Side Information Enhancement via Texture and Motion Activity Analysis in Distributed Video Coding

Xianming Liu[a], Debin Zhao[a] and Siwei Ma[b], Wen Gao[b]

[a]School of Computer Science and Technology,Harbin Institute of Technology, Harbin,150001,P.R. China;
[b]Institute of Digital Media,Peking University,Beijing,P.R. China

ABSTRACT

This paper investigates how to exploit the artifacts constraint to enhance the quality of side information (SI) in distributed video coding (DVC). The idea originates from the observation that there are usually some regions shown as artifacts in the interpolated SI, which seriously degrade the overall performance of the DVC system, especially when the GOP size is large. Encouraged by the previous work about the backward channel in the literature, we propose a simple and effective DVC scheme based on artifact detection and removal. On one hand, texture analysis via dynamic range and local entropy is utilized to determine artifact regions in spatial domain. On the other hand, motion activity evaluation is used to determine artifact regions in temporal domain, and temporal local entropy difference is further employed to eliminate the effect of other factors like brightness variance. After the artifact blocks are detected, we re-code them in intra mode and replace the co-located ones at the decoder side so that the quality of SI is enhanced. Experimental results demonstrate that the proposed SI quality enhancement technique is effective.

Keywords: Distributed video coding, side information, artifact detection, texture and motion activity analysis

1. INTRODUCTION

Distributed Video Coding (DVC) is a new video paradigm behind which the main idea is to shift the complexity from encoder to decoder. Intra-frame-encoding and inter-frame decoding is performed in DVC in order to achieve this target. It has been proved that distributed source coding can achieve the same coding efficiency as jointly encoding in theory [1][2]. But recent results reveal that DVC still can not achieve the comparable coding performance to hybrid video coding. One major reason stems from the fact that accurate motion information is hard to be achieved at the decoder [3].

In DVC, side information (SI) plays a very important role, as it directly determines the rate-distortion performance of the system. One mainly used SI generating method is based on motion-compensated temporal interpolation (MCTI) [3]. Since the decoder does not have direct access to the current frame, motion vectors have to be estimated at the decoder side based on the underlying translation motion model. When the translation model does not accord with the true object motions, motion vector estimation errors might be introduced. These errors usually result in some low quality regions in generated SI which seriously degrade the coding efficiency of DVC system, especially when the GOP size is large. Although some complex post-process steps, e.g., overlapped block motion compensation (OBMC)[4][5] or spatial motion vector smoothing [6], have been used to improve the quality of interpolated frames, the performance is still far from satisfactory.

Observing from the generated SI, we find that low quality regions are usually shown as artifacts such as blockiness, blurring and ringing. Motivated by this observation, we propose a simple and effective quality enhancement technology based on texture and motion activity analysis for SI in this paper. The proposed work can be summarized in three steps. First, for the side information generated in MCTI combined with OBMC, we utilize texture analysis and motion activity evaluation to detect blocks with artifacts, and further refine the candidates of artifact blocks by local entropy difference to get more accurate detecting results. Then, the map information of blocks with artifact is transmitted to the encoder on a backward channel [7] and the encoder

Visual Communications and Image Processing 2010, edited by Pascal Frossard,
Houqiang Li, Feng Wu, Bernd Girod, Shipeng Li, Guo Wei, Proc. of SPIE Vol. 7744,
77442E · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.863452

re-encodes these blocks in H.264 intra mode. At last, the re-coded blocks are achieved at the decoder side to replace the corresponding ones in side information.

The rest of this paper is organized as follows. Section 2 presents the overall structure of the proposed DVC system with quality enhancement for SI. Section 3 discusses the proposed artifact-detection strategy, which includes texture and motion activity analysis in detail. Experimental results are presented in Section 4. Section 5 concludes the paper.

