Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

# A network-aware error-resilient method using prioritized intra refresh for wireless video communications

Han-jie MA[†], Fan ZHOU[†‡], Rong-xin JIANG, Yao-wu CHEN
(*Institute of Advanced Digital Technology and Instrumentation, Zhejiang University, Hangzhou 310027, China*)
[†]E-mail: hanjie.ma@gmail.com; fanzhou@mail.bme.zju.edu.cn

Received Sept. 13, 2008; Revision accepted Feb. 23, 2009; Crosschecked June 10, 2009

**Abstract:** We propose a novel prioritized intra refresh method for the wireless video communication. The proposed method considers the characteristics of the human visual system, the error-sensitivity of the bitstream, and the state of the time-varying wireless channel jointly. An expected perceptual distortion model was used to adjust the intra refresh rate adaptively. This model consists of the perceptual weight map based on an attention model, the bit error probability map based on bitstream size, and the dynamic channel state information (CSI). Experimental results indicate that, compared with other intra refresh methods that consider only the content of the video or the CSI, the proposed method improves the average peak signal-to-noise ratio (PSNR) of the whole frame by about 0.5 dB, and improves the average PSNR of the attention-area by about 0.8 dB.

**Key words:** Intra refresh, Distortion, Perceptual weight, Error-sensitivity, Channel state information (CSI)
**doi:**10.1631/jzus.A0820660 **Document code:** A **CLC number:** TN919.8

## INTRODUCTION

The robust real-time video application over the wireless channel is still a challenging problem due to the time-varying, bit errors and packet erasure characteristics of the wireless links. To solve this problem, many error-resilient video coding mechanisms have been proposed, such as layered coding with transport prioritization (LCTP), multiple-description coding (MDC), joint source and channel coding (JSCC), forward error concealment (FEC), automatic repeat quest (ARQ), and data partition (Wang *et al*., 2000).

The intra refresh method is a fairly efficient error-resilient video coding mechanism to improve the quality of the video transmission over the wireless channels. This method requires no modification to the bitstream syntax and almost no extra computation. Currently, the popular intra refresh algorithms include random refresh, periodical refresh, and motion information based refresh (Haskell and Messerschmitt, 1992; Cote and Kossentini, 1999; Shu and Chau, 2008). These algorithms perform well in the error resilience, but not well in the coding efficiency. Hence, how to achieve the trade-off between the coding efficiency and the protection of error propagations becomes a challenge. Several strategies are utilized to solve this problem. In (Liao and Villasenor, 2000), an adaptive intra refresh algorithm based on an error-sensitivity metric accumulated at the encoder was proposed. Frossard and Verscheure (2001) improved this method using the adaptive MPEG-2 information structuring (AMIS) mechanism, especially for MPEG-2 video streams. The rate-distortion mode was also used to improve the performance of the intra refresh method. Cote *et al*.(2000) adopted the end-to-end rate-distortion model (E2ERD) to optimize the video coding mode selection. He *et al*.(2002) derived an analytic solution for adaptive intra refresh considering the source and channel distortion jointly. A rate distortion model based intra refresh method with the potential distortion tracking was proposed in (Zhang *et al*., 2004). In (Chen *et al*.,

‡ Corresponding author

2007), an attention-based adaptive intra refresh method for the error-prone video transmission was proposed, which is based on the E2ERD model using the subjective human vision property. This method takes into account several human vision factors, such as intensity, color, and orientation, to obtain a better subjective quality. The E2ERD model based intra refresh depends on the feedback information from the decoder end, which is not suitable for real-time and limited-bandwidth applications.

In this paper, a network-aware prioritized intra refresh method based on expected perceptual distortion for wireless video communications is proposed. The subjective human vision properties, the error-sensitivity of the bitstream, and the characteristics of the wireless channel are jointly considered to calculate the expected perceptual distortion. Then the intra refresh rate is adaptively adjusted according to each macroblock's expected perceptual distortion. The perceptual content in the video input is weighted by an attention model. A bit error probability map is calculated as the metric of the error-sensitivity of the bitstream. The channel state information (CSI) such as the bit error ratio (BER) or packet lost ratio (PLR) is also taken into consideration. As a result, the proposed intra refresh method obtains a satisfying performance in both the subjective and the objective video qualities.

