AUTOMATIC AND ADAPTIVE NETWORK-AWARE MACROBLOCK INTRA REFRESH FOR ERROR-RESILIENT H.264/AVC VIDEO CODING

Paulo Nunes[1], *Luis Ducla Soares*[2], *Fernando Pereira*[3]
[1,2]Instituto Superior de Ciências do Trabalho e da Empresa – Instituto de Telecomunicações
[3]Instituto Superior Técnico – Instituto de Telecomunicações
[1]pjln@lx.it.pt, [2]lds@lx.it.pt, [3]fp@lx.it.pt

ABSTRACT

In this paper, an automatic and adaptive network-aware macroblock Intra coding refresh method is proposed. It adaptively selects the amount of gracefully forced Intra macroblocks and the amount of cyclic Intra refresh (CIR) macroblocks based on the actual network error conditions, in terms of packet loss rate, and the target encoding bit rate. With the proposed method, the error robustness of H.264/AVC bitstreams can be significantly increased by efficiently taking into account the actual rate-distortion impact of Intra coding macroblock mode decisions, while simultaneously guaranteeing that errors do not propagate endlessly by selecting an adequate amount of CIR macroblocks per frame according to the network packet loss rate and the encoding target bit rate.

Index Terms— Error resilience, macroblock mode decision, Intra coding refresh, H.264/AVC video coding

1. INTRODUCTION

Due to the increasing demand for universal access to video content, the range of networks being used to deploy video services is still increasing as well as the amount of content being transmitted. Since many of these networks are error-prone (e.g., wireless networks and the Internet), appropriate error resilience techniques are necessary to make these video services efficiently available to the users with an acceptable quality.

Error resilience techniques are usually seen as playing a role at the decoder side of the communication chain. However, by using preventive error resilience techniques at the encoder side, which involve the intelligent design of the encoder, it is also possible to make the task of the decoder much easier in terms of dealing with errors. In fact, the performance of the decoder can greatly vary depending on the amount of error resilience help provided in the bitstream generated by the encoder. This is especially true in the case of video encoders that rely on predictive (Inter) coding to remove temporal redundancy, such as those specified in the H.264/AVC standard [1].

A very effective method to make the bitstreams generated by these video encoders more robust to errors, without any specific knowledge of what is being done at the decoder in terms of error resilience, is to use one of the Intra coding refresh schemes available in the literature [2,3,4]. These schemes can avoid long error propagation caused by channel errors by selectively coding in Intra mode different parts of the video content at different time instants. As a result, the fast quality decay usually associated with long term propagation of channel errors is avoided, improving the overall subjective impact at the decoder.

The work presented was developed within VISNET II, a European Network of Excellence (http://www.visnet-noe.org) from the DG INFSO Networked Media Unit, and was also supported by Fundação para a Ciência e a Tecnologia (Portugal).

The main problem of these Intra coding refresh schemes available in the literature is that they do not take into account the rate-distortion (RD) cost of their Intra coding decisions, which means that when these schemes are used they can significantly reduce the coding efficiency. In [5], this problem was solved by integrating the Intra coding refresh decisions in the rate control module of an H.264/AVC video encoder [6,7], thus allowing coding efficiency to be effectively combined with error robustness. More specifically, in [5], the RD cost of coding macroblocks (MBs) using the Intra and Inter modes is compared. If the cost of Intra coding is only 'slightly' larger than the cost of Inter coding, then the coding mode is forced to Intra. This reduces error propagation and improves the robustness to transmission errors, by increasing the amount of Intra coded MBs in the bitstream, without significantly increasing the RD cost.

However, the scheme proposed in [5] has one important limitation, which is the fact that the varying error characteristics of the network being used are not taken into account. This limitation implies that the amount of Intra coding refresh being used by the encoder may not be suited to a wide range of underlying network error characteristics. In [8], the authors have tackled this limitation by also considering the feedback from the network about its current error characteristics. This allows the encoder to dynamically adapt the amount of Intra coding refresh to the current state of the network and needed error robustness, thus improving the decoded video quality even further. Additionally, to guarantee that all MBs are eventually refreshed and errors do not propagate indefinitely in some of them, which could negatively affect the subjective impact at the decoder, a cyclic Intra refresh (CIR) scheme should also be concurrently applied. The problem of selecting the amount of cyclically Intra refreshed MBs for different network conditions was, however, not addressed in [8].

