Error Resilience Analysis of Multi-Hypothesis Motion Compensated Prediction for Video Coding

1. Research Team

Project Leader:	C.-C. Jay Kuo, *Electrical Engineering*
Other Faculty:	Prof. Chris Kyriakakis, *Electrical Engineering* Prof. Shrikanth Narayanan, *Electrical Engineering* Prof. Antonio Ortega, *Electrical Engineering*
Post Doc(s):	Dr. Qing Li, Dr. Meiyin Shen, Dr. Chia-Hung Yeh
Graduate Students:	Min-Kuan Chang, Yu Hu, Wei-Ying Kung, Chi-Hung Kuo, KiTae Nahm, Jingliang Peng, Xiaoming Sun, Xiaosong Zhou
Industrial Partner(s):	Conexant, Skyworks, EverFocus

2. Statement of Project Goals

The goal of our research is to study various error resilient coding techniques at both the encoder and the decoder ends, and find good approaches to achieve robust video transmission under practical constraints.

3. Project Role in Support of IMSC Strategic Plan

Techniques of continuous media streaming play an important role in IMSC strategic plan, which aims at bringing the immersive experience to users via wired or wireless channels. It is unavoidable to have packet loss during the transmission process so that error resilient coding and error concealment techniques have to be integrated with audio/video compression schemes. Our work is carried out jointly with quite a few other faculty members in IMSC due to the multi-disciplinary nature of the proposed research.

4. Discussion of Methodology Used

In a multi-hypothesis motion compensated prediction (MHMCP) coder, a macroblock (MB) can contain more than one motion vectors (MVs). Each MV specifies a reference MB, which is called a hypothesis. Then, the target MB is estimated by linearly combining all hypotheses. By carefully selecting the hypotheses and their weighting coefficients, a coding gain can be achieved. Although MHMCP was originally proposed to improve coding efficiency, it has been observed recently that it can also enhance the error resilient property of compressed video.

5. Short Description of Achievements in Previous Years

We studied next generation techniques for coding and delivery of multimedia data over the Internet and wireless networks. In particular, we studied the rate-distortion performance of MHMCP in error prone environment and then performed extensive experiments to confirm the analytical results. Several design guidelines for the MHMCP coder are drawn based on the analytical and experimental results. In addition, we have proposed an adaptation approach to further improve the performance at the expense of a feedback channel. Moreover, we have demonstrated its strength on the error resilient capability over the random intra refreshing method.

5a. Detail of Accomplishments During the Past Year

<u>5.1 Problem Formulation</u>

Suppose that a frame is corrupted during transmission. Let us consider the effect of transmission errors in an MHMCP coder. The l^{th} frame after the erroneous one contains the propagation error, which can be written as

$$\varepsilon_l = \sum_{i=1}^{n} h_i \varepsilon_{l-i} \qquad (1)$$

where n is the number of hypotheses and h_i is the i^{th} hypothesis coefficient. This relation can be repeatedly applied to measure the error propagation effect quantitatively. It is obvious that the hypothesis number n and hypothesis coefficients h_i 's ($i=1,\ldots,n$) determine the amount of the propagating error. We will discuss the effect of each parameter while fixing the other.

<u>5.2 Effect of Hypothesis Number</u>

In an error-prone environment, it is desirable to employ MHMCP rather than single hypothesis MCP to reduce propagation errors. However, MHMCP needs more bits to represent the additional motion information. To get a better understanding of the trade-off between error resilience and coding efficiency, we discuss these two issues in section 5.2.1 and 5.2.2, respectively. Then, the rate-distortion property will be discussed in section 5.2.3. Finally, in section 5.2.4, we will compare the error propagation effect on random intra-MB refreshing (IR).

5.2.1 Impact on Propagation Error

Let us consider the occurrence of a single burst error. From Equation (1), we have

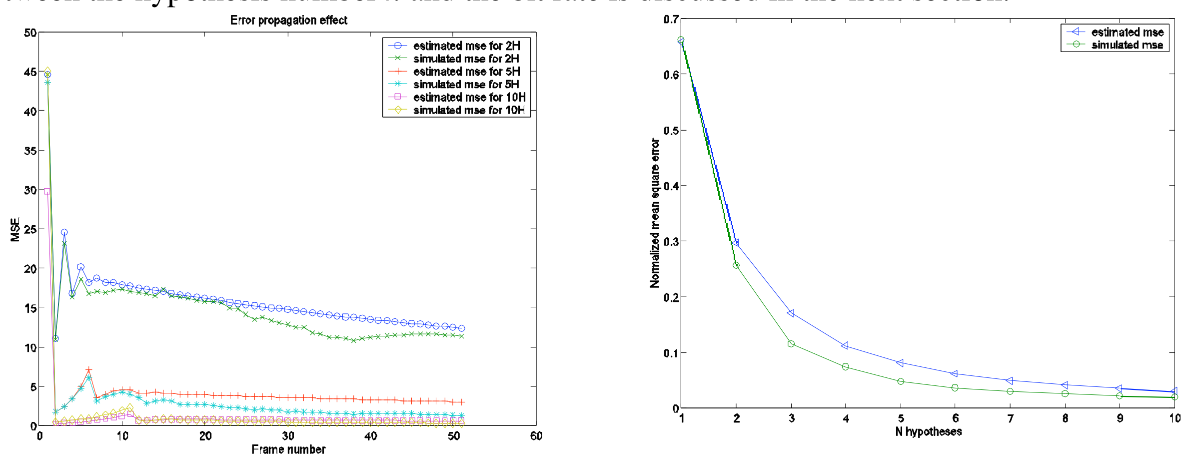
$$\varepsilon_l = \begin{cases} \frac{1}{n}\left(1+\frac{1}{n}\right)^{l-1}\varepsilon_0 \\ \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{l-i} \end{cases}, \qquad (2)$$

where ε_0 denotes the initial concealed error. Transmission errors tend to attenuate as they propagate as a result of low pass filtering operations in the prediction loop such as half-pixel motion compensation. By adopting the loop filter model, we can approximate the mean square error (MSE) of the l-th frame after the corrupted frame by MSE_l = $\_{l}^{2}/(1+\_l)$ [1], where _ is a parameter that describes the effectiveness of the loop filter to attenuate errors.

The H.264 reference codec of version JM6.1e has been modified to provide the MHMCP functionality. We have tested the effect of a burst error with three hypothesis numbers (n=2,5,10), respectively. Both estimated and simulated MSEs are shown in Figure1(a). Note that the estimated MSEs match the simulated MSEs well. Two observations can be made. First, the propagation error decreases as the hypothesis number increases. Second, the error converges to a non-zero value, i.e. 2/(n+1).

To further investigate the relationship between MSE and the hypothesis number n, Figure 1 (b) depicts the average MSEs after a burst error, where the averaging range is 100 frames in this simulation. In this figure, the average MSEs are normalized with respect to the initial MSE ($\_{0}$) of the concealed frame. We see that the estimated result and the simulated result follow a similar trend, although the estimated MSEs are slightly higher than the simulated ones.

As more hypotheses are used, the error propagation effect is more effectively suppressed. However, a large n requires more bits to encode additional motion vectors. The relationship between the hypothesis number n and the bit rate is discussed in the next section.



(a) MSE as a function of the frame number (b) Normalized MSE v.s. hypothesis number

Figure 1 The impact of different hypothesis numbers on the propagation error.

5.2.2 Impact on Bit Rates

The hypothesis number n affects the bit rate in two aspects. On one hand, the use of more hypotheses leads to a smaller prediction error, thus saving bits for residual coding. On the other hand, a larger number of hypotheses need additional bits for reference frame indexing and the coding of motion vectors. In general, the bit rate for the motion information is linearly proportional to n.

Assume that $\_{l-i}$ in Equation.1 is a Gaussian random variable with variance $\_{i}^{2}$. Then, the error variance of $\_{l}$ can be expressed as

$$\square^2_{MH} = \sum_{i=1}^{n} (h_i^2 \square^2_i + h_i \sum_{j=1, j \square i}^{n} \square_{i,j} h_j \square_i \square_j), \quad (3)$$

where $\_{i,j}$ is a correlation coefficient between the i-th and the j-th hypotheses. It is difficult to derive a closed-form formula for Equation.3. However, we can simplify the problem using a reasonable assumption. That is, the error variance of each hypothesis is the same and the correlation coefficients are the same. In terms of mathematics, we have $\_^{2}_{i=}\ \_^{2}$ and $\_{i,j} = \_$ for all i and j. Then, it can be shown that the error variance $\_^{2}_{MH}$ in Equ.3 has the minimum value

$$\square^{2}{}_{\min} = \frac{1+\square(n\square 1)}{n}.$$

We have the following observations. First, $\_^{2}_{min}$ approaches $\_\_^{2}$ as n goes to infinity, and does not vanish if _ is not zero. Second, it is desirable to choose hypotheses with low correlations to reduce the error variance.

The bit rate reduction due to the smaller prediction error can be written as

$$\frac{1}{2}\log_2\frac{\square^{2}{}_{\min}}{\square^{2}} = \frac{1}{2}\log_2\frac{1+\square(n\square 1)}{n},$$

in the unit of bits per pixel as compared to the single hypothesis (SH) case. Then, the overall change of the bit rate can be expressed as the sum of the reduced bit rate due to the smaller prediction error and the increased bit rate due to additional motion information, given by

$$\square R = \frac{1}{2}\log_2\frac{1+\square(n\square 1)}{n} + \frac{nc}{256}, \qquad (4)$$

where c is a constant to represent the average number of bits to represent a motion vector number. Simulation results given in Figure 2 confirm the fact that this formula approximates the experimental data well.

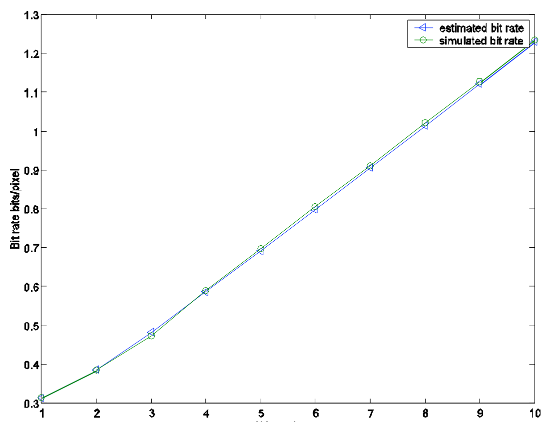


Figure 2 The plot of the bit rate (bits per pixel) as a function of the hypothesis number.

5.2.3 Rate-distortion Analysis

The optimal value of n is dependent on the correlation coefficients between hypotheses, since smaller _ leads to better coding efficiency with the same distortion. In the real-world situation with a relative high value of _, two hypotheses are sufficient for a good rate-distortion performance. However, a larger value of n is helpful in suppressing propagation errors. Here, we should select the best value of n by considering both coding efficiency and error resilience.

As mentioned before, we have modified the H.264 codec to include the MH option. We employ the Lagrangian method for greedy rate-distortion optimization. Specifically, given the previously encoded MBs, we minimize the Lagrangian cost of a current MB, given by

$$J = D_s + D_c + \lambda(R + \Delta R),$$

where $\lambda = 0.85Q^2$ [2], D_s is the source distortion caused by the quantization of prediction residuals, D_c is the channel distortion caused by the concealment of erroneous MBs and the accumulated propagation from previous reference frames, R is the bit rate for the original SH, and ΔR is the increased bit rate due to the MH option.

The value of D_s can be approximated as

$$D_s \approx \frac{1 + \square(n - 1)}{n}\overline{D_s},$$

where $\overline{D_s}$ is the experimental distortion for SH. Also, let p_l denote the frame loss rate of the previous l-th frame. Then, D_c can be expressed as

$$D_c = \sum_{i=0}^{n} p_i MSE_i$$

where denotes the mean square error of the i-th MB. Note also that ΔR is given in Equation. (4).

Then, we can easily obtain the best R-D pair if we have the R-D pair for SH, which has been extensively studied and standardized. Figure 3 plots estimated and simulated R-D curves after a burst error. The hypothesis number n varies from 1 to 10, and each R-D point corresponds to one particular n. The estimation results are close to simulation results, which indicate that the derived formulas can be effectively used for the error-resilient R-D optimized coder.

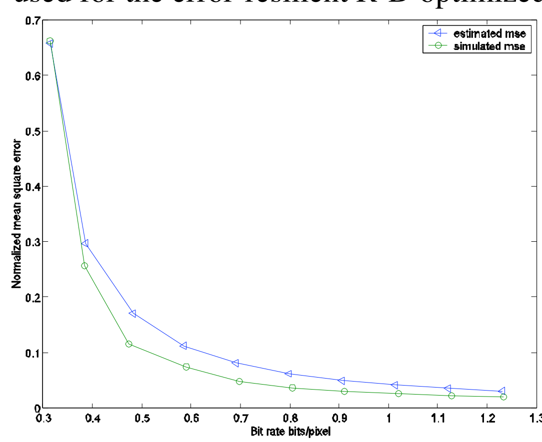


Figure 3 The normalized distortion after a burst error.

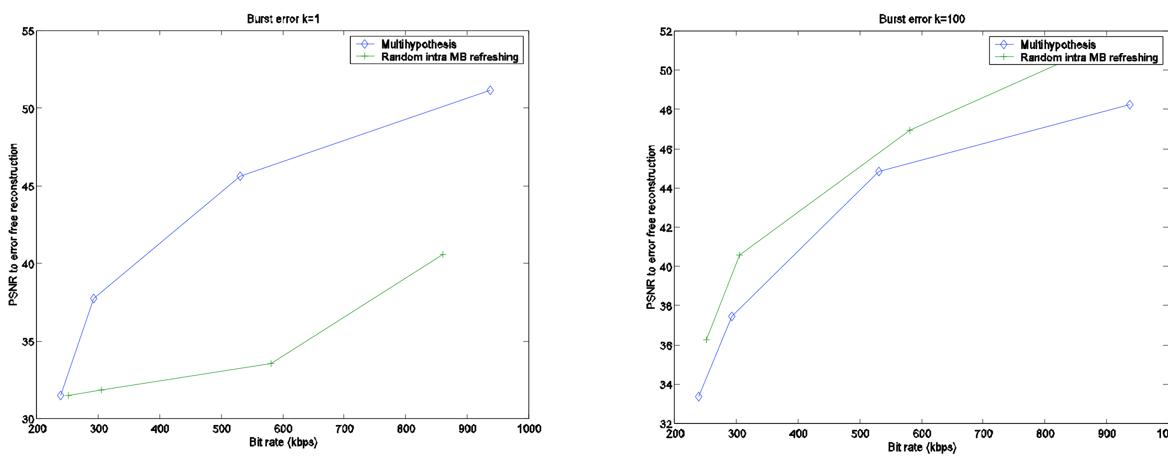
5.2.4 Comparison with Random intra-MB Refreshing

Random intra-MB refreshing (IR) [3] is a well-known error resilient tool used in the video encoder to stop error propagation. In this section, we compare its error resilient capability with that of MHMCP, which leads to some guidelines for the error-resilient MHMCP design.

In IR, the amount of propagation error can be controlled by varying the refreshing rate $0 < p < 1$, which means the p percent of MBs in a frame should be intra-coded. Unlike $\square_l$ in MHMCP, which converges to a non-zero value, $\_l$ in IR converges to 0. Therefore, in the long run, IR is better than MHMCP since IR can stop the error entirely. However, MHMCP suppresses error propagation more rapidly.

In Figure 4, we compare MHMCP and IR in terms of the PSNR of the reconstructed video and the bit rates. Various hypothesis number 1,2,5,10 with the corresponding long term memory buffer size 1,2,5,10 for MHMCP were evaluated. For comparison, IR with refreshing rates 2/99,10/99, 50/99 and 90/99 were also simulated. To achieve fair comparison, other encoding parameters such as quantization parameter are set to be the same in both cases. Every marked PSNR-BitRate point in Figure 4 was obtained by averaging 30 runs of a random burst error.

The PSNR value in Figure 4 (a) is averaged in the short term (i.e. only the first frame after an erroneous one) while the PSNR value in Figure4 (b) is averaged in the long term (consisting of 100 frames after an erroneous frame). We see that the *n*-hypothesis scheme can achieve the same or better performance than the random intra-MB refreshing scheme, if I frames are inserted periodically, where the proper length of the period can be determined by the hypothesis number *n* and the refreshing rate *p*.



(a) Short term error suppression effect (b) Long term error suppression effect

Figure 4. Error suppression comparison between MHMCP (with various n values) and IR after a burst error.

In Figure 5, we compare the rate-distortion performances of MHMCP and IR in a high packet loss rate environment, where the packet loss rate is set to 10%. The encoding parameters are set to be the same as that for Figure 5. It is observed that MHMCP with a small value of *n* is preferred in low bit rates, while IR outperforms MHMCP in high bit rates. The larger *n* can suppress error propagation more effectively but demands a high coding rate. Other error resilient tools such as IR could achieve better performance in high bit rates. Consequently, in the rate-distortion sense, MHMCP is preferred only in the low bit rate environment with a small value of *n*.

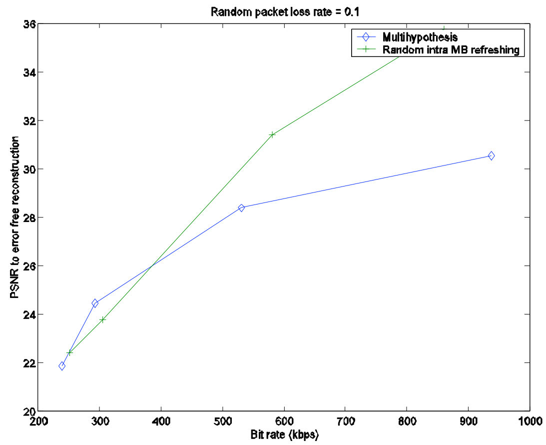


Figure 5 Error resilience comparison between MHMCP and IR in a high packet loss environment

5.3 Effect of Hypothesis Coefficients

It was shown that MHMCP performs efficiently in the R-D sense, only when the hypothesis number is small. In the following work, we investigate the impact of hypothesis coefficients on the R-D performance while fixing the hypothesis number to 3.

5.3.1 Impact on Propagation Error

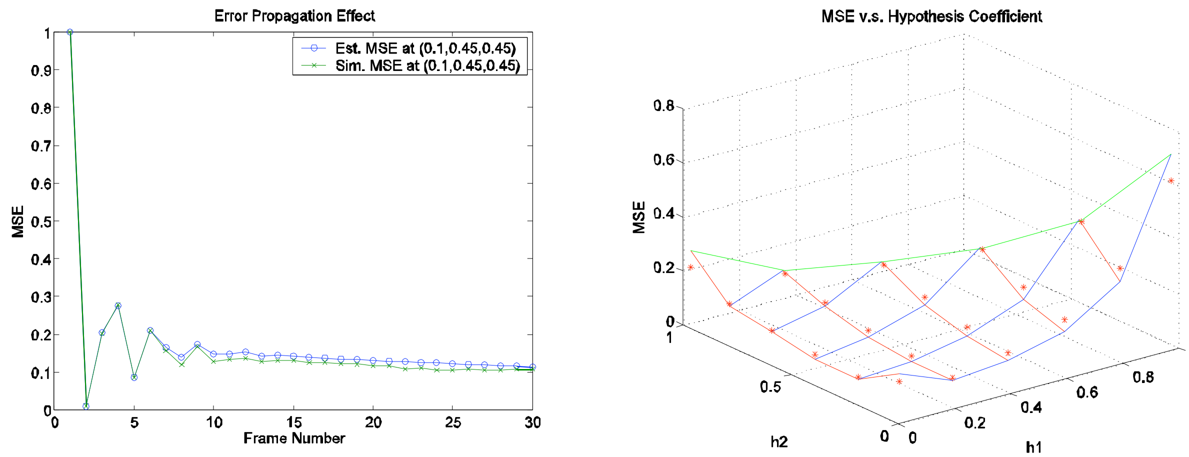
Let σ_0 denote the initial error in the 0-th frame. Then, from Equation (1), we can derive its error propagation to the following frames, given by

$$\sigma_l = \frac{1 + (1 - h_1 + h_3)\sum_{n=0}^{l} \alpha^{l-n}\beta^{n} + h_3 \sum_{n=0}^{l-1} \alpha^{l-1-n}\beta^{n}}{2 - h_1 + h_3}\sigma_0,$$

where $\gamma = \sqrt{(1-h_1)^2 + 4h_3}$, $\alpha = \frac{h_1 - 1 + \gamma}{2}$, and $\beta = \frac{h_1 - 1 - \gamma}{2}$.

We have tested the effect of a burst error with several combinations of hypothesis coefficients and confirmed that the estimated MSEs match the simulated MSEs well in every combination. Figure 6(a) demonstrates one example of error propagation effect when $(h_1, h_2, h_3) = (0.1, 0.45, 0.45)$. Note that the propagation error oscillates with decreasing amplitude and finally converges to a non-zero value.

To further investigate the relationship between MSE and the hypothesis coefficients, Figure 6(b) depicts the average MSEs after a burst error, where the averaging range is 100 frames. The average MSEs are normalized with respect to the initial MSE σ_0^2 of the concealed frame. The x- and y-axis in Figure 6(b) denote h_1, and h_2, respectively. Note that h_3 is given by 1- h_1- h_2. The red stars represent the estimated MSE values, while the lines represent the simulated MSE values. The largest MSE is observed when $(h_1, h_2, h_3) = (1,0,0)$, which is the conventional single-hypothesis motion compensated prediction (SHMCP). Also, the other two extreme cases, $(h_1, h_2, h_3) = (0,1,0)$ and $(0,0,1)$ provide relatively large MSEs. This indicates that MHMCP alleviates the propagation errors by mixing more than one prediction blocks. The minimum MSE is achieved around the point $(h_1, h_2, h_3) = (0,0.18,0.82)$. However, the hypothesis coefficients also affect the prediction performance and hence the bit rate. The minimum MSE point results in a high bit rate due to less efficient motion prediction with zero value of h_1. The effects of hypothesis coefficients on the bit rate are discussed in the next section.



(a) Comparison of the estimated MSE and the simulated MSE

(b) Normalized MSE vs. hypothesis coefficients

Figure 6 The impact of different hypothesis coefficients on the propagation error.

5.3.2 Impact on Bit Rate

Assume that the prediction error of the i-th hypothesis is given by a Gaussian random variable with variance σ_i^2 $(i=1,2,3)$. Then, the overall prediction error of MHMCP can be expressed as Equation 3 with $n=3$. We can simplify the problem by assuming that $\rho_{i,j} = \rho$ for all i and j. Furthermore, we found experimentally from training sequences that $\sigma_2^2 = 1.18\ \sigma_1^2$ and $\sigma_3^2 = 1.385\ \sigma_1^2$.

Then, the overall change of the bit rate, as compared to the conventional SHMCP, can be expressed as the sum of the reduced bit rate due to the smaller prediction error and the increased bit rate due to two additional motion vectors,

$$\Delta R = \frac{1}{2}\log_2 \frac{\sigma^2_{MH}}{\sigma^2_1} + \frac{2c}{256}, \tag{5}$$

where c is the average number of bits to represent an additional motion vector as well as the corresponding reference frame number. Parameter c depends on the coding method of motion information. In this work, we predicatively encode the motion vectors to exploit the correlation between motion vectors. First, the motion vector of each hypothesis is predicted from that of the corresponding hypothesis for the previous MB. Then, the predicted motion vectors of hypotheses 2 and 3 are further predicted from that of hypothesis 1. In our simulation, c is about 13.

Figure 7 compares the simulated bit rates with the estimated bit rates. The red stars represent the estimated bit rates, whereas the lines represent the simulated bit rates. We can see that the formula in (5) approximates the experimental data well.

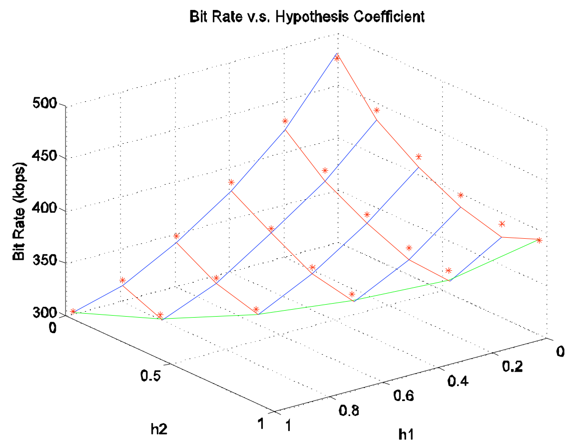


Figure 7 The bit rate (bits per pixel) as a function of the hypothesis coefficients.

5.3.3 Rate-distortion Analysis

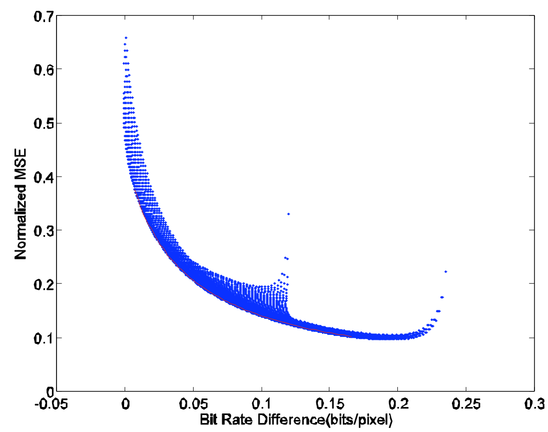


Figure 8 The rate-distortion performance after a burst error.

Using the MSE and bit rate models, we can plot the estimated R-D points in Figure 8. Each blue point corresponds to a combination of (h_1, h_2, h_3). The three spikes are observed in the three extreme cases: $(h_1, h_2, h_3) = (1,0,0)$, and $(0,1,0)$ and $(0,0,1)$. We can see that the choice of hypothesis coefficients affects the R-D performance considerably. From these data, we found that the optimum R-D points on the convex hull have similar h_2 values, which are concentrated around 1/3. The red curve correspond to the combinations, in which h_2 is fixed to 1/3. This indicates that for a video codec with triple-hypothesis motion compensation, we can fix h_2 to 1/3 and change only h_1 and h_3 to meet the overall rate or distortion requirement. The adaptation of the hypothesis coefficients can be made at the sequence or the frame level. However, due to varying video content and channel fluctuation, the performance of MHMCP can be further improved if the hypothesis coefficients are adapted at the MB level.

As section 5.2.3, we minimize the Lagrangian cost of each MB. The cost is evaluated for various combinations of hypothesis coefficients, and the optimal combination yielding the lowest cost is selected, where h_i varies from 0 to 1 with a step 0.2. Then, the index for the optimum combination is encoded as the header information for the MB.

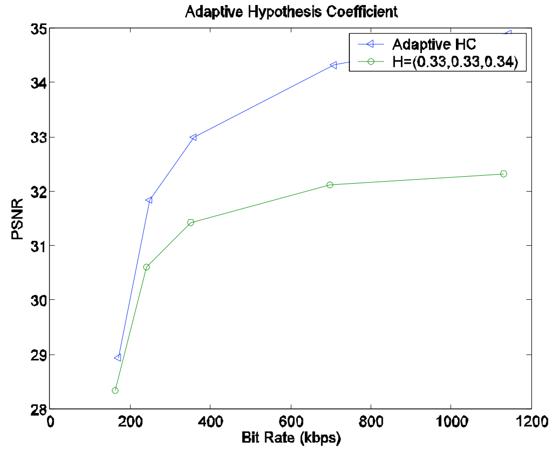
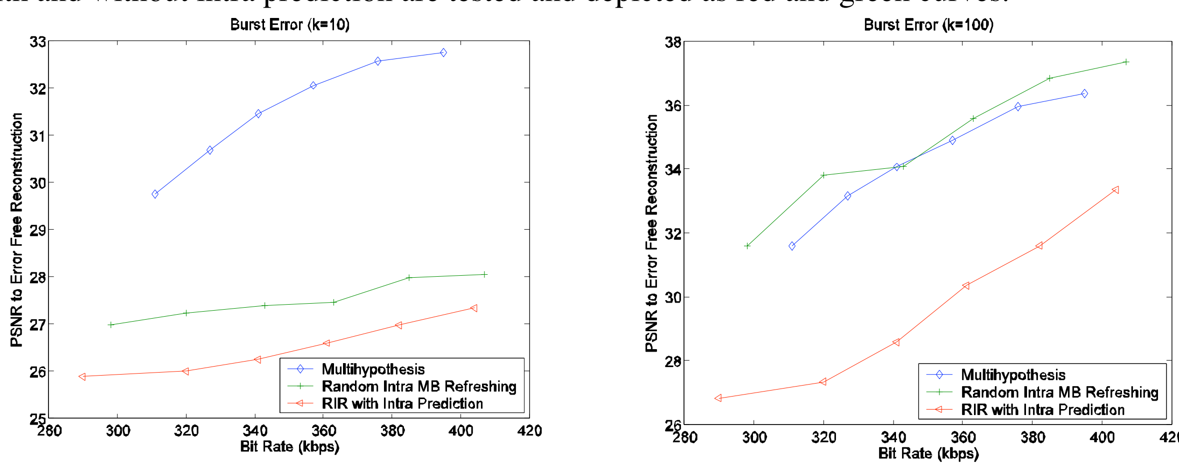


Figure 9. The rate-distortion performance after a burst error

Figure 9 shows the PSNR performance of the proposed MB level adaptation scheme. We can see that the adaptive scheme provides up to 2.5 dB better performances than the fixed scheme where $(h_1, h_2, h_3) = (0.33, 0.33, 0.34)$.

5.3.4 Comparison with Random intra-MB Refreshing

In Figure 10, we compare MHMCP and IR in terms of the PSNR of the reconstructed video and the bit rates. For MHMCP, we fix h_2 to 0.33 and change h_1 from 0.1 to 0.6. For comparison, random intra refreshing algorithm is also simulated. The refreshing rates are selected so that it generates the similar bit rate as MHMCP. To achieve fair comparison, the encoding parameters such as quantization parameter are set to be the same in both cases. The H.264 standard enforces the intra prediction, which reduces bit rates but deteriorates error resilient capability. Both IR with and without intra prediction are tested and depicted as red and green curves.



(a) Short term error suppression effect (b) Long term error suppression effect

Figure 10. Error suppression comparison between MHMCP (with various h_i values) and IR after a burst error.

The PSNR values in Figure10 (a) are averaged in the short term (i.e. only the first 10 frames after an erroneous one) while the PSNR values in Figure 10 (b) are averaged in the long term (consisting of 100 frames after an erroneous frame). MHMCP significantly outperforms both IR algorithms in short term. In the long term, MHMCP still performs better than IR with intra prediction, but achieves lower PSNR values than IR without intra prediction. Therefore, MHMCP could be considered as a good error resilient tool in H.264 codec. Also, even for H.263 and MPEG4 codecs without the enforced intra prediction, MHMCP may have its advantage provided that I-frames are regularly inserted.

6. Other Relevant Work Being Conducted and How this Project is Different

One application is to adopt the multi-hypothesis (MH) technique for error concealment at the decoder. In [4], we developed a multi-hypothesis error concealment method with adaptively chosen weighting coefficients to lower concealment errors as well as suppress propagating errors. An iterative MH error concealment algorithm for the latest standard H.26L was proposed in [5]. It is worthwhile to point out that the above methods work well even without MHMCP in the encoder. Actually, they did not address the error resilience of MHMCP at the encoder side. Lin and Wang [6] discussed the error resilient property of 2-hypothesis motion compensated prediction. However, a thorough analysis of the error resilience property of MHMCP is still lacking up to now.

7. Plan for the Next Year

Our plan for the next year includes the following:

- Continue to improve the functionalities of existing audio/video/graphic coding techniques and develop their error resilient schemes.
- Design effective compression for 3D animated meshes.
- Develop TCP-friendly rate adaptation and packet forwarding schemes for Differentiated Service (DS) networks.

8. Expected Milestones and Deliverables

We have modified the H.264 codec to include the MH option and the MB level adaptation. Several conference papers have been published on this topic. There are several long-term challenging issues worth our further study. They include: quality of service (QoS) of multimedia transmission over error-prone IP networks, coding of 3D animated meshes, seamless transition between cells (or called handoff) in 2.5G and 3G cellular systems, content and networking security, etc. We will continue to investigate these problems in the next five years and seek some major breakthrough in engineering science. The milestone chart is given below.

9. Member Company Benefits

We presented research results to several companies, including Conexant, Skyworks, and EverFocus. They show strong interests in our research work and the high quality of students. Some summer internship opportunities and other funded research projects were offered by these companies as a result of our demos.

10. References

[1] B. Girod and N. Farber, "Wireless video", in Compressed video over networks, M.-T. Sun Ed., Marcel Dekker, 2000.

[2] M. Flierl, T. Wiegand and B. Girod, "Rate-distortion optimization for video compression", IEEE Signal Processing Magazine, vol. 15, no. 6, pp.74-90, Nov., 1998.

[3] "Video Coding for Low Bitrate Communication", 1998, ITU-T Recommendation H.263.

[4] W.-Y. Kung, C.-S. Kim and C.-C. Jay Kuo, "A dynamic error concealment for video transmission over noisy channels", IEEE GlobeCom, Nov. 2002.

[5] Y. O. Park, C.-S. Kim and S.-U. Lee, "Multi-hypothesis error concealment algorithm for h.26l video", IEEE ICIP, Sep. 2003.

[6] S. Lin and Y. Wang, "Error resilience property of multihypothesis motion-compensated prediction", IEEE ICIP 2002, Jun. 2002.