FACTA UNIVERSITATIS (NIŠ)
SER.: ELEC. ENERG. vol. 20, no. 1, April 2007, 71-83

Image Quality Evaluation: JPEG 2000 Versus Intra-only H.264/AVC High Profile

Sang-Gyu Cho, Zoran Bojković, Dragorad Milovanović, Jungsik Lee, and Jae-Jeong Hwang

Abstract: The objective of this work is to provide image quality evaluation for intra-only H.264/AVC High Profile (HP) standard versus JPEG2000 standard. Here, we review the structure of the two standards and the coding algorithms in the context of subjective and objective assessments. Simulations were performed on a test set of monochrome and color image. As a result of simulations, we observed that the subjective and objective image quality of H.264/AVC is superior to JPEG2000, except the blocking artifact which is inherent, since it consists of block transform rather than whole image transform. Thus, we propose a unified measurement system to properly define image quality.

Keywords: Image coding quality, subjective/objective assessment, JPEG2000 standard, H.264 standard.

1 Introduction

Modern image compression techniques offer the possibility to store or transmit the vast amount of data necessary to represent digital images in an efficient and robust way [1]. With increasing use of multimedia technologies, image compression requires higher performance as well as new features. To address this need, some standards have been developed. For example, JPEG 2000 standard is intended not only to provide rate distortion and subjective image quality performance superior to existing standards, but also to provide features and functionality that current standards can either not address efficiently or in many cases cannot address at all [2].

Manuscript received December 14, 2006.

S. Cho, J. Lee and J. Hwang are with School of Electronic and Information Engineering, Kunsan National University Kunsan, Chonbuk 573-701, Korea. Z. Bojković and D. Milovanović are with Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia (e-mail: zoranb@sf.sf.bg.ac.yu and dragoam@gmail.com).

Different from its preceding generation, JPEG, which is a Discrete Cosine Transform (DCT) based Huffman coder, JPEG 2000 adopts a wavelet-based arithmetic coder for better efficiency. Actually, JPEG 2000 not only enhances the compression but also includes many new features, such as quality scalability, resolution scalability, region of interest (ROI), lossy and lossless in unified framework [3].

In 2003, the new video coding standard H.264/Advanced Video Coding (AVC) was finalized and it provides a new way of still image coding [4]. The improvement in coding performance comes mainly from the prediction part. Unlike the previous standards, prediction must be always performed before texture coding for both inter and intra macroblocks. Intra prediction significantly improves the coding performance of H.264/AVC intra frame coder [5]. Core techniques and architectures in the two image coding standards are shown in Table 1.

Table 1. Core techniques adopted in two different image coding standards under the basic coding architecture.

Standard	Prediction	Transform	Quantization	Entropy coding
JPEG 2000	N/A	Frame-based 2-D DWT	Scalar visual weighting	Bitplane Context-based Adaptive Binary Arithmetic Coding and Truncation
H.264/AVC Intra Coder	$4\times4/16\times16$ luma, 8×8 chroma	4×4 2-D DCT	Scalar nonuniform quantization	Exp-Golom Coding and Context-based Adaptive VLC/Binary Arithmetic Coding

In order to specify, evaluate and compare video communication systems, it is necessary to determine the quality of the video image to the viewer. Measuring visual quality is a difficult and often imprecise task because there are so many factors that can affect the result. Visual quality is inherently subjective and is influenced by many factors that make it difficult to obtain a completely accurate measure of quality. For example, a viewer's opinion of visual quality can depend very much on the task at hand, such as passively watching a DVD movie, actively participating in a video conference, communicating using sign language or trying to identify a person in a surveillance image scene.

Also, measuring visual quality using objective criteria gives accurate results, but as yet there are no objective measurement system that completely reproduces the subjective experience of a human observer watching video display. Our perception of a visual scene is formed by a complex interaction between the components of the *Human Visual System* (HVS) in the eye and the brain. The perception of visual quality is influenced by spatial fidelity and temporal fidelity. However, a viewer's opinion of quality is also affected by other factors such as the viewing environment, the observer's state of mind and the extent to which the observer interacts with visual scene. Other important influences on perceived quality include visual attention. All of these factors make it very difficult to measure visual quality accurately and quantitavely [6].