This paper proposes a fully automatic network-aware MB Intra coding refresh for error-resilient H.264/AVC video coding, which also dynamically adjusts the amount of cyclically Intra refreshed MBs according to the network conditions.

The rest of the paper is organized as follows. Section 2 briefly describes the authors' previous work on which the novel proposed approach is based, while Section 3 describes the proposed solution itself. Section 4 presents some relevant performance results for the proposed scheme in comparison with appropriate alternatives and, finally, Section 5 concludes the paper.

2. EFFICIENT INTRA CODING REFRESH

The improved network-aware MB Intra coding refresh method proposed in this paper is based on the concepts introduced in [5], which were already briefly introduced in Section 1. The main idea proposed in [5] is that, under error-prone conditions, Intra coding refresh should always be used whenever it costs only slightly more than Inter coding. Since several Intra modes are possible, the one with the lowest RD cost should be chosen. By doing so, the generated bitstream will be more robust to channel errors, without spending much additional rate on Intra coded MBs, which would reduce the decoded video quality when no channel errors exist.

To briefly review the scheme proposed in [5], the MB-level mode decision architecture shown in Figure 1, which is based on the RD optimization (RDO) technique [9], should be considered.

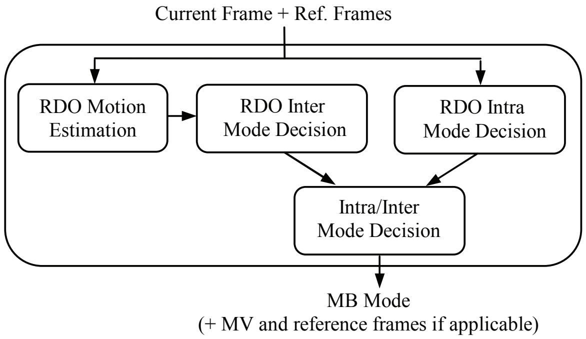

Figure 1 – Architecture of the RDO-based error resilient MB Intra/Inter mode decision [5].

In [5], to decide the best mode to encode a given MB, the best Intra and Inter modes have first to be determined, which is done by using RDO in both cases. The best Intra mode is the one that corresponds to the lowest Intra RD cost, J_{INTRA}, among the set of all possible Intra modes. The best Inter mode is the one that corresponds to the lowest Inter RD cost, J_{INTER}, among the set of all possible Inter modes.

When the best Intra mode has a lower RD cost than the best Inter mode for a given MB, which corresponds to having J_{INTRA} smaller than J_{INTER}, the Intra coding mode is used. These MBs are not considered to be forced Intra MBs since they would have been Intra coded anyway if regular RDO would have been used.

On the other hand, when the best Intra mode has a higher RD cost than the best Inter mode for a given MB, which corresponds to having J_{INTRA} larger than J_{INTER}, the ratio of these two RD cost values is compared to an input control parameter α_{RD} in order to decide which mode will be used. This input parameter specifies the tolerable RD cost increase for replacing an Inter coded MB with an Intra coded one, thus increasing the error resilience; the decision is made as follows:

$$if\left(J_{INTRA}/J_{INTER} \le \alpha_{RD}\right) \text{ the Intra mode is selected.} \qquad (1)$$

$$if\left(J_{INTRA}/J_{INTER} > \alpha_{RD}\right) \text{ the Inter mode is selected.} \qquad (2)$$

In the special case of $\alpha_{RD} = 1$, no particular coding mode is favored in an RD sense and the Intra mode will only be chosen if it leads to a lower RD cost than the Inter mode. However, for $\alpha_{RD} > 1$, the Intra mode is more or less favored over the Inter mode, in order to 'buy' error resilience where it is 'cheaper' using a clever 'buying' strategy. Therefore, the α_{RD} parameter makes it possible to control the amount of gracefully forced Intra coded MBs in the bitstream. The MBs that end up being forced to Intra mode are those for which the RD costs of Intra and Inter modes are closest, which typically correspond to MBs with high Inter RD cost and, therefore, would also be difficult to conceal at the decoder if lost, since their prediction is not simple.

This method is not enough by itself to guarantee that all MBs are frequently refreshed and, therefore, can still lead to long term error propagation for some MBs in the video sequence. To avoid this, a CIR scheme should also be applied concurrently, independently of the adopted α_{RD} parameter. In [5], the choice of the α_{RD} parameter value and the amount of CIR MBs per frame were not addressed, but this represents an important aspect that needs to be solved, since using inadequate values for the current network conditions can lead to a negative subjective impact at the decoder.

