ABSTRACT

We study the scenario of pixel-domain distributed video coding for noisy transmission environments and propose a method to allocate the available rate between source coding and channel coding to generate a robust video stream. Having observed in experiments the uncertainty of the source and the channel coding rate, we model them as random variables via offline training, estimate the decoding failure probability and calculate the mean end-to-end distortion. Adaptive quantization is performed for each slice to minimize its mean end-to-end distortion. With this joint source-channel rate allocation, we compare the robustness of two coding prototypes, namely distributed video coding and distributed video coding with forward error correction. According to our experimental results, under same total bit budget, the distributed video coding only scheme proves more robust than the latter one and the gain is up to 1 dB in PSNR.

Index Terms— Distributed video coding, joint source-channel coding, error-resilience

1. INTRODUCTION

In quite a few practical codecs [1–3] for distributed video coding (DVC), channel codes are employed to achieve data compression. Now that channel coding is adopted in DVC, it is quite straightforward to increase the coding rate to resist against the noise over the transmission channel. In this way, channel codes are employed for the dual purpose of both source coding and channel coding. This approach is referred to as DVC in the following. In comparison, a conventional way of transmitting videos over noisy channels lies in the separation of source coding and channel coding. Specifically speaking, source coding is performed first and forward error correction (FEC) is then applied independently to resist channel noise. This scheme is referred to as DVC+FEC.

Both options, DVC and DVC+FEC, prove more robust than conventional video coding. The former option is taken in SLEP [1] as well as in PRISM [2], whereas the latter one is used in layered Wyner-Ziv video coding [3]. But which option is more efficient, DVC or DVC+FEC? This is the question we try to answer in this paper.

The first work addressing this issue in the context of distributed source coding (DSC) was done by Hua et al., who compared the robustness of the two schemes, namely DSC without protection and DSC protected using FEC, confining the discussion to binary data [4]. They concluded that the former scheme is more robust. Now we extend this work to video coding and try to find out by how much DVC can outperform DVC+FEC in noisy video transmission environments.

Two constraints, namely low delay and unavailability of a feedback channel, are set in this work, to compare the robustness of DVC and DVC+FEC effectively. The two schemes would perform equally well using typical error control schemes such as retransmission or hybrid-ARQ, if delay was not an issue and the decoder could send requests to the encoder for retransmission or more FEC bits via a feedback channel. More specifically, these two constraints mean that we cannot be sure of a successful decoding. The best we can do is to allocate the bit budget between source and channel coding such that the mean end-to-end distortion is minimized.

For a fair comparison, joint source-channel rate control is applied to both DVC and DVC+FEC. Moreover, the two schemes are compared under two different circumstances. First, they are tested under the assumption that the transmission channel state information is perfectly known. Second, they are tested with channel mismatch. In this scenario, the encoder performs rate allocation assuming the channel is good, before the resulting video stream is transmitted over a channel which is worse than expected.

2. SLICE-WISE DVC

First, we develop a pixel-domain Wyner-Ziv video codec based on the work of Girod et al. [1]. Then, we enable macroblock-wise mode decision with Wyner-Ziv (WZ) mode, skip mode and intra mode, following previous work [5, 6]. A certain number of WZ macroblocks (MBs), typically eight, are collected in raster scan order and put together in a slice. MBs in intra and skip modes are packetized separately from WZ MBs. They are protected using FEC against the chan-
channel noise. And we assume that they are error-free during the transmission in order to focus on the performance comparison between DVC and DVC+FEC for WZ MBs. The discussion in Section 3 and 4 is restricted to WZ MBs.

Four uniform scalar quantizers are available and each leads to a different number of bit-planes for the quantization indices that need to be coded. Herein, we use $N_{BP}$ ($N_{BP} \in \{1, 2, 3 \text{ and } 4\}$) to represent the number of bit-planes, which identifies the quantizer used for a slice. Rate control is carried out via adaptive quantization, i.e., adaptive decision on $N_{BP}$ for each slice.

When it comes to error resilience of DVC, the encoder is supposed to be less complexity-constrained, thus motion estimation is typically switched on at the encoder for higher coding efficiency [1–3]. Our DVC codec supports motion estimation, too. Every eight by eight block ends up with a motion vector in quarter pel accuracy, which is transmitted to the decoder and helps generate the side information for decoding.

The Slepian-Wolf coding is implemented using rate compatible punctured turbo codes (RCPTC). There are 32 different rates in our codec. Symbol-based decoding is implemented by trellis merging [7].

3. DECODING FAILURE PROBABILITY

The difficulty in joint source-channel rate control for DVC lies in the uncertainty of sufficient source coding and channel coding rate. Because of this uncertainty, no perfect rate allocation can be achieved and every slice undergoes a certain decoding failure probability. Nevertheless, the mean end-to-end distortion can be minimized if we can figure out how the rate allocation influences the decoding failure probability. For this purpose, we model sufficient source and channel rate as random variables. For the convenience of presentation, we list up primary symbols and abbreviations in Table 1:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{dvc}$</td>
<td>decoding failure probability for DVC</td>
</tr>
<tr>
<td>$p_{fec}$</td>
<td>decoding failure probability for DVC+FEC</td>
</tr>
<tr>
<td>$D_S$</td>
<td>source coding distortion</td>
</tr>
<tr>
<td>$D_c$</td>
<td>channel-induced distortion</td>
</tr>
<tr>
<td>$D_e$</td>
<td>mean end-to-end distortion</td>
</tr>
<tr>
<td>$S$</td>
<td>sufficient source coding rate</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>sufficient channel-source rate ratio</td>
</tr>
<tr>
<td>$R^*$</td>
<td>optimal source rate for DVC+FEC</td>
</tr>
</tbody>
</table>

3.1. Sufficient source coding rate

We refer to the minimum rate that allows successful source decoding as the sufficient source coding rate and model it as a Gaussian random variable $S$. We determine its probability density function $f_S(s)$ using training data, following the approach in [8]. $S$ is different from slice to slice and increases generally with the conditional entropy $H(X|Y)$ of the quantized pixel $X$ given its side information $Y$ [9]. Among the training data, we collect all the samples with similar conditional entropy and train a Gaussian random variable for the source coding rate. Therefore, the mean and variance of the random variable $S$ is a function of the conditional entropy $H(X|Y)$.

3.2. Sufficient channel-source rate ratio

The sufficient channel-source rate ratio $\Gamma$ is defined as the minimum ratio of the bit rate used for FEC and that for source coding which enables successful FEC decoding. This parameter characterizes how strong the source needs to be protected by FEC. Theoretically speaking, $\Gamma$ is $(1-C)/C$, given a discrete channel with channel capacity $C$, so that the decoding error probability approaches zero. In practice, however, especially with short codewords, $\Gamma$ for a given channel state is not deterministic but exhibits statistical properties. This is why we treat $\Gamma$ as a Gaussian random variable and estimate the parameters of its probability density function $f_\Gamma(\gamma)$ using training data. This training is based on the AWGN channel model and BPSK modulation. The resulting distribution of $\Gamma$ depends on the SNR of the AWGN channel.

One issue worth mentioning is the definition of the sufficient channel-source rate ratio for the DVC scheme. This is not straightforward, because there is no explicit FEC. Given a certain channel state, we first train the total bit rate that is required to ensure error-free decoding. Then we take the difference between the total bit rate and the sufficient source rate as the channel coding rate so that we can define equivalently the sufficient channel-source rate ratio for DVC.

We have observed in the training process that the distributions of the sufficient channel-source rate ratio for DVC and DVC+FEC are similar to each other. Therefore, we use a single random variable $\Gamma$ for both schemes.

3.3. Decoding failure probability

The decoding failure probability for a slice $X$ can be calculated, if the conditional entropy $H(X|Y)$, the AWGN channel state and the bit budget $BB$ are given. For each optional $N_{BP}$, $H(X|Y)$ is calculated and $f_S(s)$ is determined. In parallel, $f_\Gamma(\gamma)$ is determined given the channel state.

In DVC+FEC, the decoding is successful, if both the FEC decoding and DVC decoding are successful. Assuming that the source rate for DVC+FEC is $R$, we calculate the decoding failure probability as:

$$p_{fec} = 1 - \int_0^R f_S(s)ds - \int_0^{BB} f_\Gamma(\gamma)df_\Gamma(\gamma).$$

Here, the source coding rate $R$ is determined by minimizing the decoding failure probability:

$$R^* = \arg \min_R p_{fec}(R).$$

In comparison, it is not necessary to specify the source coding rate explicitly in the scheme DVC, because source coding and channel coding are implemented together. Therefore, the decoding failure probability depends only on BB:

$$p_{dvc} = 1 - \int_0^{BB} (\int_0^s f_\Gamma(\gamma)df_\Gamma(\gamma)) f_S(s)ds.$$
By comparing (1) and (3), we find out that \( p_{\text{fec}} \) is always larger than or equal to \( p_{\text{dvc}} \), because
\[
\int_0^R f_S(s) ds \cdot \int_0^{BB} f_T(\gamma) d\gamma \leq \int_0^R (\int_0^{BB} f_T(\gamma) d\gamma) f_S(s) ds,
\]
and this fundamentally justifies why DVC is more robust than DVC+FEC under the assumption that \( f_T(\gamma) \) is identical to DVC and DVC+FEC. Intuitively speaking, DVC+FEC requires an explicit decision upon the source rate, which might be inadequate, thus ends up with a higher decoding failure probability.

4. JOINT SOURCE-CHANNEL RATE CONTROL

We set a bit budget BB for each slice and allocate bits between source coding and channel coding by determining \( N_{BP} \) adaptively according to the source statistics and the transmission channel state. \( N_{BP} \) influences the decoding failure probability and the mean end-to-end distortion. We decide on \( N_{BP} \) such that the mean end-to-end distortion is minimized.

4.1. DVC+FEC

As shown in Figure 1, a slice is extracted from a video frame and quantized. Then, source coding is performed using turbo encoder 1. Systematic bits are discarded and parity bits are forwarded to turbo encoder 2. Here, the parity bits for DVC are treated as systematic bits and FEC is applied to them. Then, both the systematic bits and the parity bits of turbo encoder 2 are transmitted over a noisy channel, assumed to be AWGN, to the decoder. Next, turbo decoder 2 performs binary decoding and forwards the output to source decoding. Turbo decoder 1 takes the parity bits for DVC along with the available side information and performs symbol-based decoding. Finally, the cyclic redundancy check (CRC) for every macroblock is performed. If the CRC is fine, the original pixel values are reconstructed using the decoded symbols and the side information; otherwise, only the side information is used for reconstruction.

The mean end-to-end distortion \( D_c \) of a slice depends on its \( N_{BP} \) and is estimated in the following steps:

First, the decoding failure probability \( p \) is estimated. For a certain \( N_{BP} \), the conditional entropy of the quantized symbols given the side information is calculated, thus the source rate distribution \( f_S(s) \) is determined. In parallel, the distribution of the channel-source rate ratio \( f_T(\gamma) \) is determined according to the available channel state information available.

4.2. DVC

In this scheme, as illustrated in Figure 2, only one turbo coder is employed and it is used not only to achieve compression but also to resist channel noise. The parity bits are transmitted through the noisy channel to the decoder, whereas the systematic bits are discarded. The turbo decoder runs symbol-based decoding based on the side information and the received parity bits. The remaining steps, including CRC and reconstruction, are the same as in DVC+FEC.

Fig. 2. Distributed source coding without FEC

All the steps for adaptive quantization can be carried over from the scheme DVC+FEC, only the decoding failure probability is different. Now \( p_{\text{dvc}} \) is calculated using (3). This can lead to different decisions on \( N_{BP} \) compared to the scheme DVC+FEC.

5. EXPERIMENTAL RESULTS

5.1. Decoding failure probability

Figure 3 shows an example of decoding failure probability. As discussed in Section 3, the decoding failure probability depends on \( H(X|Y) \), BB and the transmission channel state. Generally speaking, the larger the conditional entropy, the higher the decoding failure probability, as shown in Figure 3(a) and the larger the BB, the lower the decoding failure probability, as shown in Figure 3(b). This is true for both DVC and DVC+FEC. In comparison, however, the decoding failure probability in DVC is always smaller than or equal to that in DVC+FEC.
5.2. End-to-end distortion with accurate channel state information

Here, it is assumed that the encoder knows perfectly the channel state information (CSI), thus can perform adaptive quantization optimally for each slice. The joint source-channel rate control targets at 500 kbps for the foreman sequence and 300 kbps for the salesman sequence. The reconstructed video quality of the two schemes (“DVC: csi known” and “DVC+FEC: csi known”) are presented in Figure 4. They perform almost equally well. Nevertheless, the DVC+FEC is always a bit worse. After all, it has a higher decoding failure probability, thus ends up with a lower video quality.

5.3. End-to-end distortion with channel mismatch

In this simulation, the encoder performs rate control assuming that the SNR of the AWGN channel is 4 dB, when the resulting bit stream is transmitted under worse channel states. The performance comparison between the two schemes (“DVC: csi mismatch” and “DVC+FEC: csi mismatch”) is presented in Figure 4. The video quality in both schemes drops if the real channel condition is worse than the encoder expects. Comparatively speaking, DVC is significantly more robust and outperforms DVC+FEC by up to 1 dB.

6. CONCLUSIONS AND FUTURE WORK

Under noisy transmission environments and with joint source-channel rate control, DVC and DVC+FEC exhibit similar performance when given accurate channel state information. With channel mismatch, however, the former one outperforms the latter one significantly. Therefore, we can conclude that DVC is in general more robust than DVC+FEC. Nevertheless, FEC is an indispensable protection for intra MBs and skip MBs. Moreover, it is worth investigating how much DVC+FEC can improve if DVC decoding and FEC decoding are performed jointly. Another topic for future work is how to achieve the best error-resilience for the transform-domain DVC.

7. REFERENCES