# Perceptual Intra Video Encoder for High-Quality High-Definition Content

Miguel Martínez-Rach, Otoniel López-Granado, Pablo Piñol, Manuel P. Malumbres
Miguel Hernández University
Avda. Universidad s/n
Elche, Alicante, 03202, Spain
{mmrach,otoniel,pablop,mels}@umh.es

***Abstract:*** This paper presents a perceptually enhanced intra-mode video encoder based on the Contrast Sensitivity Function (CSF) with a gracefully quality degradation as compression rate increases. We have performed an evaluation of our proposal in terms of perceptual quality, memory consumption and complexity, comparing it with H.264/AVC intra, Motion-JPEG2000 and Motion-SPIHT. Evaluation results show that the proposed encoder is highly competitive especially when coding high definition video formats at high video quality levels (i.e. low compression rates). This would be of interest for those high quality media applications and services with constrained real-time and power processing demands.

## 1. Introduction

Video compression has been an extremely successful technology that has found application across many areas of television production, from content acquisition to transmission. The large volumes of data created with today's High Definition video signals have tested traditional coding schemes and it is now timely that we take advantage of the many advanced and newly developed coding techniques that deliver significantly improved coding efficiencies.

Currently, most of the popular video compression technologies operate in both Intra and Inter coding modes. Intra mode compression operates in a frame-by-frame basis, while Inter mode achieves compression working with a Group Of Pictures (GOP) at a time. Inter mode compression is able to achieve high coding efficiency over Intra mode schemes when picture content of adjacent frames is quite similar. However, under certain conditions, such as fast camera zooms and pans, high intensity motion (sports, animation, etc.), still camera flash lights and strobe lights as well as other short duration production effects, the correlation of adjacent frames is severely reduced and results in a visibly reduced picture quality or at worst, blocking artifacts.

Most of the television content productions require recordings in HD to maintain high quality of picture even though the usual final transmission is in SD (standard definition) format. At video content production stages, digital video processing applications require fast frame random access to perform an undefined number of real-time decompressing-editing-compressing interactive operations without a significant loss of original video content quality. Intra-frame coding is desirable as well in many other applications like video archiving, high-quality high-resolution medical and satellite video sequences, applications requiring simple and fast real-time encoding like video-conference and video surveillance systems [1], and Digital Video Recording systems (DVR), where the user equipment is usually not as powerful as the head-end equipment.

In [2] an experimental study was performed with H.264/AVC and JPEG2000 in order to determine the benefits of using inter frame encoding versus intra frame encoding for Digital Cinema applications. Their results draw that the coding efficiency advantages of inter frame coding are significantly reduced for film content at the data rates and quality levels required by digital cinema. This indicates that the benefit of inter frame coding is questionable, because it is computationally much more complex, creates data access complications due to the dependencies among frames and in general demands more resources. For lower resolutions their experiments confirm that inter frame coding was more efficient than intra frame coding. These results provide a justification for using JPEG2000, or other intra frame coding methods, for coding digital cinema content.

So, for all the applications mentioned above, a very interesting option to encode high-quality high-definition video content is the use of Intra coding systems, since they (1) efficiently exploit the spatial redundancies of each video sequence frame, (2) exhibit reduced complexity in the design of the encoding/decoding engines, (3) achieve fast random access capability by decoding only the selected frame, (4) have great error resilience behavior by limiting error propagation to the frame boundaries, (5) are easily portable to parallel processing architectures, i.e. multicore CPUs, and (6) have low coding/decoding delays, what it is of special interest for real-time applications.

In this work, we propose an enhanced perceptual Intra encoder suited for high-quality high-definition applications that is able to perform a very fast encoding (and decoding) with low demands of computational resources (processing power and memory).

The rest of the paper is organized as follows. In Section 2 we describe the proposed perceptual intra video encoder focusing on the perceptual CSF-based quantizer module. In Section 3 we performed several experiments comparing the behavior of our perceptual intra encoder against other popular intra codecs. Finally, in section 4 some conclusions are drawn.

