A joint encoder–decoder error control framework for stereoscopic video coding

Xinguang Xiang,*, Debin Zhao, Qiang Wang, Siwei Ma, Wen Gao

A very popular and widely accepted technique, the boundary matching algorithm (BMA), is proposed in [3] to select an optimal MV to substitute for the lost one. In [4], a technique which combines the overlapped motion compensation and the side match criterion is proposed. Chen et al. [5] propose a refined boundary matching algorithm (RBMA) to conceal different regions of a lost block with different motion vectors. In [6], a spatio-temporal boundary matching algorithm (STBMA) exploits both spatial and temporal smoothness properties of video signals is proposed to reconstruct the lost MV. On the other hand, error-resistant video coding is performed at the encoder side to suppress error propagations caused by transmission errors. Inserting more intra-coded macroblocks is an effective method to solve the problem but with the penalty on coding efficiency. Thus, how to adaptively insert intra-blocks to achieve better trade-off between coding efficiency and suppression of error propagations has also been widely exploited [2–6]. The simplest technique is temporal replacement (TR) [2], which utilizes the zero motion vector (MV) to reconstruct a lost macroblock (MB).
These techniques [2–12] are all designed for single-view video coding, so they will be inefficient if directly applied to SSVC, as the inter-view correlation is not considered and exploited. To our knowledge, several works [13–20] have been reported on error control for SSVC [13–15] or multi-view video coding [16–20]. In [13], an error concealment algorithm based on the projective transformation model is proposed to improve reconstruction results of lost blocks in stereoscopic images. A frame loss error concealment algorithm for SSVC is proposed in [14]. Tan et al. [15] propose an error-resilient layered framework for SSVC. Since stereoscopic video is a special two-view based multi-view video, some error control strategies [16–20] for multi-view video coding can also be utilized for SSVC. In [16], a robust transmission scheme of multi-view video streams is proposed with flexible macroblock ordering and systematic LT codes. Chung et al. [17] conceal a lost block of multi-view video by choosing and combining the best candidate blocks in the temporally adjacent frames or the inter-view frames at the same time instance. A novel frame loss error concealment algorithm which utilizes motion information of pictures from other views is proposed in [18]. In [19], Fecker et al. simultaneously use information from surrounding image parts, temporally preceding and succeeding frames, and neighboring camera views for extrapolating known image samples into the lost area. An efficient error concealment algorithm with three error concealment modes for multi-view video sequences is proposed in [20]. All these techniques [13–20] are exploited only at the encoder side or at the decoder side, and the joint encoder–decoder error control for SSVC is further needed to get better error robustness performance.

In this paper, a joint encoder–decoder error control framework is proposed for SSVC. By extending our previous works [21, 22], error-resilient source coding, transmission network conditions, and error concealment are jointly considered to achieve better error robustness performance for SSVC. The proposed joint encoder–decoder error control framework includes two parts: an error concealment algorithm at the decoder side and a rate–distortion optimized error resilience algorithm at the encoder side. At the decoder side, an overlapped block motion and disparity compensation (OBMDC) based error concealment scheme is proposed for SSVC. The proposed error concealment algorithm adaptively utilizes the inter-view correlations and the temporal correlations to fill up the lost stereoscopic video contents. At the encoder side, a rate–distortion optimized error resilience algorithm for SSVC is presented. In the rate–distortion optimized error resilience algorithm, we first propose that inter-view refreshment should be utilized for SSVC to suppress error propagations. Second, an end-to-end distortion model is derived for SSVC which jointly considers network conditions, inter-view refreshment, and error concealment tools used at the SSVC decoder side. Finally, based on the derived end-to-end distortion model, the rate–distortion optimized error resilience algorithm is presented to adaptively select inter-view, inter- or intra-coding for SSVC.

The rest of this paper is organized as follows: Section 2 illustrates the whole structure of the proposed joint encoder–decoder error control framework for SSVC. The OBMDC based stereoscopic error concealment algorithm is detailed in Section 3. Section 4 describes the proposed rate–distortion optimized error resilience algorithm for SSVC. Section 5 reports and analyzes the experimental results. Finally, the conclusions are drawn in Section 6.

2. The structure of proposed joint encoder–decoder error control framework for SSVC

Fig. 1 shows the structure of proposed joint encoder–decoder error control framework for SSVC. At the stereoscopic video decoder side, an OBMDC based error concealment module, which adaptively utilizes inter-view correlations and temporal correlations, is proposed to fill up the lost stereoscopic video contents and suppress error propagations. At the stereoscopic video encoder side, a rate–distortion optimized error resilience scheme module is proposed to achieve better trade-off between coding efficiency and suppression of error propagations. Note that, in order to achieve better error robustness performance, an error concealment module is simulated at the encoder side. Thus the rate–distortion optimized error resilience scheme also considers the error concealment scheme used at the decoder side. The error concealment module and the rate–distortion optimized error resilience module compose the joint encoder–decoder error control framework for SSVC. In addition, some unequal error protection techniques can be utilized in channel coding.