The most widely used objective quality measure is Peak Signal-to-Noise Ratio (PSNR). It can be calculated easily and quickly and is therefore a very popular quality measure, widely used to compare the quality of original and reconstructed video images. The PSNR measure suffers from a number of limitations. PSNR requires unimpaired original image for comparison but this may not be available in every case and it may not be easy to verify that an original image has perfect fidelity. It, moreover, does not correlate well with subjective video quality measures such as those defined in ITU-R BT.500-11. For a given image or image sequence, high PSNR usually indicates high quality and low PSNR usually indicates low quality. However, a particular value of PSNR does not necessarily equate to an absolute subjective quality.

In this work, two image standards, i.e., intra-only H.264/AVC High Profile (HP) and JPEG 2000, are compared from the point of view of image quality evaluation. Section 2 describes image quality evaluation tools, while simulation results are provided in Section 3. Finally, conclusions are drawn in Section 4.

2 Image Quality Evaluation Tools

To evaluate image quality, two methods, i.e, subjective and objective quality assessment will be used. H.264/AVC intra frame coder and JPEG 2000 standard are compared.

2.1 Subjective quality assessment

Overview of subjective test methods defined in ITU-R Rec. BT.500-11 is shown in Figure 1.

Characteristics of reference based and reference free tests are shown in Table 2 [6].

Table 2. Reference based and reference free tests characteristics.

	Reference	Stimulus	Assessment time
DSIS	need	double	short (54 60s)
SDSCE	need	double	very long (30 60 min)
DSCQS	need	double	short (54 60s)
SS	no need	single	short (41 47s)
SC	no need	single	short (54 60s)
SSCQE	no need	single	very long (30 60 min)

In categorical judgment, observers assign an image or image sequence to one

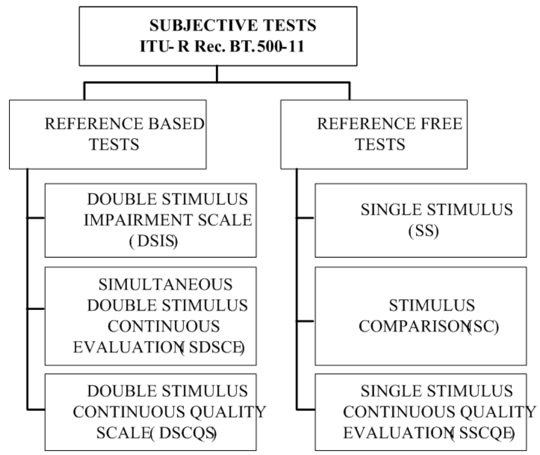


Fig. 1. Subjective test methods overview.

of a set of categories that typically are defined in semantic terms. The categories may reflect judgments of whether or not an attribute is detected. Categorical scales that assess image quality and image impairment have been used most often. The ITU-R Rec. BT.500-11 specifies 5-point (excellent-good-fair-poor-bad) quality / impairment (imperceptible perceptible but not annoying annoying very annoying) rating scales and 7-point (much better - better -slightly better - the same - slightly worth - worse - much worse) comparison scale.

2.2 Objective quality assessment

The model based measurement procedures for objective quality assessment is shown in Figure 2.

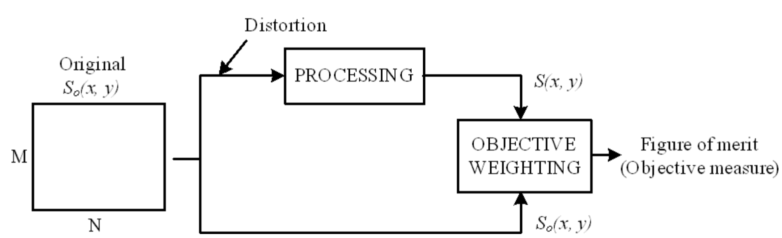


Fig. 2. Image quality objective measures.

