A Sequential Motion Compensation Refinement Technique for Distributed video coding of Wyner-Ziv frames

A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, H.K. Arachchi

School of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, UK
E-mail: [adikari.adikari, anil.fernando, Rajitha.weerakkody, eesrhhk ]@brunel.ac.uk

ABSTRACT

Distributed video coding (DVC) is one of the fastest growing coding techniques in the current signal processing world. It has its outstanding low complex encoder which is essential for some applications in the consumer market. Driven by the above motive, in this paper we propose a novel side information refinement technique for Wyner-Ziv video frames in DVC. When a bit plane is decoded, it is used to replace the corresponding bit plane in the side information and used for motion estimation and compensation to up date the side information. Then the information for the next bit plane is extracted from the updated side information and is sent to the turbo decoder as in a conventional DVC decoder. This process is repeated until all bit planes are extracted. This technique can be employed with almost all the side information generation techniques from simplest conventional pixel interpolation to advanced techniques such as bi-directional motion estimation and spatial motion smoothing (BiMESS) [8]. Simulation results show that over 1.4dB PSNR gain can be obtained with the proposed algorithm over the BiMESS smoothing algorithm at the same bit rate.

Index Terms – Distributed Video coding, Video Coding, Signal Processing, Video Codecs

1. INTRODUCTION

In the near future, there will be a very high demand for applications such as video sensor networks for security surveillance, monitoring of disaster zones, and design of realistic entertainment systems and some domestic applications like monitoring children and the elderly by the guardians and video conferencing requirement in mobile communications. Traditional video coding algorithms such as H.264, MPEG-2, and MPEG-4 will not be suitable for such applications as the encoder of these codecs are very computationally expensive.

Distributed video coding (DVC) is known to be a technique used to reduce the processing complexity of the encoder, leading to low cost implementation, while majority of the computations are taken over by the decoder. This is a solution for the requirement of capturing video streams at multiple remote sites and then decoding and processing at a central office. This type of approach is the best solution for the applications mentioned earlier. Wyner-Ziv [2] approach is a feasible solution for this, which uses the correlations of the original frame with a frame (side information) predicted at the decoder using the available information. We only need to transmit the parity information through the channel. Then, we make use of this parity information with side information to decode the original information.

In DVC, the decoder needs to exploit the inherent temporal and spatial correlations of the picture information present within an image and between images in a continuous sequence. The primary difference in this approach in comparison with other traditional compression methods is in the complexity balance of the encoder and the decoder as mentioned earlier. Non DVC approaches involved a complex encoder structure and an easy to implement-low cost decoder which was motivated by applications such as broadcasting, video on demand, and video streaming. However, in contrast to it, majority of the applications listed above are dominated by a number of video capturing points located remotely and ready to upload information to be processed at a central processor. This inevitably drives the attention towards a simple low cost yet competitive encoder implementation. DVC is the best known solution for this issue so far we managed to achieve. The major contributor of low complexity is its ability to shift the work load of exploiting the source statistics to the decoder. The encoder side advantages include low requirement of memory, computational capacity and power which are generally scarce resources at the remote sites.

The exploitation of the source correlation statistics at the decoder in DVC is by means of generating a corresponding secondary information series that has a statistical dependency with the original data, which is known as statistically dependent side information. In order to generate side information in the decoder side where all the information are yet with the encoder, a sequence of selected original frames is generally passed to the decoder over the channel using an existing compression techniques and are called ‘key-frames’. The frequency of key-frame transmission could vary upon the DVC implementation strategy. The missing frames are estimated at the decoder using basic interpolation techniques or more complex and accurate motion prediction methods. This kind of initial estimations are the side information for a particular frame.
is assumed that the side information so generated is a form of the original sequence subjected to noise while transmission. The identification of the statistical distribution of this noise or the introduced errors in the estimation is a part of the ongoing research activity. If the frame that has to be transmitted over the channel is turbo coded, the above side information represents the corrupted version of the above systematic bits. Then together with the parity bits that are transmitted over the channel it is possible to decode the original systematic bits since side information is known to the decoder. At the encoder this parity bits are generated by passing the original video sequence through a set of shift registers and logic gates. Furthermore this parity bit sequence is generally subjected to puncturing, of which the rate is determined by the channel bandwidth requirement. However the main reason for puncturing is to control the bit rate. On the other hand puncturing means lose of information that results a low quality of image reconstructed at the decoder. Side information generated in the decoder has a vital role in the quality of the reconstructed image under a given bandwidth, since higher the quality of the side information lower the number of parity bits required by the decoder to decode the systematic bits with minimum decoding errors. The number of bits transmitted over the channel to represent each pixel (bpp) and the closeness of the reconstructed image at the decoder to the original frame held back at the encoder (PSNR) are the common measures of the goodness of the video codec implementation. The theoretical base and the guidelines for Distributed Source Coding were set by Slepian-Wolf [1] and the current work in this field is based on the work by Wyner-Ziv [2]. Based on this concept, several turbo coded DVC codecs have been proposed. In this paper, we propose a novel side information refinement technique using sequential motion compensation in order to modify the side information with the decoded information. Then, the relationship between 

