We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

13

Error Resilient Video Coding Techniques Based on the Minimization of End-to-End Distortions

Wen-Nung Lie[1] and Zhi-Wei Gao[2]
[1]*National Chung Cheng University, Chia-Yi, Taiwan*
[2]*TECO Group Research Institute, Taipei, Taiwan*

1. Introduction

Due to successful development of video compression techniques, the huge amount of redundancies contained in raw videos can be removed effectively. This makes it possible to transmit videos over bandwidth-limited channels. The applications can be video conferencing, distance learning, streaming videos on Internet and mobile phones, digital high quality TV, …, etc.

Generally speaking, redundancies in videos can be categorized into different types: spatial, temporal and statistical redundancies. Operations of discrete cosine transform (DCT) and quantization are used to remove the spatial redundancy and motion estimation and compensation (ME/MC) are used to remove the temporal redundancy. Entropy coder is, on the other hand, for removing the last type of redundancy.

However, highly compressed videos are often fragile to noises. Once compressed video bit streams encounter any kinds of errors in transmission, the quality of the reconstructed videos at decoder side degrades seriously. The reason is that error propagation in spatial or temporal direction occurs when the decoder decodes erroneous bit streams to reconstruct videos. In order to enhance the quality of the decoded videos at decoder side, two common techniques are often adopted: one is to generate robust compressed video bit streams at encoder side, known as the "error resilience"; the other is to conceal errors in the reconstructed videos at decoder side, known as the "error concealment". Both techniques are capable of substantially improving the quality of the decoded videos in error-prone transmission environments.

This chapter will focus on the issue of error resilient video coding. The main concept of error resilient coding is to increase the robustness of the compressed videos at the expense of extra bit rates; that is, inserting redundancies important to video error recovery. The coding efficiency (i.e., PSNR/bit-rate) unavoidably lowered down, whereas a lower PSNR degradation at decoder side can be achieved in case of severe channel errors. Researchers often faced a problem of how to schedule the overall bit resources such that error resiliency capability can be maximized.

Among the algorithms developed for video error resiliency, end-to-end distortion, which measures the difference between the raw video data and that finally obtained before display at decoder side (possibly with channel errors and according error concealment), was recently popularly adopted as the criterion for optimization (minimization). Here, in this book chapter, we first propose an algorithm of error-resilient motion estimation and mode decision by considering end-to-end distortions for H.264/AVC standard. Then, this algorithm is extended for the enhancement of H.264-based multi-hypothesis coding (MHC) at a given hypothesis-weighting vector, which was traditionally proposed with its good rate-distortion performance in noise-prone channels. Finally, based on the availability of motion vectors, an adaptive hypothesis-weighting algorithm is proposed to make error resiliency adaptive to video contents, frame by frame.

2. Modelling of end-to-end distortions

End-to-end distortion, the distortion between the raw data and the one reconstructed at decoder side (possibly incurred with channel errors), has been adopted as an optimized criterion in applications such as intra/inter mode decision (Chang et al., 2005; Cote & Kosentini, 1999; Leontaris & Cosman, 2004; Zhan et al., 2000) and motion estimation (ME) (Harmanci & Tekal, 2005; Wiegand et al., 2000; Yang & Rose, 2005). We first model the end-to-end distortions to include (He et al, 2002): source distortion D_s, incurred by n_q, and channel distortion D_c, incurred by n_c. Here, n_q is related to the quantization noise and n_c to the incompleteness of error concealment and motion compensation (or, error propagation). A pictorial illustration of our model is depicted in Fig.1, where the subscript n is the frame index, f_n represents the original video signal, $\hat{f}_n$ represents the encoded video (i.e., decoded video without errors), and $\tilde{f}_n$ represents the video reconstructed at the decoder side in presence of channel noise n_c.

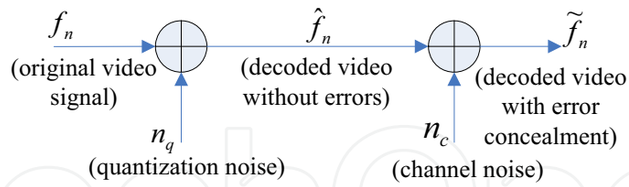

Fig. 1. The modelling of end-to-end distortion.

Via the above model, it is ready to observe that the end-to-end distortion $D_{end\text{-}to\text{-}end}$ can be formulated as

$$\begin{aligned} D_{end-to-end} &= E\left\{(f_n - \tilde{f}_n)^2\right\} \\ &\cong E\left\{(f_n - \hat{f}_n)^2 + (\hat{f}_n - \tilde{f}_n)^2\right\} = D_s + D_c \end{aligned} \tag{1}$$

where $D_s = E\left\{(f_n - \hat{f}_n)^2\right\}$ and $D_c = E\left\{(\hat{f}_n - \tilde{f}_n)^2\right\}$. (2)

Obviously, a cross product term is ignored in deriving Eq.(1). This model was first verified on the H.263 test platform (He et al., 2002) and then proved to have an averaged deviation of only 0.863% on H.264/AVC platform (Lie et al, 2006). Based on this fact, $D_{end\text{-}to\text{-}end}$ can then be reduced by minimizing D_s and D_c separately.

However, reducing D_s may conflict with reducing D_c at a specified constant bit-rate (since a lower D_s causes a higher source bit-rate and hence, a lower channel bit-rate and larger D_c). Therefore, the original problem of minimizing end-to-end distortions becomes a typical problem of multi-objective optimization.

3. Minimizing end-to-end distortions for error resilient motion estimation (ERME)

One criterion to optimize the error resiliency of the transmitted videos in error-prone environments is to consider the end-to-end distortions under a given bit rate budget. The difficulty in estimating end-to-end distotions is mainly on knowing well the effect of error propagation caused by the loss of motion vectors (MVs) and the strategy of error concealment at decoder. Generally, the applicable methods up to now can be roughly categorized into simulation-based and model-based.

For simulation-based methods, a set of error patterns are generated according to channel conditions (e.g., signal-to-noise ratio (SNR), bit error rate (BER), or packet loss rate (PLR)) and coding parameters, such as period of resynchronization, intra-refreshing rate, etc., are tuned accordingly to minimize the end-to-end distortions. The drawback of the simulation-based approaches is the high computing load, which makes them only suitable for off-line video streaming applications (e.g., video transcoding (Reyes et al., 2000; Xia et al., 2004). On the other hand, for model-fitting approaches (He et al., 2002; Stuhlmuller, et al., 2000; Eisenberg et al., 2006), effects of error propagation are described in terms of mathematical models. Model parameters are then determined in accordance with the video contents by collecting a small set of training data. Hence, evaluation of coding parameters becomes a process of model interpolation, which would reduce the computational complexity substantially.

To suppress error propagation without sacrificing much coding efficiency, methods of searching MVs in criteria other than the traditional least SAD (sum of absolute difference) have ever been proposed. For example, Wiegand et al. (2000) adopted the end-to-end distortion as the optimizing criterion in searching MVs. This kind of algorithms could be classified as the Error Resilient Motion Estimation (ERME). However, their method generates/simulates the error patterns and analyzes the possible error propagation by constructing trees whose sizes will grow substantially for a large GOP (group of picture) size. This restricts their practical applications, in views of both memory requirements and computing loads. We need an effective but simple procedure to estimate the end-to-end distortions during the computationally intensive ME process.

An alternative method in estimating end-to-end distortions, well known as the ROPE (Recursive Optimal Per-pixel Estimate) (Zhan et al, 2000), was developed for intra/inter mode decision of each coded frame. They differently modeled error propagation caused by lost MVs by exploiting a statistical approach. Following that, the ROPE algorithm was

applied to motion estimation (Yang & Rose, 2005; Harmanci & Tekal, 2005) for increasing the performance in error resiliency. Essentially, the end-to-end distortions are decomposed into three sources: error propagation, error concealment, and quantization. To estimate the quantization errors, a series of processes like DCT/IDCT, quantization and inverse quantization, should be performed for each candidate under evaluation. This causes no problems in mode decision where only two modes (intra/inter) are considered, but certainly incurs prohibitively high computational loads for motion estimation (Yang & Rose, 2005) (normally over one thousand of MV candidates).

Extending the prior concept and avoiding the high computational complexity in estimating the end-to-end distortions, a new optimization algorithm is proposed for ERME in this chapter. The new optimizing criterion of our proposed algorithm consists of two conflicting objective functions; specifically, one of them relates to the enhancement of error resiliency and the other to the increase of coding efficiency. Hence, finding a MV that minimizes this criterion becomes a problem known as the multi-objective optimization (Ringuest, 1992). In this situation, a solution that minimizes all objective functions simultaneously does not always exist when the objective functions conflict with each other. Instead, a constrained optimization method can be developed to find a solution that compromises among these conflicting objective functions. Applying this concept, MVs thus found is capable of compromising between error resiliency and coding efficiency. Moreover, since the computing procedure of our proposed algorithm does not include the terms relating to the quantization errors, the computational complexity is not as high as that of Yang & Rose (2005).

To derive error resilient MVs, the channel distortion D_c should be modelled first. The search criterion is defined below:

$$(\Delta x^*, \Delta y^*) = \arg\min_{(\Delta x, \Delta y) \in S} \left\{ \sum_{i=1}^{B} E\left\{ \left(\hat{f}_n^i(\Delta x, \Delta y) - \tilde{f}_n^i(\Delta x, \Delta y) \right)^2 \right\} \right\} \tag{3}$$

where the subscript n is the frame index, Δx and Δy are the horizontal and vertical components of the MV, S is the feasible set for motion vectors, B is the number of pixels in a block, both $\hat{f}_n^i(\Delta x, \Delta y)$ and $\tilde{f}_n^i(\Delta x, \Delta y)$ represent the i^{th} pixel of a block motion-compensated by using MV= $(\Delta x, \Delta y)$, and $E\{.\}$ is the expectation operator. Note that $\hat{f}_n^i(\Delta x, \Delta y)$ represents the pixel value correctly decoded, wheras $\tilde{f}_n^i(\Delta x, \Delta y)$ is the pixel value obtained by considering erroneous reconstruction.