3. PROPOSED NETWORK-AWARE INTRA CODING REFRESH METHOD

Based on the ideas described in Section 2, this paper proposes a fully automatic and adaptive method to select simultaneously, for the actual network error conditions and target bit rate: i) the amount of gracefully forced Intra MBs, through the specification of the α_{RD} parameter; and ii) the amount of CIR MBs per frame. The method is fully automatic meaning that no user intervention is needed; the method is adaptive meaning that it assumes some feedback from the network is available, namely in terms of the average packet loss rate (PLR), and it reacts and adapts to it. The following sections describe the proposed method.

3.1 Intra refresh with network-aware α_{RD} selection

When information about the network error conditions is available at the encoder, the α_{RD} parameter can be adaptively computed, based on these error conditions (i.e., the PLR) and other encoding characteristics (i.e., the target bit rate r_b). This approach has been followed in [8], where the authors have developed the following mapping function to adaptively compute the α_{RD} parameter:

$$\alpha_{RD} = f_{\alpha}(PLR, r_b) = 1 + \left(a_0 \cdot r_b + b_0\right) \cdot \left(1 - e^{-c_0 \cdot PLR}\right) \qquad (3)$$

with $a_0 = 0.83 \times 10^{-6}$, $b_0 = 0.97$, and $c_0 = 0.90$.

Through (3), the encoder is able to adjust the amount of Intra refresh according to the network error conditions and the available bit rate. This Intra refresh method typically increases the Intra refresh for the more complex MBs, which are those typically more difficult to conceal. The main problem of this approach is that it does not guarantee that all MBs in the scene are refreshed. This is clearly illustrated in Figure 2 for the *Foreman* sequence, where the right image represents the relative amount of MB Intra refresh along the sequence (lighter blocks mean more Intra refresh). As it can be seen, with this Intra refresh scheme some MBs are never refreshed, which can lead to errors propagating indefinitely along time in these MB positions (dark blocks in Figure 2).

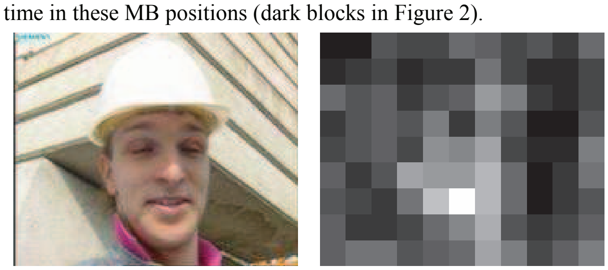

Figure 2 – Relative amount of Intra refresh for the MBs of the Foreman sequence (QCIF, 15Hz, 128 kbit/s and $\alpha_{RD} = 1.1$).

3.2 Intra refresh with network-aware α_{RD} and CIR selection

The main drawback of the scheme described in the previous section can be alleviated by introducing some additional CIR MBs per frame to guarantee that all MB positions are refreshed with a minimum periodicity. This requirement raises the problem of how to adaptively select an adequate amount of CIR MBs that is sufficiently high to avoid long term error propagation without penalizing too much the encoder RD performance.

In this paper, the adequate α_{RD} value and the number of CIR MBs per frame are decided separately, using a different model for each of these two error resilience parameters. For the α_{RD} selection, the model in (3) is used. As for the selection of the number of CIR MBs, it was verified after exhaustive testing that the optimal amount of CIR MBs tends to increase linearly with the bit rate r_b, for a given PLR, but tends to increase exponentially with the PLR, for a given bit rate. Based on these observations, the following model is proposed here for the selection of the amount of CIR MBs per frame:

$$CIR = f_{CIR}(PLR, r_b) = (a_1 \cdot r_b + b_1) \cdot e^{c_1 \cdot PLR} \quad (4)$$

where a_1, b_1 and c_1 are the model parameters that need to be estimated. These parameters have been determined by non-linear curve fitting (the Levenberg-Marquardt method) of the optimal amount of CIR MBs per frame, experimentally determined for a set of representative test sequences, encoding bit rate ranges and packet loss rates (see Section 4). The estimated parameters were $a_1 = 12.97 \times 10^{-6}$, $b_1 = -0.13$, and $c_1 = 0.24$.

Figure 3 shows the proposed model as well as the experimental data for the *Mobile and Calendar* test sequence. As can be seen, a simple linear model would not have represented well the experimental data.

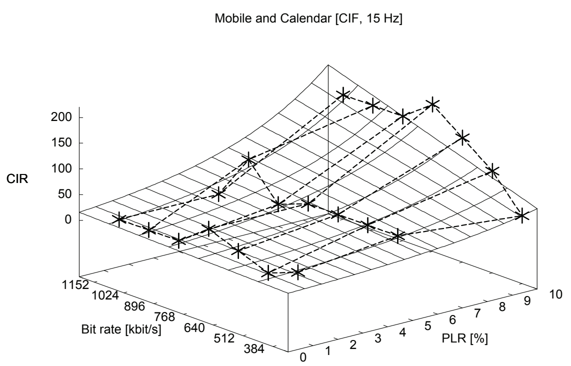

Figure 3 – Optimal amount of CIR MBs per frame versus PLR and bit rate for the Mobile and Calendar sequence.

The CIR order is randomly defined once before encoding, as specified in [7], to avoid the subjectively disturbing effect of performing sequential (e.g., raster scan) refresh. The determined order is then cyclically followed with the computed number of MBs being refreshed in each frame.

Therefore, the proposed network-aware MB Intra coding refresh (NIR) scheme can be briefly described by the following steps in terms of encoder operation:

1. Obtain the PLR value through network feedback.
2. Compute the number of CIR MBs to be used per frame, by using the proposed f_{CIR} function defined by (4) and rounding it to the nearest integer.
3. Compute the α_{RD} value by using the proposed f_α function defined by (3).
4. For each MB in a frame, check if it should be forced to Intra mode according to the CIR order and the determined number of CIR MBs per frame; if not, perform Intra/Inter mode decision using the α_{RD} value computed in Step 3; encode the MB with selected mode.
5. At the end of the frame, check if a new network feedback report has arrived; if yes, go back to Step 1; if not, go back to Step 4.

The definition of when the network reports are issued is out of the scope of this work since this will depend on how the network protocols are configured and the varying characteristics of the network itself [10].

Notice that independently selecting the α_{RD} value and the amount of CIR MBs, while they are likely interdependent, can lead to chosen values that do not correspond to the optimal (α_{RD}, CIR) pair. However, it has been verified after extensive experimentation that the proposed independent selection process is still robust in the sense that the chosen values are typically close enough to the optimal pair and, therefore, the overall performance is not dramatically penalized.

4. PERFORMANCE EVALUATION

To evaluate the performance of the proposed NIR scheme, it has been compared in similar conditions to a reference Intra refresh scheme, which basically corresponds to the network-aware version with the cyclic Intra refresh scheme of the JM 13.2 reference software [7]. In the reference scheme, the optimal number of CIR MBs per frame is selected manually for the considered network conditions, while in the proposed solution the selection of the amount of CIR MBs per frame and the α_{RD} parameter are done fully automatically. For the proposed and reference schemes, the *Mother and Daughter*, the *Foreman*, and the *Mobile and Calendar* video sequences have been encoded using the H.264/AVC Baseline Profile (with JM 13.2 [7]). The used test conditions are summarized in Table 1. For QCIF, each frame was divided in 3 slices, while for CIF each frame was divided in 6 slices. In both cases, each slice consists of 3 MB rows. After encoding, each slice was mapped to an RTP packet for network transmission [10].

Table 1 – Test conditions

Video test sequence	*Mother and Daughter*	*Foreman*	*Mobile and Calendar*
Spatial resolution	QCIF	QCIF	CIF
Frame rate (Hz)	10	10	15
Bit rate (kbit/s)	24 – 64	48 – 128	384 – 1152

For the reference scheme, the number of cyclically Intra refreshed MBs per frame was chosen for each PLR and bit rate, such that the decoded video quality would be the best possible. This was done manually by performing an exhaustive set of tests using many different amounts of CIR MBs per frame and then choosing the one that leads to the highest decoded average PSNR value, obtained by averaging over 50 different error patterns. For the QCIF video sequences, the possible values for the number of cyclically Intra refreshed MBs were chosen from the representative set {0, 5, 11, 22, 33, …, 99}, while for the CIF video sequences the representative set consisted of {0, 22, 44, 66, …, 396}.

To simulate the network conditions, three different PLRs were considered: 1%, 5% and 10%. Since each slice is mapped to one RTP packet, each lost packet will correspond to a lost video slice. Packet losses are considered independent and identically distributed. For each one of the studied PLRs, each coded bitstream has been corrupted and then decoded 50 times (i.e., corresponding to 50 different error patterns or runs), while applying the default error concealment technique implemented in the H.264/AVC JM 13.2 reference software [7,11]. The presented results correspond to PSNR averages of these 50 different runs for the luminance component (PSNR Y).

For the conditions mentioned above, PSNR Y results are shown in Table 2, Table 3, and Table 4 for the *Mother and Daughter*, *Foreman*, and *Mobile and Calendar* video sequences, respectively. In these tables, NIR refers to the proposed network-aware Intra coding refresh scheme and JM refers to the reference technique (winning cases appear in bold). In addition, OPT corresponds to the manual selection of the best (α_{RD}, CIR) pair.

Table 2 – PSNR results for the Mother and Daughter sequence

Bit rate (kbit/s)	PSNR Y (dB)								
	PLR = 1%			*PLR* = 5%			*PLR* = 10%		
	JM	NIR	OPT	JM	NIR	OPT	JM	NIR	OPT
24	34.67	34.62	35.21	31.91	**32.48**	33.32	30.68	**31.79**	32.23
32	36.01	**36.42**	36.65	32.59	**34.06**	34.57	31.12	**32.87**	33.16
40	36.57	**37.26**	37.36	33.44	**34.91**	35.40	32.02	**33.61**	34.03
48	36.97	**38.00**	38.14	33.92	**35.75**	35.87	32.68	**34.28**	34.47
56	37.77	**38.68**	38.69	34.49	**36.33**	36.54	33.14	**34.46**	35.14
64	37.92	**39.34**	39.34	34.94	**36.84**	37.03	33.60	**35.13**	35.55

Table 3 – PSNR results for the Foreman sequence

Bit rate (kbit/s)	PSNR Y (dB)								
	PLR = 1%			*PLR* = 5%			*PLR* = 10%		
	JM	NIR	OPT	JM	NIR	OPT	JM	NIR	OPT
48	30.02	29.92	30.32	25.94	**27.01**	27.01	23.89	**24.44**	24.88
64	31.06	**31.79**	31.83	26.96	**28.59**	28.59	25.14	**25.76**	26.22
80	32.05	**32.71**	32.73	28.08	**29.30**	29.30	26.88	26.24	26.96
96	32.63	**32.90**	33.69	29.06	**29.75**	29.93	28.27	27.12	28.34
112	33.34	**33.70**	34.34	29.89	**30.01**	30.50	29.20	27.90	29.28
128	33.74	**34.57**	35.05	30.79	**30.98**	31.32	30.01	28.77	30.08

Table 4 – PSNR results for the Mobile and Calendar sequence

Bit rate (kbit/s)	PSNR Y (dB)								
	PLR = 1%			*PLR* = 5%			*PLR* = 10%		
	JM	NIR	OPT	JM	NIR	OPT	JM	NIR	OPT
384	25.42	**25.62**	25.75	21.49	**21.98**	21.98	19.17	**19.75**	19.75
512	26.32	**26.79**	26.79	22.62	**22.76**	23.04	20.41	**20.83**	20.83
640	27.12	**27.45**	27.60	23.42	**23.42**	23.86	21.40	**21.66**	21.78
768	27.84	**28.35**	28.46	24.07	**24.40**	24.50	22.23	**22.34**	22.44
896	28.37	**28.80**	29.30	24.58	24.46	25.14	22.71	22.45	23.00
1024	28.78	**29.27**	29.78	25.13	**25.21**	25.63	23.23	23.00	23.33
1152	29.25	**29.88**	30.37	25.47	**25.65**	26.15	23.58	23.07	23.79

To help the reader to better read the gains obtained with the proposed technique, the results obtained for the *Mother and Daughter* sequence are also shown in chart form in Figure 4, for both JM and NIR. Similar plots for the *Foreman* and *the Mobile and Calendar* sequences are not shown due to space restrictions, but the trends are similar.

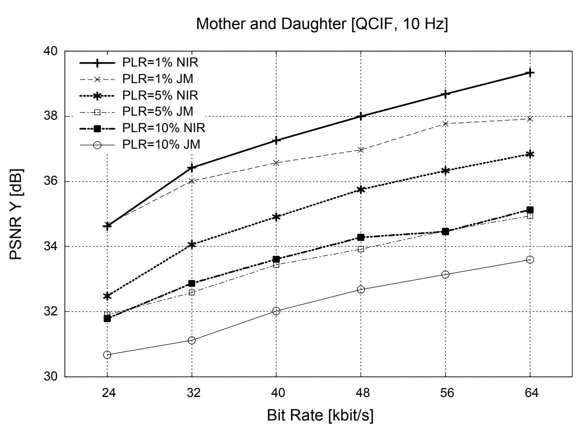

Figure 4 – PSNR results for the Mother and Daughter sequence.