\[ \begin{align*}
Y_m(i, j) &= g(X_{\text{coded}}(i, j), X_{\text{coded}}(i-1, j), \ldots, X_{\text{coded}}(i, j)), \\
X_{\text{coded}+1}(i, j), \ldots, X_{\text{coded}+(N_1-1)}(i, j), X_{\text{coded}+N}(i, j)) \\
g() &= \text{a function to describe the motion compensated prediction is done using } N_2 \text{ past reference frames and } N_1 \text{ future frames.}
\end{align*} \]

(4)

Then, the relationship between \( Y_m(i, j) \) and \( X_m(i, j) \) can be modeled with a noise term, \( n_m(i, j) \) as shown in equation 5. It can be shown that the noise term \( n_m(i, j) \) can be approximated to an additive stationary white noise signal, if the motion estimation is accurate. For most of the cases, this noise process can either be modeled using a Gaussian or using a Laplace probability distribution. Based on the above concept, several algorithms for Wyner-Ziv coding have been proposed recently. Aaron et al. has proposed a turbo coded based Wyner-Ziv codec for motion video [3,4]. They used simple frame interpolation [3] and motion interpolation and extrapolation techniques [4] to predict the side information. Based on this turbo based Wyner-Ziv codec, several side information prediction algorithms were proposed. Tagliasacchi et al. has proposed a motion compensated temporal filtering technique [5]. Ishwar et al. [6] presented an information theoretic study of video codecs based on source coding with side information. Natario et al. proposed an algorithm to generate side information based on motion field smoothing to provide improved performance [7]. Ascenso et al., presented a scheme using motion interpolation to derive the side information [8]. They used forward and bidirectional motion estimation and a spatial motion smoothing algorithm to generate the side information. They also proposed a motion refinement algorithm using weighted motion estimation to further improve the side information [9]. So far, this algorithm is the best available pixel domain Wyner-Ziv codec algorithm in the literature. Later on, we will make use of this algorithm to compare our results. Several other literature have also been reported [10,11,12,13]. It is clear that all these literature mainly considered on the side information improvement in the turbo Wyner-Ziv codec.

2. RELATED WORK

According to the Stepen-Wolf theorem, two statistically dependent signals \( X \) and \( Y \) can be separately encoded and still be jointly decoded at the decoder with an arbitrarily small error probability if the following conditions are satisfied.

\[ R_X \geq H(X / Y) \]  
\[ R_Y \geq H(Y / X) \]  
\[ R_X + R_Y \geq H(X, Y) \]

(1)  
(2)  
(3)

Where, \( H(X,Y) \) is the joint entropy of \( X \) and \( Y \), \( H(X/Y) \) and \( H(Y/X) \) are their conditional entropies, while \( R_X \), \( R_Y \) are the rates of \( X \) and \( Y \) respectively.

Let’s assume that \( X_m(i, j) \) is the current Wyner-Ziv frame and \( Y_m(i, j) \) is the correlated side information for the current frame generated based on Equation (4).

\[ Y_m(i, j) = g(X_{\text{coded}}(i, j), X_{\text{coded}}(i-1, j), \ldots, X_{\text{coded}}(i, j)), \\
X_{\text{coded}+1}(i, j), \ldots, X_{\text{coded}+(N_1-1)}(i, j), X_{\text{coded}+N}(i, j)) \\
g() \text{ is a function to describe the motion compensated prediction is done using } N_2 \text{ past reference frames and } N_1 \text{ future frames.}
\]