2. THE FRAMEWORK OF PROPOSED SCHEME

Fig. 1 illustrates the framework of the proposed DVC system based on artifact detection and removal, in which each input frame is independently encoded as either intra-frame (Key frame) or Wyner-ziv frame (WZ frame). Key frames are encoded and decoded using the conventional H.264 intra codec; The DCT-based Wyner-ziv coding scheme described in [3] is adopted for WZ frames. For a WZ frame, a DCT transform is first applied. The resulting transform coefficients are uniformly quantized with 2^M intervals. The quantized values are then grouped together to form bit planes which are input into the turbo encoder. Side information is generated at the decoder side from adjacent reconstructed frames by MCTI with OBMC. The core of our scheme, highlighted in Fig. 1, is first to detect blocks with artifacts in generated side information and then to construct a binary map information, where each entry in the map indicates whether the block is artifact region or not. The map information of all blocks of one frame is transmitted to encoder on a backward channel. Note that the backward channel is assumed to be error-free, which is consistent with the assumption of the feedback channel. A simple entropy coding algorithm is used in order to efficiently code this information. The detected blocks are then re-coded in intra mode and transmitted to the decoder to replace the corresponding ones in original SI. Consequently, artifacts are removed and original SI is refined. After that, the refined side information in conjunction with the received parity bits is used to reconstruct the original frame. If the decoder can not reliably decode the original stream, it requires additional parity bits from the encoder through feedback until an acceptably probability of bit error is guaranteed. There is a tradeoff between the number of blocks to replace and the quality of improved side information.

As mentioned above, the core of our system is the artifact detection module. In next section, we will elaborate on the proposed artifact detection technology.

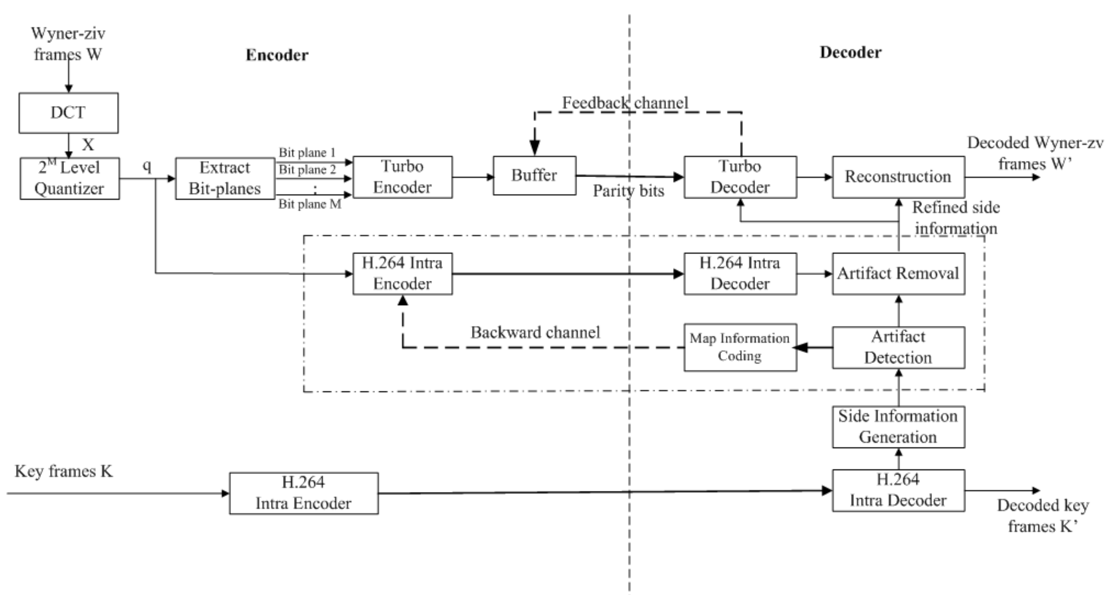

Figure 1. Structure of the proposed DVC scheme

3. ARTIFACT DETECTION STRATEGY

Motion inconsistence, camera noises and noises brought by flickering of light usually lead to artifacts in side information. It is expected that the side information quality can be improved by the removal of artifacts. The artifact detection in SI has two characteristics: 1) the current frame is not available in detection, so it is a no-reference problem; 2) the temporal information of adjacent frames can be used to refine the detection. For 1), we propose to utilize texture analysis in spatial domain to measure whether a block is an artifact one or not. For 2), we propose to employ motion activity analysis in temporal domain to determine which blocks are artifact ones and to exploit temporal local entropy difference to eliminate the effect of brightness variance. In a nutshell, the candidate artifact blocks are determined by spatio-temporal measurements.