The most widely used objective methods are Mean Square Error (MSE) and

Peak Signal-to-Noise Ratio (PSNR). The former is given by

$$MSE = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M}(S(x,y) - S_0(x,y))^2}{N \times M} \tag{1}$$

where $S(x,y)$ and $S_0(x,y)$ denote the reconstructed and original sample, respectively. The latter is given

$$PSNR = 10\log_{10}\frac{S_{max}^2}{\sigma_{diff}^2} \quad (\mathrm{dB}) \tag{2}$$

where denotes the differential power and is the power of the peak amplitude (8 bits→255). This objective quality measure is a measure for how much the individual pixels differ from the original image. The higher the PSNR, the higher the objective quality and the more the image resembles the original. This does not necessarily mean that the subjective image quality is also higher.

Also, the other methods for objective quality assessment like *Delta*, *Blurring*, *Blocking*, *Structural SIMilarity* (SSIM) index, as well as Video *Quality Metric* (VQM) should be taken into account, too.

The value of the *Delta metric* is the mean difference of the color components in the corresponding points of image, i.e,

$$Delta = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M}(S(x,y) - S_0(x,y))}{N \times M} \tag{3}$$

Blurring metric allows to compare power of blurring of two images. If value of the metric for the first picture is greater than for the second, it means that the second picture is more blurred than first. *Blocking metric* was created to image visual measure of blocking. More bright areas correspond to greater blocking artifact [7].

Structural similarity (SSIM) *index* is based on measuring three components: luminance similarity, contrast similarity and structural similarity. The system diagram of SSIM index is shown in Figure 3.

The luminance of each signal x and y is compared. It is estimated as the mean intensity

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

The luminance comparison function $l(x,y)$ is then a function of μ_x and μ_y. The contrast comparison $c(x,y)$ is then the comparison of σ_x and σ_y. The signal contrast is given by

$$\sigma_x = \left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right]^{\frac{1}{2}} \tag{5}$$

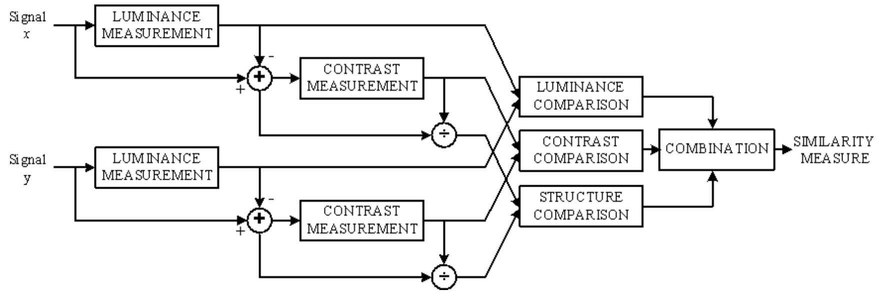


Fig. 3. Diagram of the structural similarity (SSIM) measurement system.

The structure comparison $s(x,y)$ is conducted on these normalized signals $(x-\mu_x)/\sigma_x$ and $(y-\mu_y)/\sigma_y$. Then, the three components are combined to yield an overall similarity measure

$$SSIM(x,y) = f(l(x,y)\,c(x,y)\,s(x,y)) \tag{6}$$

where $f(\cdot)$ is the combination function. More bright areas correspond to greater difference [8].

In *Video Quality Metric* (VQM), Discrete Cosine Transform (DCT) is used to correspond to human perception. Figure 4 is an overview of VQM system diagram [9].

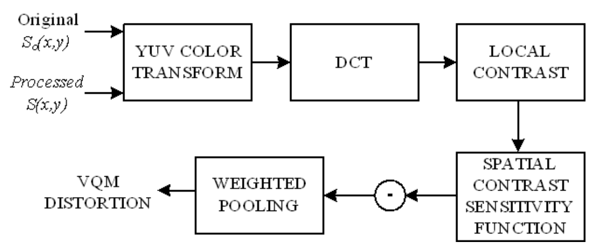


Fig. 4. Diagram of VQM measurement system.