## PROPOSED NETWORK-AWARE PRIORITIZED INTRA REFRESH METHOD

### System model of the proposed intra refresh method

Frossard and Verscheure (2001) used the mean luminance difference (MLD) to measure the distortion corresponding to the simplest metric correlated with human perception. The MLD of macroblock $B(i, j, n)$ is defined as follows:

$$D_{\text{mb}}(i, j, n) = \frac{1}{256}\left|B_{\text{recon}}(i, j, n) - B_{\text{dam}}(i, j, n)\right|, \quad (1)$$

where $(i, j, n)$ denotes the macroblock position in the video frame $n$, $B_{\text{recon}}(i, j, n)$ denotes the correctly reconstructed macroblock at the encoder, and $B_{\text{dam}}(i, j, n)$ denotes the corresponding damaged macroblock. Assuming that the frame-copy error-concealment technique is used at the decoder side, the corresponding damaged macroblock can be approximately replaced by the correctly reconstructed macroblock with the same position in the previous frame. Thus, the MLD of macroblock $B(i, j, n)$ can be modified as follows:

$$D_{\text{mb}}(i, j, n) = \frac{1}{256}\left|B_{\text{recon}}(i, j, n) - B_{\text{recon}}(i, j, n-1)\right|. \quad (2)$$

The distortion due to error propagation can be weighted by the corresponding loss probability matrix (Frossard and Verscheure, 2001). Hence, the expected distortion can be given by

$$D_{\text{e\_mb}}(i, j, n) = D_{\text{mb}}(i, j, n)P_{\text{mb}}(i, j, n), \quad (3)$$

where $P_{\text{mb}}(i, j, n)$ represents the probability for macroblock $B(i, j, n)$ suffering from bit errors.

However, video distortion in the above model does not correlate well with the human perception of video quality. Human usually pays different attentions to different regions depending on the perceptual acuteness of each region. Therefore, attention based perceptual weight should be added to improve the perceptual distortion metrics. Let $W_{p\_\text{mb}}(i, j, n)$ denote the perceptual attention weight of macroblock $B(i, j, n)$. Then the expected perceptual distortion is defined as follows:

$$D_{\text{e}p\_\text{mb}}(i, j, n) = W_{p\_\text{mb}}(i, j, n)D_{\text{mb}}(i, j, n)P_{\text{mb}}(i, j, n). \quad (4)$$

Macroblocks should be forced intra-mode encoded when the accumulation of the expected perceptual distortion at position $(i, j)$ reaches a given threshold $\Delta_n$ (Frossard and Verscheure, 2001). Then the condition for a macroblock $B(i, j, n)$ to be intra-coded in frame can be simplified by

$$\sum_{m=n_0}^{n} D_{\text{ep\_mb}}(i, j, m) \ge \Delta_n, \quad (5)$$

where $n_0$ denotes the last intra-coded video frame number.

Fig.1 shows the proposed framework of the video encoder using the expected perceptual distortion model in a typical wireless video communication system. It can be seen that, the perceptual weight map is calculated from the content of the current input video frame, the bit error probability map is computed from the bitstreams of the encoder output, and the CSI is obtained from the transceiver at the physical layer of the wireless channel. These factors are combined to calculate the expected perceptual distortion of the current input video frame. Then the video encoder adaptively chooses forced intra-mode coding for each macroblock according to the predicted perceptual distortion to improve video quality.

