FGS-MR: MPEG4 Fine Grained Scalable Multi-Resolution Layered Video Encoding

Siddhartha Chattopadhyay, Suchendra M. Bhandarkar, Kang Li
Department of Computer Science,
The University of Georgia, Athens, GA, 30602-7404, USA
siddh@cs.uga.edu suchi@cs.uga.edu kangli@cs.uga.edu

ABSTRACT
The MPEG-4 Fine Grained Scalability (FGS) profile aims at scalable video encoding, in order to ensure efficient video streaming in networks with fluctuating bandwidth. In order to allow very low bit rate streaming, the Base Layer of an FGS video is encoded at a very low bit rate, resulting in very low video quality. In this paper, we propose FGS-MR, which uses content aware multi-resolution video frames to obtain better video quality for a target bit rate, compared to existing MPEG-4 FGS Base Layer video encoding schemes. FGS-MR is an integrated approach that requires only encoder side modification, and is transparent to the decoder. In addition, FGS-MR can be used with any existing MPEG-4 codec which supports FGS, since it entails "smart" video preprocessing and does not involve any components from the MPEG-4 compression pipeline. FGS-MR is a mask based technique. We have demonstrated an unsupervised algorithm to automatically create the mask from a given video sequence.

Categories and Subject Descriptors
H.5.1 [**Information Systems**]: Information Interfaces and Presentation-*Multimedia Information Systems*.

General Terms
Algorithms, Standardization

Keywords
Scalable Video, MPEG-4 FGS, Multi-Resolution

1. INTRODUCTION
Streaming video technology is an integral part of Internet video service providers such as Yahoo, MSN, AOL and many others. Since Internet bandwidth availability is dynamic, it is essential to dynamically adapt the bit rate of video streaming. The state-of-the-art MPEG-4 video encoding pipeline uses layer-based encoding, termed as the fine grained scalability (FGS) [3] [8] [9] [13] profile. MPEG-4 FGS encodes the video into a Base Layer and an Enhancement Layer. The Base Layer bit rate is the minimum bit rate at which the video can be streamed. However, this bit rate may still be too high for networks with lower bandwidths. The Base Layer may be encoded in even lower bit rates to allow streaming to these low bandwidth networks; however, this inevitably leads to a drastic reduction in video quality to the point that the visual information is almost useless. This happens because MPEG-4 codecs employ algorithms to truncate the DCT coefficients, or equivalently, render the video frames at a uniformly low resolution, which sharply reduces the visual quality. This is undesirable, since some spatial regions of the video might require very high resolution, whereas in some other spatial regions of the video frame, very low resolution may suffice.

In this paper, we propose an additional dimension for video adaptation; multi-resolution encoding for MPEG-4 fine grained scalability profile (FGS-MR) of the FGS Base Layer, in order to obtain improved rate-distortion performance. FGS-MR preprocesses each frame of the video to generate a multi-resolution representation, which alters the DCT coefficients such that the bit rate is reduced, yet the desired visually important features of the video are preserved.

FGS-MR is a mask based approach, where the mask defines regions of interest in a video. Given a frame of the video, FGS-MR creates high resolution around the mask, and low resolution in regions not covered by the mask, such that the overall rate-distortion performance is good. The multiple resolution representation ensures that the DCT coefficients of the video frames are altered such that the bit rate of the resulting MPEG-4 encoded video is reduced.

There exist technologies to selectively enhance the quality of the spatial regions of the video frame while streaming at a constrained bandwidth. MPEG-4 selective enhancement [11] is used in the enhancement layer of MPEG-4 FGS in order to stream better quality of video within selected image regions. However, MPEG-4 selective enhancement does not provide quality improvement for the Base Layer. In order to improve the quality of the Base Layer, MPEG-4 FGS uses adaptive quantization (FGS-AQ) [12]. FGS-AQ quantizes each 8 X 8 DCT block differently according to its relevance in improving the video quality.

The proposed FGS-MR technique in this paper gives significantly better rate-distortion performance compared to FGS-AQ, by using a pixel level multi-resolution video frame representation. Thus, given a desired video quality (e.g. in PSNR), FGS-MR can achieve lower bit rate video encoding

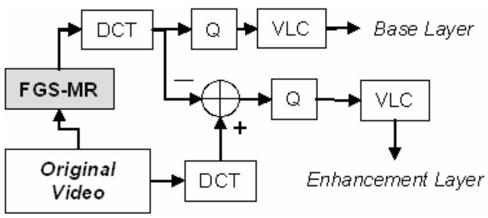

Figure 1: A basic MPEG-4 FGS scalable video encoding pipeline, with the proposed FGS-MR step added (in the shaded box).

compared to FGS-AQ. Consequently, FGS-MR consumes less power at the network interface of the device receiving the streaming video data. FGS-MR pre-processes the video in a manner that is transparent to the decoder; FGS-AQ, on the other hand, requires additional parameters for the decoder to decode each frame.

The use of a *mask* for FGS-MR operations opens up tremendous opportunities to create very interesting applications. The MR-mask can be generated automatically based on a variety of spatial features such as edges, contours and other objects of interest. One can also design various temporal feature-based masks based on moving objects in a scene. Thus, FGS-MR is more than just a tool; it is a whole new approach to intelligent, content based scalable video encoding. FGS-MR provides the potential for highlighting objects of interest (such as faces and advertisements) in streaming videos, streaming reasonable quality video in mobile phones with very low available bandwidths, amongst many other applications.

The rest of the paper is organized as follows: In Section 2, a background of the existing technologies for MPEG-4 layered encoding of video, and MPEG-4FGS-AQ, is provided. Section 3 describes the proposed FGS-MR technique in detail. Detailed quantitative and qualitative comparisons of the proposed FGS-MR Base Layer video encoding with the existing MPEG-4 FGS-AQ are described in Section 4. Finally, conclusions and potential future works are discussed in Section 5.

2. BACKGROUND

FGS-MR is essentially an improved method for encoding the Base Layer for MPEG-4 layered video encoding; hence, it is essential to first describe the MPEG-4 layered video encoding scheme, and then describe the MPEG-4 FGS adaptive quantization scheme (FGS-AQ) that enables selective enhancement of the Base Layer for MPEG-4 FGS.

2.1 Fine Grained Scalability (FGS)

FGS separates the video frames into two layers, which are referred to as the Base Layer and the Enhancement Layer (**Figure 1**, minus the shaded box). The Base Layer is encoded at the minimum bit rate available to the video streaming network. The Enhancement Layer is obtained by encoding the difference between the original DCT coefficients and the coarsely quantized Base Layer coefficients in a bit-plane fashion [8], [9], [10]. The Enhancement Layer can be truncated at any bit position and can provide fine granularity of control of the reconstructed video quality, which is proportional to the number of bits actually decoded.

Given the standard FGS pipeline as described above, FGS-MR is a pre-processing step at the very beginning of the Base Layer encoding pipeline (**Figure 1**, shaded box). The Base Layer, since it is encoded at the minimum bit rate, is often the most significant layer that can be streamed across to the client when the bandwidth drops down to a specified minimum value. This makes it critical that the Base Layer retains good visual quality in the semantically important regions of the video for a specified bit rate. MPEG-4 uses adaptive quantization in its Fine Grained Scalability Base Layer encoding (FGS-AQ) scheme to assign more bits to DCT coefficients of the blocks corresponding to the desired regions that need to be enhanced.

It must be noted that FGS-MR is actually an improvement over FGS-AQ in many aspects, as will be shown later in the paper. However, it is first necessary to understand FGS-AQ, in order to be convinced that FGS-MR is indeed better. In the next sub-section, we describe the FGS-AQ scheme briefly.

2.2 MPEG-4 FGS-AQ

FGS-AQ is MPEG's solution to obtain varying video quality within the frames of the Base Layer video in FGS, in order to better utilize the available bit rate. FGS-AQ [12] is achieved via a quantization matrix that defines different quantization step sizes for the different transform coefficients within a block (prior to performing entropy coding on these coefficients). For example, the low frequency DCT coefficients normally contribute more to the visual quality; consequently smaller step sizes are used for quantizing them. Adaptive quantization can also be controlled from one macro-block to another via a quantization factor whose value varies from one macro-block to the other. These adaptive quantization tools have been employed successfully in the MPEG-2 and MPEG-4 (Base Layer) standards [8] [12]. However, the aim of FGS-AQ is not to improve the rate-distortion performance, but rather to improve the visual quality of the resulting video. As a result, the rate-distortion performance of an FGS encoder that uses FGS-AQ may actually degrade due to the overhead needed for transmitting the FGS-AQ parameters.

One of the main aims of the proposed FGS-MR technique is to actually obtain better rate-distortion performance for the Base Layer encoding compared to the FGS-AQ technique. The by-products of using FGS-MR are its transparency to the existing MPEG-4 codecs, and also not requiring any additional codecs for decoding multi-resolution video frames. In the next section, we describe the proposed FGS-MR technique in detail.

3. PROPOSED APPROACH: FGS-MR

Multi-Resolution FGS (FGS-MR) is based on the fundamental observation that applying a low pass filter on the color-space of an image is equivalent to DCT coefficient truncation in the corresponding DCT space of the image [6]. This observation has been utilized to use low pass filters with different cutoffs in different regions of the video, such that the overall bit rate of the video is low, yet important parts of the video are encoded at good resolution. This process is called Multi-Resolution (MR) representation of the video frames. The semantically/visually

(a) Original (5.12 Mbps)

(b) MR-Mask

(c) FGS-AQ
PSNR=22.77 dB (0.17 Mbps)

(d) FGS-MR
PSNR = 26.5 dB (0.17 Mbps)

Figure 2: Comparison of video quality with proposed Multi-Resolution (FGS-MR) Base Layer encoding with MPEG-4 Adaptive Quantization (FGS-AQ) Base Layer video encoding technique; (a) The original 320 X 240 frame; original AVI video is encoded at 5.12 Mbps (b) The MR-Mask with added padding. Canny parameters: Lower = 75, Upper = 120, determined empirically (c) Video frame after FGS-AQ for Base Layer; target bit rate = 0.17 Mbps; PSNR = 22.77 dB. Blocks corresponding the edges in MR-mask have been enhanced (d) MR-Frame; bit rate = 0.17 Mbps; PSNR = 26.5 dB.

important regions of the video frame are defined by an MR-Mask. Thus, FGS-MR is a two step process:

(i) Create suitable MR-Mask which defines regions of semantic/visual interest within each frame of the video.

(ii) Use the MR-Mask to generate a multi-resolution representation (MR-Frame) of the original video frame. This MR-Frame is further fed to the standard MPEG-4 compression pipeline.

In the next section, we first describe an edge based method to create MR-mask automatically from a given video frame. Next, we describe an effective method for creating the multi-resolution MR-Frame corresponding to the MR-Mask.

3.1 MR-Mask creation

The MR-Mask is essentially a weight mask, with each pixel having a floating point value lying between, and inclusive of, 0 and 1. The mask is created based on the fundamental observation that the edges of an image form important features for the semantic/visual understanding of the image contents. The edges are detected from the image by using a Laplacian operator for detecting the second derivatives of the image intensity values. We have used a Laplacian method based algorithm called Canny edge detector [2]. A detailed description of the Canny edge detection algorithm is beyond the scope of this paper. Once the edges are detected, the MR-Mask is created such that the edges are given the weights of 1, the nearby regions within a certain radius (termed as padding) given weights less than 1, and the rest of the non-edge regions weights of 0. An example MR-Mask is given in **Figure 2b**. The parameters of the Canny edge detector have been determined empirically.

Since the MR-Mask requires content-based analysis of the video scene, one may argue that a natural method of encoding would be to use MPEG-4 object based encoding [1][15]. However, FGS-MR is based on edges, and not regions, whereas MPEG-4 object encoding defines objects by regions. Of course, other masks may be created which are based on regions. However, MPEG-4 FGS Base Layer encoding does not support object based encoding.

3.2 MR-Frames based on the MR-Mask

After the creation of the MR-Mask, the original video frame, V_O, is processed to form a multi-resolution frame, V_F, termed as the MR-Frame representation. This MR-Frame, V_F, is in turn fed to the standard MPEG-4 compression pipeline to form the Base Layer (**Figure 1**, including the shaded box).

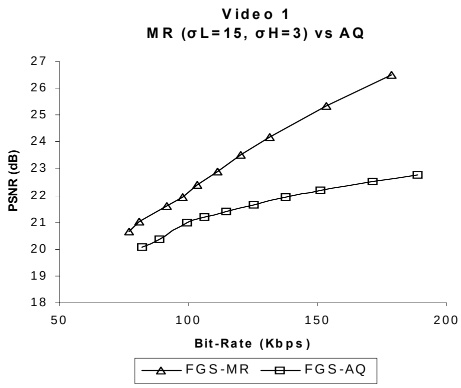

Figure 3: Rate-distortion comparison for *Video 1*

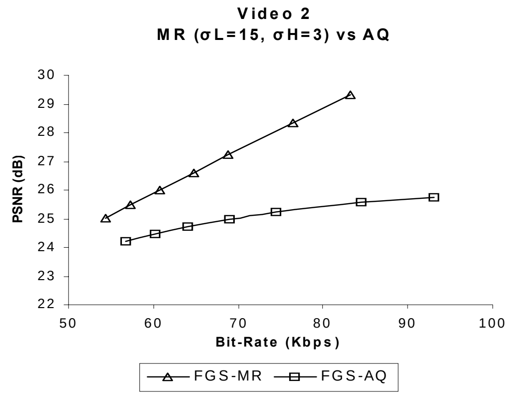

Figure 4: Rate-distortion comparison for *Video 2*

The simplest method of creating MR video frames is to render a weighted combination of two video frames of different resolutions. The original video frame, V_O, is used to render two video frames, V_H and V_L, such that V_H is a high resolution rendering and V_L is a low resolution rendering of the same video frame. We assume that a Gaussian filter, with parameter σ, (denoted as $G(\sigma)$) representing the standard deviation, is used as a representative low pass filter.

V_L can be obtained by convolving V_O with a Gaussian filter $G(\sigma_L)$; similarly, V_H can be obtained by convolving V_O with a Gaussian filter $G(\sigma_H)$. Keeping $\sigma_L > \sigma_H$ ensures that V_L is more smoothed compared to V_H; in other words, V_L is a lower resolution frame compared to V_H. In order to combine the two resolutions, a mask weight matrix **W**, corresponding to the MR-Mask, is used. An intermediate video frame, V_I, is created from the two video frames, V_H and V_L, and the weight matrix **W**, and is given by:

$$V_I = (\mathbf{I} - \mathbf{W})V_L + \mathbf{W}V_H \qquad (1)$$

where **I** is the matrix with all entries as 1. The intermediate video frame, V_I, is a multi-resolution frame representation of the original video frame, V_O. However, V_I contains abrupt changes in resolution, which is not pleasing to the eye. Another smoothing iteration is performed on the intermediate frame V_I with a Gaussian filter $G(\sigma_I)$, to yield the final frame V_F as a smoothed version of V_I.

The final video frame V_F is now processed with further spatial and temporal compression techniques defined in MPEG-4 to create a complete video as a spatially and temporally compressed sequence of MR video frames. The video quality and encoded bit rate depend on the following four parameters; MR-Parameters σ_L, σ_H, σ_I and MR-Mask **W**. Whereas σ_L, σ_H, σ_I are bounded scalar quantities which control the bit rate, the weight matrix **W** controls the quality of the encoded video frame (it also controls, to a small extent, the encoded bit rate).

3.3 Computing the MR-Parameters σ_L, σ_H, σ_I

The rate-distortion performance of FGS-MR depends significantly on the MR-parameters, σ_L, σ_H and σ_I. The optimal values of the MR-parameters can be computed by optimizing a suitable figure of merit function. We have devised a figure of merit function δ, which is a scaled comparison of Quality (Q) with the obtained compression ratio (C) of a video sequence. The metric δ is designed such that the higher its value is, the better is the rate-distortion performance. We have defined δ as

$$\delta = Q/C \qquad (2)$$

where $Q = 2^{PSNR(\sigma L,\ \sigma H,\ \sigma I)/10}$ and C = Compression Ratio (FGS-MR + MPEG-4 encoded vs MPEG-4-only encoded Base Layer). The expression Q is obtained from the standard PSNR equation: PSNR = 10log(Q), where Q is the ratio of the mean squared error and the number of pixels in the frame. The metric δ is maximized for the best quality of video obtained for the least bit rate of video; the higher the metric δ, the more *efficient* the video encoding. The values of σ_L, σ_H, σ_I which maximize δ can be potentially determined exhaustively, by enumerating all possible values of σ_L, σ_H, σ_I, and computing δ for each combination. The MR-Parameters σ_L, σ_H, σ_I are essentially the size of the Gaussian convolution masks, which take odd integral values, and are bounded in size. However, efficient methods have to be devised to find the global optimum (maximum) of δ.

We have empirically determined that using $\sigma_L = 15$, $\sigma_H = 3$, and varying σ_I to control the encoded bit rate yields good rate-distortion performance, and better power saving capabilities, with an edge mask. We confirm these claims in the next section.

4. RESULTS AND DISCUSSIONS

In this section, we first describe some preliminary results supporting the validity of FGS-MR scheme in terms of rate-distortion performance and power consumption at the client end for receiving the streaming video data. We also analyze reasons for the superior performance of FGS-MR compared to FGS-AQ. An important point to note here is that FGS-MR encoding is performed offline on a video file; as such we do not have a solution for real time FGS-MR Base Layer creation.

Due to lack of space, we have used two representative videos to show our results:

Video 1: A 30 second video of a single person walking in a well lighted room. Frame Rate: 30 fps; Frame-Size: 320 X 240

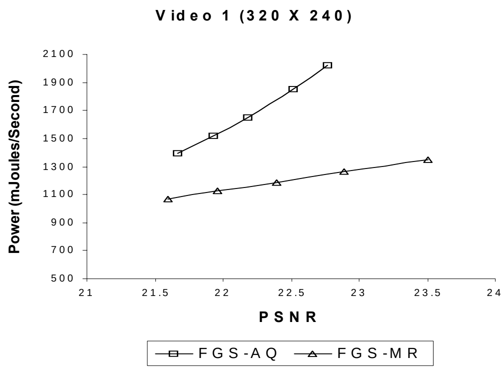

Figure 5: Power consumption at WNIC for *Video 1*

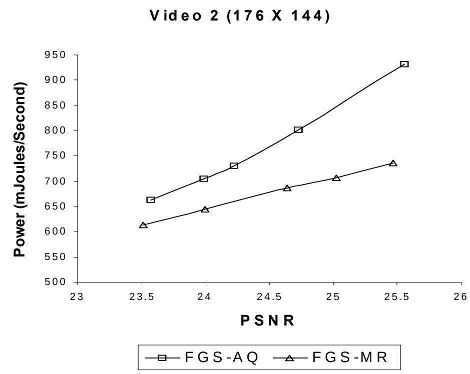

Figure 6: Power consumption at WNIC for *Video 2*

Video 2: A panning view across a room in poorly lighted room; i.e., the background is not stationary at any time. There are no moving objects in the room. Frame Rate: 30 fps; Frame Size: 176 X 144.

4.1. Rate Distortion

In order to compare the proposed FGS-MR technique with the existing MPEG FGS-AQ technique, we have compared the quality of a video sequences (PSNR) with the corresponding target bit rate. We have used a Gaussian kernel for the low pass filter. The MR-Mask is generated automatically from each video frame by using a Canny edge detector [2] with parameters $threshold_{low}$ = 75 and $threshold_{high}$ = 120. These values have been chosen empirically.

To give a subjective idea about the video quality improvement obtained by using FGS-MR as compared to FGS-AQ, **Figure 2** shows a frame of *Video 1*, encoded using FGS-AQ and FGS-MR, at the same bit rate (0.17 Mbps). The visual quality of FGS-MR (PSNR = 26.5 dB) is significantly better than that obtained using FGS-AQ (PSNR = 22.77 dB). The FGS-MR video has been obtained by using the MR-Parameters: $\sigma_L = 21$, $\sigma_H = 3$ and $\sigma_I = 3$. The MR-Parameter values have been chosen empirically in order to obtain the bit rate as that of FGS-AQ for objective comparison.

To give objective evidence of the superior quality of video obtained via FGS-MR compared to FGS-AQ, **Figure 3** shows a plot of PSNR as a function of the target bit rate for *Video 1*, by assigning $\sigma_L = 15$, $\sigma_H = 3$ (empirically chosen) and varying σ_I from 3 to 25 in **equation (1)**. Note that the PSNR values for FGS-MR are well above those obtained via MPEG-4 FGS-AQ for the entire range of bit rates (70 Kbps - 200 Kbps). **Figure 4** shows plots of PSNR vs bit rate for *Video 2*. In this case too, the PSNR values show significant improvement over the FGS-AQ method for all bit rates ranging from 56 Kbps to 100 Kbps.

The observation that higher quality video can be obtained for the same bit rate using the FGS-MR scheme can be attributed to the fact that pixel level enhancement can be performed in the FGS-MR scheme, whereas FGS-AQ does block-based enhancements. Thus, in order to enhance even a single pixel in a block, FGS-AQ enhances the whole block. Although this improves the overall PSNR slightly, the bit rate goes up. This is not the case for FGS-MR. FGS-MR enhances a single pixel; thus the DCT coefficients in the corresponding block may be enhanced less drastically. Thus effectively, for the same bit rate, the quality of video obtained is better for FGS-MR compared to FGS-AQ.

In conclusion, FGS-MR delivers superior rate-distortion performance compared to FGS-AQ for MPEG-4 FGS Base Layer encoding.

4.2. Resource Consumption

The fact that higher PSNR values can be obtained by using FGS-MR for the same target bit rate can also be interpreted as the ability of FGS-MR to obtain lower bit rates of video for the same video quality, compared to FGS-AQ. Using this observation, one can also infer that, for the same video quality, FGS-MR encoded Base Layer video streaming consumes less power at the wireless network interface card (WNIC) compared to the FGS-AQ encoded Base Layer.

To show power consumption reduction due to lower bit rate video encoding, we have used the following simple power consumption model: For a video of time duration *T*, data size *S* and given available bandwidth *B*, the energy used by the WNIC is given by

$$E_{network} = E_R \bullet S/B + E_S \bullet (T - S/B) \quad (3)$$

Dividing **equation (2)** with the time duration T, we get the power model

$$P_{network} = E_R \bullet b/B + E_S \bullet (1 - b/B) \quad (4)$$

where $P_{network}$ is the power consumed (mJoules/second) at the WNIC of the mobile device and *b* is the bit rate of the streaming video. E_R is the energy used by the WNIC during data reception and E_S is the energy used by the WNIC when it is sleeping and not receiving data. Using **equation (4)**, we compute the WNIC power utilization for reception of the *Video 1* and *Video 2*. We have used energy usage data (E_R and E_S) from [7] [14] to obtain the power usage for data reception in m-joules/sec.

In order to compare power usage patterns of FGS-AQ and FGS-MR schemes, we plot power usage at the WNIC, using **equation (4)**, as a function of the video quality for *Video 1* (**Figure 5**). The graph in **Figure 5** shows that, in order to stream the same quality of video (assuming no frame dropping, packet loss etc), FGS-MR encoded Base Layer video consumes less power at the WNIC compared to FGS-AQ encoded Base Layer Video. A similar plot for *Video 2* (**Figure 6**) confirms our claim.

Thus, from these preliminary results, we can conclude that FGS-MR is better than FGS-AQ both in terms of rate-distortion performance and power consumption for data reception at the WNIC.

5. CONCLUSIONS

A novel multi-resolution Base Layer encoding technique for MPEG-4 fine grained scalability (FGS-MR) video encoding has been presented. Results show that the rate-distortion performance of the proposed FGS-MR technique is significantly better compared to the existing MPEG-4 Adaptive Quantization technique for FGS Base Layer encoding (FGS-AQ). In addition, FGS-MR is transparent to the decoder; whereas FGS-AQ, on the other hand, requires special AQ parameters at the decoder end.

FGS-MR is based on the fundamental observation that applying a low-pass filter in the color-space of an image is equivalent to DCT coefficient truncation at the DCT space of the image. FGS-MR works in the color-space of the video, as compared to FGS-AQ, which works in the DCT space of the video frames. As a result, FGS-MR can be integrated seamlessly into existing MPEG-4 FGS codecs which do not need to have the basic adaptive quantization modules in them. Since FGS-MR does not require additional parameters to be passed during the streaming process to the decoder (as opposed to FGS-AQ), it is easy to use with existing FGS codecs.

FGS-MR is more than just a tool; it is a whole new approach to intelligent, content based scalable video encoding. Since FGS-MR is inherently parametric, a potential future research endeavor will be to compute the parameters automatically from the given video. In addition, many new types of masks may be used, such as motion layer masks, tracking masks, feature masks for human faces and other objects of interest in the video.

6. REFERENCES

[1] Bertini, M., Cucchiara, R., Del Bimbo, A. and Prati, A., *Object-based and Event-based Semantic Video Adaptation*, in Proceedings of IAPR International Conference on Pattern Recognition (ICPR 2004), vol. 4, Cambridge, UK, pp. 987-990, Aug, 23-26, 2004.

[2] Canny, J., *A computational approach to edge detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, pp. 679--698, 1986.

[3] Coding of Audio-Visual Objects, Part-2 Visual, Amendment 4: *Streaming Video Profile*, ISO/IEC 14 496-2/FPDAM4, July 2000.

[4] Davies, E., *Machine Vision: Theory, Algorithms and Practicalities*, Academic Press, pp 42 - 44, 1990.

[5] Geusebroek J.-M., Smeulders, A. W. M., and Weijer, J. van de, *Fast Anisotropic Gauss Filtering*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 3, march 2001.

[6] Gonzalez, R.C. and Woods, R.E., *Digital Image Processing*, Addison-Wesley, Reading, MA, 1992.

[7] Havinga, P. J. M, *Mobile Multimedia Systems*, PhD thesis, Univ. of Twente, Feb 2000.

[8] Li, W., *Overview of Fine Granularity Scalability in MPEG-4 Video Standard*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, pp. 301-317, Mar. 2001.

[9] Radha, H., van der Schaar, M., and Chen, Y., *The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP*, IEEE Transactions on Multimedia, vol. 3, pp. 53–68, Mar. 2001.

[10] Richardson, I. E. G. *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*, Wiley, 2004.

[11] Schaar, M. van der, and Lin, Y.-T., *Content-based selective enhancement for streaming video*, in Proc. IEEE International conference on Image Processing, Vol. 2, pp. 977–980, 2001.

[12] Schaar, M. van der, Chen, Y., and Radha, H., *Adaptive Quantization Modes for Fine-Granular Scalability*, Contrib. to 48th MPEG Meeting, m4938, July 1999.

[13] *Syntax of fine granularity scalability for video coding*, ISO/IECJTC1/SC29/WG11, MPEG99/M4792, July 1999.

[14] Stemm, M., Gauthier, P., Harada, D., and Katz, R. H., *Reducing power consumption of network interfaces in hand-held devices*, In Proc. 3rd Intl. Workshop on Mobile Multimedia Communications, September 1996.

[15] Wang, H., Schuster, G. M. and Katsaggelos, A. K., *Rate-distortion optimal bit allocation scheme for object-based video coding*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 9, pp. 1113-1123, Sept. 2005.