

WCAM: CONTENT-BASED CODING AND SECURING OF MOTION JPEG 2000 AND H.264 FOR WIRELESS SURVEILLANCE VIDEO TRANSMISSION

J. Meessen^a, C. Parisot^a, D. Agrafiotis^b, C. Le Barz^c, Y. Sadourny^c, J.-F. Delaigle^a, D. Bull^b and D. Nicholson^c

^aMultitel, Belgium, {Jerome.meessen, Christophe.parisot, delaigle}@multitel.be

^bDept. of Electrical & Electronic Engineering, University of Bristol, U.K., {d.agrafiotis, dave.bull}@bristol.ac.uk

^cThales Communication, France, {didier.nicholson, cedric.lebarz}@fr.thalesgroup.com

Keywords: Segmentation, Region Of Interest, Motion JPEG 2000, MPEG-4 AVC / H.264, selective encryption.

Abstract

This paper presents the integrated method of the IST WCAM project for content-based Motion JPEG 2000 and H.264 video coding. The idea is to link statistical segmentation with the video coders so as to guarantee high visual quality for semantically relevant objects while meeting the wireless bandwidth constraints. In the case of Motion JPEG 2000 coding, the segmentation results are also used for selective encryption.

WCAM's contributions to this challenging problem and ongoing work are presented with current results. Possible extensions are discussed.

1 Introduction

In January 2004, the IST WCAM project started with as goal to study, develop and validate a wireless, seamless and secured end-to-end networked audio-visual system [4]. This project focused on the technology convergence between video surveillance and multimedia content distribution over the Internet. Therefore, the video content is encoded in emerging formats, Motion JPEG 2000 (MJ2) and MPEG-4 AVC/H.264, and transmitted through wireless LAN to different types of decoding platforms like PDA's and Set Top Boxes. While robust wireless transmission is taken into account, the video content is also secured using a Digital Right Management (DRM) system.

To reach its scientific objectives, WCAM has implemented video processing tools such as real-time H.264 coder and decoder, real-time Motion JPEG 2000 coder, a scalable Motion JPEG 2000 decoder as well as a PDA client application supporting these formats and metadata streaming.

This paper addresses the critical video surveillance issue tackled by WCAM: how to guarantee high visual quality for the semantically relevant objects while respecting the real-time and wireless channel bandwidth constraints.

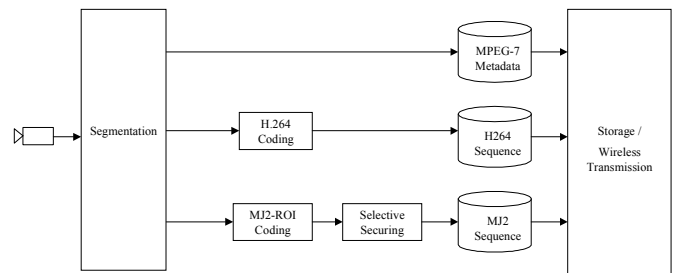


Figure 1. WCAM's coding system

Basically, when the camera is fixed, we propose to link automatic video analysis tools, such as segmentation, to the MJ2 and H.264 coding. The object of semantic importance are then localised in each frame and can be coded with higher quality compared to their background. Moreover, in the case of MJ2, the output of the video analysis is also exploited to selectively secure objects of the video sequence. The coding system is illustrated on Figure 1. We review here the scientific contributions of WCAM and the ongoing work related to that challenging problem.

The paper is organised as follows. In section 2, the proposed automatic scene analysis method is reviewed including a brief description of the chosen metadata format. Section 3 presents the results of object-based coding with MJ2 while the selective securing of the video content is discussed in Section 4. Our current work on automatic Region-Of-Interest (ROI) coding with H.264 is introduced in Section 5 before concluding in Section 6.

2 Automatic scene analysis and description

The goal of the scene analysis module is to detect and follow regions of interest (ROI) of the video stream in

order to both generate metadata describing the relevant events for the surveillance application and provide information for the coding modules optimization.