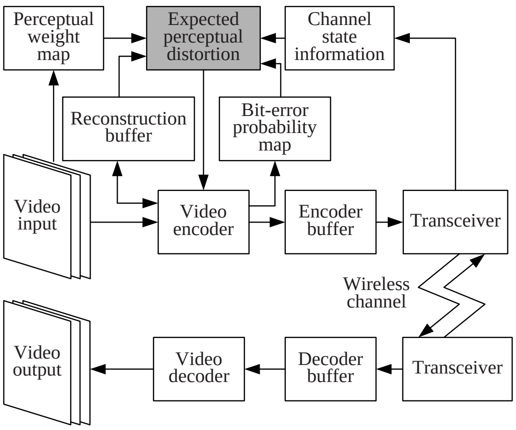



**Fig.1 Overview of the video encoder using the expected perceptual distortion model in a typical wireless video communication system**

## Perceptual weight map based on attention model

Our perceptual weight map is calculated using a perceptual attention model that considers the three factors simultaneously: motion, skin tone, and fovea. These factors are fused with different weights to compute the final attention weight map.

Motion is an important focus in video communications and the motion areas always receive more attention. Thus, it should have high perceptual weights in the perceptual weight map. Spatial and temporal mean absolute difference (MAD) is calculated based on macroblocks to evaluate the motion information of the current frame (Liu *et al*., 2007). The motion weight of macroblocks is adaptively changed according to the global motion of the current frame because more attention should be paid to fast moving regions.

The presence of human is another important focus in video communications. Human face and hands are the most attractive parts of the human content. Thus, the skin regions should be paid more attention to and have a high perceptual weight. Since the skin areas are sometimes still and cannot be detected using only the motion factor, the skin tone model proposed in (Hsu *et al*., 2002) has been adopted using macroblocks. The macroblocks based algorithm is more effective because the high accuracy of pixels in face detection is not required in our attention model.

Fovea also affects human perceptual attention (Lee *et al*., 2002) and people usually focus on the center region of the video scene. Hence, the center region should have a higher weight than other regions in the video scene. Motion activities and human presence in the peripheral region are captured by the higher weight placed upon the motion and skin tone model.

Let $W_{p\_\mathrm{mb}}(i, j, n)$ denote the perceptual weight of macroblock $B(i, j, n)$, $p_\mathrm{s}$ the value of the skin factor, $p_\mathrm{m}$ the value of the motion factor, $p_\mathrm{fov}$ the value of the fovea effect factor, and $w_\mathrm{s}$, $w_\mathrm{m}$, $w_\mathrm{fov}$ their weights respectively. Then we have

$$W_{p\_\mathrm{mb}}(i,j,n) = w_\mathrm{s} p_\mathrm{s}(i,j,n) + w_\mathrm{m} p_\mathrm{m}(i,j,n) + w_\mathrm{fov} p_\mathrm{fov}(i,j,n). \quad (6)$$

Fig.2 shows the perceptual weight maps of the 54th frame of the Foreman sequence (QCIF) and the

(a)

(b)

**Fig.2 Perceptual weight maps of the test sequences**
(a) Foreman; (b) Paris

73rd frame of the Paris sequence (CIF) provided by our attention model. The Foreman sequence is a 'centered' sequence; the perceptual attention model performs well. The perceptual weights placed upon the center regions of chair and table are a deficiency of the Paris sequence with our model.

**Bitstream size based bit error probability map**

Based on the variable-length codes (VLCs), when one macroblock is corrupted, the successive macroblocks will be dropped until the next resynchronization point. Assuming that the whole frame is coded into one slice, each slice has a constant bit size by using a constant bit-rate control mechanism at the encoder side. Then each slice has the same bitstream size and the decoder can be correctly synchronized at the start of each slice when bit errors occur. Let $S(i, j, n)$ denote the bitstream size of macroblock $B(i, j, n)$, $L_{head}$ the bitstream length of the slice head, $L(i, j, n)$ the bitstream length between the start of the slice and macroblock $B(i, j, n)$, and then we have

$$L(i,j,n) = L_{head} + \sum_{x=0}^{i}\sum_{y=0}^{j} S(x,y,n). \quad (7)$$

If one or more errors occur in this slice before the macroblock $B(i, j, n)$, the probability of this condition can be defined as follows (Liao and Villasenor, 2000):

