are scattered across the bit-stream, such that certain parts of the bit-stream cannot be decoded without the preceding part. Regarding the quality scalability, SPIHT gives a bit-precision embedded bit-stream that enables truncation at any position, but on the other hand, it is harder to achieve resolution scalability due to the mentioned cross-subband dependencies. Although it is possible to reorder the bit-stream during encoding to achieve resolution and spatial progression on each bit plane, there is a penalty on the bit budget since some bits obtained by testing the significance of insignificant sets would be redundant. Therefore, in order to achieve resolution scalability for SPIHT we use transcoding technique in compressed domain, but to make complexity comparable to parsing as it is used in JPEG2000, and therefore to make comparison fairer, arithmetic coding is not performed. This binary-coded version of SPIHT gives just slightly worse results than its arithmetic-coded version in terms of *PSNR*, which is an important advantage of this compression method.

*JPEG* - Apart from the baseline mode of the JPEG standard, progressive mode supports progression by quality with successive approximation and spectral selection methods, in which DCT coefficients are encoded in bit-plane by bit-plane scans or are partitioned into successive scans [2]. In our tests only the first method was used. Resolution scalability (called hierarchical refinement in JPEG) can be achieved by compression of multiple images that correspond to different resolution levels. Since it reduces the overall performance considerably, this particular feature was not tested.

## 4. TEST CONDITIONS AND METHODOLOGY

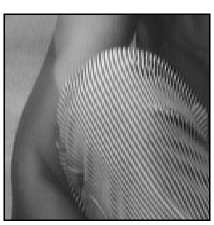
In our tests, images are compressed without a priori knowledge of the decoding bit rates. This simulates the scenario where the full scalability is sought (e.g. multi-user wireless system with adaptive bandwidth allocation), or generally when the environment of the future usage of the image is unknown (e.g. fitting the compressed image to available size on a storage unit). This test measures the performance in terms of PSNR for a selected range of resolutions and quality layers. The test image was compressed once and the obtained bit-stream was used for extracting images of reduced resolution and quality. In the following discussion we will denote resolution layers as follows: original resolution as 'full', first lower as '1/4' or 'quarter' (each image dimension halved), the next lower as '1/16', and so on. For each resolution layer, the image was decoded on various bit rates, from the minimum to the maximum rate in steps of 0.01 bpp. This resulted in the rate-distortion curve for a particular resolution layer. Although the information encompassed within obtained curve is invaluable, it is generally difficult to form a final objective decision on the overall performance of the particular coder for some application. As a solution, we propose an empirical method that measures the overall performance and obtains a single value - scalability performance measure $\sigma$. The measure is expressed with the relation:

$$\sigma = 10\log_{10}\left(255^2 \Big/ \sum_i w_i d\left(b_i\right)\right) \ [\mathrm{dB}], \quad \text{where } \sum_i w_i = 1,$$

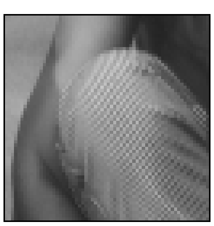
where $b_i$ are the selected output bit rates, $d(b_i)$ distortion of selected coder expressed with *MSE* and $w_i$ are the weights of the given bit rates. Normalisation to the maximum possible distortion and logarithmic representation allow easier comparison of the results. The rationale behind weighting is that if there exists some a priori knowledge on the particular application, specifically if the probability of a user requesting a certain bit rate is known, then the distribution of weights represents the probability density function (pdf) of requests. The case where this a priori knowledge is not available is represented with a flat pdf, meaning that each bit rate is equally probable and is assigned the same weight. In that way, $\sigma$ is the mean distortion that the users experience. To test our method for a particular application we devised a simple wireless communication scenario where bandwidth capacity of the receiver device is related to a received signal level that is transmitted from uniformly configured array of base stations. Additionally, we assume that the highest bandwidth corresponds to the highest quality compressed bit-stream (maximum bit rate of our target rates), and the lowest bandwidth to the lowest quality bit-stream (0.01 bpp). In this situation only a small portion of receivers are near these extremes and the majority of them are grouped around some mid-level quality. We modelled pdf function for such scenario as a piecewise linear with the maximum at 75% of the quality range, and in the results this model is denoted as *wireless* while the flat pdf is denoted as *flat*.