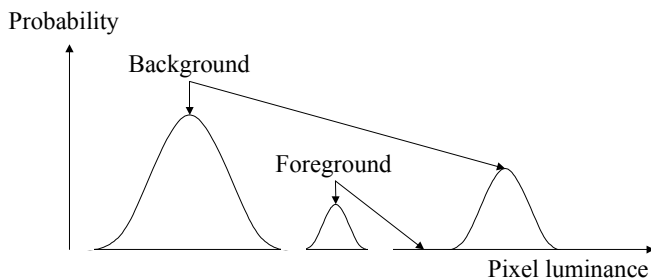


Figure 2. Mixture of Gaussian for pixels' state modelling

As we have described in [8], the automatic ROI extraction module of WCAM is based on a real-time statistical segmentation. The algorithm is based on a mixture of Gaussians modelling of the luminance for each pixel of the background as illustrated on Figure 2 [10],[4],[2]. The main advantage of this technique is that it can automatically deal with backgrounds that have multiple states like cyclic states such as blinking lights, grass and trees moving in the wind, acquisition noise etc. Furthermore, the background model update is unsupervised when the scene conditions are changing, e.g. lighting conditions in outdoor applications, increase or decrease of the background states number for each pixel independently...

The results of the segmentation are further analysed to detect high-level events, such as ‘abandoned luggage’ or ‘intrusion in a restricted area’. For each frame, the statistical segmentation process produces a set of spatially independent blobs (e.g. Related segmentation on Figure 3). Those blobs are analysed in terms of shape, texture, position and temporal coherency in order to robustly extract abnormal behaviour events. In the case of intrusion detection for example, we verify that the size of the detected blobs correspond to mobiles we are looking for (e.g. people) and an alert is sent only if blobs of the same size can be found in consecutive frames with coherent displacements (e.g. speed range of people). As another example, an abandoned luggage can be defined as a segmented blob with no movement and constant texture for a long period (e.g. ten seconds in a corridor). An abandoned luggage can be detected even if it is sometimes occluded by walkers as soon as we are able to recognize its shape and texture during non-occluding periods. As often, the more accurate the event definition is, the lower the false alarm rate will be. We also highly recommend using a camera calibration when it is possible since it will allow automatically adapting the size and speed of events we are looking for to any image position. When such target events are detected, the CAP ‘Common Alert Protocol’, international standard for alarm signalling, is used to warn the client [1].

The segmentation results together with this high-level content are described in an MPEG-7 XML file, which is transmitted to the client and/or enables offline content-based retrieval, though this is not directly addressed by WCAM [13].

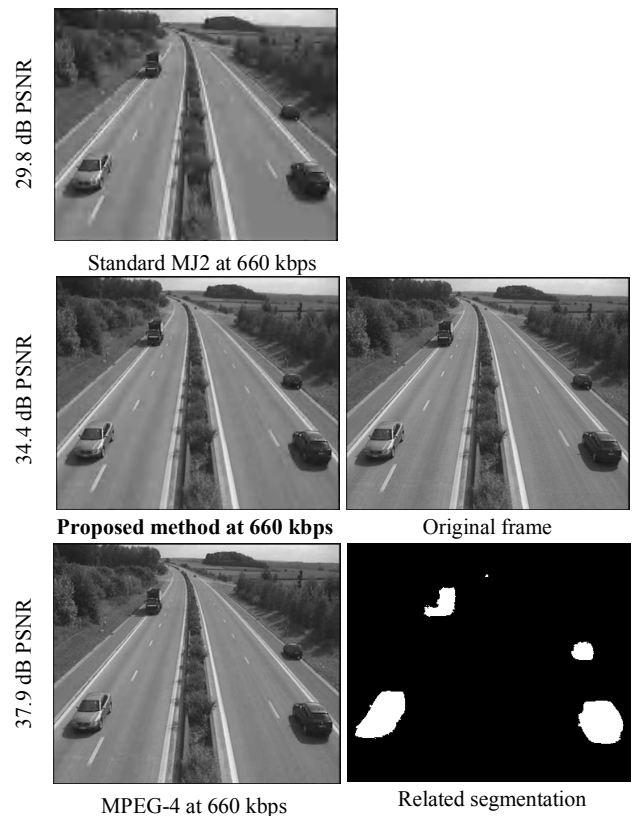


Figure 3. Results of the smart coding when separately transmitting the regions of interest and the estimated background.

3 Smart Motion JPEG 2000 coding

The segmentation outputs for each frame a binary mask representing the mobile objects that are of interest for the user. The smart MJ2 coding consists in using this mask as ROI definition by the coder. The specific JPEG 2000 ROI coding method has been described in [8]. It combines the standard *max-shift* with a local allocation control. The achieved performances are identical to the *max-shift* method only, while allowing rate control for both foreground and background.