$$P(i-1,j,n) = 1-(1-p_{BER})^{L(i-1,j,n)}, \quad (8)$$

where $p_{BER}$ denotes the BER of the wireless channel. Set $B(i_e, j_e, n)$ to be the last macroblock of the slice, and the distribution of bit errors for the macroblock is independent of each other. Then the probability that bit errors occur at macroblock $B(i, j, n)$ can be given by

$$\begin{aligned} P_{mb}(i,j,n) &= \frac{S(i,j,n)}{L(i_e,j_e,n)} P(i_e,j_e,n) \\ &= \frac{S(i,j,n)}{L(i_e,j_e,n)} [1-(1-p_{BER})^{L(i_e,j_e,n)}]. \end{aligned} \quad (9)$$

From Eq.(9), it can be seen that the bit error probability of macroblock $B(i, j, n)$ can be described by the bitstream size of the macroblock and the BER of the channel. Fig.3 shows the macroblock bit error probability maps of the 54th frame of the Foreman sequence (QCIF) and the 73rd frame of the Paris sequence (CIF).

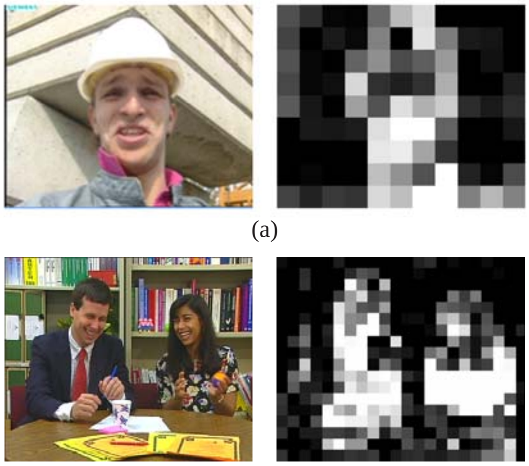


**Fig.3 Bit error probability maps of the test sequences**
(a) Foreman; (b) Paris

**CSI for wireless video communications**

Eq.(9) shows that the bit error probability of macroblocks also depends on the CSI of the wireless channel, such as BER. The CSI can be obtained from the physical layer or dynamically measured through the real-time control protocol (RTCP) while bringing no delays into the transmission.

Wireless channels typically suffer from multipath and shadowing effects that lead to loss or erasure of bits and packets. Assuming the wireless channel is a Rayleigh fading channel, we can estimate it using the two-state Gilbert-Elliot's Markov model, as illustrated in Fig.4 (Gilbert, 1960): it is assumed that the channel has a good state and a bad state and the transition of the corresponding state is $\boldsymbol{H} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$, where $p$ represents the probability of good state changing to bad state, and $q$ represents the probability of bad state changing to good state. This model was used in our experiments to simulate the burst bit errors and packet loss in wireless video communications.

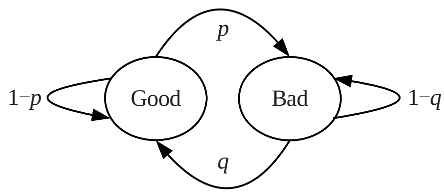



**Fig.4 Two-state Gilbert model for wireless channel state**

## SIMULATION AND DISCUSSION

### Simulation setup

Some experiments have been carried out to verify the performance of the proposed intra refresh method. The Foreman (QCIF, 176×144) and Paris (CIF, 352×288) test sequences were utilized and encoded using H.264 reference software JM13.2. The simulation was run by using 5000 coded pictures with I frames every 100 pictures; the pictures were encoded with 9 slices per I frame, 3 slices per P frame, and no B frames. For the Foreman (QCIF) sequence, 64, 128, 192 and 256 kbps target bitrates were used. For the Paris (CIF) sequence, 256, 512, 768 and 1024 kbps target bit-rates were used. At the decoder side, the error concealment based on frame copy was used. It can be enabled in the configuration file of JM13.2.