## 2. Perceptual Intra Video Encoder

During the last years, image and video encoders have included much of the knowledge of our Human Visual System (HVS) in order to obtain a better perceptual quality of the compressed sequences. The most widely used characteristic is the contrast adaptability of the HVS, because HVS is more sensitive to contrast than to absolute luminance [3]. The Contrast Sensitivity Function (CSF) relates the spatial frequency with the contrast sensitivity to determine the HVS sensitivity level.

We propose a perceptual intra video encoder (PM-LTW) which it is inspired in the tree-based wavelet image coder proposed in [4]. The basic idea of our encoder proposal is very simple: after computing a dyadic wavelet transform over the source image [5], wavelet coefficients are quantized by means of our perceptual CSF-based quantizer, then a symbol map (zero-trees) is built and entropy encoded, and finally the significant coefficient bits are raw encoded. In the following subsections we will detail the CSF function to be used and the proposed perceptual CSF-based quantizer.

### 2.1. *Contrast sensitivity function*

Most of HVS-models account for the varying sensitivity over spatial frequency, color, and the inhibiting effects of strong local contrasts or activity, called masking. One of the initial HVS stages is the visual sensitivity as a function of spatial frequency that is described by the CSF.

A closed form model of the CSF [6] for luminance images is given by:

$$H(f) = 2.6(0.0192 + 0.114f)e^{-(0.114f)^{1.1}} \qquad (1)$$

where spatial frequency is $f = \left(f_x^2 + f_y^2\right)^{1/2}$ and it is measured in cycles/degree ($f_x$ and $f_y$, are the horizontal and vertical spatial frequencies). Usually, spatial frequency is also measured in cycles per optical degree (cpd), which makes the CSF independent of the viewing distance.

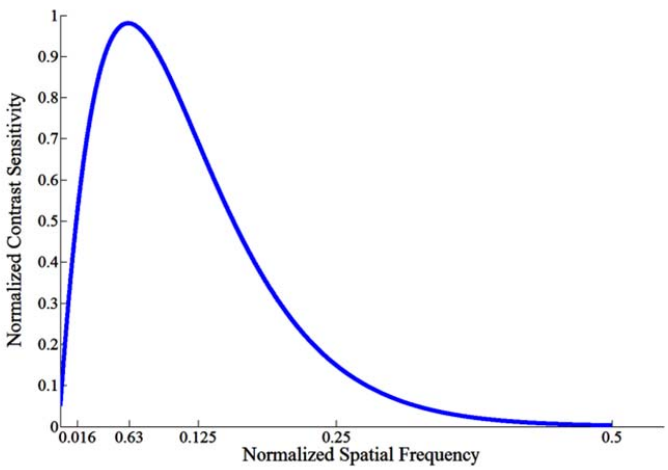


**Figure 1:** CSF function

Figure 1 depicts the CSF curve obtained with equation (1). It characterizes luminance sensitivity as a function of normalized spatial frequency. As it can be seen, the CSF behaves as a bandpass filter, which is most sensitive to normalized spatial frequencies between 0.025 and 0.125 and less sensitive to very low and very high frequencies. CSF curves exist for chrominance as well. However, unlike luminance stimuli, human sensitivity to chrominance stimuli is relatively uniform across spatial frequency.

We have selected the CSF-based encoding approach since it is simple, effective, and widely used in other wavelet-based image encoders where its benefits were clearly stated [7][8][9][10]. Also, as many other works, in [6] authors demonstrated that the MSE cannot reliably predict the difference of the perceived quality of two images. So, by means of psychovisual experiments, they proved that the aforementioned CSF model applies to wavelet coefficients a perceptual equalization that would help to reduce the visible artifacts introduced by the lossy coding stage. So, this was the main reason that leads us to adopt this model in our study.

## 2.2. Perceptual CSF-based quantizer

In order to properly apply the CSF function to the DWT coefficients, the mapping between frequency and the CSF-weighting value applied to each wavelet coefficient is a key issue. As wavelet based codecs perform multi-resolution signal decomposition, the easiest approach is to find a unique weighting value for each wavelet frequency subband.

If further decompositions at the frequency domain are done, for example by the use of packet wavelets, a finer association could be done between frequency and CSF weights [11].

The most common way to implement the CSF curve is using an Invariant Scaling Factor Weighting (ISFW) [12]. This approach can be applied in two ways depending on the stage of the codec where it will be applied.