Equation (3) is actually not suitable for practical applications since a series of DCT/IDCT and quantization processes is required in computing $\hat{f}_n^i(\Delta x, \Delta y)$ and $\tilde{f}_n^i(\Delta x, \Delta y)$. Approximation of computations is necessary to make it mathematically tractable. Here, Eq.(3) is modified by changing the squared term into an absolute term:

$$(\Delta x^*, \Delta y^*) = \arg\min_{(\Delta x, \Delta y) \in S} \left\{ \sum_{i=1}^{B} E\left\{ \left| \hat{f}_n^i(\Delta x, \Delta y) - \tilde{f}_n^i(\Delta x, \Delta y) \right| \right\} \right\} \tag{4}$$

Based on the inequality $E\{|\hat{x}-\tilde{x}|\} \geq |E\{\hat{x}-\tilde{x}\}|$, Eq.(5) will hold.

$$\begin{aligned} &\min_{(\Delta x,\Delta y)\in S}\left\{\sum_{i=1}^{B} E\left\{\left|\hat{f}_n^i(\Delta x,\Delta y)-\tilde{f}_n^i(\Delta x,\Delta y)\right|\right\}\right\} \\ &\geq \min_{(\Delta x,\Delta y)\in S}\left\{\sum_{i=1}^{B}\left|E\left\{\hat{f}_n^i(\Delta x,\Delta y)-\tilde{f}_n^i(\Delta x,\Delta y)\right\}\right|\right\} \\ &= \min_{(\Delta x,\Delta y)\in S}\left\{\sum_{i=1}^{B}\left|\hat{f}_n^i(\Delta x,\Delta y)-E\left\{\tilde{f}_n^i(\Delta x,\Delta y)\right\}\right|\right\} \end{aligned} \tag{5}$$

Note that the first term in summation is considered to be deterministic with respect to the $E\{.\}$ operator, due to its independence from the channel conditions. The term $E\left\{\tilde{f}_n^i(\Delta x,\Delta y)\right\}$ in Eq.(5) can be easily estimated by (Zhan et al., 2000):

$$E\left\{\tilde{f}_n^i(\Delta x,\Delta y)\right\} = (1-p_e)\left(E\left\{\tilde{f}_{n-\alpha}^j\right\} + r_n^i(\Delta x,\Delta y)\right) + p_e \cdot E\left\{\tilde{f}_{n-1}^i\right\} \tag{6}$$

where p_e is the error probability of a considered pixel, $r_n^i(\Delta x,\Delta y)$ is the residual produced after motion prediction (with MV= $(\Delta x,\Delta y)$), $E\left\{\tilde{f}_{n-\alpha}^j\right\}$ is the pixel value compensated from the j^{th} pixel (pointed to by MV= $(\Delta x,\Delta y)$) of the $(n\text{-}\alpha)^{\text{th}}$ frame, α is a positive number (accounting for the long-term memory prediction in H.264/AVC), and $E\left\{\tilde{f}_{n-1}^i\right\}$ is the pixel value recovered by adopting zero-motion as the scheme of error concealment.

According to the coding principle, $\hat{f}_n^i(\Delta x,\Delta y)$ can be expressed as:

$$\hat{f}_n^i(\Delta x,\Delta y) = \hat{f}_{n-\alpha}^j + r_n^i(\Delta x,\Delta y), \tag{7}$$

where $\hat{f}_{n-\alpha}^j$ is the pixel value motion-compensated from the j-th pixel of the $(n\text{-}\alpha)^{\text{th}}$ frame. Substituting Eqs.(6) and (7) into Eq.(5), we change the optimization problem in Eq. (4) to become:

$$\begin{aligned} &(\Delta x^*,\Delta y^*) \\ &= \arg\min_{(\Delta x,\Delta y)\in S}\left\{\sum_{i=1}^{B}\left|\begin{array}{l}\left(\hat{f}_{n-\alpha}^j + r_n^i(\Delta x,\Delta y)\right) - p_e E\left\{\tilde{f}_{n-1}^i\right\} \\ -(1-p_e)\left(E\left\{\tilde{f}_{n-\alpha}^j\right\} + r_n^i(\Delta x,\Delta y)\right)\end{array}\right|\right\} \end{aligned} \tag{8}$$

Equation (8) is actually not a good criterion, due to its prohibitively high complexity in computing $r_n^i(\Delta x,\Delta y)$. To yield $r_n^i(\Delta x,\Delta y)$, the processes of DCT, quantization, and IDCT need to be performed for each MV candidate $(\Delta x,\Delta y)$. This is also the reason why algorithms provided in Yang & Rose (2005) and Harmanci & Tekal (2005) are impractical in view of computing complexity. Equation (8) can be rewritten to be

$$(\Delta x^*, \Delta y^*) = \arg\min_{(\Delta x, \Delta y) \in S} \left\{ \sum_{i=1}^{B} \left| \left(\hat{f}_{n-\alpha}^{j} \right) - \left((1-p_e) \cdot E\left\{ \tilde{f}_{n-\alpha}^{j} \right\} + p_e \cdot E\left\{ \tilde{f}_{n-1}^{i} \right\} \right) \right| + \left(p_e \cdot r_n^i(\Delta x, \Delta y) \right) \right\} \quad (9)$$

Considering that $r_n^i(\Delta x, \Delta y)$ is the prediction residual (expectedly smaller than the pixel reconstructions $\hat{f}_{n-\alpha}^{j}$, $E\left\{\tilde{f}_{n-\alpha}^{j}\right\}$, and $E\left\{\tilde{f}_{n-1}^{i}\right\}$ and $p_e < 1 - p_e < 1.0$ (for $p_e < 0.5$), the term $p_e \cdot r_n^i(\Delta x, \Delta y)$ can be ignored, with respect to the other three terms, reducing Eq.(9) into Eq.(10):

$$(\Delta x^*, \Delta y^*) = \arg\min_{(\Delta x, \Delta y) \in S} \left\{ \sum_{i=1}^{B} \left| \left(\hat{f}_{n-\alpha}^{j} \right) - \left((1-p_e) \cdot E\left\{ \tilde{f}_{n-\alpha}^{j} \right\} + p_e \cdot E\left\{ \tilde{f}_{n-1}^{i} \right\} \right) \right| \right\} \quad (10)$$

Note that, the ignorance of $p_e \cdot r_n^i(\Delta x, \Delta y)$ eases the computation of channel distortions significantly. In Eq.(10), the first term $\hat{f}_{n-\alpha}^{j}$ is readily available when encoding the n^{th} frame, while the second term (also called the first moment of $\tilde{f}_{n-\alpha}^{j}$) can be derived by using the technique proposed in Zhan et al. (2000). According to Eq.(6), the first moment of $\tilde{f}_n^i$ (i.e., $E\left\{\tilde{f}_n^i\right\}$) can be recursively updated after its motion vector $(\Delta x, \Delta y)^*$ and residual r_n^i are figured out. This guarantees the availability of $E\left\{\tilde{f}_{n-\alpha}^{j}\right\}$ and $E\left\{\tilde{f}_{n-1}^{i}\right\}$ on evaluating Eq.(10) for the n^{th} frame. Clearly, the extra computations required in computing Eq.(10) come from the updating of $E\left\{\tilde{f}_n^i\right\}$.

The quantity to be optimized in Eq.(10) approximates the channel distortion described in Eq.(3). Hence, we can have a constraint on the source distortion D_s when optimizing the channel distortion D_c. That is,

$$\min_{(\Delta x, \Delta y) \in S} \left\{ \sum_{i=1}^{B} \left| \left(\hat{f}_{n-\alpha}^{j} \right) - \left((1-p_e) \cdot E\left\{ \tilde{f}_{n-\alpha}^{j} \right\} + p_e \cdot E\left\{ \tilde{f}_{n-1}^{i} \right\} \right) \right| \right\} \quad \text{subject to} \quad D_s^{\min} < \sum_{i=1}^{B} \left\{ (f_n^i - \hat{f}_n^i)^2 \right\} \le \sigma_1^2 \quad (11)$$

where $D_s^{\min}$ is the achievable lower bound when considering to optimize D_s only and σ_1^2 is a selected threshold. It is known that constraining the source distortion is equivalent to limiting the quantization noise, i.e., keeping the quantization parameter (QP) below a value. Accordingly, the motion-prediction residues should be kept low to control the resulting bit rate to a targeted value (since a small QP will increase the bit rate). Hence, we change the above inequality condition to

$$\gamma^{\min} < \sum_{i=1}^{B} \left(f_n^i - \hat{f}_{n-\alpha}^{j(\boldsymbol{w},\alpha)} \right)^2 \le Thd1 \tag{12}$$

where $(\boldsymbol{w},\alpha)$ stands for a MV candidate $\boldsymbol{w}$ that refers to the (n-α)$^{\text{th}}$ frame, $j(\boldsymbol{w},\alpha)$ represents the pixel j pointed to by $\boldsymbol{w}$ in frame-(n-α), $f_n^i - \hat{f}_{n-\alpha}^{j(\boldsymbol{w},\alpha)}$ is the motion-prediction residue for pixel i of the frame n, $\gamma^{\min}$ represents the smallest residual power that can be achievable, and *Thd1* is a selected threshold. This change of constraint obviously eases the evaluation process significantly since the computation of $\hat{f}_{n-\alpha}^{j(\boldsymbol{w},\alpha)}$ is in no need of IDCT for a given $(\boldsymbol{w},\alpha)$, while the computation of $\hat{f}_n^i$ requires r_n^i (Eq.(7)), which needs DCT, quantization, and IDCT for each MV candidate $(\boldsymbol{w},\alpha)$.

Replacing the constraint of Eq.(11) with the inequality in Eq.(12), we have the following constrained optimization problem:

$$\underset{(\boldsymbol{w},\alpha)}{Min}\{EP_{(\boldsymbol{w},\alpha)}\} \qquad \text{subject to} \quad CR_{(\boldsymbol{w},\alpha)} \le Thd1 \tag{13}$$

where

$$EP_{(\boldsymbol{w},\alpha)} = \sum_{i=1}^{B} \left| \hat{f}_{n-\alpha}^{j(\boldsymbol{w},\alpha)} - \left((1-p_e) \cdot E\left\{ \tilde{f}_{n-\alpha}^{j(\boldsymbol{w},\alpha)} \right\} + p_e \cdot E\left\{ \tilde{f}_{n-1}^{i} \right\} \right) \right| \tag{14}$$

$$CR_{(\boldsymbol{w},\alpha)} = \sum_{i=1}^{B} \left(f_n^i - \hat{f}_{n-\alpha}^{j(\boldsymbol{w},\alpha)} \right)^2 \tag{15}$$

In another viewpoint, $EP_{(\boldsymbol{w},\alpha)}$ measures the level of error propagation caused by channel distortion and $CR_{(\boldsymbol{w},\alpha)}$ reflects the power of the motion-prediction residuals relating to source distortion, for a given $(\boldsymbol{w},\alpha)$. By adjusting *Thd1*, different levels of compromise between $EP_{(\boldsymbol{w},\alpha)}$ and $CR_{(\boldsymbol{w},\alpha)}$ can be achieved.