3.1 Spatial texture analysis

Artifact detection without current frame is intrinsically difficult, because the distinction between image features and artifacts is often ambiguous. A major category of useful information is the texture including in the image itself. We utilize texture analysis to distinguish artifacts from local image characteristics. Generally speaking, the block with a spread histogram indicates that it contains more information and a narrow histogram less information. As illustrated in Fig. 2, the block A, which is regarded as a detail region, has a spread histogram; whereas the block C, which is regarded as a smooth region, has a narrow histogram. This can be characterized by the dynamic range (DR) of a block's histogram, which is described as:

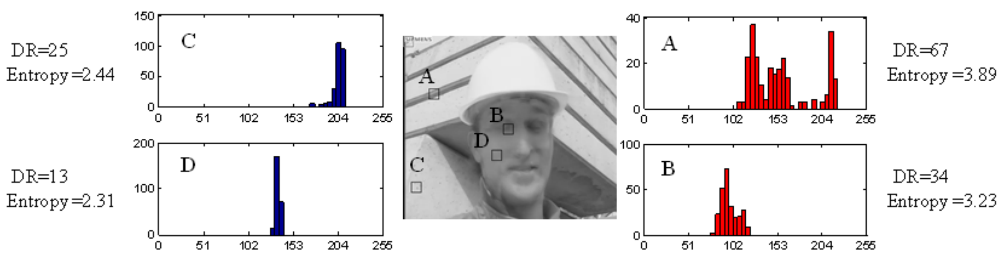

Figure 2. Histograms for four kinds of region in SI. (A) highly texture region;(B) strong artifact region;(C) fine texture region;(D)mild artifact region

$$DR = \sum_{i=0}^{255} F(i) \tag{1}$$

where

$$F(i) = \begin{cases} 1, H(i) > 0 \\ 0, H(i) = 0 \end{cases} \tag{2}$$

$H(i)$ is the number of pixels whose values are equal to i. Note that $\sum_{i=0}^{255} H(i) = N$, N is the total number of pixels in the macroblock. One SI frame can be divided into detail regions and smooth regions according to the DR.

Artifacts reduce the variation of intensities, thus decrease the entropy values. In both detail regions and smooth regions, blocks with lower local entropy are probably the artifact ones. As shown in Fig. 2, the entropy value of the strong artifact region B is lower than that of highly texture region A. The similar comparison result can be observed between the mild artifact region D and the fine texture region C. Thus, local entropy can be used as an artifacts detection measure to further classify the detail and smooth regions. The local entropy of a block is defined as follows according to the Shannon information theory:

$$E = -\sum_{i=0}^{255} P(i) \log_2 P(i) \tag{3}$$

where $P(i)$ is the probability density function (pdf) of each pixel's luminance value. Once the histogram is obtained, $P(i)$ can be computed by Eq.(4):

$$P(i) = \frac{H(i)}{N} \tag{4}$$

where $i \in [0, 255]$ denotes the bins of histogram.

As stated above, for a block W in SI, a coarse classification based on DR and local entropy is given in Fig. 3, T_{DR}, $T_{Entropy1}$ and $T_{Entropy2}$ are thresholds which are empirically selected.

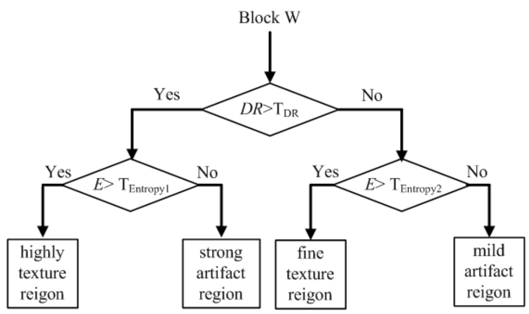

Figure 3. Flow chart of the coarse classification

3.2 Temporal motion activity analysis

It is not sufficient to only use texture for artifact detection, because the structure of coding artifacts could be exactly the same as that of object details. To get more accurate detecting results, temporal motion activity analysis is employed together with spatial texture analysis. The MCTI method, which is based on the translation motion model, performs motion estimation on the forward and the backward frames to obtain the motion vectors (MVs) and then generates the interpolated SI by scaling the obtained MVs. When the translation model is not satisfied, artifacts might occur. Consequently, it is reasonable for us to determine whether a current block is an artifact one according to the motion consistency between forward and backward frames. SAD (Sum of Absolute Differences) is calculated for a block W in SI as a measure.

$$SAD(W) = \sum_{(i,j)\in W} | \quad f_{n+1}(i - dx, j - dy) - \quad f_{n-1}(i + dx, j + dy)| \tag{5}$$

where f_{n+1} represents the backward frame and f_{n-1} represents the forward one, (dx, dy) is the scaled MV for the block of W.