When combined with an appropriate RTP packetisation, the ROI coding solution also enables efficient prioritisation at the network level.

Moreover, as we have shown in [8], the statistical nature of the chosen segmentation algorithm can be further exploited. Indeed, the segmentation easily provides an

estimation of the background at each time. It just requires getting the most probable value of each pixel. The transmission of the ROI's and this estimated background in two separate MJ2 streams, with different frame rates, leads to remarkable delivery bandwidth reduction coupled with a complexity reduction at both the encoder and decoder. This is illustrated on Figure 3 with the MJ2 *speedway* CIF reference sequence. For this sequence at 660kBps, the PSNR gain was of more than 4dB compared to standard MJ2 performance. The PSNR gets then comparable to the one obtained with MPEG-4 at the same rate. Furthermore, the subjective quality is also improved since compression error fluctuations and acquisition noise are removed for the background.

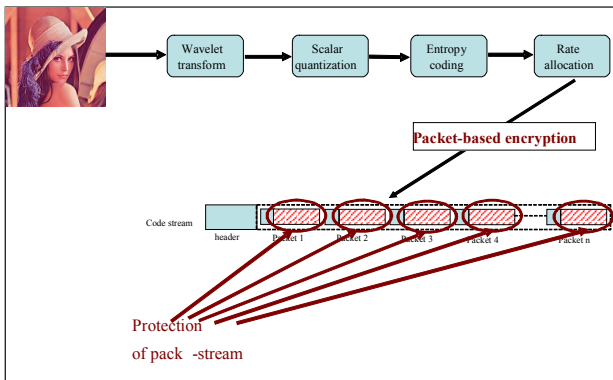


Figure 4. Packet based encryption principle

4 Automatic content-based video securing

In WCAM, video data is protected at content level rather than during transmission, which ensures the protection and confidentiality of the images along the whole distribution chain [10]. The encryption keys are managed by a Digital Rights Management (DRM) system [12].

The content encryption tools are compliant with the JPSEC upcoming ISO standard [4]. Moreover, they are selective, which allows ensuring confidentiality of data while encrypting only parts of it: the ciphered, compressed data is still ordered as it was in the original, unprotected image data. JPEG 2000 compliant viewers are able to decode our JPSEC images, but all the encrypted parts will appear scrambled.

The basic encrypted units we handle are JPEG 2000 packet bodies, as depicted on Figure 4. Each packet holds data related to a given set of resolution level, quality layer, component and precinct, which provides an excellent flexibility when choosing the parts of the image to be protected. An automated encryption scheme is obtained when directly using the results of the segmentation process described in Section 2. All the packets containing some ROI data are selected and protected. The rest of the image can possibly be protected using another set of keys. This

way, the whole image is only accessible to authorised users, but the ROI has another protection layer, which allows performing access control at the user level.

Another protection method was also developed in the WCAM project, which enables a more accurate protection of the ROI [2]. This scrambling technique adds a pseudo-random noise to selected parts of a JPEG 2000 codestream, at the wavelet coefficients level. As for the previous method, the image can only be de-scrambled with a set of keys: in this case, they drive the pseudo-random number generator (PRNG).

The access control relies on the secrecy of the encryption keys for both protection methods, which is handled by the DRM system.

5 ROI coding with H.264

In all standard hybrid video coders (MPEG-2, H.263, MPEG -4, H.264) the number of bits required for coding each macroblock in a video frame depends on the level of activity in the macroblock, the effectiveness of the prediction, and the quantisation step size. The latter one is controlled by the quantisation parameter (QP). The value of QP controls the amount of compression and corresponding fidelity reduction for each macroblock. Uniform (or constant) quality coding assigns the same value of QP to all frames and hence to all macroblocks in a coded video leading to large fluctuations of the output bit rate. Rate controlled coding assigns a portion of the pre-defined bit budget to each frame/macroblock with the aim of maximising the average quality of the whole video/frame. This type of coding leads to very limited fluctuations of the output bit rate at the expense of variable quality for the reconstructed frames. In both cases all macroblocks / regions are effectively treated as if they have the same importance (priority).