Traditionally, the optimal solution of Eq.(13) can be found via a full or a fast search on all possible $(\boldsymbol{w},\alpha)$'s. Note that our algorithm would not change the nature of full/fast search in the traditional ME process, but to provide a more proper criterion in selecting MVs that are expected to achieve better compromise between error resiliency and coding efficiency.

4. Minimizing end-to-end distortions for error resilient mode decision (ERMD)

In comparison with H.263 and MPEG-1/2/4 standards, mode decision (MD) for both intra and inter-coded frames (here, we only focus on the inter-coding case due to its close relation to the ME topic) is one of the most distinctive and evolving features that make significant progress in coding efficiency. Essentially, a MB is divided into sub-blocks of variable sizes (e.g., 16×16, 16×8, 8×16, 8×8, 4×8, 8×4, and 4×4 pixels), each of which is associated with an estimated MV (via the ERME method previously discussed). A cost is then measured with each combination (or, called a mode) of block partition for a MB. Mode decision is thus to

determine a partition, together with the estimation of associated MVs, that minimizes a selected cost function for each MB. The traditional cost function in H.264/AVC is based on the so-called SATD (Sum of Absolute Transform Difference). It is shown (Stuhlmuller et al., 2000) that for error-free transmission, mode selection in a Lagrangian rate-distortion framework is capable of enhancing the coding efficiency effectively.

Essentially, different goals would require different optimizing criterion or cost functions. It is straightforward to motivate us that both ME and MD based on modified criteria would enhance error resiliency for error-prone transmission.

In H.264/AVC, the cost (i.e., SATD) defined for Lagrangian optimization framework is related to the residuals after motion compensation, or, the difference between the motion-predicted MB and the original data. Instead of that, the end-to-end distortion defined in Eq. (1) is incorporated into the algorithm of Lagrangian optimization. More precisely, after finding MVs for each mode candidate by using the ERME algorithm, the following optimization problem is to be solved:

$$\min_{m \in \boldsymbol{M}} D_s(m) + D_c(m) + \lambda R(m), \tag{16}$$

where $\boldsymbol{M}$ is the set of mode candidates m's allowable in H.264/AVC, D_s and D_c represent the source and channel distortion as defined in Eq.(1), R is the bit rate needed to encode a MB given m, and λ is a Lagrangian multiplier.

D_s can be computed directly from the original data and the reconstructed data at local decoder. To evaluate D_c, cases of inter- and intra-coded MBs are to be separately discussed.

For intra-coded MBs, channel distortion $D_{C,i}^{I}$ totally results from incompleteness of error concealment. Equation (17) below formulates this observation:

$$D_{C,i}^{I,n} = p_e \cdot E\left\{ \left(\hat{f}_n^i - \tilde{f}_{n-1}^i \right)^2 \right\} \tag{17}$$

where i is the pixel index in an MB, n is the frame index, p_e is the error probability relating to the transmission environment, and $E\{.\}$ is the expectation operator. Equation (17) can be further arranged to yield Eq.(18):

$$\begin{aligned} D_{C,i}^{I,n} &= p_e \cdot E\left\{ \left(\hat{f}_n^i - \hat{f}_{n-1}^i + \hat{f}_{n-1}^i - \tilde{f}_{n-1}^i \right)^2 \right\} \\ &\le p_e \cdot E\left\{ \left(\left| \hat{f}_n^i - \hat{f}_{n-1}^i \right| + \left| \hat{f}_{n-1}^i - \tilde{f}_{n-1}^i \right| \right)^2 \right\} \\ &\le p_e \cdot E\left\{ \left(\hat{f}_n^i - \hat{f}_{n-1}^i \right)^2 \right\} + 2p_e \cdot E\left\{ \left| \hat{f}_n^i - \hat{f}_{n-1}^i \right\| \hat{f}_{n-1}^i - \tilde{f}_{n-1}^i \right| \right\} \\ &\quad + p_e \cdot E\left\{ \left(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i \right)^2 \right\} \end{aligned} \tag{18}$$

Note that $\left|\hat{f}_n^i - \hat{f}_{n-1}^i\right|$ stands for the frame difference after encoding, which is independent of the channel condition (hence, deterministic with respect to the $E\{.\}$ operator). Hence, Eq. (18) is reduced to

$$\begin{aligned} D_{C,i}^{I,n} &\le p_e \cdot \left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)^2 + 2p_e \cdot \left|\hat{f}_n^i - \hat{f}_{n-1}^i\right| \cdot \left|\hat{f}_{n-1}^i - E\left\{\tilde{f}_{n-1}^i\right\}\right| + p_e \cdot E\left\{\left(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i\right)^2\right\} \\ &= p_e \cdot \left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)^2 + 2p_e \cdot \left|\hat{f}_n^i - \hat{f}_{n-1}^i\right| \cdot \left|\hat{f}_{n-1}^i - E\left\{\tilde{f}_{n-1}^i\right\}\right| + p_e \cdot D_{C,i}^{n-1} \end{aligned}, \tag{19}$$

where $D_{C,i}^{n-1}$ represents channel distortion of the pixel *i* in frame (*n*-1).

For inter-coded MBs, distortions resulting from error propagation should be taken into consideration. First, let us consider the case in which MVs and residuals $\hat{e}_n^i$ are received correctly. In this case, the decoded pixel $\tilde{f}_n^i$ will be

$$\tilde{f}_n^i = \hat{e}_n^i + \tilde{f}_{n-\alpha}^k, \tag{20}$$

where $\tilde{f}_{n-\alpha}^k$ stands for the pixel value compensated from the *k*-th pixel of frame *n-a*. If errors occur, the error concealment procedure will replace the erroneous MB $\tilde{f}_n^i$ with $\tilde{f}_{n-1}^i$ (zero-motion recovery is assumed). Combining this observation with Eq.(20), channel distortion of a pixel *i* in an inter-coded MB can be formulated as

$$D_{C,i}^{P,n} = p_e \cdot E\left\{\left(\hat{f}_n^i - \tilde{f}_{n-1}^i\right)^2\right\} + (1-p_e) \cdot E\left\{\left[\hat{f}_n^i - \left(\hat{e}_n^i + \tilde{f}_{n-\alpha}^k\right)\right]^2\right\}. \tag{21}$$

Similar to the treatment of the term in Eq. (17), Eq. (21) can be arranged to yield

$$\begin{aligned} D_{C,i}^{P,n} &\le p_e \cdot E\left\{\left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)^2\right\} + 2p_e \cdot \left|\hat{f}_n^i - \hat{f}_{n-1}^i\right| \cdot \left|\hat{f}_{n-1}^i - E\left\{\tilde{f}_{n-1}^i\right\}\right| \\ &\quad + p_e \cdot E\left\{\left(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i\right)^2\right\} + (1-p_e) \cdot E\left\{\left(\hat{f}_{n-\alpha}^k - \tilde{f}_{n-\alpha}^k\right)^2\right\} \\ &= p_e \cdot \left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)^2 + 2p_e \cdot \left|\hat{f}_n^i - \hat{f}_{n-1}^i\right| \cdot \left|\hat{f}_{n-1}^i - E\left\{\tilde{f}_{n-1}^i\right\}\right| + p_e \cdot D_{C,i}^{n-1} + (1-p_e) \cdot D_{C,k}^{n-\alpha} \end{aligned} \tag{22}$$

where $D_{C,k}^{n-\alpha}$ stands for the channel distortion of the *k*-th pixel in frame (*n-a*).

Based on Eqs. (19) and (22), we are able to estimate the channel distortion term in Eq. (16) for each considered mode m. When applying Eq. (19) and (22), it is necessary to estimate $E\left\{\tilde{f}_{n-1}^i\right\}$, $D_{C,i}^{n-1}$, and $D_{C,k}^{n-\alpha}$. The technique in Zhan et al. (2000) can be used to recursively calculate $E\left\{\tilde{f}_{n-\alpha}^i\right\}$, $\alpha = 1,2,3,\cdots$. A general formula to evaluate $E\left\{\tilde{f}_n^i\right\}$, in case of intra-coded MBs, is given in Eq. (23), whereas Eq. (24) is adopted in case of inter-coded MBs.

$$E\left\{\tilde{f}_n^i\right\} = (1-p_e) \cdot \hat{f}_n^i + p_e \cdot E\left\{\tilde{f}_{n-1}^i\right\} \tag{23}$$

$$E\left\{\tilde{f}_n^i\right\}=(1-p_e)\cdot\left(\hat{e}_n^i+E\left\{\tilde{f}_{n-\alpha}^k\right\}\right)+p_e\cdot E\left\{\tilde{f}_{n-1}^i\right\} \tag{24}$$

For n=0, we set $E\left\{\tilde{f}_0^i\right\}=\hat{f}_0^i$, i.e., no channel errors are assumed. On the other hand, the term $D_{C,i}^{n-1}$, and $D_{C,k}^{n-\alpha}$ are available when processing frame n, they are computed by using Eq. (19) and Eq. (22), depending on the encoding type (I or P) of the MB considered.

5. Multi-hypothesis coding based on minimization of end-to-end distortions

The so-called Multi-Hypothesis Coding (MHC) was proposed to find out more than one motion compensated MB, called hypothesis, from different reference frames and combine these hypotheses via weighting coefficients to form a predicted MB. It is verified (Sullivan, 1993; Flierl et al., 1998; Flierl et al., 2000) that by choosing the number of hypotheses and the weighting coefficients, the multi-hypothesis technique improves coding efficiency further when compared with the single hypothesis technique. Theoretical discussions about the rate-distortion performance of the multi-hypothesis technique can be found in literature (Flierl et al., 2002; Girod, 2000).