Computing the PSNR values for the comparison on lower resolution layers is not as straightforward as it is in original resolution case where the reference is the uncompressed image. Given the uncompressed image, the most important question is how to obtain the low resolution reference images for resolution scalability performance evaluations. The main difficulty is the fact that when the reduced resolution image is obtained directly from lower wavelet subbands, it results in the artifacts which are usually not present in the standard image resizing techniques. Firstly, the choice of the wavelet filter and subsampling pattern results in different pixel shifts of the reduced resolution image. To solve this problem we adjusted subsampling pattern to comply with the pixel shift of both compared compression methods. For the filter to be used for resizing in our experiments, we designed Hamming-window based, linear-phase filter with cutoff frequency that corresponds to the desired reduction in resolution (hereafter denoted as "reference filter"). Such filtered, downsampled and appropriately shifted image was then used in comparisons as the reference image. However, the difference in frequency response between our reference filter and wavelet used in image compression method is another cause for the difference between reference image and decoded lower resolution image. The frequency response of the wavelet that is used in both of the examined compression methods – Cohen, Daubechies, Feauveau 9/7 filter pair (or shorter CDF9/7) from [4], shows that aliasing introduced with subsampling is much more present than in the reference filter, which has response closer to the ideal that causes no aliasing. To justify our approach we performed several subjective tests, and concluded that for some images with areas that contain periodic high frequency content, the aliasing was highly noticeable on lower resolution images obtained with CDF9/7, compared to those obtained with our reference filter. We also compared our filter with MPEG half-band filter (B) from [5]. For example, Fig. 2. shows the visual difference between using these filters. The one-dimensional frequency content representation graph of the images, shown on the right, proves that the difference between CDF9/7 and our filter is due to aliasing. The MPEG half-band filter causes considerable image blurring since its response strongly filters the higher frequency components, as can be deduced from the graph.

However, since the result of resizing is highly subjective (e.g. under certain conditions aliasing is present even in the human visual system), results of comparison for the lower resolution layers should be taken only as an indicator of a possible outcome of a subjective comparison. However, if the obtained rate-distortion curve lies closely to the one obtained when efficient compression method is applied directly on the reference image, this would be an indication that scalability is efficiently accomplished. For that purpose we used the arithmetic version of SPIHT of which the obtained results are displayed together with the results from compared scalable methods (Fig. 3.b). One might argue here that the logical choice for reference filter would be CDF9/7 since that one is used in both JPEG2000 and SPIHT, but here the general comparison system, which is impartial regarding the used compression method, is sought. It should be also noted that the problem of selecting the correct conditions for scalable image and video compression comparison is still unsolved issue and effort is being made for devising a standard evaluation methods [6].

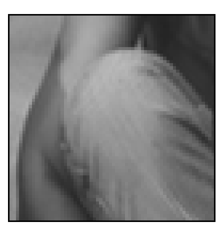


a) Original



b) CDF9/7

c) Reference filter



d) MPEG (B) half-band filter

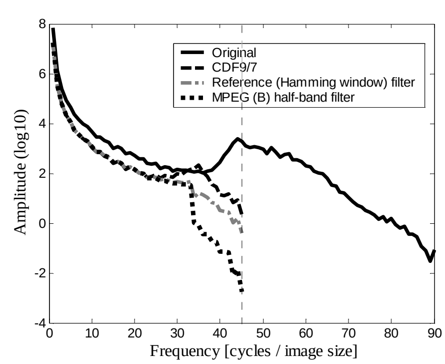



**Fig. 2.** The effect of aliasing on reduced resolution representation (1/4)

## 5. TEST RESULTS

Fig. 3. shows the results of comparison for two selected coders, for two highest levels of resolution. It is important to note here that the bit rate that is expressed as the number of bits per pixel do not correspond to the number of pixels in particular resolution but to the number of pixels in the original image. It can be seen that JPEG2000 and SPIHT perform similarly on the selected range of output bit rates, with slight advantage of JPEG2000 on higher bit rates. As it was expected, JPEG performs significantly worse on all bit-rates. Displayed PSNR values for JPEG2000 are with header size included, and for that reason the considerable gain of SPIHT compared to the JPEG2000 is present on the lowest bit rates. To demonstrate the importance of the selection of target bit rates for JPEG2000 one extreme case is displayed on the graph on Fig. 3.a). In this case the compressed bit-stream consists only of one quality layer, and truncation of bit-stream on any rate lower than the final rate produces considerably sub-optimal results.