Out of regard for the fact that larger SAD does not necessarily lead to smaller similarity due to brightness change, as depicted in Fig. 4. The temporal local entropy difference (TLED) of the current block W is also taken into account to rationalize our criterion. The TLED value $E^W_{n-1,n}$ and $E^W_{n,n+1}$ are calculated for W and the reference blocks in forward and backward reference frames respectively. If Eq.(6) is satisfied, the block W will be regarded as an artifact one.

$$SAD(W) > T_{temporal} \&\& \frac{E^W_{n-1,n} + E^W_{n,n+1}}{2} > T_{TLED} \tag{6}$$

Finally, blocks detected by texture and motion activity analysis are regarded as artifact ones. In order to assess the efficiency of the proposed artifact detection strategy, some experimental remarked results are given. In Fig. 5, (a) and (c) are two frames in SI, (b) and (d) are their corresponding artifacts detection results.

255	255	255	255
255	255	255	255
255	255	255	255
255	255	255	255

253	253	253	253
253	253	253	253
253	253	253	253
253	253	253	253

Figure 4. Small brightness change makes large SAD

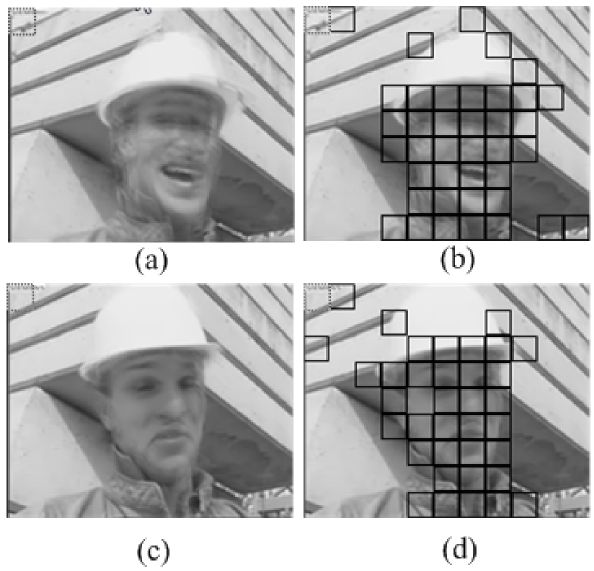

Figure 5. (a)The 14^{th} frame in SI generated by MCTI (b) the corresponding image with artifact blocks masked for the 14^{th} frame (c)The 70^{th} frame in SI generated by MCTI (d)the corresponding image with artifact blocks masked for the 70^{th} frame

4. EXPERIMENTAL RESULTS

In this section, experimental results are provided to demonstrate the performance of the proposed SI enhancement scheme. Results of two test sequences including *Foreman* and *Hall*(QCIF, 30Hz, 4:2:0) are presented. In each sequence, 100 frames are selected and the GOP structure is IWWWI, where Key frames are encoded with H.264 intra mode and WZ frames are encoded with the proposed TDWZ codec. By removing artifact blocks, the quality of SI improves significantly. Fig. 6 illustrates the objective quality comparison between original SI generated by MCTI and the SI refined in our method for the *Foreman* sequence. And Fig. 7 depicts the subjective comparison results of the 14^{th} and the 70^{th} frames in SI. It can be clearly observed the artifact regions are almost removed and the quality of SI is enhanced.

However, it should be noted that the quality of SI is improved at the expense of additional bits, which include the bits used to transmit map information to the encoder side and the bits used to re-code artifact blocks. As a consequence, we further test the efficiency of our scheme in terms of overall performance. We change the quality of SI for WZ frame according to corresponding Key frames by quantizing, and the quantization parameters we chose are 24, 28, 32 and 36. At each point, the codec will choose the best performance by adjusting the coding bit plane number of WZ coefficients. The overall bits include two parts: additional bits and bits used to encode all frames including Key frames and WZ frames. Fig. 8 gives the overall rate-distortion performance. There are two other interpolation algorithms to compare:

1) Motion-compensated temporal interpolation (MCTI) [3];

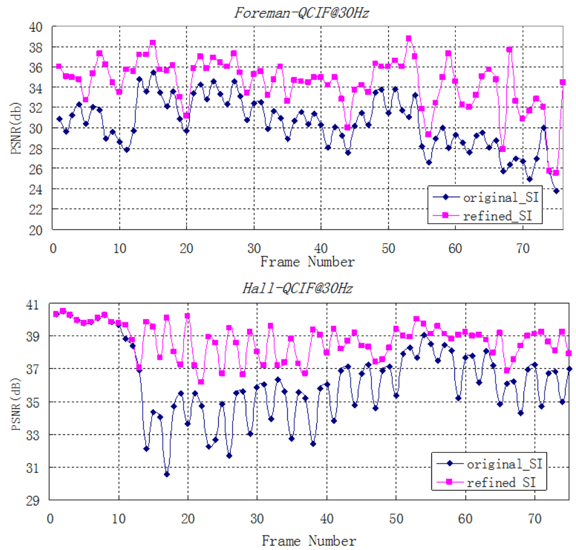

Figure 6. Objective quality comparision for original and refined SI of *Foreman* and *Hall* (QP = 24)