(4)

\[ Y_m(i, j) = X_m(i, j) + n_m(i, j) \]

(5)
3. PROPOSED TECHNIQUE

The proposed technique is based on sequential motion estimation and compensation to update the side information. Figure 1 illustrates the architecture of the codec with the proposed side information generation and refinement functions. Initially when there is no decoded information available to update the side information, bi-directional motion estimation and spatial motion smoothing (BiMESS) is used with adjacent key frames to generate the first version of side information [8]. Thus, the turbo decoder uses the output of BiMESS to decode the most significant bit plane. Since the side information naturally has some errors the turbo decoder requests parity bits from the buffer until it receives enough parity bits to decode the original systematic bits with a predetermined error rate. As in the other literature, it is assumed that the decoder has the complete knowledge about the probability of errors. When the first bit plane is extracted, it is used to replace the existing first bit plane in the side information. Then the modified side information is bi-directionally motion estimated and compensated to update the side information. It is clear that this version of the side information is more accurate than the previous side information as the decoded first bit plane has been intelligently used for motion compensation. The second bit plane of the updated side information is used as the systematic bits for the second bit plane for turbo decoding. Then, the decoded second bit plane is used to modify the current side information as in the previous case. The modified side information now is bi-directionally motion estimated and compensated to update the side information again. Then the third bit plane is extracted from the modified side information. This process is continued until all bit planes are decoded. Finally, when all bit planes are decoded, the most updated side information is once more motion compensated before the reconstruction process. This step is important to improve the reconstruction process with the available latest decoded information. Since each decoded bit plane has been used to refine side information in the above process, the accuracy of the side information increases continuously resulting a lower number of parity bits sent via the channel to the decoder to achieve the same subjective image quality. Once all the required bitplanes are decoded, they are sent to the reconstruction function together with the latest version of the side information in order to reconstruct the final Wyner-Ziv frame as mentioned above.

4. SIMULATION RESULTS

In our simulations, the odd frames are the Wyner-Ziv frames which go through the interfame encoder to generate the parity sequence. The even frames are directly passed to the decoder as key frames using JPEG2000. The Wyner-Ziv frames are first passed through the $2^M$ level quantizer of which the level M is an independently varied parameter based on the expected quality of output and the available channel bandwidth. Next, the Slepian-Wolf based encoder incorporates the bit plane extractor and then the turbo encoder. Two rate $\frac{1}{2}$ component encoders with constraint length 4 (K=M+1) and a generator polynomial of [17 10] in octal form are used as the turbo encoder. A S-random interleaver is used in front of the bottom constituent encoder. Wyner-Ziv frame rate is assumed to be 15 frames per second and different quantization levels are used to achieve different bit rates, i.e. $2^M \in \{2,4,8,16\}$. In the motion compensation the search range is set to 8 and the block size is 4x4. All other simulation parameters are set as in [8] and are used for all simulations in this paper. We have considered several video sequences available in the public domain and Figure 2 presents a performance comparison for the “foreman” video sequence. The obtained results are compared against the best available pixel domain Wyner-Ziv codec from the literature [8] in the same test conditions. Results clearly show that there is a significant PSNR gain up to 1.4dB with the proposed algorithm over BiMESS [8]. It also shows that the proposed algorithm performs well at higher bit rates. This is due to the fast convergence of the side information when more bit planes are available. H.263+ intraframe coding and H.263+ interframe coding with a pattern of I-B-I-B- are also presented for the comparison. It shows that the gap between the Wyner-Ziv codec which is at its early ages and the H.263+ (I-B-I-B-) codec has been narrowed significantly. We have tested the proposed technique for several other video sequences and observed similar performance improvement.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a novel side information refinement technique for DVC using sequential motion compensation that makes use of the decoded information. Simulation results clearly show that the proposed algorithm can get a significant PSNR gain up to 1.4dB compared to the full motion refinement algorithm [8] at the same bit rate. FMR algorithm [8] needs more computational power than our proposed algorithms since FMR algorithm needs to do a weighted motion compensation for each bit plane. Furthermore, it needs to apply the reconstruction process for each decoded bit plane. Therefore, proposed codec is more suitable for DVC applications since it has a significant PSNR gain at a lower computational cost. We are currently concentrating on making use of the color space further to refine the motion compensation to improve the proposed algorithm.
6. REFERENCES