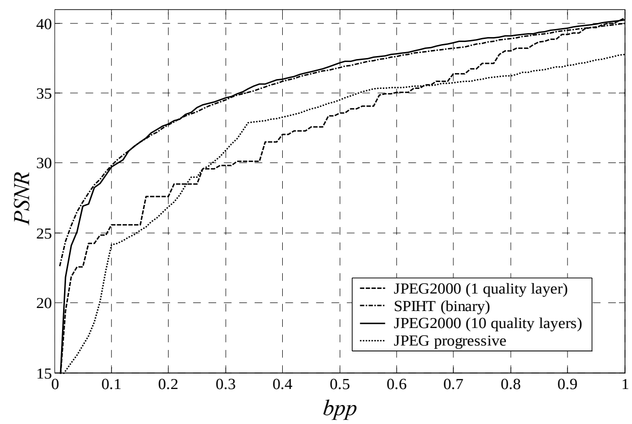
Comparison graph for 1/4 resolution shows similar behaviour, with even less difference between compared methods. It is also interesting to observe that in SPIHT more bits are spent on lower resolution layers than it is in JPEG2000. This is deducible from the value of the highest bit rate after which the rate-distortion curve for 1/4 resolution is flat, meaning there is no more bits in the bit-stream that describe corresponding resolution level, which is ~0.6 bpp in the case of JPEG2000 and ~0.7 bpp in the case of SPIHT. We observed the same behaviour for other test images. The discrepancy on the high bit rates that is noticeable between rate-distortion curve obtained with arithmetic SPIHT on reference image and curves obtained with two compared methods can be attributed to the difference between employed filters. Table 1. confirms the closeness of results for SPIHT and JPEG2000, and inferior performance of JPEG.

## 6. CONCLUSION

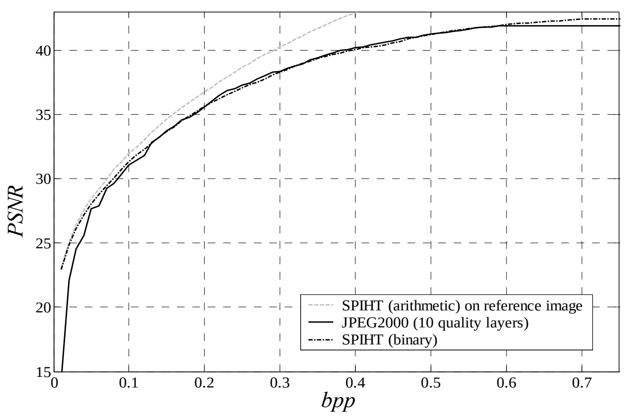
A comparison of scalability functionalities in still image coding methods was presented. In order to make objective comparison easier, a scalability performance measure that expresses performance with one value was devised. The problem of finding lower resolution reference image was addressed and a method for comparison of the performance on lower resolution layers was proposed. The test showed similar performance of SPIHT and JPEG2000, while JPEG performed considerably worse. However, JPEG2000 represents the best technique of all the evaluated since only that one provides true scalability for all dimensions in one bit-stream.

**Table 1**. Scalability performance measure $\sigma$ [dB]

| | | JPEG2000 | | SPIHT | | JPEG | |
|---|---|---|---|---|---|---|---|
| | | Goldhill | Lena | Goldhill | Lena | Goldhill | Lena |
| *flat* | full res | 30.87 | **33.36** | **30.92** | 32.69 | 26.15 | 26.04 |
| | 1/4 res. | 33.37 | **35.15** | **33.75** | 34.07 | - | - |
| | 1/16 res. | 36.25 | **37.30** | **37.47** | 35.66 | - | - |
| | 1/64 res. | 38.14 | **38.71** | **39.41** | 37.05 | - | - |
| *wireless* | full res | **33.29** | 36.64 | 32.98 | **36.81** | 30.79 | 32.23 |
| | 1/4 res. | **37.17** | **39.90** | 37.16 | 39.70 | - | - |
| | 1/16 res. | 39.97 | **40.55** | **40.68** | 40.37 | - | - |
| | 1/64 res. | 40.1 | **39.67** | **40.31** | 39.50 | - | - |




a) Full resolution




b) 1/4 resolution

**Fig. 3.** Comparison of scalability performance for JPEG, JPEG2000 and SPIHT for test image "Lena"

**Acknowledgment:** We wish to acknowledge support provided by the European IST Project BUSMAN, under grant nr. IST BUSMAN-2001-35152.

## References:


[1] M.D. Adams, "The JPEG-2000 Still Image Compression Standard", ISO/IEC JTC 1/SC 29/WG1/N2412, Dec. 2002.

[2] D. Taubman, M.W. Marcellin, *JPEG2000 Image Compression: Fundamentals, Standards and Practice*, Kluwer Academic Publishers, 2002.

[3] A. Said, W.A. Pearlman, "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.6, no. 3, pp .243-250, June 1996.

[4] A. Cohen, I. Daubechies, J.C. Feauveau, "Biorthogonal Bases of Compactly Supported Wavelets", *Communications on Pure and Applied Mathematics*, vol. 45, no. 5, pp. 485-560, May 1992.

[5] W. Li, J.-R. Ohm, M. van der Schaar, H. Jiang, S. Li, "MPEG-4 Video Verification Model version 19.0", M10431 in *Contribution to 67th MPEG Mtg.*, Hawaii, December 2003.

[6] J.-R. Ohm, K. Hanke, "Principles for evaluation of scalable wavelet coding technology", M8207 in *Contribution to 59th MPEG Mtg.*, Jeju Island, March 2002.