Error resiliency property of the multi-hypothesis coding technique has also been discussed (Lin & Wang, 2002; Kung et al., 2006), where the effect of temporal error propagation after burst errors is modeled. Specifically, in Kung et al. (2006), the proposed model is applied at encoder to decide determine the hypothesis-weighting to minimize propagation errors. However, their model restricted to a single burst error, which is not feasible in practical situation.

In this chapter, we try to apply the technique of end-to-end distortion optimization for multi-hypothesis video coding to further enhance the robustness of the transmitted video. The application is two folds: 1) finding error resilient MVs for a given set of hypothesis-weighting coefficients, similarly as in Section 3, and 2) adapting hypothesis-weighting coefficients to video contents. Both the above two techniques (ERME for a given hypothesis-weighting vector and adaptive hypothesis-weighting) can be integrated for further enhancing the error resiliency of the transmitted videos.

Rate-distortion theorem tells us that minimization of D_s can be alternatively achieved by minimizing the power of the residual signal. In case of multi-hypothesis coding technique, it is the power of the signal $f-\boldsymbol{h}^T\hat{\boldsymbol{c}}$ that needs to be minimized, where $\hat{\boldsymbol{c}}$ is a column vector composed of N hypotheses, and $\boldsymbol{h}$ is an $N\times 1$ weighting vector. For the channel distortion D_c, it is formulated similarly as in Section 3. Hence, finding MVs that minimize the end-to-end distortions for a given $\boldsymbol{h}$ can be conducted similarly as in Section 3.

In this case, error resilient motion estimation for MHC can be formulated as finding :

$$\begin{aligned}&(\Delta\mathbf{x}^*,\Delta\mathbf{y}^*)=\\&\arg\min_{(\Delta\mathbf{x},\Delta\mathbf{y})\in\mathbf{S}}\left\{\sum_{i=1}^{p}E\left\{\left(\hat{f}_n^i(\Delta\mathbf{x},\Delta\mathbf{y})-\tilde{f}_n^i(\Delta\mathbf{x},\Delta\mathbf{y})\right)^2\right\}\right\},\end{aligned} \tag{25}$$

where **Δx** and **Δy** denote the vectors of x and y components, respectively, of MVs for producing hypotheses (hence, the dimension of **Δx** and **Δy** equals the number of hypotheses, *N*), *i* is the pixel position index, **S** is the feasible region where MVs is evaluated and *p* is the number of pixels in a MB. Notice the similarity between Eq.(3) and Eq.(25). The only difference comes from the fact that $\hat{f}_n$ and $\widetilde{f}_n$ are now function of *N* MVs, i.e., $(\mathbf{\Delta x}, \mathbf{\Delta y})_{N\times 2}$, for multi-hypothesis coding.

We will not waste space to derive similar formulas as Eqs.(4)~(6) for ERME of MHC, but go directly to the one similar to Eq.(7). In accordance with the principle of MHC, it implies

$$\hat{f}_n^{\,i}(\mathbf{\Delta x}, \mathbf{\Delta y}) = \sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) + \hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y}), \tag{26}$$

where h_k represents the weighting coefficient applicable to the *k*-th hypothesis, $\hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k)$ is the hypothesis obtained by using the MV whose *x*- and *y*-components are $(\Delta x_k, \Delta y_k)$ from the prediction frame with time index $n-k$, and $\hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y})$ is the quantized residual (prediction error). It is assumed that different hypotheses come from different frames. By assuming that each lost or corrupted MB is recovered via zero-motion replacement from the previous frame, $E\{\widetilde{f}_n(\mathbf{\Delta x}, \mathbf{\Delta y})\}$ (similarly as in Eq. (4)), with the knowledge of the error probability p_e, can be estimated below (Sun & Reibman, 2001):

$$\begin{aligned} &E\left\{\widetilde{f}_n^{\,i}(\mathbf{\Delta x}, \mathbf{\Delta y})\right\} = \\ &(1-p_e)\left\{\left(\sum_{k=1}^{N} h_k E\left\{\widetilde{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k)\right\}\right) + \hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y})\right\} + p_e \cdot E\left\{\widetilde{f}_{n-1}^{\,i}\right\} \end{aligned} \tag{27}$$

It should be noted that in Eq. (27), even current residual data is correctively received, the hypothesis prediction source may be erroneous due to error propagation and incompleteness of error concealment. Substituting Eq. (26) and Eq. (27) into Eq (25), we get

$$\begin{aligned} &(\mathbf{\Delta x}^*, \mathbf{\Delta y}^*) \\ &= \arg\min_{(\mathbf{\Delta x}, \mathbf{\Delta y})\in \mathbf{S}} \left\{ \sum_{i=1}^{p} \left| \begin{array}{l} \sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) + \hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y}) - p_e E\left\{\widetilde{f}_{n-1}^{\,i}\right\} \\ -(1-p_e)\left(\sum_{k=1}^{N} h_k E\left\{\widetilde{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k)\right\} + \hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y})\right) \end{array} \right| \right\} \end{aligned} \tag{28}$$

Again, based on the triangular inequality $|A+B| \le |A| + |B|$ and the fact that all h_k's sum to 1.0, we change Eq. (28) to:

$$\begin{aligned} &(\mathbf{\Delta x}^*, \mathbf{\Delta y}^*) \\ &= \arg\min_{(\mathbf{\Delta x}, \mathbf{\Delta y})\in \mathbf{S}} \left\{ \sum_{i=1}^{p} \left(\sum_{k=1}^{N} h_k \left| \begin{array}{l} \left(\hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) + \hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y}) \right. \\ \left. -(1-p_e)\left(E\left\{\widetilde{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k)\right\} + \hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y})\right) - p_e E\left\{\widetilde{f}_{n-1}^{\,i}\right\} \right) \end{array} \right| \right) \right\} \end{aligned} \tag{29}$$

Note that MV's obtained from Eq. (29) will be sub-optimal with respect to that obtained from Eq. (28) due to a higher end-to-end distortion. However, this arrangement reduces the computing complexity since it divides the original optimization problem into N sub-problems which can be solved individually.

Similarly as in Section 3, considering that the error probability p_e is commonly less than 0.2 and that the magnitude of $r_n^i(\mathbf{\Delta x}, \mathbf{\Delta y})$ is usually smaller than the hypothesis signal $\hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k)$, we ignore the term $p_e \hat{r}_n^i(\mathbf{\Delta x}, \mathbf{\Delta y})$ and rearrange Eq. (29) to be:

$$\begin{aligned}&(\mathbf{\Delta x}^*, \mathbf{\Delta y}^*)\\ &= \arg \min_{(\mathbf{\Delta x},\ \mathbf{\Delta y}) \in \mathbf{S}} \left\{ \sum_{i=1}^{p} \left[\sum_{k=1}^{N} h_k \left| \left(\hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) - (1-p_e) E\left\{ \tilde{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right\} - p_e E\left\{ \tilde{f}_{n-1}^{i} \right\} \right) \right| \right] \right\}\end{aligned} \tag{30}$$

Expressing the power (variance) of the prediction residual signals in MHC as:

$$\sigma_n^2(\mathbf{\Delta x}, \mathbf{\Delta y}) = \sum_{i=1}^{p} \left(f_n^{\ i} - \sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right)^2, \tag{31}$$

the following constrained optimization problem is formulated to better compromise between coding efficiency and error resiliency:

$$\min_{(\mathbf{\Delta x},\ \mathbf{\Delta y}) \in \mathbf{S}} \left\{ \sum_{i=1}^{p} \left[\sum_{k=1}^{N} h_k \left| \left(\hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) - (1-p_e) E\left\{ \tilde{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right\} - p_e E\left\{ \tilde{f}_{n-1}^{i} \right\} \right) \right| \right] \right\} \tag{32}$$

$$\text{Subject to : } \sigma_n^2(\mathbf{\Delta x}, \mathbf{\Delta y}) = \sum_{i=1}^{p} \left(f_n^{\ i} - \sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right)^2 < T$$

Obviously, the threshold T determines the tradeoff between coding efficiency and error resiliency. It is not straightforward enough to see from Eq. (32) how the T impacts upon finding error resilient MV for each hypothesis. For more clarity, Eq. (31) is re-written as:

$$\sigma_n^2(\mathbf{\Delta x}, \mathbf{\Delta y}) = \sum_{i=1}^{p} \left[\begin{aligned} &\sum_{k=1}^{N} h_k^2 \left(f_n^{\ i} - \hat{f}_{n-k}^{i,j} \right)^2 \\ &+ 2 \sum_{k=1,\ k \neq q,}^{N} \sum_{q=1}^{N} h_k h_q \left(f_n^{\ i} - \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right) \left(f_n^{\ i} - \hat{f}_{n-q}^{i,j}(\Delta x_q, \Delta y_q) \right) \end{aligned} \right]. \tag{33}$$

Define

$$\delta_k^2(\Delta x_k, \Delta y_k) \equiv \sum \left(f_n^{\ i} - \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right)^2 \tag{34}$$

which represents the autocorrelation (or, variance) of the motion residual signal and is always lager than the cross-correlation term (a Gaussian residual signal of zero-mean is

assumed). Therefore, for a given hypothesis number k and for each $q = 1 \sim N$, $q \neq k$, the following inequality will hold,

$$h_k^2 \delta_k^2 \geq \sum_{i=1}^{p} \left[h_k h_q \left(f_n^{\ i} - \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right) \left(f_n^{\ i} - \hat{f}_{n-q}^{i,j}(\Delta x_q, \Delta y_q) \right) \right] \tag{35}$$

Hence, $\sigma_n^2(\mathbf{\Delta x}, \mathbf{\Delta y})$ in Eq. (33) can be expressed in terms of $\delta_k^2(\Delta x_k, \Delta y_k)'s, k = 1 \sim N$:

$$\sigma_n^2(\mathbf{\Delta x}, \mathbf{\Delta y}) \leq \sum_{k=1}^{N} h_k^2 \delta_k^2 + 2 \sum_{k=1}^{N} h_k^2 (N-k) \delta_k^2 = \sum_{k=1}^{N} h_k^2 \delta_k^2 (1 + 2N - 2k) \tag{36}$$

Hence, a constraint T on $\sigma_n^2(\mathbf{\Delta x}, \mathbf{\Delta y})$ can be decomposed into separate constraints $T_k's$ on $\delta_k^2(\Delta x_k, \Delta y_k)'s, k = 1 \sim N$. In other words, there exists a mapping between (T_1, T_2, …T_N) and T via

$$T = \sum_{k=1}^{N} h_k^2 T_k (1 + 2N - 2k), \tag{37}$$

which is similar to Eq. (36). We can specify the value of the total threshold T indirectly via individual $T_k's$ to tradeoff between coding efficiency and error resiliency separately for each hypothesis. That is, we divide the original constrained optimization problem into N sub-problem, which is much simpler.