Figure 7. Subjective quality comparision for original and refined SI of foreman (QP = 24)(a)The 14^{th} frame in SI generated by MCTI (b) the corresponding refined result for the 14^{th} frame (c)The 70^{th} frame in SI generated by MCTI (d)the corresponding refined result for the 70^{th} frame

2) MCTI with spatial smoothing (MCTI+SS), for which we choose the better result between two strategy: BiME+OBMC [5] and BiME+AWMF [6]. For *Foreman* sequence, MCTI+SS is BiME+AWMF; for *Hall* sequence, MCTI+SS is BiME+OBMC.

As illustrated, our method can improve up to 1dB for *Foreman* sequence and 0.8dB for *Hall* sequence at the high bit-rates compared with MCTI+SS. It means our method especially works well at the high bit-rates since the advantages of quality enhancement for SI remarkably outweigh the overhead caused by additional bit streams.

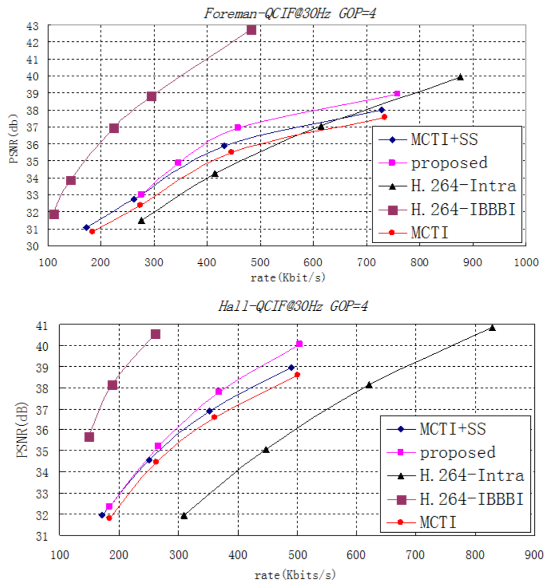

Figure 8. R-D curves of the overall performance

5. CONCLUSION

In this paper, we present a novel SI enhancement scheme through artifact detection and removal in distributed video coding. Our method considers both spatial and temporal information to detect the blocks with artifacts. After side information is generated as in prior work, blocks in interpolated frame are classified into likely and unlikely to exhibit artifacts via texture and motion activity analysis. For those blocks most likely to include artifacts, intra coded data is sent explicitly to replace the corresponding ones so that the quality of SI is enhanced. The method presented in this paper is promising in that it constrains the artifacts in SI and leads to the improvement of overall performance in DVC system.

6. ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation 60736043 and National Basic Research Program of China (973 Program, 2009CB320905). The authors would like to thank the anonymous reviewers for their constructive suggestions which helped us improving our manuscript.

7. REFERENCES

[1] D.Slepian and J. K. Wolf, “Noiseless coding of correlated information sources”, IEEE Trans. Information Theory, Vol.19, pp.471-480, July.1973.
[2] A.Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder, ” IEEE Trans. Information Theory, Vol. 22, pp.1-10, Jan.1976.
[3] B. Girod, A. Aaron, S. Rane, and D. R. Monedero, “Distributed video coding,” Proc. IEEE, Vol. 93, no.1, pp.71-83, Jan.2005.
[4] J. Zhai, K. Yu, J. Li, and S. Li, “A low complexity motion compensated frame interpolation method,” in Proc. ISCAS, May 2005, vol.5, pp.4927-4930.
[5] B.D.Choi, J.W.Han, C.S. Kim, and S.J. Ko, “Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation,” IEEE Trans. Circuits Syst. Video Technol, vol. 17, No. 4, pp. 407-416, Apr. 2007.
[6] J. Ascenso, C. Brites, and F. Pereira, “Improving Frame Interpolation with Spatial Motion Smoothing for

Pixel Domain Distributed Video Coding", EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, Slovak, July 2005.
[7] L.Liu, Z.Li, and E.J.Delp, "Efficient and Low-Complexity Surveillance Video Compression Using Backward-Channel Aware Compression," IEEE Trans. Circuits Syst. Video Technol,Vol.19, No.4, Apr.2009.