The more detailed structure of proposed error-resilient stereoscopic video coding framework is further shown in Fig. 2. Although
DCP may bring error propagations as well as MCP in SSVC, stereoscopic videos can be protected from error propagations of the current view by appropriate DCP. Thus, the inter-view refreshment, which means that the stereoscopic encoder increases or decreases inter-view predicted blocks according to some criteria, is proposed to suppress error propagations for SSVC. The proposed rate–distortion optimized error resilience algorithm jointly considers inter-view refreshment, transmission network conditions, and error concealment tools used at the SSVC decoder side. Based on the rate–distortion optimized error resilience algorithm, the encoder can adaptively select inter-view, inter- or intra-coding for SSVC.

3. Overlapped block motion and disparity compensation based error concealment

The prediction structure of SSVC is shown in Fig. 3, where the left view is predicted only by MCP, called independent view, and the right view is predicted by both MCP and DCP, called inter-view predicted view.

Since the inter-view predicted view is predicted by both MCP and DCP, the error concealment scheme for the inter-view predicted view is different from that for the independent view. We propose that the neighboring available disparity vectors (DV) should be utilized to reconstruct the lost MBs as well as the neighboring available MVs. Then, a novel error concealment method is proposed based on overlapped block motion and disparity compensation (OBMDC), whose weights are determined by the boundary matching algorithm and viewpoints. Fig. 4 shows the flowchart of proposed error concealment algorithm for SSVC.

3.1. Selection of the replacing prediction vector

In SSVC, blocks in the inter-view predicted view may be predicted from the independent view with DVs by utilizing the inter-view correlations. Furthermore, the inter-view correlations can be utilized for error concealment. Error concealment techniques for the traditional single-view coding usually use the MVs of the lost MB's neighboring MBs to reconstruct the lost MB. However, in stereoscopic video coding, utilizing a correct disparity vector (DV) to reconstruct a lost MB usually can get better error concealment performance than utilizing an incorrect MV. Furthermore, when the inter-view correlations are stronger than temporal correlations, even utilizing a correct MV cannot reconstruct a MB.

---

**Fig. 2.** The more detailed structure of proposed error-resilient stereoscopic video coding framework.

**Fig. 3.** The prediction structure of SSVC.

**Fig. 4.** The flowchart of proposed error concealment algorithm.
better lost MB than utilizing a correct DV. Therefore we should utilize not only the MVs but also the DVs of the lost MB's neighboring MBs as candidate prediction vectors to reconstruct the lost MB.

A $16 \times 16$ MB can be divided into sub-blocks with different sizes for motion estimation or disparity estimation, among which usually the minimum sub-block size is $4 \times 4$. Hence a neighboring MB can provide at most four nearest MVs or DVs. As shown in Fig. 5, $MV^{1}_{T}, MV^{1}_{L}, MV^{1}_{TL}, MV^{2}_{T}, MV^{2}_{L}, MV^{2}_{BL}$ and $MV^{2}_{TR}$ (the maximum value of $i$ is 4) denote the nearest MVs of the top, top-left, top-right, bottom, bottom-left and bottom-right blocks, respectively if the blocks are temporally predicted. Similarly $DV^{1}_{T}, DV^{1}_{L}, DV^{1}_{TL}, DV^{2}_{T}, DV^{2}_{L}, DV^{2}_{BL}$ and $DV^{2}_{TR}$ (the maximum value of $i$ is 4) denote the nearest DVs of the top, top-left, top-right, bottom, bottom-left and bottom-right blocks, respectively if the blocks are predicted from the independent view.

We utilize the available neighboring prediction vectors as candidate prediction vectors for the lost MB. For each candidate prediction vector, we can reconstruct a candidate replacing MB from the reference frames utilizing motion compensation or disparity compensation. We utilize the boundary matching algorithm (BMA) [3] to choose an appropriate motion or disparity vector which is used as the motion or disparity vector of the lost MB from these candidates. The cost function of BMA is defined as the absolute difference between the external boundary of the lost macroblock in the current frame and the internal boundary of the replacing macroblock in the reference frame, and it is formulated as follows:

$$\text{Cost}_{\text{BMA}} = \sum_{x=x_{0}}^{x_{0}+15} \sum_{y=y_{0}}^{y_{0}+15} \left[ w_{T} \cdot |P(x, y) - P'(x + \delta, y)| + w_{B} \cdot |P(x, y_{0} + 16) - P'(x + \delta, y_{0} + 15 + \delta_t)| + \sum_{y'=y_{0}+1}^{y_{0}+15} |w_{L} \cdot |P(x_{0} - 1, y') - P'(x_{0} + \delta, y + \delta_t)| + \sum_{y'=y_{0}+1}^{y_{0}+15} |w_{R} \cdot |P(x_{0} + 16, y') - P'(x_{0} + 15 + \delta, y + \delta_t)| \right]$$