To simulate the burst bit errors in wireless channels, a two-state Markov chain was used with the state space $\boldsymbol{H}=\begin{bmatrix}0.1 & 0.9\\ 0.9 & 0.1\end{bmatrix}$, $p_{\mathrm{BER}}=10^{-3}$ and $10^{-2}$ dB (Yajnik *et al*., 1999). When the video data were transported over wireless channels, the bitstreams of one slice were packed into one packet and all frames were subjected to bit errors.

### Simulation results

Simulations were carried out based on four different intra refresh methods as follows:

(1) AIR: Adaptive Intra Refresh method suggested in MPEG-4 (International Organization for Standardization, 2001).

(2) PIR: Perceptual weight map based Intra Refresh method.

(3) BIR: Bit error probability map based Intra Refresh method.

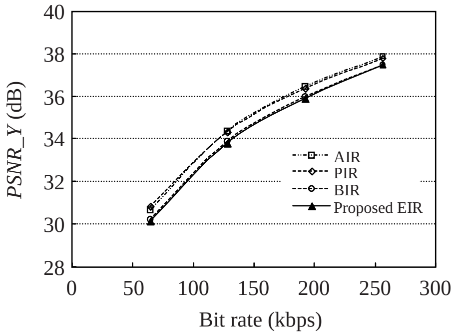
(4) EIR: Our proposed network-aware prioritized Intra Refresh method based on Expected perceptual distortion.

Fig.5 shows the four methods' subjective results of the 81st frame of the Foreman (QCIF) sequence and the 571st frame of the Paris (CIF) sequence. It can be observed that the subjective video quality of the proposed intra refresh method is better than the other three methods' subjective video quality.

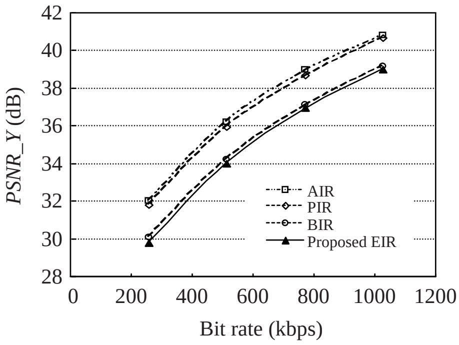
The peak signal-to-noise ratio (PSNR) of luminance components was further used for the performance comparison of the objective video quality. Fig.6 shows the average PSNR of the whole frame area of the Foreman (QCIF) and Paris (CIF) sequences encoded with four intra refresh methods when the video data were transported over a no-error channel. It is observed that the PSNR performances of the AIR and PIR methods are better than those of the BIR and proposed EIR methods. The reason is that the AIR and PIR methods do not consider the weight of the channel state and the macroblocks are less frequently forced intra-mode coded than the BIR and EIR methods. As a result, the coding efficiencies of the AIR and PIR methods are higher and the video qualities are also better than those of the BIR and EIR methods when there is no bit error affecting the transported video data.

(a) (b) (c) (d)

**Fig.5 Subjective results of the Foreman (QCIF) and Paris (CIF) sequences of the four methods**
(a) AIR; (b) PIR; (c) BIR; (d) Proposed EIR

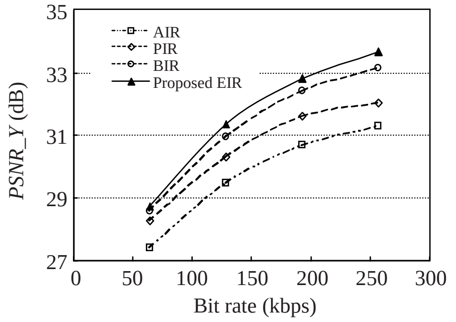



(a)