Region of Interest (ROI) coding aims at providing preferential treatment for those regions of the video frame, which are deemed to be of higher importance to the viewer. In our case, the ROI corresponds to the objects automatically extracted by the analysis tools described in Section 2. Preferential treatment can mean maintaining a minimum level of quality for the ROI, or assigning a larger proportion of a pre-defined bit budget to it compared to other less important regions.

Here, we focus on how to add ROI functionality to the H.264 JM rate control method [7][14]. We describe a modified rate control method based on that of the JM H.264 encoder, which can accommodate ROI coding given a segmentation mask. Adding ROI functionality to a hybrid video coder such as H.264 requires adding some form of control/influence over the macroblock QP allocation process. The rate control method adopted by the

JM consists of up to 3 control layers, namely the GOP layer, the frame layer and, if specified, the basic unit layer. These layers are illustrated on Figure 5. Herein we describe ROI modifications for the GOP and frame layer.

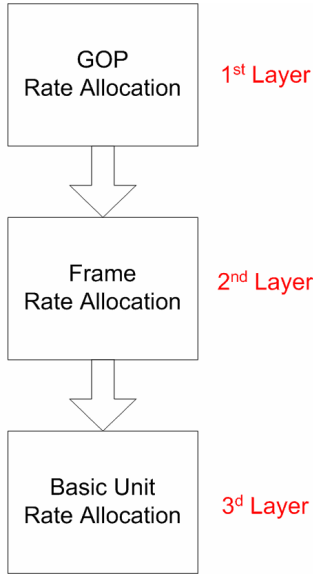


Figure 5: The rate control layers of the JM encoder.

5.1. GOP layer

In the 1st rate control layer a specific bit budget is allocated to the remaining pictures within a group of pictures (GOP) based on the coding rate and the occupancy of the virtual buffer employed for the rate regulation. The occupancy of the buffer is updated after coding each picture. From a ROI point of view what is important about the GOP layer rate control is that it assigns a QP - $QP_i(I)$ - to the first IDR and stored picture (P) of each (i^{th}) GOP based on the average QP assigned to the P pictures of the previous GOP as well as the respective $QP_{i-1}(I)$ allocation made at the beginning of the previous GOP. We modify this process so that the QP allocation considers past QP allocations for regions of similar priority. In other words, given that a ROI has been detected at the first frame of the new GOP and that a ROI was present during the previous GOP, $QP_i(I)$ is calculated separately for the MBs of the ROI and for those of the background.

5.2. Frame layer

At the 2nd layer the frame bit budget allocation takes place. We focus on the case of P frames. Having calculated a target buffer level for each remaining P picture in the current GOP (after coding the first two pictures) frame bits are allocated based on the target buffer level, the frame rate, the available channel bandwidth, and the actual buffer occupancy. Additionally frame bits are allocated based on the number of bits remaining for the GOP. The actual target bits for the j^{th} frame of the i^{th} GOP - $T_i(j)$ - are a weighted combination of these two allocation

processes. For ROI coding we distribute the target frame bits among the different regions in a way that reflects their relative importance. Additionally the quantisation step calculation is now done separately for the ROI and background using the ROI and background target bits respectively. The quantisation step calculation requires prediction of the mean absolute difference (MAD) which again we do separately for the ROI and background based on the actual MAD of the ROI/background in the previous frame. More specifically the ROI bit allocation is done in two stages. At the first stage bits are allocated based on the normalized size and normalized MAD of the ROI and background as follows:

$$RoIT_{i,sm}(j) = (w_s S_{RoI} + w_{MAD} MAD_{RoI}) \times (T_i(j) - m_{h,i}(j)) \quad (1)$$

$$BcgdT_{i,sm}(j) = (w_s S_{Bcgd} + w_{MAD} MAD_{Bcgd}) \times (T_i(j) - m_{h,i}(j)) \quad (2)$$

where S_{RoI} and S_{Bcgd} are the normalized size of the ROI and background respectively, MAD_{RoI} and MAD_{Bcgd} are the normalized MAD values and w_s and w_{MAD} are the weights of the two factors indicating how much they influence the bit allocation process ($w_s + w_{MAD} = 1$). m_h is the number of header (including motion) bits. Once the initial allocation has taken place, we employ a priority constant P_{RoI} (ranging from 0 to 1) to specify the proportion of the background texture bits - $BcgdT_{i,sm}(j)$ in (2) - that will be allocated to the ROI as follows:

$$RoIT_i(j) = RoIT_{i,sm}(j) + (P_{RoI} \times BcgdT_{i,sm}(j)) \quad (3)$$

$$BcgdT_i(j) = BcgdT_{i,sm}(j) - (P_{RoI} \times BcgdT_{i,sm}(j)) \quad (4)$$

where $RoIT$ and $BcgdT$ are the bits finally allocated to the ROI and background respectively. The quadratic equation suggested in [7] is then employed with separate coefficients for ROI and background ($RoIc_1$, $RoIc_2$, $Bcgdc1$, $Bcgdc2$), separate MAD values ($RoI\tilde{\sigma}_i(j)$, $Bcgd\tilde{\sigma}_i(j)$) and the previously allocated bits, in order to find the quantization parameters for the ROI and background:

$$RoIT_i(j) = RoIc_1 \times \frac{RoI\tilde{\sigma}_i(j)}{RoIQ_{step,i}(j)} + RoIc_2 \times \frac{RoI\tilde{\sigma}_i(j)}{RoIQ_{step,i}^2(j)} + m_{h,i}(j) \quad (5)$$

$$BcgdT_i(j) = Bcgdc1 \times \frac{Bcgd\tilde{\sigma}_i(j)}{BcgdQ_{step,i}(j)} + Bcgdc2 \times \frac{Bcgd\tilde{\sigma}_i(j)}{BcgdQ_{step,i}^2(j)} + m_{h,i}(j) \quad (6)$$

The quantization parameters for both ROI and background are then bounded by the same conditions used in the original method for smooth quality variations between frames. Below example results are given for the ‘mobile’ test sequence. The ROI and background are shown in Figure 6. PSNR versus bit-rate graphs for different priorities are given in Figure 7.

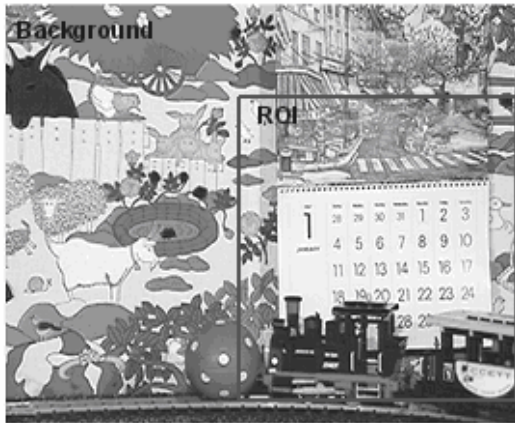


Figure 6. ROI and background for the 'Mobile' sequence.

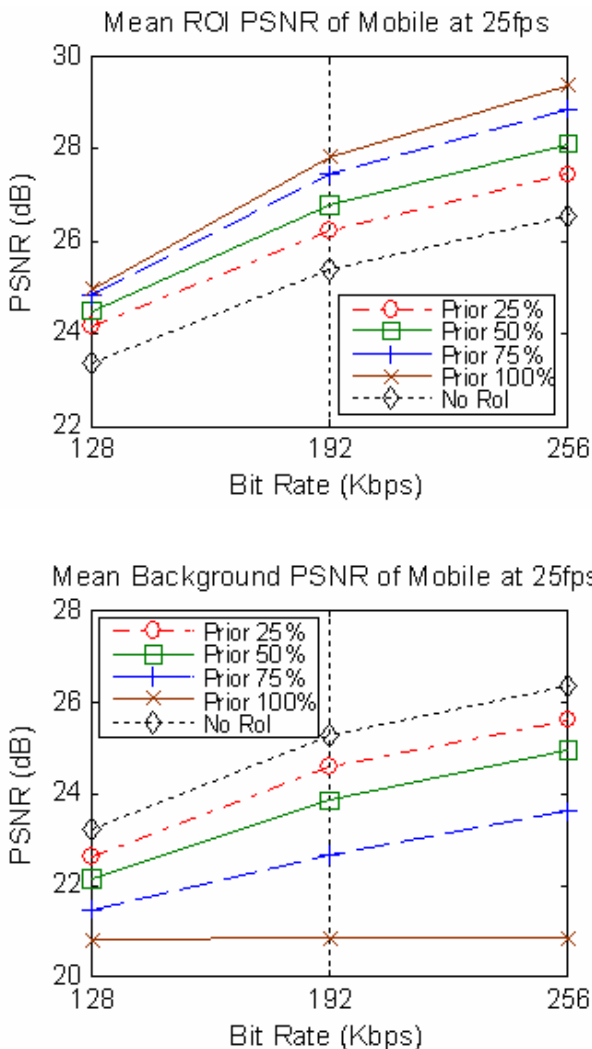


Figure 7. Mean ROI & Background PSNR for a number of priority constants.