The proposed error resilient motion estimation algorithm for a given weighting vector $\boldsymbol{h}$ is now summarized as follows. First, set a constraint $\delta_k^2(\Delta x_k, \Delta y_k) \leq T_k$ for each reference frame f_{n-k} for motion estimation and then choose among the MVs, which satisfy the above constraint, the one that minimizes

$$\sum_{i=1}^{p} \left| \left(\hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) - (1 - p_e) E\left\{ \tilde{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k) \right\} - p_e E\left\{ \tilde{f}_{n-1}^{i} \right\} \right) \right| \tag{38}$$

The above procedure is performed for each hypothesis k, k= 1~N. Finally, generate the motion prediction residual by subtracting the weighted hypothesis $\sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j}(\Delta x_k, \Delta y_k)$ from the original video data.

6. Adaptive multi-hypothesis coding

In Section 5, error resilient motion estimation for a given set of hypothesis-weighting coefficients $\boldsymbol{h}$ is discussed. That is, $\boldsymbol{h}$ remains constant along the whole video. Now, in this section, we explore the advantage of varying the weighting coefficients $\boldsymbol{h}$, frame by frame, to further enhance error resiliency of the transmitted videos, according to the channel packet loss rate and video contents.

Before illustrating how to estimate the optimal $\boldsymbol{h}$ for error-prone video transmission, the power of the channel distortion signal is derived similarly as in Section 5:

$$\begin{aligned}\sigma_{\tilde{n}\tilde{n}}^2 &= \sum_{i=1}^{p} E\left\{\left(\hat{f}_n^i - \tilde{f}_n^i\right)^2\right\} \\ &= \sum_{i=1}^{p}\left(p_e \cdot E\left\{\left(\hat{f}_n^i - \tilde{f}_{n-1}^i\right)^2\right\} + (1-p_e)\cdot E\left\{\left[\hat{f}_n^i - \left(\hat{r}_n^i + \sum_{k=1}^{N} h_k \tilde{f}_{n-k}^{i,j}\right)\right]^2\right\}\right)\end{aligned} \tag{39}$$

The i^{th} term in summation can be reformulated as follows.

$$\begin{aligned}& p_e \cdot E\left\{\left(\hat{f}_n^i + \hat{f}_{n-1}^i - \hat{f}_{n-1}^i - \tilde{f}_{n-1}^i\right)^2\right\} + (1-p_e)\cdot E\left\{\left[\sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j} - \sum_{k=1}^{N} h_k \tilde{f}_{n-k}^{i,j}\right]^2\right\} \\ &= p_e \cdot E\left\{\left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)^2\right\} + p_e \cdot E\left\{\left(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i\right)^2\right\} + 2p_e \cdot E\left\{\left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)\left(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i\right)\right\} \\ &\quad + (1-p_e)\cdot E\left\{\left[\sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j} - \sum_{k=1}^{N} h_k \tilde{f}_{n-k}^{i,j}\right]^2\right\} \\ &= p_e \cdot \left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)^2 + p_e \cdot E\left\{\left(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i\right)^2\right\} + 2p_e \cdot \left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)\left(\hat{f}_{n-1}^i - E\left\{\tilde{f}_{n-1}^i\right\}\right) \\ &\quad + (1-p_e)\cdot E\left\{\left[\sum_{k=1}^{N} h_k \hat{f}_{n-k}^{i,j} - \sum_{k=1}^{N} h_k \tilde{f}_{n-k}^{i,j}\right]^2\right\}\end{aligned} \tag{40}$$

It is further assumed that $\left(\hat{f}_{n-k}^i - \tilde{f}_{n-k}^i\right)$ and $\left(\hat{f}_{n-q}^i - \tilde{f}_{n-q}^i\right)$, $k \neq q$, are independent. Then, the channel distortion power in Eq.(39) can be approximated as:

$$\sigma_{\tilde{n}\tilde{n}}^2 \cong \sum_{i=1}^{p}\left(\begin{array}{c} p_e \cdot \left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)^2 + p_e \cdot E\left\{\left(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i\right)^2\right\} + 2p_e \cdot \left(\hat{f}_n^i - \hat{f}_{n-1}^i\right)\left(\hat{f}_{n-1}^i - E\left\{\tilde{f}_{n-1}^i\right\}\right) \\ + (1-p_e)\cdot \sum_{k=1}^{N} h_k^2 E\left\{\left(\hat{f}_{n-k}^{i,j} - \tilde{f}_{n-k}^{i,j}\right)^2\right\} \end{array}\right) \tag{41}$$

Some notes about Eq. (41) are emphasized below. First, for given MVs, the first moment $E\left\{\tilde{f}_{n-1}^i\right\}$ can be estimated by using Eq. (27). Secondly, Eq. (41) is a recursive formula, meaning that the channel distortion power of the previous N frames, i.e., $E\left\{\left(\hat{f}_{n-k}^i - \tilde{f}_{n-k}^i\right)^2\right\}$, $k = 1 \sim N$, are used in estimating the channel distortion power of the current frame. Finally, if a strategy other than zero-motion is adopted for error concealment, the first two terms in Eq. (41) will change accordingly and may result in a lower channel distortion power. However, discussions about finding proper concealment strategies that are able to minimize channel distortion power are beyond the scope of this chapter.

The last term in Eq. (41) reveals that it is possible to find one combination of h_k's such that the channel distortion power is minimized. This problem can be formulated as follows:

$$\begin{aligned} &\text{minimize} \quad \left(1/2\right) \boldsymbol{h}^T \mathbf{D} \boldsymbol{h} \\ &\text{subject to} \quad \mathbf{1}^T \boldsymbol{h} = 1 \quad \text{and} \quad \boldsymbol{h} \succeq 0 \end{aligned} \tag{42}$$

where **D** is an $N \times N$ diagonal matrix whose elements on diagonal, denoted as $\sigma^2_{\tilde{n}_k \tilde{n}_k}$, is

$$\sum_{b=1}^{M} \sum_{i=1}^{p} E\left\{ \left(\hat{f}^{b,i,j}_{n-k} - \tilde{f}^{b,i,j}_{n-k} \right)^2 \right\}, \tag{43}$$

where b is the MB index and M is the total number of MBs in a video frame.

The problem in Eq. (43) is a convex optimization problem and the Karush-Kuhn-Tucker (KKT) conditions can be applied to find a solution of Eq. (42). Here, we consider the case of N = 3 for better understanding:

$$\begin{bmatrix} \sigma^2_{\tilde{n}_1 \tilde{n}_1} & 0 & 0 \\ 0 & \sigma^2_{\tilde{n}_2 \tilde{n}_2} & 0 \\ 0 & 0 & \sigma^2_{\tilde{n}_3 \tilde{n}_3} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} + v \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{44}$$

where v is a Lagrange multipliers. Equation (44) is used to estimate the optimal value of h_k, k = 1~3, with a constraint of $h_1 + h_2 + h_3 = 1.0$.

$$h_k = \frac{\sigma^2_{\tilde{n}_i \tilde{n}_i} \sigma^2_{\tilde{n}_l \tilde{n}_l}}{\sigma^2_{\tilde{n}_1 \tilde{n}_1} \sigma^2_{\tilde{n}_2 \tilde{n}_2} + \sigma^2_{\tilde{n}_1 \tilde{n}_1} \sigma^2_{\tilde{n}_3 \tilde{n}_3} + \sigma^2_{\tilde{n}_2 \tilde{n}_2} \sigma^2_{\tilde{n}_3 \tilde{n}_3}}, \quad \text{for } i \neq k \neq l \tag{45}$$

Note that the determination of optimal ***h*** is based on the availability of MVs for all the MBs in a frame. That is, multi-hypothesis motion estimation based on the conventional algorithm is performed first, then Eq. (45) (as well as other related formula) is used to figure out the optimal hypothesis-weighting coefficients ***h****, and finally these N hypotheses are combined to estimate the prediction. By the way, the overhead for transmitting ***h*** information is less and can be ignored.

7. Experimental results

7.1 Experiments for ERME and ERMD

The proposed ERME and ERMD algorithms are implemented on H.264/AVC test model JM 9.3 with rate control being enabled to meet channel bandwidth constraint. The number of reference frame and the search range for motion estimation are 3 and ±16 pixels, respectively. There are a total of 100 CIF frames for each test sequence. The performance of the proposed ERME algorithm is verified first by encoding the first frame as I picture and the rest as P pictures without intra-coded MBs therein. To evaluate the proposed ERME algorithm, mode decision for MBs in P pictures is purposely disabled, that is, only the block size of 16x16 pixels is considered. The setting of the threshold *Thd1* in Eq. (13) needs to be explained further. Theoretically, by adjusting *Thd1*, one can explore the trading behaviours

between error resiliency and coding efficiency of MVs. For the purpose of analysis, let $Thd1 = \alpha \cdot \hat{\sigma}^2$, where α is a pre-set constant and $\hat{\sigma}^2$ is derived via dynamic MB analysis. For better understanding, we let $\hat{\sigma}^2$ be equal to $\gamma^{\min}$ (previously stated in Eq. (12)), the smallest residual energy that can be achievable via conventional motion estimation process. Henceforth, α is selected to be larger than 1. It follows that a larger α implies a larger sacrifice on matching optimality (in a metric of square errors), or, the coding efficiency. Note that *Thd1* is dynamically varied on the MB base to account for the variation in video contents.

Figure 2 illustrates the PSNR performance at varying α and transmission bit rates. Curves marked with "alpha_*x*" represent the results obtained by setting α to *x*, e.g., "alpha_10" means $\alpha = 10$. On the other hand, curves annotated with "alpha_*" represent the results obtained by using conventional ME method (i.e., no consideration on error resiliency). It is assumed that each transmitted packet contains a slice which is composed of a row of consecutive MBs. When errors occur during transmission, the method of zero motion is adopted for error concealment at receiver, i.e., the damaged MB is replaced with that at the same location in the previous frame. Different error patterns for each packet loss rate (5% or 15%) are simulated to obtain an ensemble average PSNR of the reconstructed video.