The main feature of VQM are:

- YUV Color transform
- DCT transform. This step separates incoming images into different spatial frequency components.
- Convert each DCT coefficients to local contrast (LC) using following equation

$$LC(i,j) = DCT(i,j)\left(\frac{DC}{1024}\right)^{0.65}\frac{1}{DC} \tag{7}$$

DC is the DC component of each block. For 8 bits image, 1024 is mean DCT value. The best parameter for fitting psychophysics data is 0.65. After this step, most values lie between $[-1,\ 1]$.

- The $LC(i,j)$ are converted to just-noticeable differences by multiplying each DCT coefficient by its corresponding entry in the human spatial contrast sensitivity function (SCSF) matrix.
- Weighted pooling of mean and maximum distortion.

$$VQM = Mean\_dist + 0.005 \times Max\_dist \tag{8}$$

where

$$diff = S_0(x,y) - S(x,y) \tag{9}$$
$$Mean\_dist = 1000 \times mean(mean(abs(diff))) \tag{10}$$
$$Max\_dis = 1000 \times maximum(maximum(abs(diff))) \tag{11}$$

Maximum distortion weight parameter 0.005 is chosen based on several primitive psychophysics experiments. Parameter 1000 is the standardization ratio [9].

3 Simulation Results

In order to evaluate image quality using H.264/AVC intra frame coder, the comparison is carried out with JPEG 2000 coder. The simulation results are shown in both subjective and objective quality assessments. The software used for H.264/AVC is the Reference Software JM10.1v developed by the VCEG JVT, while for JPEG 2000, JasPer.1.701.0v. The input sequence is Lena with the size 512×512, while the used format is 4:0:0 for gray and 4:2:0 for color.

The configuration of H.264/AVC encoder is as follows: High Profile (HP) intra coding, with bit rate less than 64 kbits per image for monochrome and less than 136 kbits per image for color. We used 8×8 transform mode enabling adaptive choice between 4×4/8×8 transform and prediction modes. Also, the simulation is carried out with CABAC, loop filter disabled, and the rate distortion R(D) optimization enabled.

As for the configuration of JPEG2000 encoder, 9/7 filter (lossy default) wavelet transform with 5 levels of wavelet decomposition was used. Also, we have single layer mode and R(D) optimization of packet for a given target rate, as well as embedded block coding with optimized truncation scheme (EBCOT). Because of their specific characteristics (Table 2), two subjective quality methods are chosen: DSCQS and SC.

3.1 Double stimulus continuous quality scale (DSCQS)

The double-stimulus method is thought to be especially useful when it is not possible to provide test stimulus test conditions that exhibit the full image of quality. The method is cyclic in that the processor is used to view a pair of pictures, each from the same source, but one is the process under examination and the other one directly from the source. It is used to assess the quality of both. The method requires the assessment of two versions of each test picture. One of each pair of test pictures is unimpaired while the other presentation might or might not contain an impairment. The unimpaired picture is included to serve as a reference. But the observers are not told which is the reference picture. In the series of tests, the position of the reference picture is changed in pseudo-random fashion. The observers are simply asked to assess the overall picture quality of each presentation by inserting a mark on a vertical scale. The vertical scales are printed in pairs to accommodate the double presentation of each test picture. The scales provide a continuous rating system to avoid quantizing errors, but they are divided into five equal lengths which correspond to the model ITU-R five-point quality scale. The associated terms categorizing the different levels are the score as those normally used.

Test image *Lena* is encoded by two standards and the reconstructed images are displayed for the subjective assessment using the DSCQS. The size of compressed bitstream is 35.4 kbits per image. As shown in Figure 5, five quality scales are marked on the five-step quality chart for reference and reconstructed image.

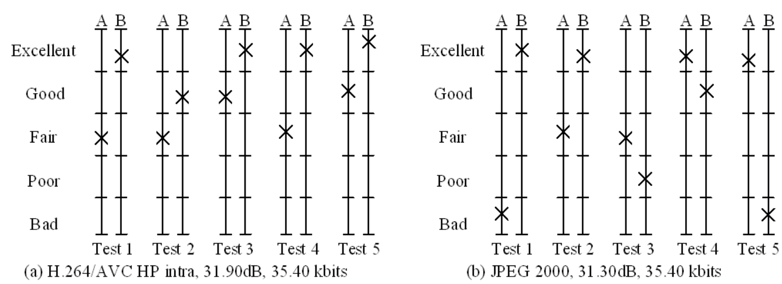


Fig. 5. DSCQS quality scales of Lena image with (a) H.264/AVC HP intra coding and (b) JPEG 2000 coding.

The pairs of assessment (reference and test) for each test condition are converted from measurement of length on the score sheet to normalize scores in the image 0 to 100. Then, the differences between the assessment of the reference and the test condition are calculated. It can be concluded that in this case H.264/AVC HP intra coding is better.

3.2 Stimulus comparison (SC)

In stimulus comparison methods, two images or sequences are displayed and the viewer provides an index of the relation between the two presentations. This method yields a distribution of judgements across scale categories for each condition pair. The way that responses are analyzed depends on the judgement made (e.g. difference) and the information required (e.g. just noticeable differences, ranks of conditions, distances among conditions, etc.). The main features of this method are :

- no reference available,
- images (sequences from two different quality stages are displayed simultaneously or consecutively (if possible for all combinations),
- assessor provides an index of the comparison of the two actual images (or sequences), e.g., (A) and (B),
- comparison is made on the basis of the comparison scale.

Example of SC method is shown in Figure 6.

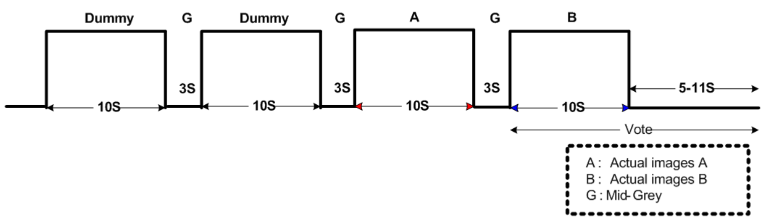


Fig. 6. Example of SC method.

The subjective quality assessment using stimulus comparison (SC) scales of Lena sequence with H.264/AVC HP intra coding and JPEG 2000 coding testing 35.40 kbits per image is shown Table 3.

Table 3. SC scales of Lena image using H.264/AVC HP intra-coding and JPEG2000 with 35.40 kbits per image.

Assessor	H.264/AVC HP intra	JPEG 2000
1	+1	−1
2	+1	−3
3	+3	−1
4	+2	−2
5	+1	−1

It can be seen that H.264/AVC HP intra coding gives better results than JPEG 2000 standard coding when the SC method of subjective quality assessment is used.

3.3 Objective quality assessment

The success of a coding standard relies primarily on its improved coding efficiency when compared to other existing or emerging coding standards. Although several studies, [10–12] have focused on coding performance comparisons of JPEG 2000 with its predecessors in the still image coding, little is known about the relative performance of JPEG 2000 in terms of coding efficiency compared to its competition in the area of video coding.

On the other side, H.264/AVC has been shown to provide a major breakthrough with regard to compression efficiency [13]. But again, most of the published studies were intended to evaluate the overall coding performance of H.264/AVC in comparison to prior video coding standards [14].

We investigate the intra coding performance of H.264/AVC High Profile (HP) in comparison with JPEG 2000. The test images are Lena, monochrome and color with resolution 512$\times$512 samples. This can be considered as a good representation of the typical spectrum of imagery targeted by a still image coding standard.

Figure7 shows Lena monochrome images comparing H.264/AVC HP intra coding (a) to JPEG-2000 coding (b), both with 35.40 kbits per image. It can be seen that the image (b) shows the blurring distortion and loss of details. Image (a) is better than (b), but there are still blocking artifacts in it.

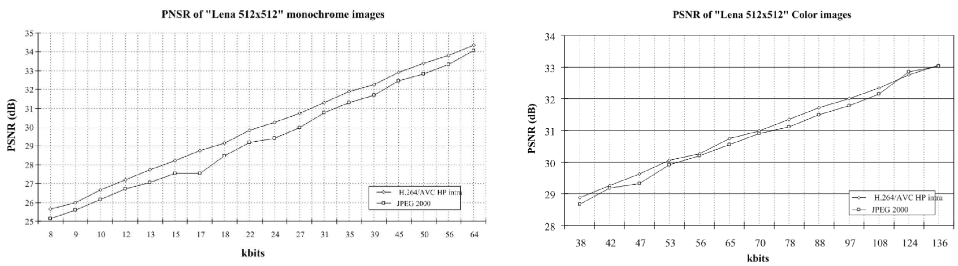
(a) H.264/AVC HP intra, 31.90 dB, 35.40 kbits

(b) JPEG 2000, 31.30 dB, 35.40 kbits

Fig. 7. Lena monochrome images comparing H.264/AVC HP intra coding (a) and JPEG 2000 coding (b) with 35.40 kbits per image.

Figure 8(a) shows the rate distortion graphs as the outcome of our coding experiment for Lena monochrome image comparing H.264/AVC HP intra coding to JPEG 2000 on each kbits per image. Lena image is taken at less than 64 kbits per image. The peak signal-to-noise ratio (PSNR) of H.264/AVC HP intra coding is higher than JPEG 2000 for about between 0.15 dB to 0.76 dB at all of the tested bit rates. It means that there is an average gain in favor of the H.264/AVC HP intra coding.

Figure8(b) presents the results of comparing the performance of H.264/AVC HP intra coding and JPEG 2000 for Lena color image at less than 136 kbits per image. As it can be observed, there is a certain advantage in favor of H.264/AVC HP intra coding from 0.02 dB to 0.30 dB.



(a) (b)

Fig. 8. (a) Lena monochrome image, and (b) Lena color image comparing of H.264/AVC HP Intra coding and JPEG 2000 coding on each kbits per image.

In case of color image, the PSNR (dB) gain is smaller compared to monochrome image. There are at least two reasons for that (1) the invoked intra prediction in H.264/AVC HP and (2) the AC chroma coefficients of that component for the whole macroblock are set to zero [15, 16].

The comparison results using other methods (Delta, Blurring, Blocking, SSIM index and VQM) for objective quality assessment considering Lena 512×512 monochrome and color images are shown in Table 4 and Table 5, respectively.

From Table 4, we can see that most of the measurement metrics (Delta, Blurring and VQM) are better for H.264/AVC HP intra coding comparing to JPEG 2000. Only Blocking and SSIM index elements are higher. For example, Blocking element of H.264/AVC HP intra coding is about 1.66 times higher than in the case of using JPEG2000.

The results in Table 5 are presented for the Y component of Lena color image. Delta, Blocking and VQM elements of JPEG 2000 are more better than H.264/AVC HP intra coding elements. On the other hand, PSNR, Blurring and SSIM index elements are better than in the H.264/AVC HP intra coding case. The Blocking element of H.264/AVC HP intra coding is more than two times of JPEG 2000.

Table 4. Objective quality assessment of *Lena* 512×512 monochrome image using H.264/AVC HP intra coding and JPEG 2000 coding with about 35.40 kbits per image.

Measurement metric	H.264/AVC HP intra	JPEG 2000
PSNR [dB]	31.90	31.30
Delta	1.4821	1.6077
Blurring	3.4926	3.8018
Blocking	14.4867	8.7091
SSIM index	0.950	0.9418
VQM	0.6808	0.7615

Table 5. Objective quality assessment of with Lena 512×512 color image Y component using H.264/AVC HP intra coding and JPEG 2000 coding with about 38 kbits per image.

Measurement metric	H.264/AVC HP intra (Y component)	JPEG 2000 (Y component)
PSNR [dB]	28.87	28.68
Delta	4.6835	4.4735
Blurring	9.6339	10.9133
Blocking	16.4012	7.5528
SSIM index	0.8523	0.8603
VQM	1.9374	1.9373

4 Concluding Remarks

In this work, in order to evaluate image quality two methods, i.e., subjective and objective quality assessment were used in H.264/AVC HP intra frame coder and JPEG 2000. At first, two subjective quality assessment methods were used Double Stimulus Continuous Quality Scale (DSCQS) and Stimulus Comparison (SC). Lena image is encoded using 35.40 kbits per image. It can be seen that H.264/AVC HP intra coding gives better results than JPEG 2000 standard coding for both DSCQS and SC methods. After that, we used monochrome and color Lena images. We obtained that comparing to JPEG 2000, there is an advantage in favor of H.264/AVC HP intra coding from the point of view of PSNR in the range of 0.15dB to 0.76dB for monochrome image and in the range of 0.02dB to 0.30dB for the color image.

Finally, the comparison results using other methods for objective quality assessment are better for H.264/AVC HP intra coding in the case of PSNR, Delta, Blurring and VQM for monochrome image. On the other hand, in the case of using color image Y component, they are better for H.264/AVC HP intra coding compared to JPEG 2000 in the case of PSNR, Blurring and SSIM index measurement metrics.

References

[1] T. Sikora, “Trends and perspectives in image and video coding,” *Proc. IEEE*, vol. 93, pp. 6–17, 2005.

[2] C. Christopoulos, A. Skodras, and T. Ebrahimi, “The JPEG 2000 still image coding system: an overview,” *IEEE Trans. on Consumer Eletronics*, vol. 46, pp. 1103–1127, 2000.

[3] *JPEG 2000 Part*, ISO/IEC JTC1/SC 29/WG 1 Std., Mar. 2000.

[4] J. V. Team, *Recommendation and Final Draft International Standard of Joint Video Specification*, ITU-T Std., May 2003.

[5] Y.-W.Huang, B.-Y.Hsieh, T. Chen, and L.G.Chen, “Analysis, fast algorithm, and vlsi architecture design for H.264/AVC intra frame coder,” *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 15, pp. 378–401, 2005.

[6] *Methodology for the subjective assessment of the quality of television pictures*, ITU-R Std., 2002.

[7] Msu video quality measurement tool. [Online]. Available: http://www. compression.ru/video/quality_measure/info_en.html

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. on Image Proc.*, vol. 13, pp. 600–612, 2004.

[9] F. Xiao, *DCT-based video quality evaluation*, 2000, student final project Digital Video Processing (EE392J).

[10] T. Ebrahimi *et al.*, “JPEG 2000 still image coding versus other standards,” *Proc. SPIE.4115*, pp. 446–454, 2000.

[11] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, “Performance comparison: H.26L intra coding vs. JPEG 2000,” *Doc. JVT-D039*, July 2002.

[12] *JPEG 2000 still image compression standard*, ISO/IEC JTC1/SC29/WG1 Std. N2412, Dec. 2002.

[13] G. Sullivan *et al.*, *Doc.JVT-L050*, JVT of ISO/IEC MPEG ITU-T VCEG Draft of version 3 of H.264/AVC, July 2004.

[14] T. Wiegand *et al.*, “Rate-constrained coder control and comparison of video coding standards,” *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 13, pp. 688–703, 2003.

[15] K. P. Lim, G. Sullivan, and T. Wiegand, *Text description of joint model reference encoding methods and decoding concealment methods*, Std. Doc. JVT-K049, Mar. 2004.

[16] IainE.G.Richardson, *H.264 and MPEG-4 Video Compression*. New York: John Wiley, 2003.