The first one is introduced in some codecs like JPEG2000 by replacing the MSE by the CSF-Weighted MSE (WMSE). This is done in the Post-Compression Rate Distortion Optimization (PCRD-OPT) algorithm where the WMSE replaces the MSE as the cost function which drives the formation of quality layers [13].

The second one performs a scaling (or weighting) of wavelet coefficients. It can be introduced after wavelet filtering stage as a simple multiplication of wavelet coefficients at each frequency subband by the corresponding weights. We will employ this approach since it is simple (low complexity) and it leaves the other compression stages unmodified, allowing portability to other encoders, integration with different quantization schemes, or even other wavelet filters.

So, our perceptual CSF-based quantizer will be composed of two stages. First, the CSF function defined in previous subsection will be applied to all wavelet coefficients by means of a specific CSF weighting matrix. To compute the weighting matrix, we performed an ISFW implementation of the CSF.

**Table 1:** Proposed CSF Weighting matrix

| Level/Orientation | **LL** | **LH** | **HH** | **HL** |
|---|---|---|---|---|
| **L1** | 1.0 | 1.1795 | 1.0 | 1.7873 |
| **L2** | 1.0 | 3.4678 | 2.4457 | 4.8524 |
| **L3** | 1.0 | 6.2038 | 5.5842 | 6.4957 |
| **L4** | 1.0 | 6.4177 | 6.4964 | 6.1187 |
| **L5** | 1.0 | 5.1014 | 5.5254 | 4.5678 |
| **L6** | 1.0 | 3.5546 | 3.9300 | 3.1580 |

In Table I, the resulting CSF weighting matrix is shown, defining the scaling weights for each wavelet level decomposition and orientation subband. These weighting factors were directly computed from the CSF curve by normalizing its corresponding values, so that the most perceptually important frequencies are scaled with higher values, while the less important are preserved. This scaling process increases the magnitude of all wavelet coefficients, except for the LL subband that are neither scaled nor quantized in our coding algorithm. After the CSF weighting process described above, a simple uniform scalar quantization is applied to achieve the desired bit-rate.

## 3. Experimental Results

We have compared our PM-LTW proposal with Motion-JPEG2000 (Jasper 1.701.0), Motion-SPIHT (Spiht 8.01), x.264/Intra (FFmpeg version SVN-r25117, profile High, level 4.0) and H.264/AVC/Intra (High-10, JM16.1) in terms of R/D performance, coding delay and memory consumption. All evaluated encoders have been tested on an Intel Pentium Core 2 CPU at 1.8 GHz with 6GB of RAM memory, employing several wellknown video sequences with different formats like Foreman, Hall, Container, and News (QCIF and CIF), Mobile and Ducks-take-off (ITU-D1) and Pedestrian area, Station2, and Ducks-take-off (HD1080p).

Although most studies employ PSNR metric to measure video quality performance, we decided to use in our study objective quality assessment metrics and subjective tests, since our proposal includes perceptual-based encoding techniques that may not be properly evaluated by PSNR metric. There are several studies about the convenience of using other video quality assessment metrics than PSNR that better fit to human perceptual quality assessment (i.e. subjective tests results) [11][14][15][16].

One of the best behaving objective quality metrics is VIF [3], which has been proven [11][14] to have a better correlation with subjective perception than other metrics that are usually used for codec comparisons [15][16], like MSSIM [17]. The VIF metric uses statistics models of natural scenes in conjunction with distortion models in order to quantify the statistical information shared between the test and reference image.

In spite of using objective quality metrics, like VIF, running subjective tests is still required to validate the final evaluation results. So, we have arranged a simple subjective test involving 15 non-expert evaluators as suggested by ITU-BT500 and followed the guidelines found at ITU-TP.910 Recommendation [18]. The Double-stimulus Impairment Scale (DSIS) evaluation method was employed. A 5-grade scale from 0 to 1 (with 0.2 steps) was used to rate the quality of the test video sequences where 0 = bad, 0.25 = Acceptable, 0.5 = Good, 0.75 = Excellent and 1 = Visually Lossless. Although five quality levels are defined, our study will focus only on the first four levels, from “Visually lossless” to “Acceptable”.