From the figures, it is observed that with an increasing α , the coding efficiency decrease (i.e., less PSNR at a given bit rate) as expected when PSNRs are measured at encoder's local decoder (i.e., error-free scenario, (a)(b)). This is reasonable since the number of MV candidates satisfying the constraint in Eq. (12) is increased and all of them lead to residual energy larger than $\gamma^{\min}$. The sacrificed coding efficiency would lead to a gain on error resiliency, if the PSNRs are measured at receiver's decoder (i.e., error-prone scenario, (c)(d)). Take the curve of $\alpha = 2$ and $\alpha = 6$ for comparison, there exists a crossing point where the average PSNR for $\alpha = 6$ becomes better than that for $\alpha = 2$. However, this is not the similar case (no distinct crossing point exists) between $\alpha = 10$ and $\alpha = 20$.

It is observed from Fig, 3(a) that the anchor's face encounters severe distortion due to the lack of error resiliency for conventional ME algorithms. On the other hand, distortions of the anchor's face in (b) and (c) are much less than that in (a) of Fig. 3, due to the enhancement of error resiliency for increasing α . However, increasing quantization errors in other areas (e.g., words in the background) may lower down the total PSNR performance. This situation becomes clearer when D_s dominates D_c (e.g., $\alpha = 20$ not shown here).

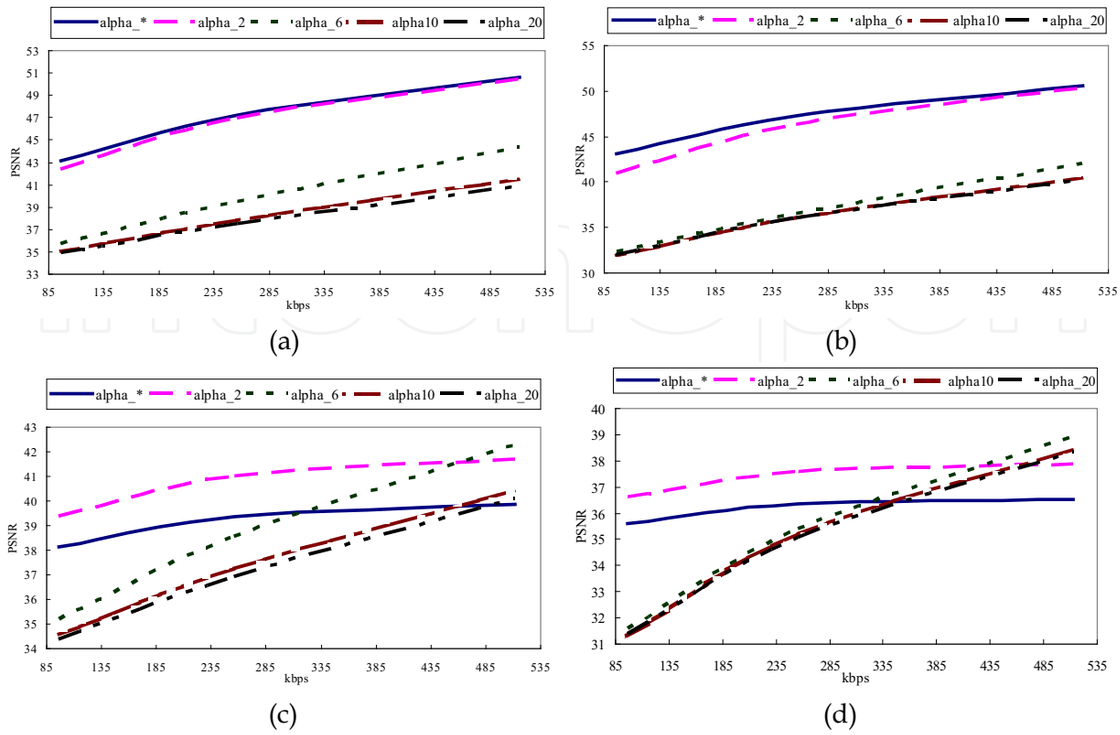

Fig. 2. Illustration of compromising error resiliency with coding efficiency of the proposed ERME algorithm. (a)(b) Reconstructed "Akiyo"at local decoder. (c)(d) Reconstructed "Akiyo"at decoder. The packet loss rate is (a)(c) 5% and (b)(d) 15%, respectively.

(a) Traditional ME (b) $\alpha = 2$ (c) $\alpha = 10$

Fig. 3. Visual quality at varying α, showing different compromises between D_s and D_c. The bit rate is 256 kbps and the packet loss rate is 5%.

Another experiment of compromising error resiliency with coding efficiency is conducted for the video "Foreman" (with high motion), which is also not shown here. In comparison with the experiments for "Akiyo" (Fig. 2), the PSNR crossing point is advanced to a lower bit rate for high motion videos. We can temporarily come to a conclusion that α can be chosen smaller (e.g., 2) for static videos while larger (e.g., ≥ 6) for high motion videos.

Remind that the choice of $\hat{\sigma}^2 = \gamma^{\min}$ is only for comparison purpose. Practically, $\hat{\sigma}^2$ should be derived by a fast and simple analysis for each MB, e.g., by summing the square of the temporal difference with respect to the previous frame. This method is certainly computationally cheaper and makes the threshold *Thd1* MB-adaptive.

Next, rate-distortion performance of our proposed ERME+ERMD algorithm is illustrated. We relax the allowable modes to include 16×16, 8×16, 16×8, 8×8, and intra 16×16. Notice that since we focus on finding MVs which can improve robustness to transmission errors of the compressed videos, the methods for compressing I-pictures here conforms to that recommended in the H.264/AVC standard. The integration of both techniques means that for each mode in $\boldsymbol{M}$, except the intra 16×16, the ERME algorithm is first applied to find the respective MVs, then the ERMD algorithm, with embodiment in term of Eq. (16), is applied to select the optimal mode among $\boldsymbol{M}$. In the following experiments, a value of $\alpha = 2$ is selected for ERME algorithm.

In Fig. 4, rate-distortion performances of several variational methods are compared. Three test sequences are chosen to represent typical videos of high motion (e.g., "Stefan"), moderate motion (e.g., "Foreman"), and static (e.g., "Akiyo") contents. In these figures, curves annotated with "MEO-MDO" represent the proposed "ERME+ERMD" algorithm, those annotated with "MEN-MDO" represent "ME+ERMD" algorithm, and curves annotated with "MEN-MDN" represent the traditional "ME+MD" algorithm adopted in H.264/JVC JM 9.3 model. For packet loss rate of 5% and 15%, the proposed ERME+ERMD improves error resiliency of the compressed video significantly (by 1~7 dB). These experiments also verify that ME+ERMD algorithm is not sufficient to support transmission robustness. The ERME algorithm is promising to provide additional support to complement this deficiency.

In Table 1, statistics of coding modes finally chosen by conventional H.264/AVC and the proposed ERME+ERMD algorithm are listed, which are obtained after gathering the related information for three different video sequences under different bit rates and a given 5% packet loss rate. It is observed that although ERME is able to prevent compressed videos from temporal error propagation, which has been verified in Fig. 2, the most efficient mode to prevent error propagation is "intra". The percentage is higher for higher bit rates. This behavior is reasonable since the additional bit rate will gain more benfits from intra-coding than from finer quantization.

7.2 Experiments for MHC

The multi-hypothesis coding algorithm is also implemented based on the H.264 video coding standard. Also, the first frame is coded as I frame and the rest are coded as P frames without any intra-coded blocks in them. It is assumed that a packet contains of a row of MBs.

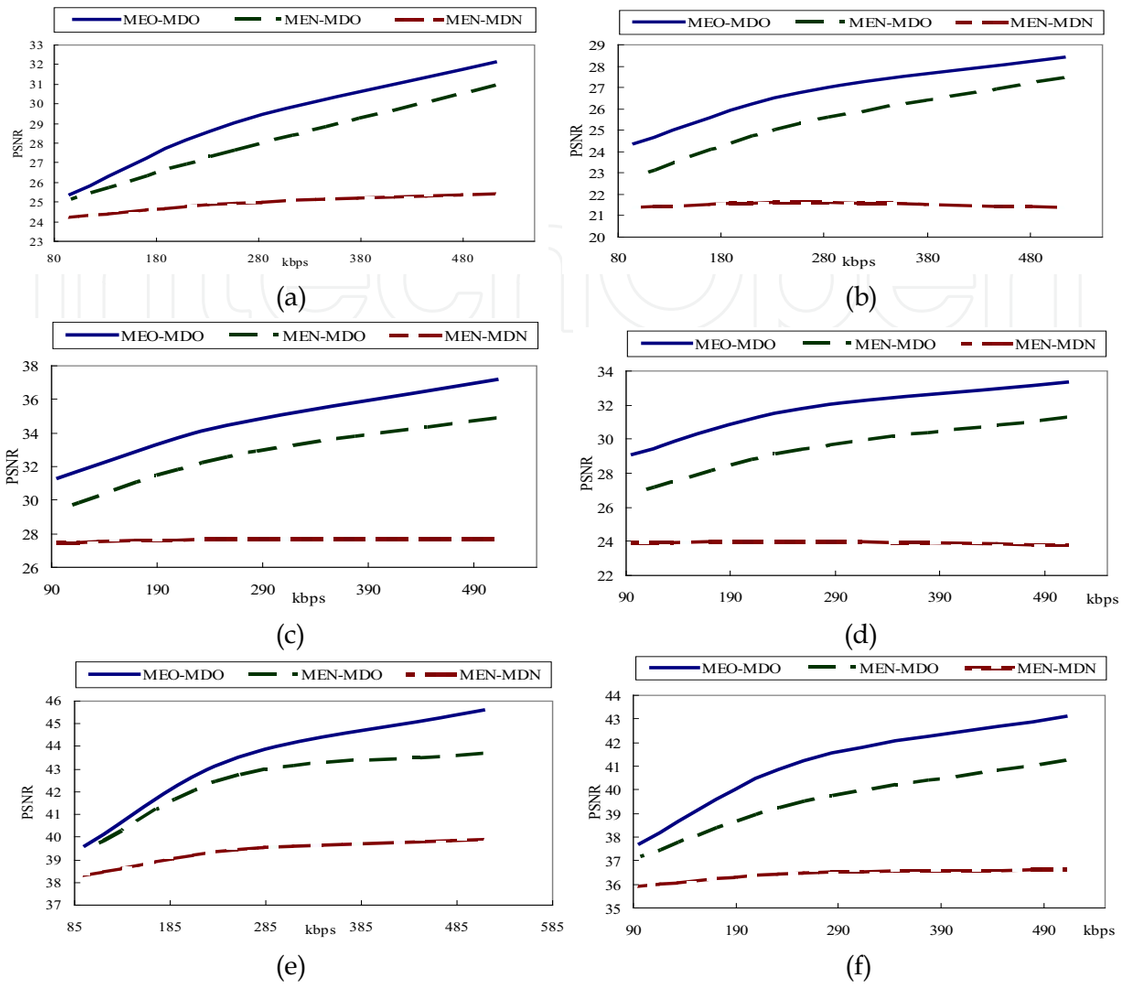

Fig. 4. Rate-distortion performances of reconstructed video at decoder for ERME+ERMD (MEO-MDO), ME+ERMD (MEN-MDO), and ME+MD (MEN-MDN) methods. (a)(b) "Stefan", (c)(d) "Forman", (e)(f) "Akiyo". (a)(c)(e) : PLR=5%, (b)(d)(f): PLR=15%.