(1)

where $(x_{0}, y_{0})$ denotes the coordinate of the top-left pixel in the lost MB, and $(\delta, \delta_t)$ denotes the candidate vector. $P$ and $P'$ denote the pixels of the current and reference frames, respectively. The subscripts $T, B, L$, and $R$ are short for top, bottom, left, and right, respectively. The weight $w_2$ is set to 1, if the top neighboring MB in current frame is available. Otherwise, $w_2$ is set to 0. So are the definitions of $w_{PL}$, $w_{P_{BR}}$, and $w_{P_{LR}}$. The motion or disparity vector which results in the smallest cost is selected. And the corresponding MB from the reference frame determined by the vector can be used to replace the lost MB.

### 3.2. OBMDC based error concealment

In the previous subsection, a replacing MB for the lost MB is reconstructed with just one motion or disparity vector which utilizes either the temporal correlations or the inter-view correlations in the sequence of images. This is based on the assumption that each pixel of one MB undergoes uniform translational movement. However, the pixels may be unstructured and the movements may be irregular in some regions. In this case, the algorithm in previous subsection will be inefficient. Thus, we further propose an overlapped block motion and disparity compensation (OBMDC) technique to utilize both the temporal correlations and the inter-view correlations to deal with this problem.

For each available neighboring prediction vectors, as shown in Fig. 5, we reconstruct a candidate replacing MB from the reference frames. In the proposed OBMDC technique, we utilize these candidate replacing MBs to produce a new replacing MB, and each pixel in the new replacing MB is determined by a weighted average of the corresponding pixels from these candidate replacing MBs.

The largest weight is assigned to the optimal replacing MB which is selected by BMA, because the MB has the best block boundary matching MB. Although error concealment can fill up the lost video and suppress error propagations to a certain extent, the output video quality is still not satisfied. In this section, we propose a rate-
distortion optimized error resilience algorithm for SSVC to deal with the problem at the encoder side. First, we propose that inter-view refreshment should be utilized for SSVC to suppress error propagations. Second, an end-to-end distortion model for SSVC is derived which jointly considers transmission network conditions, inter-view refreshment, and error concealment tools. Finally, based on the derived end-to-end distortion model, we present the rate-distortion optimized error resilience algorithm to adaptively select inter-view, inter- or intra-coding for SSVC.

4.1. Inter-view refreshment

Similarly as inserting more intra-coded MBs for traditional single-view coding [7], we propose a novel refreshment scheme which is called inter-view refreshment for SSVC. Here, inter-view refreshment means that the stereoscopic encoder increases or decreases inter-view predicted blocks under some criteria to suppress error propagations. Although intra-coding can effectively suppress error propagations, it brings a big coding efficiency loss. Fortunately, in SSVC, we find that inter-view refreshment can provide a better trade-off between coding efficiency and suppression of error propagations.

First, inter-view refreshment can provide similar suppression of error propagations as inserting more intra-coded MBs at some conditions. If the reference blocks in the independent view are transmitted correctly, then the inter-view predicted blocks in the inter-view predicted view can be protected from error propagations. Furthermore, when the network conditions of the independent view is better than that of the inter-view predicted view or some unequal error protection techniques are utilized to make the packet loss rate of the independent view lower than that of the inter-view predicted view (especially when the packet loss rate of the independent view is very low), it can be expected that the independent view will have lower error propagations, where more inter-view predicted blocks should be encouraged to suppress error propagations. However, DCP may still bring error propagations as well as MCP in SSVC. Thus, when the network conditions of the independent view is worse, more intra-coding blocks are encouraged to be coded for the independent view and the DCP proportion for the inter-view predicted view should be reduced. When the network conditions of the two views are almost the same, smart decisions should be taken to adaptively select inter-view, inter- or intra-coding which result in a better trade-off between coding efficiency and suppression of error propagations. Therefore, inter-view refreshment can make the inter-view predicted view robust to errors. Second, inter-view refreshment is usually more efficient than inserting more intra-coded MBs in terms of coding efficiency because of the inter-view correlations existing in stereoscopic video. Thus, inter-view refreshment can provide a better trade-off between coding efficiency and suppression of error propagations.

Based on the above analysis, inter-view refreshment is an effective error-resilient coding technique for SSVC. Furthermore, SSVC with inter-view refreshment should adapt to error-prone network conditions to achieve better error resilience performance. For the inter-view predicted view, in contrast to traditional single-view error-resilient video coding, not only the network conditions (e.g., packet loss rate) of itself but also the network conditions of the independent view should be considered in error-resilient stereoscopic video coding. A rate-distortion optimized error resilience algorithm and the associated end-to-end distortion model are needed to realize the above coding strategy.

4.2. End-to-end distortion model for SSVC

From the above discussion, we know that inter-view refreshment can improve error resilience performance. However, how to realize inter-view refreshment efficiently is still a problem. In error-resilient single-view hybrid video coding, the end-to-end models [9–12,23] are often utilized to insert more intra-coded MBs. Similarly, if the end-to-end distortion distribution of the stereoscopic video can be estimated by an appropriate end-to-end distortion estimation algorithm, inter-view refreshment can be realized to get a better trade-off between coding efficiency and suppression of error propagations. In this subsection, we derive an end-to-end distortion model for SSVC to estimate the distortion distribution. In our derived end-to-end distortion model, the transmission network conditions for stereoscopic video, inter-view refreshment, and error concealment tools at the SSVC decoder side are jointly considered.

For pixel i in frame n at viewpoint view which references pixel j in frame ref(view, n), let $f_{view}^{i}$ denote the original value of the pixel i, and let $\tilde{f}_{view}^{i}$ and $\tilde{f}_{ref}^{i}$ denote the reconstructed values in the encoder and decoder, respectively. ref(view, n) represents the prediction reference frame, and it may come from the independent view or the inter-view predicted view decided by view, n, and coding mode. Let $f_{view}^{i}$ be the reconstructed residue in the encoder, thus

$$f_{view}^{i} = f_{ref}^{i} + f_{view}^{i}$$

When the pixel i is lost, it copies from pixel k in frame ec_ref(view, n), where ec_ref(view, n) represents the error concealment reference frame, and it also may come from the independent view or the inter-view predicted view decided by view, n, and the used error concealment algorithm. Let $p_{view}$ be the transmission error rate for viewpoint view. Then we can represent $f_{view}^{i}$ as

$$f_{view}^{i} = \begin{cases} \tilde{f}_{view}^{i} & \text{w.p. } 1 - p_{view} \\ \tilde{f}_{ec_{ref}}^{i} & \text{w.p. } p_{view} \end{cases}$$

Thus, we can derive the end-to-end distortion $d(view, n, i)$ for stereoscopic video as

$$d(view, n, i) = \mathbb{E}\left\{ (f_{view}^{i} - f_{view}^{i})^2 \right\}$$

$$= (1 - p_{view})\mathbb{E}\left\{ (f_{view}^{i} - f_{view}^{i} + f_{view}^{i})^2 \right\}$$

$$+ p_{view}\mathbb{E}\left\{ (f_{view}^{i} - \tilde{f}_{ec_{ref}}^{i})^2 \right\} \approx (1 - p_{view})$$

$$\times \mathbb{E}\left\{ (f_{view}^{i} - \tilde{f}_{view}^{i})^2 \right\} + (1 - p_{view})$$

$$\times \mathbb{E}\left\{ (\tilde{f}_{view}^{i} - \tilde{f}_{ec_{ref}}^{i})^2 \right\}$$

$$= (1 - p_{view})d_{view}(view, n, i) + (1 - p_{view})$$

$$\times d_{view}(view, n, i) + p_{view}d_{view}(ec_{ref}(view, n, i))$$

where $d_{view}(view, n, i)$, $d_{view}(ref(view, n, j))$, and $d_{view}(ec_{ref}(view, n, i))$ denote the source distortion, the error-propagation distortion from the reference frame, and the error concealment distortion, respectively. For the independent view, the source distortion comes from MCP and intra-coding, while for the inter-view predicted view, the source distortion comes from DCP, MCP and intra-coding. Obviously, the source distortion $d_{view}(view, n, i)$ can be estimated at the encoder side, so we need to calculate the error-propagation distortion $d_{view}(ref(view, n, j))$ and the error concealment distortion $d_{view}(ec_{ref}(view, n, i))$.

In order to calculate the error-propagation distortion and the error concealment distortion, an error concealment module is added at the encoder side. Especially, in the derived end-to-end model for SSVC, error concealment algorithms for stereoscopic video coding which utilize inter-view correlations are considered.
Since DCP is utilized for inter-view prediction, the error-propagation distortion model for the independent view and the inter-view predicted view are different. For the independent view, the error-propagation distortion only comes from itself, while for the inter-view predicted view, the error-propagation distortion not only comes from itself, but also comes from the independent view. We derive the error-propagation distortion from the reference frame \( \text{d}_{\text{ep}}(\text{view}, n, i) \) as

\[
d_{\text{ep}}(\text{view}, n, i) = E\left\{ \left( \hat{f}_{\text{view}, n} - \hat{f}_{\text{ref}(\text{view}), n} \right)^2 \right\} = (1 - p_{\text{view}}) E\left\{ \left( \hat{f}_{\text{view}, n} - \hat{f}_{\text{ref}(\text{view}), n} \right)^2 \right\} + p_{\text{view}} E\left\{ \left( \hat{f}_{\text{view}, n} - \hat{f}_{\text{ec}(\text{view}), n} \right)^2 \right\} \\
\approx (1 - p_{\text{view}}) E\left\{ \left( \hat{f}_{\text{ref}(\text{view}), n} - \hat{f}_{\text{ref}(\text{view}), n} \right)^2 \right\} + p_{\text{view}} E\left\{ \left( \hat{f}_{\text{view}, n} - \hat{f}_{\text{ec}(\text{view}), n} \right)^2 \right\}
\]

where \( \text{d}_{\text{ep}}(\text{view}, n, i) \) denotes the distortion between the reconstructed pixel values and the reconstructed error concealment pixel values in the encoder, which can be calculated at the encoder side. Thus, the error-propagation distortion of frame \( n \) can be derived from the error-propagation distortions of its reference frame and its error-concealment reference frame. We utilize a block-level distortion map to store error-propagation distortions of frames in each view. Error-propagation distortion of the first frame in each view is set to 0, because they are usually intra-frames. Then the error-propagation distortions of the following frames can be calculated frame by frame by formula (6).

The error concealment distortion is determined by the error concealment scheme at the decoder side. The error concealment distortion models for the independent view and the inter-view predicted view are different. For the independent view, the error concealment distortion only comes from itself, while for the inter-view predicted view, the error concealment distortion not only comes from itself, but also comes from the independent view. The error concealment distortion \( \text{d}_{\text{ec}}(\text{view}, n, \text{ec}_{\text{ref}(\text{view}), n}, i) \) can be derived as

\[
d_{\text{ec}}(\text{view}, n, i) = E\left\{ \left( \hat{f}_{\text{view}, n} - \hat{f}_{\text{ec}(\text{view}), n} \right)^2 \right\} \\
= E\left\{ \left( \hat{f}_{\text{view}, n} - \hat{f}_{\text{ec}(\text{view}), n} \right)^2 \right\} + E\left\{ \left( \hat{f}_{\text{ec}(\text{view}), n} - \hat{f}_{\text{ec}(\text{view}), n} \right)^2 \right\} \\
\approx E\left\{ \left( \hat{f}_{\text{view}, n} - \hat{f}_{\text{ec}(\text{view}), n} \right)^2 \right\} + E\left\{ \left( \hat{f}_{\text{ec}(\text{view}), n} - \hat{f}_{\text{ec}(\text{view}), n} \right)^2 \right\} \\
= \text{d}_{\text{ec},\text{ref}}(\text{view}, n, \text{ec}_{\text{ref}(\text{view}), n}, i) + \text{d}_{\text{ep}}(\text{ec}_{\text{ref}(\text{view}), n}, i)
\]

where \( \text{d}_{\text{ec},\text{ref}}(\text{view}, n, \text{ec}_{\text{ref}(\text{view}), n}, i) \) denotes the distortion between the original pixel values and the reconstructed error concealment pixel values in the encoder, which can also be calculated at the encoder side. \( \text{d}_{\text{ep}}(\text{ec}_{\text{ref}(\text{view}), n}, i) \) can be calculated by formula (6). Thus the error concealment distortion \( \text{d}_{\text{ec}}(\text{view}, n, i) \) can be calculated at the encoder side.

Note that the three distortion elements of the proposed end-to-end distortion model can be calculated at the encoder side. Thus the derived end-to-end distortion for stereoscopic video can be utilized for rate–distortion optimization.

In the derived end-to-end distortion model for SSVC, not only the error-propagation distortion caused by MCP is considered, but also the error-propagation distortion caused by DCP is considered. Thus with the estimation of the end-to-end distortion distribution for SSVC, inter-view refreshment can be utilized for error-resilient stereoscopic video coding. The derived end-to-end distortion model also considers transmission network conditions of stereoscopic video, i.e., different packet loss rates for different views. Furthermore, the derived end-to-end model can adapt to error concealment algorithms for stereoscopic video coding in the distortion calculation.

4.3. Rate–distortion optimized error resilience algorithm for SSVC

In this subsection, based on the derived end-to-end distortion model, the rate–distortion optimized error resilience algorithm is proposed to adaptively select inter-view, inter- or intra-coding for SSVC over error–prone networks.

The Lagrange multiplier indicates the relation between the distortion and the rate. Based on the derivation of Lagrange multiplier for single-view hybrid video coding [23], the source probability distribution can be approximated as uniform within each quantization interval, i.e., for each quantization interval \( \Delta \), the source distortion \( D_{\Delta} \) conforms to

\[
D_{\Delta} = \frac{\Delta^2}{12}
\]

and the rate \( R(\Delta) \) conforms to

\[
R(\Delta) = \frac{1}{2} \log_2 \left( \frac{\beta}{\Delta^2/12} \right)
\]

where \( \beta \) denotes a constant depending on the variance of the source.

In SSVC, the derivation of the new Lagrange multiplier for stereoscopic video coding over error–prone networks should consider both the packet loss distortion and the prediction structure of SSVC. Thus, according to (5) and (8), we can obtain

\[
D_{\Delta} = (1 - p_{\text{view}}) \frac{\Delta^2}{12} + (1 - p_{\text{view}}) D_{\text{ep}} + p_{\text{view}} D_{\text{ec}}
\]

where the error-propagation distortion \( D_{\text{ep}} \) and the error concealment distortion \( D_{\text{ec}} \) are both independent of the quantization interval \( \Delta \) of the current frame.

Combining (9) and (10), we can derive the new Lagrange multiplier

\[
\lambda_{\text{view}} = -\frac{dD_{\text{view}}}{dR} = -\frac{dD_{\Delta}}{dR} = (1 - p_{\text{view}}) \frac{2}{12} + (1 - p_{\text{view}}) D_{\text{ec}} = (1 - p_{\text{view}}) \lambda_0
\]

for SSVC, where \( p_{\text{view}} \) denotes the packet loss rate and \( \lambda_0 \) is the Lagrange multiplier used in H.264 or H.263. Notice that each view has its own \( \lambda_{\text{view}} \) decided by its packet loss rate \( p_{\text{view}} \).

Based on the derived end-to-end distortion model and the Lagrange multiplier, we propose a rate–distortion optimization algorithm to select the optimal coding mode for error–resilient stereoscopic video coding. According to (5) and (11), supposing \( \rho_i \) denotes prediction vectors for the macroblock \( m \) of the frame \( n \)
at viewpoint view, the optimal coding mode $o'(view, n, m)$ can be decided by

$$o'(view, n, m) = \arg \min_{o \in O} \left( (1 - p_{view})(D_o(view, n, m, o) + D_{ref}(view, n, p)) + p_{view}D_{cc}(er_{ref}(view, n, m) + \lambda_{view}R) \right)$$

where $O$ denotes the set of all candidate coding modes including intra-, inter- and inter-view coding.

The proposed rate–distortion optimized error resilience algorithm for SSVC considers not only the source distortion, but also the error-propagation distortion caused by DCP and MCP. Especially, when the encoder side can simulate the error concealment scheme at the decoder side, the error-propagation distortion can be estimated more accurately. Furthermore, the SSVC error concealment scheme, which utilizes the inter-view correlations among the views, is also considered in the rate–distortion optimized error resilience algorithm. Thus the proposed algorithm can achieve better trade-off between coding efficiency and suppression of error propagations for error-resilient stereoscopic video coding.

5. Experimental results and analysis

We utilize the H.264/AVC reference software JM 10.0 [24] to simulate a stereoscopic video coding system, as shown in Fig. 3. The sequence of the left view is encoded independently with MCP, and the sequence of the right view is predicted with both MCP and DCP. Then they are transmitted to the decoder respectively. The video sequences Ballroom and Race1 are encoded using the IPPP GOP structure for 240 frames within each view. Each row of macroblocks composes a slice and is transmitted in a separate packet. The packet loss rates (PLR) at 5%, 10%, and 20% [25] are tested in experiments. We assume that the first I frame in a GOP in each view is error free for all tests.

<table>
<thead>
<tr>
<th>Video sequence</th>
<th>Ballroom</th>
<th>Race1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1 = 5$, $w_2 = 4$, and $w_3 = 3$</td>
<td>30.07</td>
<td>31.20</td>
</tr>
<tr>
<td>$w_1 = 1$, $w_2 = 0$, and $w_3 = 0$</td>
<td>29.50</td>
<td>30.76</td>
</tr>
</tbody>
</table>

5.1. Experiments on the error concealment scheme for SSVC

Firstly, we assume that the independent view sequence is transmitted correctly and the inter-view predicted view sequence is transmitted with packet loss. Under the condition, there is no error propagations from the independent view to the inter-view predicted view, thus we can test the performance of our proposed stereoscopic error concealment scheme for the inter-view predicted view. We compare our proposed stereoscopic error concealment scheme (SEC) with temporal replacement (TR) method, the error concealment method of JM (JM) [24], and the spatio-temporal boundary matching algorithm (STBMA) [6] which only perform error concealment process in the temporal direction of a single-view. The QP is set to 28 for the tests.

For SEC, the weights $w_1$, $w_2$, and $w_3$ in formula (2) are set to 5, 4, and 3 empirically. Note that, when the weights $w_1$, $w_2$, and $w_3$ in formula (2) are set to 1, 0, and 0, the proposed OBMDC algorithm does not play a role in the SEC scheme. Table 2 shows the PSNR (dB) performance comparison of SEC for the inter-view predicted view with the two given weights. The PLR is set to 20%. The results indicate that the proposed OBMDC algorithm can utilize inter-view correlations to improve the error concealment quality.

Fig. 6 shows the PSNR performance results of different error concealment schemes for the inter-view predicted view under the given conditions. SEC has 3.28–6.20 dB error concealment performance improvement than TR, 1.10–2.21 dB improvement than JM, and 0.35–2.18 dB improvement than STBMA. Fig. 7 shows the error concealment results of sequence Race1, where SEC scheme can result in better subjective quality than the three other schemes. The results indicate that the proposed SEC scheme can utilize inter-view correlations to improve the error concealment quality for the inter-view predicted view.

For both the independent and inter-view predicted view sequences are transmitted with the same PLR. Table 3 shows the PSNR performance comparison of the four different methods. The QP is set to 28 for the test. When PLR = 20%, SEC has 0.79 dB average performance improvement than JM, especially the average gain is about 1.19 dB in the inter-view predicted view sequence. Compared to STBMA with PLR = 20%, SEC has 0.63 dB average performance improvement, especially the average gain is about 0.87 dB in the inter-view predicted view sequence.

5.2. Experiments on the rate–distortion optimized error resilience scheme for SSVC

First, we evaluate the accuracy of the proposed end-to-end distortion model for SSVC on the simulation SSVC system. Suppose the frame-level end-to-end distortion $D(n)$ is defined as the average of pixel-level distortions, we can obtain
\[ D(n) = \frac{1}{P} \sum_{i=1}^{P} D(n, i) \]  

(13)

where \( P \) is the number of pixels in the frame \( n \). \( D(n, i) \) is the end-to-end distortion of the pixel \( i \). Thus, we can obtain the proposed end-to-end distortion \( D_{\text{model}} \) at the encoder side and the actual end-to-end distortion \( D_{\text{actual}} \) at the decoder. Also, we can obtain the source distortion \( D_{\text{source}} \) at the encoder side. We define the estimation error \( E_D \) between \( D_T \) and \( D_{\text{actual}} \) as

\[ E_D(T) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (D_T(n) - D_{\text{actual}}(n))^2} \]  

(14)

where \( N \) is the number of frames in the sequence. \( D_T \) can be \( D_{\text{model}} \) or \( D_{\text{source}} \).

Table 4 shows the estimation error between the estimated end-to-end distortion and the actual distortion at the decoder side. In the test, both the independent and inter-view predicted view sequences are transmitted with PLR = 10\% and the QP is set to 28. The results indicate that the proposed end-to-end distortion estimated at the encoder side is very close to the actual distortion at the decoder side. And the proposed end-to-end distortion model can be utilized to substitute source distortion model for error-resilient stereoscopic video coding.

Then, in order to test the performance of our proposed stereoscopic rate–distortion optimized error resilience algorithm (SRDO-
ER), the simulation SSVC system serves as anchor, while the error concealment method JM is applied at the decoder side. We compare the random intra-update (RIU) algorithm which has been adopted in [24] with SRDOER. Under the conditions, for SRDOER, the error concealment module at the encoder side is the same as JM at the decoder side.

We also test the error-resilient performance of SRDOER algorithm on the inter-view predicted view. The independent view sequence is transmitted error free and the inter-view predicted view sequence is transmitted under different packet loss rates. Table 5 shows the decoded PSNR performance comparison for the inter-view predicted view at the decoder side under the given conditions. The proposed algorithm has 1.22–2.13 dB error-resilient performance improvement than the anchor and 0.29–1.06 dB improvement than RIU. Fig. 8 shows decoded PSNR-bitrates performance comparison between SRDOER and RIU for the inter-view predicted view with PLR = 20%. For similar error-resilient performance, SRDOER has 46.63% average bit-rate savings for Ballroom and 32.48% average bit-rate savings for Race1 compared to RIU. The results indicate that the proposed SRDOER algorithm can provide a better trade-off between coding efficiency and

<table>
<thead>
<tr>
<th>Video sequence</th>
<th>Scheme</th>
<th>Packet loss rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>Race1</td>
<td>Anchor</td>
<td>33.82</td>
</tr>
<tr>
<td></td>
<td>RIU</td>
<td>34.62</td>
</tr>
<tr>
<td></td>
<td>SRDOER</td>
<td>35.29</td>
</tr>
<tr>
<td>Ballroom</td>
<td>Anchor</td>
<td>33.10</td>
</tr>
<tr>
<td></td>
<td>RIU</td>
<td>33.47</td>
</tr>
<tr>
<td></td>
<td>SRDOER</td>
<td>34.32</td>
</tr>
</tbody>
</table>

Table 5
Average error-resilient PSNR (dB) performance comparison for the inter-view predicted view (QP = 28).

Table 6
Average error-resilient PSNR (dB) performance comparison for stereoscopic video at different packet loss rates (QP = 28).

<table>
<thead>
<tr>
<th>Video sequence</th>
<th>Scheme</th>
<th>Packet loss rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Independent view</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>Race1</td>
<td>Anchor</td>
<td>33.17</td>
</tr>
<tr>
<td></td>
<td>RIU</td>
<td>33.42</td>
</tr>
<tr>
<td></td>
<td>SRDOER</td>
<td>34.94</td>
</tr>
<tr>
<td>Ballroom</td>
<td>Anchor</td>
<td>32.08</td>
</tr>
<tr>
<td></td>
<td>RIU</td>
<td>32.45</td>
</tr>
<tr>
<td></td>
<td>SRDOER</td>
<td>34.46</td>
</tr>
</tbody>
</table>

Fig. 8. Average decoded PSNR (dB) performance comparison at different bit-rate point for the inter-view predicted view (PLR = 20%, QP = 28, 32, 36, 40).
suppression of error propagations for the inter-view predicted view, thus it can be utilized for error-resistant stereoscopic video coding.

Table 6 shows the decoded PSNR performance comparison when both the independent and inter-view predicted view bitstreams are transmitted under different packet loss rates. SRDOER has 1.72–4.28 dB error-robust performance improvement than the anchor and 0.91–2.90 dB improvement than RIU. Fig. 9 shows decoded PSNR-bitrates performance comparison between SRDOER and RIU. The results show that the proposed SRDOER algorithm provides superior error robustness performance for error-resistant stereoscopic video coding.

5.3. Experiments on the joint encoder–decoder error control framework for SSVC

Table 7 shows the decoded PSNR performance comparison of the six different error control schemes for SSVC. The QP is set to 28. Anchor is the simulation stereoscopic system, two error resilience schemes (RIU and SRDOER) at the encoder side and two error concealment schemes (JM and SEC) at the decoder side are simulated on the anchor. Since SRDOER considers the stereoscopic error concealment scheme to achieve better error robustness performance, the proposed SRDOER and SEC can compose a joint encoder–decoder error control framework SRDOER + SEC for SSVC, where SEC is performed on the error concealment module at the encoder side. Note that, in the error control scheme SRDOER + JM, the error concealment scheme JM is also simulated at the encoder side for error resilience. The proposed joint encoder–decoder error control algorithm SRDOER + SEC has 1.66–3.14 dB error robustness performance improvement than RIU+JM, 0.75–2.39 dB improvement than RIU+SEC, and 0.14–0.81 dB improvement than SRDOER + JM. The results indicate that the proposed joint encoder–decoder error control framework for SSVC (SEC + SRDOER) has superior error robustness performance for SSVC.

6. Conclusions

In this paper, utilizing inter-view correlations in stereoscopic video, a joint encoder–decoder error control framework is proposed for SSVC. The proposed joint encoder–decoder error control framework jointly considers error-resilient source coding, transmission network conditions and error concealment scheme for SSVC. At the decoder side, we propose an overlapped block motion and disparity compensation based error concealment scheme which can adaptively utilize inter-view correlations and temporal correlations in stereoscopic video. At the encoder side, a rate-distortion optimized error resilience algorithm is proposed for SSVC. We first propose that inter-view refreshment should be utilized for SSVC to suppress error propagations. Then, an end-to-end distortion model for SSVC is derived, which jointly considers the transmission network conditions, inter-view refreshment, and error concealment tools at the decoder side. Finally, based on the derived end-to-end distortion model, the rate–distortion optimized error resilience algorithm is presented to achieve better trade-off between coding efficiency and suppression of error propagations for SSVC. The error concealment scheme and the rate–distortion optimized error resilience algorithm compose the joint encoder–decoder error control framework for SSVC. Simulation results show that the proposed joint encoder–decoder error control framework has superior error robustness performance for stereoscopic video transmission over error-prone networks. In future, some unequal error protection techniques will be studied in the stereoscopic error control framework.

Acknowledgments

This work was supported in part by National Science Foundation (60736043) and National Basic Research Program of China (973 Program, 2009CB320905).

References


Table 7
Average error control PSNR (dB) performance comparison for stereoscopic video at different packet loss rates (QP = 28).

<table>
<thead>
<tr>
<th>Video sequence</th>
<th>Scheme</th>
<th>Packet loss rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Independent view</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inter-view view</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>Race1</td>
<td>Anchor + JM</td>
<td>33.17</td>
</tr>
<tr>
<td></td>
<td>Anchor + SEC</td>
<td>33.81</td>
</tr>
<tr>
<td></td>
<td>RIU + JM</td>
<td>33.42</td>
</tr>
<tr>
<td></td>
<td>RIU + SEC</td>
<td>34.02</td>
</tr>
<tr>
<td></td>
<td>SRDOER + JM</td>
<td>34.94</td>
</tr>
<tr>
<td></td>
<td>SRDOER + SEC</td>
<td>35.06</td>
</tr>
<tr>
<td>Ballroom</td>
<td>Anchor + JM</td>
<td>32.08</td>
</tr>
<tr>
<td></td>
<td>Anchor + SEC</td>
<td>32.68</td>
</tr>
<tr>
<td></td>
<td>RIU + JM</td>
<td>32.45</td>
</tr>
<tr>
<td></td>
<td>RIU + SEC</td>
<td>33.11</td>
</tr>
<tr>
<td></td>
<td>SRDOER + JM</td>
<td>34.46</td>
</tr>
<tr>
<td></td>
<td>SRDOER + SEC</td>
<td>34.81</td>
</tr>
</tbody>
</table>