In order to measure the bit-rate savings of our proposal respect to the other encoders, we need to define the lower thresholds of the different quality levels by means of the VIF [3] objective quality metric and the results obtained from the subjective tests. So, through subjective testing we will map the thresholds of the different quality levels into the native VIF metric space, being able to compute the average bit-rate differences among our proposal and the one obtained by the selected encoders for each quality level.

The subjective test material is configured as follows: all the video sequences were encoded at 16 different bit-rates through the entire bit-rate range (from extremely high compression up to nearly lossless rates) with the video codecs under test.

After analyzing resulting data, the VIF value thresholds are obtained for each quality level. To establish the “Visually Lossless” lower threshold we choose among all reconstructed videos scored with quality “1” in the subjective tests, the one with the lowest bit-rate. The value of VIF quality for that bit-rate will determine the “Visually Lossless” lower threshold. For the following quality level, “Excellent”, we proceed in a similar way by selecting all the reconstructed videos scored in the range [0.75..1) the one with the lowest bit-rate to determine the VIF value that correspond with the “Excellent” lower threshold. The rest of lower thresholds are calculated in the same way.

From the objective tests raw data, we detected that the thresholds for each quality level depend on the picture resolution. For example, for the “Good” level, when picture resolution was CIF or QCIF the lower threshold corresponds to 0.80 VIF units, but at higher picture resolutions the VIF value is around 0.75 VIF units. In the same way, for small size sequences the lower threshold for the “Acceptable” level corresponds to 0.70

VIF units while for higher resolution sequences it is around 0.60 VIF units. In Table 2 we show the lower thresholds found for the different quality levels and video formats.

**Table 2:** Lower quality thresholds for quality levels and video formats

| Lower Thresholds | CIF & QCIF | ITU & HD |
|---|---|---|
| Visually Lossless | 0.93 | 0.90 |
| Excellent | 0.87 | 0.85 |
| Good | 0.80 | 0.75 |
| Acceptable | 0.70 | 0.60 |

Having fixed the VIF lower thresholds for the different video quality levels under our study, we proceed to estimate the bit-rate produced by the tested video encoders at each quality level. In figure 2 we show the VIF R/D curves for the HD1080 "Pedestrian area" video sequence with the quality levels marked with different background colors.

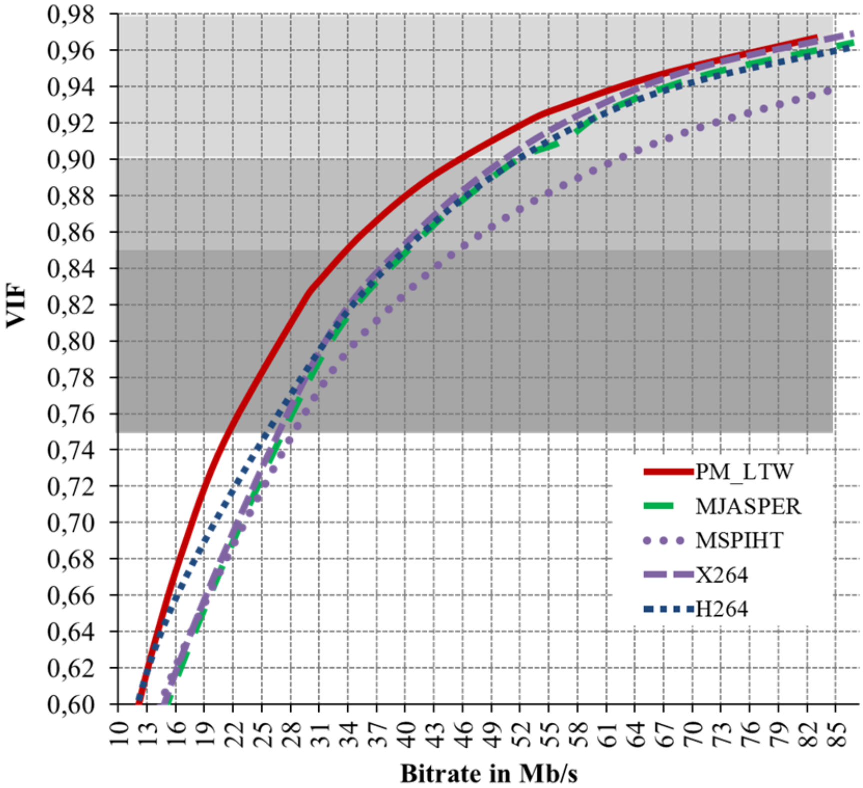


**Figure 2:** VIF R/D curves for a HD1080 sequence.

In order to estimate the average bit-rate gain of our encoder proposal at "Excellent", "Good", and "Acceptable" quality levels, we compute the average value of the bit-rate differences between the VIF curves inside the quality level. So, the final average gain of our proposal with respect other encoder at a particular quality level would be computed as the average of the gains measured from all reconstructed videos that were scored inside that quality level. However, for the "Visually Lossless" level the bitrate difference between two encoders should be measured at the threshold VIF value, since higher VIF values get the same perceptual quality, i.e. visually lossless.

Table 3 shows the relative bit-rate savings that in average can be achieved for each defined quality level. When comparing our proposal with Motion-JPEG2000 or Motion-SPIHT and regardless the sequence frame resolution and quality level, always bit-rate savings are achieved.

**Table 3:** Average PM-LTW Relative Bit-Rate Savings

| PM-LTW vs … | Format | ~Lossless | Excellent | Good | Acceptable |
|---|---|---|---|---|---|
| **M-JP2K** | HD | 5.87% | 11.21% | 14.53% | **17.22%** |
| | ITU-D1 | 11.88% | 10.33% | 9.05% | 9.02% |
| | CIF | 9.26% | 4.03% | 2.93% | 4.38% |
| | QCIF | 7.32% | 6.59% | 7.58% | 9.08% |
| **M-SPIHT** | HD | **37.59%** | 36.63% | 31.34% | 22.87% |
| | ITU-D1 | 19.84% | 18.28% | 16.32% | 14.94% |
| | CIF | 13.76% | 12.82% | 12.58% | 12.77% |
| | QCIF | 12.13% | 12.04% | 12.70% | 13.15% |
| **x.264** | HD | 12.11% | 14.09% | 17.02% | **19.42%** |
| | ITU-D1 | 16.11% | 15.41% | 14.48% | 13.98% |
| | CIF | -1.96% | -2.32% | -2.63% | -2.94% |
| | QCIF | -1.68% | -2.51% | -3.61% | -5.04% |
| **H.264** | HD | **17.86%** | 16.68% | 11.23% | 2.92% |
| | ITU-D1 | 12.80% | 6.50% | -2.31% | -9.06% |
| | CIF | -2.05% | -4.05% | -6.72% | -9.27% |
| | QCIF | -3.04% | -4.97% | -7.63% | -10.59% |

In general, the trend is that the bit-rate saving increases as the frame resolution does. For QCIF and CIF resolution, x264 and H.264/AVC give a better performance for all defined quality levels, being the bit-rate savings greater for H.264 than for x264.

Looking at ITU-D1 video resolutions, PM-LTW R/D performance increases as the quality level does. When compared to x264, M-SPIHT and M-JP2K, PM-LTW achieves lower bit-rate at all quality levels, i.e. bit-rate savings are obtained at each quality level in that frame resolution. However, the improvements with respect to H.264/AVC are only achieved at "Excellent" and "Visually Lossless" quality levels for this frame resolution.

Now, we will proceed to compare the codecs under test in terms of coding delay and memory requirements. Figure 3 shows the coding speed in frames per second obtained by the different encoders being evaluated. As shown, PM-LTW outperforms the rest of encoders for any sequence frame resolution. For the highest resolution PM-LTW is 1.08 times as fast as M-SPIHT, 2.22 times as fast as M-JASPER, 2.30 times as fast as x264 and 28.09 times as fast as H.264/AVC. The current implementation of our codec is not optimized in any sense. While comparing with M-JPEG2000 using KKDU 5.2.5. execution times of PM-LTW are faster only for the QCIF frame resolution. KKDU 5.2.5 is fully optimized including multi-thread and multicore hardware capabilities, processor intrinsics like MMX/SSE/SSE2/SIMD and fast multicomponent transform. Therefore KKDU outperforms PM-LTW in coding time, processing up to 102.13 fps in CIF resolution, 42.43 fps in ITU-D resolution and 14.0 fps in HD resolution.