The presented results show that, when the proposed fully automatic NIR scheme is used, the decoded video quality is significantly improved for the vast majority of tested conditions when compared to the reference method with a manually selected amount of CIR MBs (JM). Improvements of the NIR method can be as high as 1.90 dB for the *Mother and Daughter* sequence encoded at 64 kbit/s and a PLR of 5%. The most significant exception is for the PLR of 10% and higher bit rates (see Table 3 and Table 4). This exception is due to the fact that, for these PLR and bit rate values, the number of CIR MBs chosen with the proposed f_{CIR} is slightly different from the optimal values.

When comparing the NIR scheme to the one proposed in [8], which does not use CIR, the NIR PSNR Y values are most of the times higher than or equal to those achieved in [8]. The highest gains occur for the *Foreman* sequence encoded at 128 kbit/s and a PLR of 10% (0.90 dB), and for the *Mobile and Calendar* sequence encoded at 768 kbit/s and a PLR of 10% (0.60 dB). For the cases, where the NIR leads to lower PSNR Y values, the losses are never more than 0.49 dB, which happens for the *Mobile and Calendar* sequence encoded at 896 kbit/s and a PLR of 5%.

Notice, however, that the scheme in [8] cannot guarantee that all macroblocks will eventually be refreshed, which is a major drawback for real usage in error-prone environments. On the other hand, the one proposed in this paper can, not only overcome this drawback, but it does so fully automatically, without any user intervention.

5. FINAL REMARKS

This paper proposes a method to efficiently and fully automatically perform Intra coding refresh, while taking into account the PLR of the underlying network and the encoded bit rate. The proposed method can be used to efficiently generate error resilient H.264/AVC bitstreams that are perfectly adapted to the channel error characteristics. This is extremely important because it can mean that error resilient video transmission will be possible in environments with varying error characteristics with an improved quality, notably, when compared to the case where the MB Intra coding decisions are taken without considering the error characteristics of the network.

6. REFERENCES

[1] ISO/IEC 14496-10, "Information Technology – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding", 2005.

[2] ISO/IEC 14496-2, "Information Technology – Coding of Audio-Visual Objects – Part 2: Visual (2nd Ed.)", 2001.

[3] P. Haskell, D. Messerschmitt, "Resynchronization of Motion Compensated Video Affected by ATM Cell Loss," *Proc. ICASSP*, San Francisco, CA, Vol. 3, pp. 545-548, Mar. 1992.

[4] J. Y. Liao, J. D. Villasenor, "Adaptive Intra Block Update for Robust Transmission of H.263," *IEEE Trans. on Circ. and Sys. for Video Tech.*, Vol. 10, No. 1, pp. 30-35, Feb. 2000.

[5] P. Nunes, L. D. Soares, F. Pereira, "Error Resilient Macroblock Rate Control for H.264/AVC Video Coding", *Proc. ICIP*, San Diego, CA, USA, Oct. 2008.

[6] Z. Li, F. Pan, K. Lim, G. Feng, X. Lin, S. Rahardaj, "Adaptive Basic Unit Layer Rate Control for JVT", Doc. JVT-G012, 7th MPEG Meeting, Pattaya, Thailand, Mar. 2003.

[7] H.264/AVC Reference Software, JM 13.2, Available: http://iphome.hhi.de/suehring/tml/download/.

[8] L. D. Soares, P. Nunes, F. Pereira, "Efficient Network-Aware Macroblock Mode Decision for Error Resilient H.264/AVC Video Coding", *Proc. SPIE, Vol. 7073*, San Diego, CA, USA, Aug. 2008.

[9] T Wiegand, H. Schwarz, A. Joch, F. Kossentini, G. Sullivan, "Rate-Constrained Coder Control and Comparison of Video Coding Standards", *IEEE Trans. on Circ. and Sys. for Video Tech.*, Vol. 13, No. 7, pp. 688-703, Jul. 2003.

[10] S. Wenger "H.264/AVC Over IP", *IEEE Trans. on Circ. and Sys. for Video Tech.*, Vol. 13, No. 7, pp. 645-656, Jul. 2003.

[11] K.-P. Lim, G. Sullivan, T. Wiegand, "Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods", Doc. JVT-X101, ITU-T VCEG Geneva Meeting, Geneva, Switzerland, Jun. 2007.