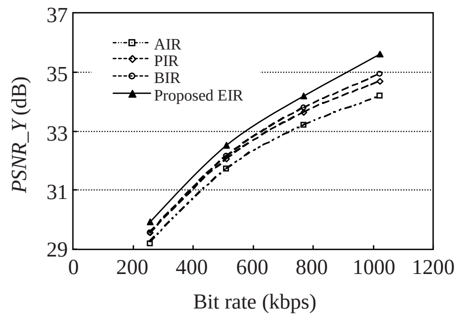



(b)

**Fig.6 Frame average *PSNR* of different intra refresh methods over no-error channels. (a) Foreman; (b) Paris**




(a)



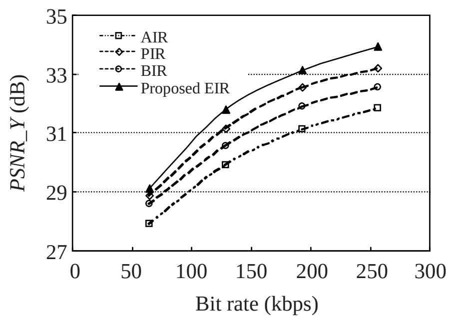

(b)

**Fig.7 Frame average *PSNR* of different intra refresh methods over wireless channels. (a) Foreman; (b) Paris**

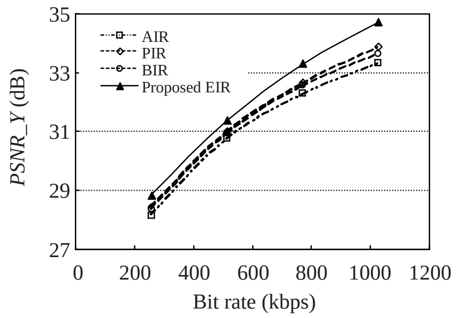
In Figs.7 and 8, the whole frame and the attention area average PSNR performances of the four methods for the Foreman (QCIF) and Paris (CIF) sequences are shown when the video data are transported over wireless channels with bit errors.

It can be seen that the performances of the AIR method are the worst in both the whole frame and the attention areas. The reason is that this method does not take the perceptual weight of the video content and the channel state into account. It uses the same intra refresh rate for all the regions of the frame and ignores the changes of the channel state. Therefore, in the AIR method, the macroblocks are forced intramode coded in a fixed order, which cannot stop the error propagation effectively. Thus, the effect of error concealment degrades and the video quality decreases.

It can be seen that the BIR method performs better in the whole frame than the PIR method does (Fig.7). It is also observed that the PIR method performs better in the attention area than the BIR method does (Fig.8). This is because the BIR method considers the bit error probability of the whole frame and the PIR emphasizes the priority of the attention area.




(a)




(b)

**Fig.8 Attention-area average *PSNR* of different intra refresh methods over wireless channels. (a) Foreman; (b) Paris**

Figs.7 and 8 also show that, compared with other three methods, the proposed EIR method has an obvious improvement in the PSNR performance when the video data are transported over wireless channels with bit errors. The improvement is especially significant in the attention areas. In the proposed EIR method, macroblocks are adaptively forced intra-mode coded according to the CSI. As a result, the intra refresh rate is changed as the channel state changes. What is more, the presented method also takes the perceptual weight of the video content into account, so it performs more effectively in the attention areas. It also can be seen that the improvement of the Paris (CIF) sequence is not as significant as that of the Foreman (QCIF) sequence. The reason is that the Foreman (QCIF) sequence is very intra-heavy to begin with, and the Paris (CIF) sequence has a large still region which is non-attention area.

## CONCLUSION

In this paper, we present a prioritized intra refresh method based on expected perceptual distortion for the wireless video communication. The proposed expected perceptual distortion model considers both the content of the video frame with the perceptual weight based attention model and the bitstream size based bit error probability map. The channel state information is also taken into account. Experimental results indicate that the proposed EIR method can achieve better performance in both the subjective and the objective video quality.

The future research may focus on (1) improving the algorithm of the perceptual weight map for other video applications and optimizing it to get a low computation cost, and (2) studying other error-resilient mechanisms such as ARQ and other FEC methods based on the proposed expected perceptual distortion model.