Regarding memory requirements, in Figure 4 we can see the maximum amount of memory (in Mbytes) required for each encoder and resolution. As it can be seen, PM-LTW requires near 4 times less memory resources than Motion-SPIHT, Motion-JPEG2000 and x.264 and up to 40 times less memory than H.264/AVC.

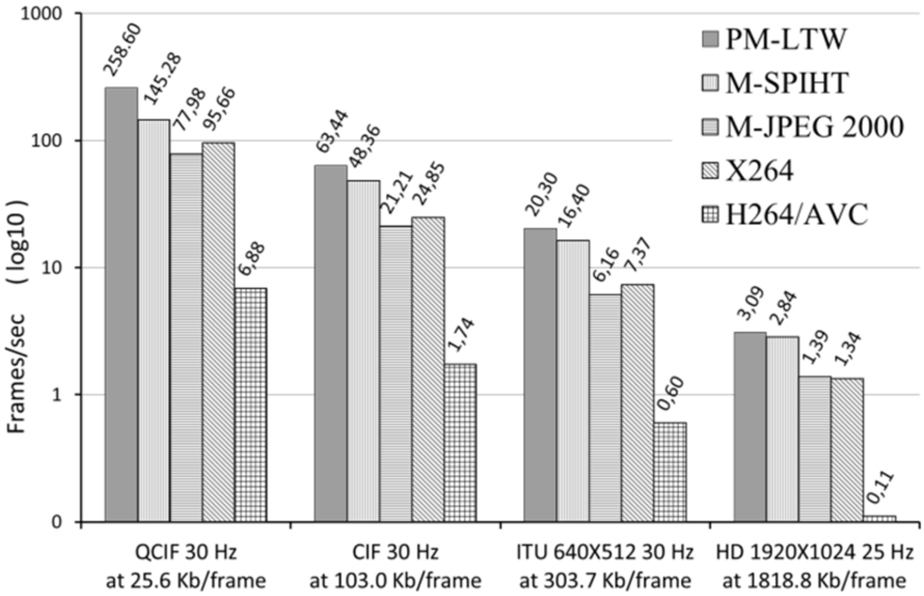


**Figure 3:** Encoder frame rate at different sequence sizes.

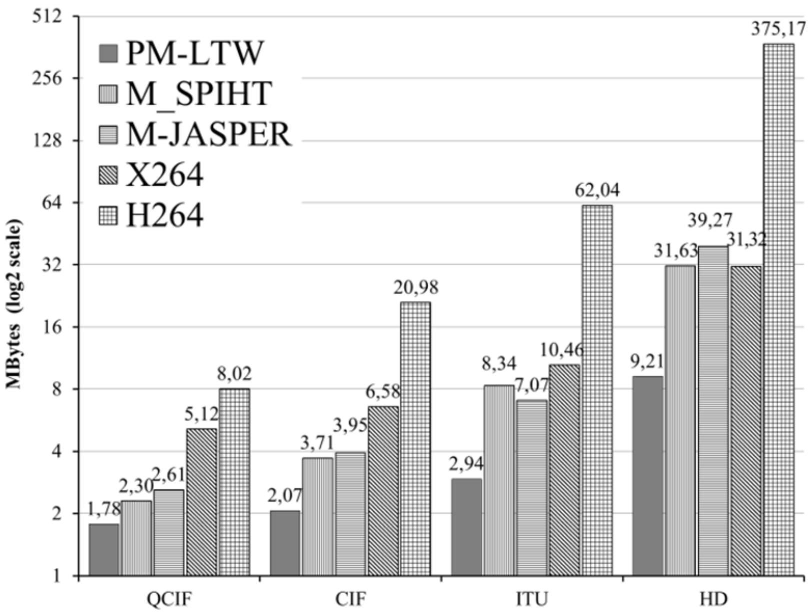


**Figure 4:** Memory requirements at different video formats.

## 4. Conclusions

Our proposed perceptual enhanced Intra encoder reveals the importance of exploiting the Contrast Sensitivity Function (CSF) behavior of the HVS by means of an accurate perceptual weighting of the wavelet coefficients, especially at high definition and high quality video formats. PM-LTW is very competitive in terms of perceptual quality being able to obtain important bit-rate savings at high quality levels; it is faster and requires less memory than the other evaluated non optimized encoders. So, in general, we have shown that bringing together the attractive advantages of intra video coding with the benefits of using perceptual encoding techniques, in a similar way as our PM-LTW encoder does, significant performance improvements would be achieved for those digital video processing applications, like the ones demanded by television and film industry to create, store and deliver high-quality high-definition video content productions.

## References


[1] Jang-Seon Ryu and Eung-Tea Kim, "Fast intra coding method of h.264 for video surveillance system," *International Journal of Computer Science and Network Security*, 7(10), 2007.

[2] Michael Smith and John Villasenor, "Intra-frame jpeg2000 vs. inter-frame compression comparison: The benefits and trade-offs for very high quality, high resolution sequences," SMPTE Technical Conference and Exhibition, Pasadena, California, October 20-23 2004.

[3] H. Rahim Sheikh, A.C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, 14(12), 2005.

[4] Oliver, J., & Malumbres, M. P, "Low-complexity multiresolution image compression using wavelet lower trees," *IEEE Transactions on Circuits and Systems for Video Technology*, 16(11), 1437–1444, 2006.

[5] Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I, "Image coding using wavelet transform," *IEEE Transaction on Image Processing*, 1(2), 205–220, 1992.

[6] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," IEEE *Transactions on Information Theory*, vol. 20, no. 4, pp. 525 – 536, Jul. 1974.

[7] A. Beegan, L. Iyer, A. Bell, V. Maher, and M. Ross, "Design and evaluation of perceptual masks for wavelet image compression," in Digital Signal Processing Workshop, 2002 and 2nd Signal Processing Education Workshop. Proceedings of 2002 IEEE 10th, Oct. 2002, pp. 88 – 93.

[8] A. Gaddipati, R. Machiraju, and R. Yagel, "Steering image generation with wawelet based perceptual metric," in Eurographics, 1997.

[9] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust, "Comparing real and synthetic images: Some ideas about metrics," in In Proc. Sixth Eurographics Workshop on Rendering, Dublin, Ireland,, 1995, pp. 82– 91.

[10] N. Moumkine, A. Tamtaoui, and A. Ait Ouahman, "Integration of the contrast sensitivity function into wavelet codec," in In Proc. Second International Symposium on Comunications, Control and Signal Processing ISCCSP, Marrakech, Morocco, March 2006.

[11] Xinbo Gao, Wen Lu, Dacheng Tao, and Xuelong Li, “Image quality assessment based on multiscale geometric analysis,” *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1409–1423, 2009.

[12] Marcus J. Nadenau, Julien Reichel, and Murat Kunt, “Wavelet-based color image compression: Exploiting the contrast sensitivity function,” *IEEE Transactions on Image Processing* 12(1), 2003.

[13] D. S. Taubman and M. W. Marcellin, JPEG2000 Image Compression Fundamentals, Standars and Practice. Springer Science+Business Media, Inc., 2002, no. ISBN: 0-7923-7519-X.

[14] M. Martinez-Rach, O. Lopez, P. Piñol, J. Oliver, and M. Malumbres, “A study of objective quality assessment metrics for video codec design and evaluation,” in Eight IEEE International Symposium on Multimedia, vol. 1, ISBN 0-7695-2746-9. San Diego, California: IEEE Computer Society, Dec 2006, pp. 517–524.

[15] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3440– 3451, 2006.

[16] Z. Wang, A. Bovik, H. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, 2004.

[17] F. De Simone, M. Ouaret, F. Dufaux, A. G. Tescher, and T. Ebrahimi, “A comparative study of jpeg2000, avc/h.264 and hdphoto,” in Proc. of Applications of Digital Image Processing XXX, San Diego, August 2007.

[18] ITU-T P.910, Subjective video quality assessment methods for multimedia applications, ITU-R Std., September 1999.