6 Conclusions

We have presented IST WCAM's results and current activities related to unsupervised semantic-based surveillance video coding in Motion JPEG 2000 and H.264. The methods allow strong compression rate while preserving a high visual quality of relevant objects. The output of the video analysis also enables selective encryption.

Though the WCAM project is now ending, many possible extensions and new challenges have raised, such as using more advanced video analysis tools or extending the method towards efficient replenishment strategy.

The described smart coding modules have been integrated in a complete real-time surveillance system. Other results of WCAM are also available online on the project web site [4].

Acknowledgements

This work has been funded by the EU commission under FP6 IST-2003-507204 project WCAM "Wireless Cameras and Audio-Visual Seamless Networking".

References

- [1] Common Alert Protocol, <http://www.incident.com/cap>
- [2] X. Desurmont, C. Chaudy, A. Bastide, C. Parisot, J.F. Delaigle, B. Macq, "Image analysis architectures and techniques for intelligent systems", *IEE proc. on Vision, Image and Signal Processing, Special issue on Intelligent Distributed Surveillance Systems*, **152**(2), pp. 224-231, April 2005.
- [3] F. Dufaux, T. Ebrahimi, "Smart video surveillance system preserving privacy", in *Proceedings of SPIE Vol. 5685 - Image and Video Communication and Processing 2005*, San Jose, CA, January 2005
- [4] T. Ebrahimi, C. Rollin, S. Wee, eds., "JPSEC Study Text for Final Draft International Standard, version 2.0," ISO/IEC JTC1/SC29/WG1N3697, July 2005
- [5] FP6 IST-2003-507204 project WCAM "Wireless Cameras and Audio-Visual Seamless Networking", <http://www.ist-wcam.org>
- [6] K. Kim, T. Horprasert, D. Harwood, L. Davis, "Codebook-based Background Subtraction and Performance Evaluation Methodology", 2003.
- [7] Zhengguo Li, Feng Pan, Keng Pang Lim, Xiao Lin and Susanto Rahardja, "Adaptive Rate Control for H.264", in *Proceedings of at IEEE International Conference on Image Processing*, October 2004.
- [8] Jerome Meessen, Christophe Parisot, Cedric Le Barz, Didier Nicholson and Jean-Francois Delaigle, "Smart Encoding for Wireless Video Surveillance," *proc. of*

- SPIE - Image and Video Communications and*
- [9] Jerome Meessen, Christophe Parisot, Xavier Desurmont and Jean-Francois Delaigle, "Scene Analysis for Reducing Motion JPEG 2000 video Surveillance Delivery Bandwidth and Complexity," *proc. of IEEE International Conference on Image Processing (ICIP 05)*, vol. I, pp. 577-580, Genova, Italy, September 2005
- [10] Y. Sadourny, V. Conan, P. Fonseca, C. Serrão, "WCAM: secured video surveillance with digital rights management", in *Proceedings of SPIE Vol. 5685 - Image and Video Communication and Processing 2005*, San Jose, CA, January 2005
- [11] C. Stauffer, W.E.L. Grimson, "Adaptive Background mixture models for real-time tracking", *Proc. of Processing 2005*, San Jose, CA, January 2005.
- IEEE Conference on Computer Vision and Pattern Recognition*, **2**, pp. 246-252, June 1999.
- [12] Siegert G., Serrão C., "An Open-Source Approach to Content Protection and Digital Rights Management in Media Distribution Systems", *8th Annual CTI Conference*, Denmark, 2004
- [13] T. Sikora, "The MPEG-7 Visual Standard for Content Description - An Overview," *IEEE Trans. on Circuits and Systems for Video Technology*, **11**(6), pp. 696-702, June 2001.
- [14] Gurry Sullivan, Thomas Wiegand, Keng-Pang Lim, "Joint Model Reference Encoding Methods and Decoding Concealment Methods", *Document JVT-1049*, San Diego, USA, September 2003.