## ACKNOWLEDGEMENTS

The authors would like to thank Albert Wang and the anonymous reviewers for their thoughtful comments and suggestions.

## References

Chen, Q.Q., Chen, Z.B., Gu, X.D., Wang, C., 2007. Attention-based adaptive intra refresh for error-prone video transmission. *IEEE Commun. Mag.*, **45**(1):52-60. [doi:10.1109/MCOM.2007.284538]

Cote, G., Kossentini, F., 1999. Optimal intra coding of blocks for robust video communication over the Internet. *Signal Process.: Image Commun.*, **15**(1-2):25-34. [doi:10.1016/S0923-5965(99)00022-3]

Cote, G., Shirani, S., Kossentini, F., 2000. Optimal mode selection and synchronization for robust video communications over error-prone networks. *IEEE J. Select. Areas Commun.*, **18**(6):952-965. [doi:10.1109/49.848249]

Frossard, P., Verscheure, O., 2001. AMISP: a complete content-based MPEG-2 error-resilient scheme. *IEEE Trans. Circuits Syst. Video Technol.*, **11**(9):989-998. [doi:10.1109/76.946516]

Gilbert, E.N., 1960. Capacity of a burst-noise channel. *Bell Syst. Techn. J.*, **39**:1253-1265.

Haskell, P., Messerschmitt, D., 1992. Resynchronization of Motion Compensated Video Affected by ATM Cell Loss. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p.545-548. [doi:10.1109/ICASSP.1992.226155]

He, Z.H., Cai, J.F., Chen, C.W., 2002. Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding. *IEEE Trans. Circuits Syst. Video Technol.*, **12**(6):61-82. [doi:10.1109/TCSVT.2002.800313]

Hsu, R.L., Abdel-Mottaleb, M., Jain, A.K., 2002. Face detection in color images. *IEEE Trans. Pattern Anal. Mach. Intell.*, **24**(5):696-706. [doi:10.1109/34.1000242]

International Organization for Standardization, 2001. Information Technology—Coding of Audio-visual Objects—Part 2: Visual. ISO/IEC 14496-2:2001, International Organization for Standardization, Geneva.

Lee, S., Pattichis, M.S., Bovik, A.C., 2002. Foveated video quality assessment. *IEEE Trans. Multim.*, **4**(1):129-132. [doi:10.1109/6046.985561]

Liao, J.Y., Villasenor, J., 2000. Adaptive intra block update for robust transmission of H.263. *IEEE Trans. Circuits Syst. Video Technol.*, **10**(1):30-35. [doi:10.1109/76.825855]

Liu, Y., Li, Z.G., Soh, Y.C., 2007. A novel rate control scheme for low delay video communication of H.264/AVC standard. *IEEE Trans. Circuits Syst. Video Technol.*, **17**(1):68-78. [doi:10.1109/TCSVT. 2006.887081]

Shu, H.Y., Chau, L.P., 2008. Intra/inter macroblock mode decision for error-resilient transcoding. *IEEE Trans. Multim.*, **10**(1):97-104. [doi:10.1109/TMM.2007.911300]

Wang, Y., Wenger, S., Wen, J.T., Katsaggelos, A.K., 2000. Error resilient video coding techniques. *IEEE Signal Process. Mag.*, **17**(4):61-82. [doi:10.1109/79.855913]

Yajnik, M., Moon, S., Kurose, J., Towsley, D., 1999. Measurement and Modeling of the Temporal Dependence in Packet Loss. 18th Annual Joint Conf. of the IEEE Computer and Communications Societies, p.345-352. [doi:10.1109/INFCOM.1999.749301]

Zhang, Y., Gao, W., Sun, H.F., Huang, Q.M., Lu, Y., 2004. Error Resilience Video Coding in H.264 Encoder with Potential Distortion Tracking. IEEE Int. Conf. on Image Processing, p.163-166. [doi:10.1109/ICIP.2004.1418715]