Bit rate = 128kbs, Packet loss rate = 5%						
Modes	Foreman`		Stefan		Mobile	
	H.264 (%)	Proposed (%)	H.264 (%)	Proposed (%)	H.264 (%)	Proposed (%)
16x16	47.6	22.3	52.3	22.1	54.1	31.9
16x8	20.5	3.4	19.1	1.8	23.2	6.2
8x16	24.0	3.0	15.9	2.3	22.3	8.6
8x8	0	0	0	0	0	0
Intra	7.9	71.3	12.7	73.8	0.4	53.3

Table 1. Statistics of coding modes chosen by H.264 and the proposed ERME+ERMD.

First, the performance of the proposed error resilient motion estimation algorithm for MHC is evaluated. According to Eq. (38), we have decomposed the multi-hypothesis problem into N single-hypothesis problem. Let the threshold T_k for each hypothesis be in a form of $x \cdot \delta_k$, where δ_k is the energy of the prediction errors based on the MV (i.e., Δx_k and Δy_k) obtained by using the conventional motion estimation algorithm and x is a scale factor. Obviously, for each hypothesis, the threshold T_k is capable of adapting to video contents block-wise rather than frame-wised. On the other hand, adjustment of x makes us able to control the degree of compromise between coding efficiency and error resiliency. It is also easy to observe that the larger scaling factor x is chosen, the larger source distortion D_s is incurred, i.e., the more coding efficiency is sacrificed. Note that the above-mentioned way of getting δ_k is only for comparison purpose. One more practical method is to let δ_k be the energy of the residuals obtained via zero-motion compensation. Note that zero-motion compensation (or direct frame-difference) incurs less computational load and is hence more feasible.

In Figs. 5 and 6, rate-distortion performances in both the error-free and error-prone transmission environments between the proposed error resilient motion estimation and the conventional H.264 motion estimation techniques, are compared. The curves with "alpha_x" represent the results obtained by varying x in $T_k = x \cdot \delta_k$ to account for different levels of tradeoffs between coding efficiency and error resiliency. The curve marked with "alpha_*" represents the conventional ME algorithm.

From the figures, it is observed that in error-free environments, the proposed algorithm always results in a poorer performance than the conventional ME algorithm. This situation is even worse when x is increased (thus T_k is increased). It is observed that with packet losses, our proposed algorithm improves videos quality at higher bit rates. A larger T_k will result in a better error resiliency (up to 1 dB). At low bit rates, since the coding efficiency is quite sacrificed, the gain in error resiliency is not enough to overcome the former loss. It is also observed that similar performances are obtained for different weighting coefficients $\boldsymbol{h}$.

Next, the performance of our adaptive multi-hypothesis coding technique is to be evaluated. The number of hypotheses is 3 and the quantization parameter (QP) remains fixed during encoding a video. The packet loss rate is set to 5% and PSNR is measured between the reconstructed frames at local decoder and the reconstructed frames at receiver's decoder. Since the accuracy of the estimated channel distortion power is essential to the determination of the hypothesis-weighting coefficients to improve error resiliency, the accuracy of Eq. (41) is evaluated first. In experiments not shown here, the average difference between the estimated (via Eq.(41)) and the measured PSNRs for the high-motion video STEFAN is no more than 3.16 %.

Figure 7 relates to the experiment results of adaptive multi-hypothesis coding based on Eqs. (42) and (45). Notice that now MVs for each MB are obtained by using the conventional motion estimation algorithm. The two-state Gilbert channel as described in Zorzi et al. (1997) is used to simulate error conditions of packet error rate = 10% and average packet-error-burst length = 18.

In Fig. 7, the curves marked with "Adaptive" represent our proposed algorithm and curves marked with "Fixed" represent the conventional algorithm (i.e., constant and even weighting coefficients). From the figure, it is observed that our proposed algorithm really enhances the error resiliency of the multi-hypothesis video coding technique by up to 1 dB.

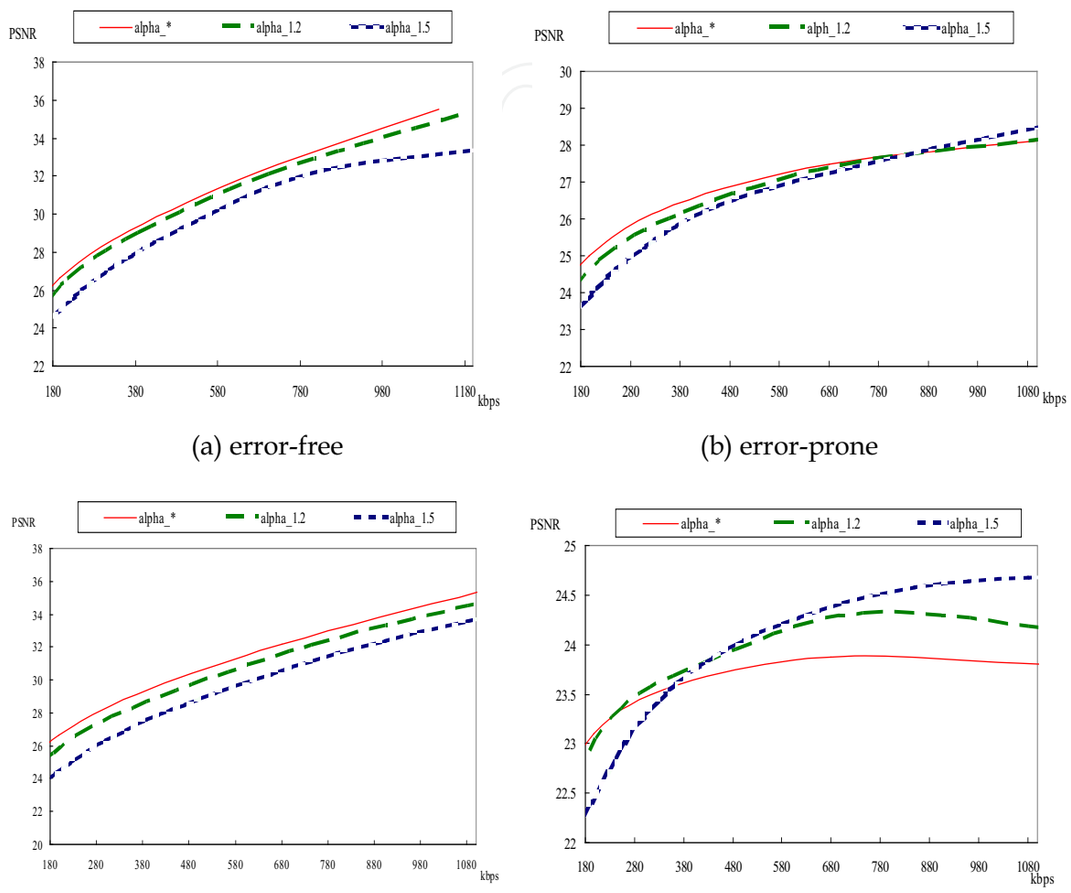

Fig. 5. Rate-distortion performance comparison with hypothesis-weighting coefficients (1/3, 1/3, 1/3) along the whole video "STEFAN". The packet loss rate for (a)(b) and (c)(d) is 5% and 15%, respectively.

8. Conclusion

In this chapter, error resilient coding techniques considering end-to-end distortions for videos transmitted on error-prone channels are discussed. The main concept of the algorithms is to decompose end-to-end distortions into two parts: source distortion relating to coding efficiency and channel distortion relating, on the other hand, to error resiliency. Most often, these two objectives are intrinsically mutual conflicting under a target bit rate. Hence it needs to make a proper compromise between them.

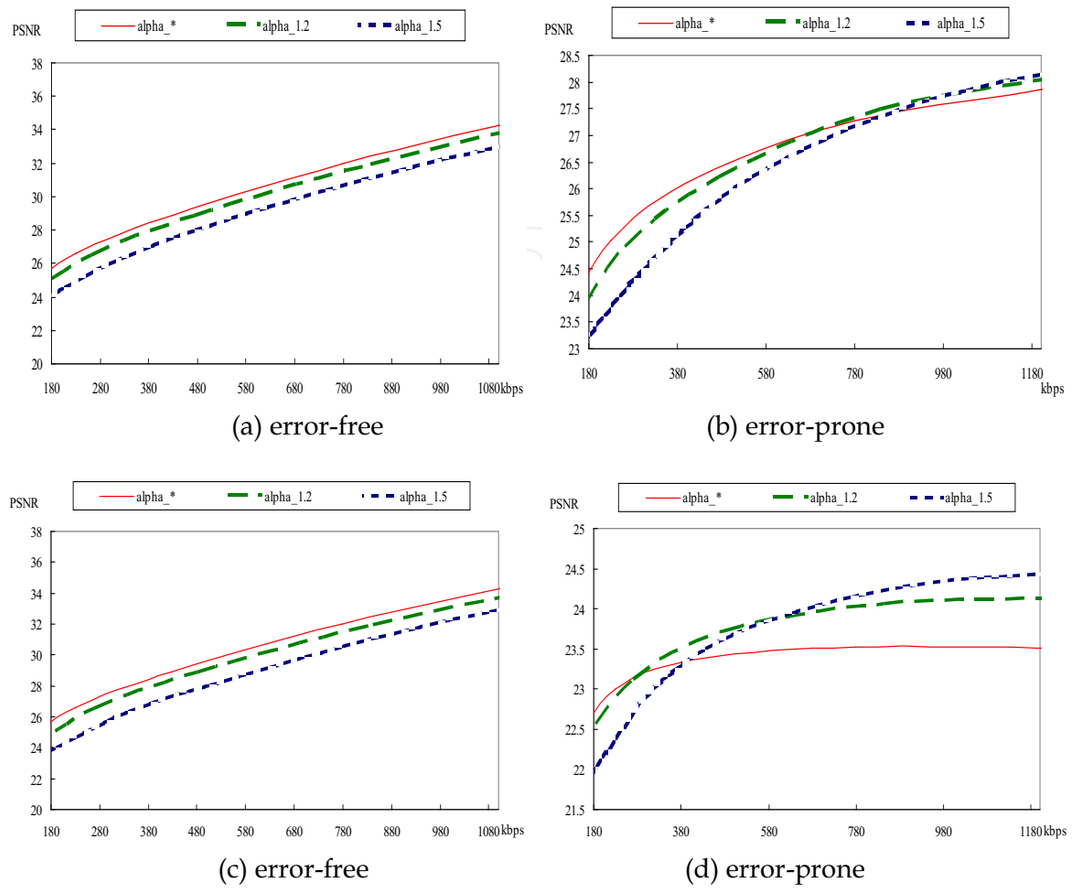

Fig. 6. Rate-distortion performance comparison with hypothesis-weighting coefficients (0.7, 0.2, 0.1) along the whole video "STEFAN". The packet loss rate for (a)(b) and (c)(d) is 5% and 15%, respectively.

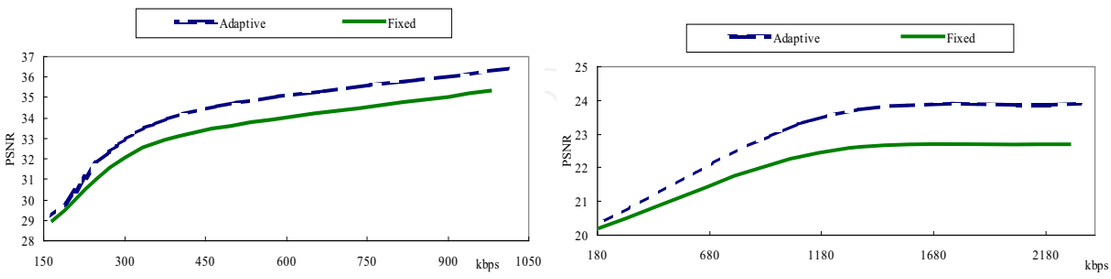

Fig. 7. Performance of the adaptive multi-hypothesis coding technique. The average packet-error-burst length is 18 and the packet-error-rate is 10 %.

In accordanc with the above concept, we propose ERME algorithms for both traditional H.264/AVC and multi-hypothesis coding architectures to suppress the temporal error propagation effectively with relatively low computational overhead. The similar concept is also applied to inter mode selection of H.264/AVC and to adaptive multi-hypothesis coding by finding out the relationship between the end-to-end distortions and the coding parameters, e.g., coding modes or hypothesis weighting coefficients before solving the optimization problem. Experiment results verify the accuracy of the proposed end-to-end distortion model in respective area.

The performance of the ERME algorithm is sunstantially affected by the constraint that controls different degrees of compromise between coding efficiency and error resiliency. How to set the constraint more accurately is a quite important issue that needs to be investigated further. Besides, integrating the ERME and adaptive hypothesis-weighting algorithms for further enhancement of error resiliency is not done yet. Note that the premise of the ERME is a given weighting $\boldsymbol{h}$, while adaptive hypothesis-weighting relies on the availability of MVs, seemingly a chicken-and-egg problem. A possible way is to compute the optimal $\boldsymbol{h}$ for frame t+1 based on the MVs of frame t by assuming that consecutive frames have strong similarity on MVs or channel distortion power, thought it might be violated due to large motion or scene change.

9. References

Boyd, Stephen & Vandenberghe, Lieven (2004). *Convex Optimization*, Cambridge University Press.

Chang, Pao-Chi; Lee, Tien-Hsu; Chen, Jhin-Bin & Tsai, Ming-Kuang (2005). Encoder-originated error resilient schemes for H.264 video coding. *Proceedings of 18th IPPR Conference on Computer Vision, Graphics and Image Processing,* pp. 406–412, Aug. 2005.

Cote, G. & Kossentini, F. (1999). Optimal intra coding of blocks for robust video communication over the Internet. *Signal Processing: Image Commu*., Sep.1999,pp. 25-34.

Eisenberg, Yiftach; Zhai, Fan; Papppas, Thrasyvoulos N.; Berry, Randall & Katsaggelos, Aggelos K. (2006). VAPOR: Variance-aware per-pixel optimal resource allocation. *IEEE Trans. on Image Processing*, Vol. 15, No 2, Feb. 2006, pp. 289-200.

Flierl, M.; Wiegand, T. & Girod, B. (2000). A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction. *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, Vol. 4067, pp. 238-249, June 2000.

Flierl, Markus; Wiegand, Thomsa & Girod, Bernd (1998). A locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction. *Proceedings of IEEE Conf. On Data Compression*, pp. 239-248, March 1998.

Flierl, Markus; Wiegand, Thomas & Girod, Bernd (2002). Rate-constrained multihypotheis prediction for motion-compensation video compression. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 12, No. 11, Nov. 2002, pp. 957-969.

Girod, Bernd (2000). Efficiency analysis of multihypothesis motion-compensated prediction for video coding. *IEEE Trans. on Image Processing*, Vol. 9, No. 2, Feb. 2000, pp.173-183.

Harmanci, Oztan & Tekal, A. Murat (2005). Stochastic frame buffers for rate distortion optimized loss resilient video communications. *Proceedings of IEEE Int'l Conf. Image Processing,* Vol. 1, pp. 789-792, Sep. 2005.

He, Zhihai; Cai, Jianfei & Chen, Chang Wen (2002). Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.12, No.6, Jun. 2002, pp. 511-523.

Kung, Wei-Ying; Kim, Chang-Su & Kuo, C. –C. Jay (2006) Analysis of multi-hypothesis motion compensation prediction for robust video transmission. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 16, No. 1, Jan. 2006, pp. 146-153.

Leontaris, A. & Cosman, P.C. (2004). Video compression for lossy packet networks with mode switching and a dual-frame buffer. *IEEE Trans. on Image Processing*, Vol. 13, July 2004, pp. 885-897.

Lie, Wen-Nung; Gao, Zhi-Wei & Liu, Tung-Lin (2006). Joint source-channel video coding based on the optimization of end-to-end distortion. *Proceedings of Pacific-Rim Symposium on Image and video Technology,* 2006.

Lin, Shunan & Wang, Yao (2002). Error resilience property of multihypothesis motion-compensated prediction. *Proceedings of IEEE International Conference on Image Processing*, Vol. 3, pp.545-548, June 2002.

Reyes, Gustavo de los; Reibman, Amy R.; Chang, Shih Fu & Chuang, Justin C.-I. (2000). Error-resilient transcoding for video over wireless channels. *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 6, June 2000, pp. 1063-1074.

Ringuest, Jeffrey L. (1992). *Multiobjective Optimization: Behavioral and Computational Considerations.* Kluwer Academic Publishers.

Stuhlmuller, Klaus; Farber, Niko; Link, Michael & Girod, Bernd (2000). Analysis of video transmission over lossy channel. *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 6, June 2000, pp. 1012-1032.

Sullivan, G. J. (1993). Multi-hypothesis motion compensation for low bit rate video coding. *Proceedings of IEEE Int'l Conf. Acoustics, Speech, and Signal processing,* Vol. 5, pp. 437-440, Apr. 1993.

Sun, M.-T. & Reibman, A.R. (2001). *Compressed Video over Network*, Marcel Dekker, Inc. New York, Basel.

Wiegand, Thomas; Farber, Niko; Stuhlmuller, Klaus & Girod, Bernd (2000). Error-resilient video transmission using long-term memory motion-compensated prediction. *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 6, June 2000, pp. 1050-1062.

Xia, Minghui; Vetro, Anthony; Sun, Huifan & Liu, Bede (2004). Rate-distortion optimized bit allocation for error resilient video transcoding. *Proceedings of IEEE Int'l Symp. Circuits and Systems,* Vol. 3, pp. 945-948, May 2004.

Yang, Hua & Rose, K. (2005). Rate-Distortion optimized motion estimation for error resilient video coding. *Proceedings of IEEE Int'l Conf. Acoustics, Speech, and Signal processing,* Vol. 2, March 2005, pp. 173-176.

Zhan, R.; Regunathan, S. L. & Rose K. (2000). Video coding with optimal intra/inter mode switching for packet loss resilience. *IEEE Journal of Selected Areas in Commun.*, Vol. 18, No. 6, June 2000, pp. 966-976.

Zorzi, M.; Rao, K.R. & Milstein, L. B. (1997). ARQ error control for fading mobile radio channels. *IEEE Trans. On Vehicle Technology.*, Vol. 46, May 1997, pp. 445-455.

Multimedia

Edited by Kazuki Nishi

ISBN 978-953-7619-87-9

Hard cover, 452 pages

Publisher InTech

Published online 01, February, 2010

Published in print edition February, 2010

Multimedia technology will play a dominant role during the 21st century and beyond, continuously changing the world. It has been embedded in every electronic system: PC, TV, audio, mobile phone, internet application, medical electronics, traffic control, building management, financial trading, plant monitoring and other various man-machine interfaces. It improves the user satisfaction and the operational safety. It can be said that no electronic systems will be possible without multimedia technology. The aim of the book is to present the state-of-the-art research, development, and implementations of multimedia systems, technologies, and applications. All chapters represent contributions from the top researchers in this field and will serve as a valuable tool for professionals in this interdisciplinary field.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen