Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 2627 – 2632

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

A Content-Adaptive Side Information Generation Method for Distributed Video Coding

Jinxiu Zhu[a,*], Li Li[a], Yongqin Shao[a], Ning Cao[b]

[a]*College of Computer & Information(Changzhou), Hohai Univ., Changzhou 213022, China*
[b]*College of Computer & Information, Hohai Univ., Nanjing 210098, China*

Abstract

In this paper, a content-adaptive method to generate side information at the block level is presented. First, motion compensated temporal interpolation (MCTI) algorithm is used between the reconstructed key frames at the decoder to acquire initial motion vectors. Second, the image is segmented and the edge of moving region is detected by obtained the residual frame between two consecutive key frames. Furthermore, hierarchical motion estimation (HME) and motion vector filter (MVF) are adopted for edge region and an adaptive motion vector filter (AMVF) is introduced in non-edge region to correct the false estimated motion vectors. The proposal is tested and compared with the results of the state-of-the-art DISCOVER codec and RD improvements on the set of test sequences are observed.

Keywords : Distributed Video Coding (DVC); Content-adaptive SI refinement; the region of interest (ROI); hierarchical motion estimation (HME); motion vector filter (MVF)

---

1. Introduction

The traditional video coding frameworks are designed with a low complexity decoder and a high complexity encoder. However, due to limited energy and computing ability, several emerging applications like wireless video surveillance and mobile camera phones cannot afford such a high complex encoder. As a consequence, a low complexity encoder with high coding efficiency is much desirable. Distributed video

---

* Corresponding author. Tel.: +86-0519-85191841;
E-mail address: zhujinxiu1972@163.com

doi:10.1016/j.proeng.2012.01.362

coding (DVC) [1] is a new coding paradigm based on distributed source coding principles with a modified complexity balance between the encoder and decoder. Side information quality is one of the decisive factors that affect the overall performances of DVC. A lot of studies have been devoted to the accuracy of estimating current frames. One of the most efficient approaches to generate the side information (SI) is through motion compensated temporal interpolation (MCTI) where the current frame is estimated based on past and future reference frames. The apparent advantage of MCTI is its conceptual simplicity, and block matching can reflect some relationship between motion and interpolated intensity values, especially when the motion accords with the translation model. However, due to the original frames are not available at the decoder, block matching may not be effective locally, what usually results in artifacts in SI. There have been several approaches to improve the quality of SI proposed: overlapped block motion compensation (OBMC) and probabilistic compensation [2], motion field smothering algorithm [3] [4], iterative side information refinement technique [5] [6], hierarchical motion estimation [7].

2. The proposed framework

The transform-domain DVC architecture we used in this paper is shown in Fig 1. In this coding scheme, every frame is divided into 4×4 non-overlapping blocks. The first frame of a GOP is defined as a key frame, that is, all the blocks in this frame are key blocks. The remaining frames of the GOP are WZ frames and the blocks in these frames are WZ blocks. Key blocks are encoded using a conventional H.264/AVC intra coder. WZ blocks are transformed, quantized, and each bit plane of the quantization indices is encoded using a channel coder (LDPC) and the parity bits are transmitted to the decoder. At the decoder, side information (SI) is initially created using MCTI from the previous and next key frames and utilized the Block Matching algorithm (BMA) for motion estimation. Generally speaking, MCTI can work efficiently in the regions with low motion, such as background, since linear motion model is usually satisfied in these regions. However, in the rest regions, especially the regions with nonlinear movement, motion vectors from BMA are often not faithful to true object motions. If the estimated block does not really move to the goal block, the quality of the interpolated frame will greatly deteriorate, so the real movement parameters of the goal block must be found in DVC. So, the initial SI is corrected by a novel content-adaptive SI refinement. Then the refined SI is fed as an input to LDPC decoder. LDPC decoder starts decoding based on the scanned DCT coefficients of the refined SI and a correlation model between the source and SI. The LDPC decoder requests more parity bits via the feedback channel until the frame is successfully decoded which is controlled by a request stopping criteria. Finally, decoded WZ frame is obtained after reconstruction.

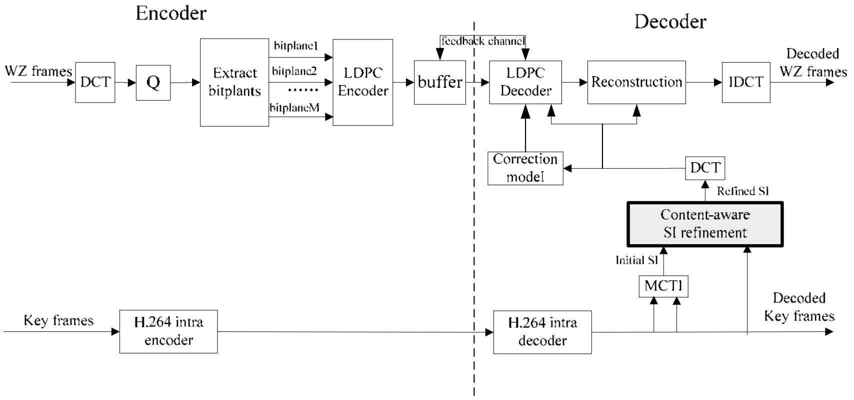


Fig. 1 The proposed DVC framework

3. Content-adaptive side information generation

To improve the quality of side information, in our proposed approach, the spatial and temporal information are both taken into consideration in the moving region detection. The process of the proposed content-adaptive side information refinement is shown in Fig 2. Firstly, it generates the initial motion vectors (MV) and SI by using the MCTI. Secondly, the image is segmented and the edge of moving region is detected by obtaining the residual frame between two consecutive key frames. For edge region, hierarchical motion estimation (HME) and motion vector filter (MVF) is adopted to acquired more accurate motion vectors. In contrast, an adaptive motion vector filter (AMVF) is introduced in non-edge region to correct the false estimated motion vectors.

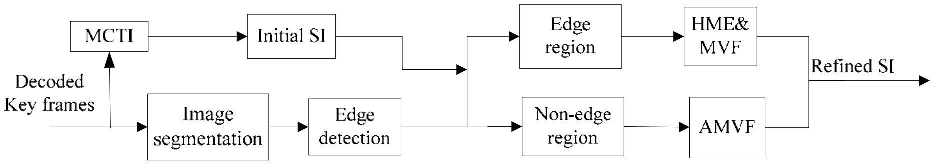


Fig. 2 Content-adaptive SI refinement

3.1. Otsu Threshold Segmentation and Edge Detection

In this paper, we choose the Otsu's Threshold [8] [9] to do image Segmentation. In Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance, defined as a weighted sum of variances of the two classes:

$$\sigma_{\omega}^{2}(t)=\omega_{A}(t)\sigma_{A}^{2}(t)+\omega_{B}(t)\sigma_{B}^{2}(t) \quad (1)$$

Weights ω_A and ω_B are the probabilities of the two classes separated by a threshold t. σ_A^2, σ_B^2 are the variances of these classes. Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance. We select the 7th and 9th frame of the Foreman sequence to illustrate the process of the image segmentation and edge detection, as shown in Fig 3.

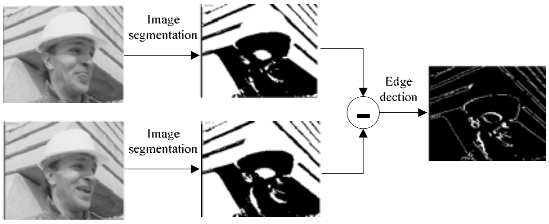


Fig. 3 The process of the image segmentation and edge detection (the 7th and 9th frame of the Foreman sequence)

3.2. Hierarchical Motion Estimation

Fig 4 shows HME algorithm of the 8×8 block as an example. The image is classified into three levels: bottom-, mid- and top-level. In bottom-level, block size is 8×8, comprising four 4×4 blocks of the mid-level at the same location. Hence each 4×4 block is called the sub-block of the 8×8 block. Likewise, a 4×4 block of the mid-level image comprises four 2×2 blocks of a top-level image. The relatively optimum motion vectors can be obtained by motion estimating in 8×8 blocks, and the motion estimation of the 2×2 blocks or 4×4 blocks do the role of the compensation to the larger ones'.

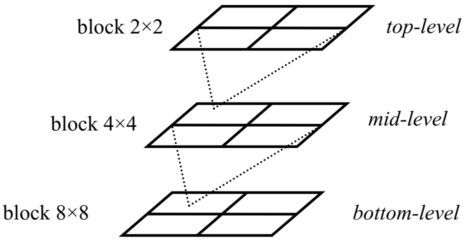


Fig. 4 Hierarchical Motion Estimation

For instance, there are two objects in the central block *O* (Fig. 5. (a)). The motion vector of the central block using traditional algorithm is presenting in Fig. 5. (b), with the new one obtained by the proposed algorithm in Fig. 5. (c); from Fig 5 we can see more precise motion vectors is obtained with the proposed method. And to some extent, the mismatch is avoided by using more predicted motion vectors, which including the central block and other three surrounding ones (Fig. 5. (a)).

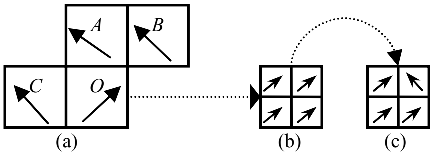


Fig. 5 Comparison: (a) block *O* and other three surrounding ones; (b) the old MVs; (c) the new MVs

3.3. Motion Vector Filter

Generally speaking, for non-edge region, such as the background, the internal motion vectors should be consistent. But due to the impact of image noise, partial matches and other factors, there are inconsistencies in some cases. Just as Fig 6 shows, block 0 has obvious different motion vector compared to its surrounding ones. Thus we cannot know the real direction of motion. Under the circumstances, vector median filter (VMF) [10] is used to amend these false motion vectors to obtain the right ones. However there is only modest spatial correlation between motion vectors after block matching, so here we use the weighted median filter [11] to process the motion vectors.

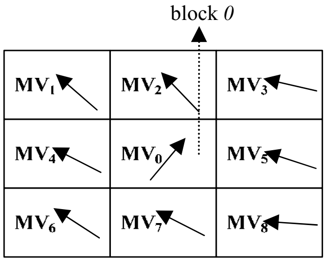


Fig. 6 Inconsistency of motion vector in partial area

In Fig 6, supposing $\mathbf{MV_0}$ is the false motion vector and needs filtering, and the eight motion vectors of surrounding blocks is $\mathbf{MV_i}$ (i=1, 2, …, 8). Hence the output of the vector filter is $\mathbf{MV_{awf}}$.

$$MV_{awf} = \arg\min_{MV_i} \sum_{i=0}^{8} \omega_i \left\| MV_j - MV_i \right\|_2 \quad (j=0, 1, \ldots, 8) \qquad (2)$$

Weight ω_i is decided by the compatibility of relevant motion vectors: take the nine motion vectors respectively as the motion vector of the current encoding block, then computing the motion vector residual R(**MV**$_{i)}$,Thus we get the ω_i .

$$\omega_i = R(\mathbf{MV}_0)/R(\mathbf{MV}_i) \tag{3}$$

$$R(\mathbf{MV}_i) = \sum_{p\in B}\left|K_{i-1}(p-MV_i)-K_{i+1}(p+MV_i)\right| \tag{4}$$

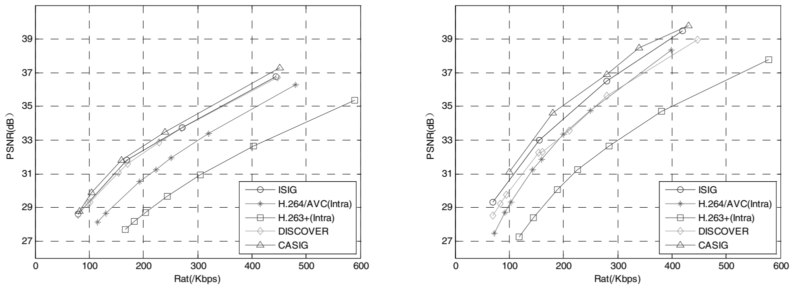
$$p = (x_i, y_i) \tag{5}$$

B represents block *0*, K_{i-1} and K_{i+1} is the two consecutive key frames, *p* is the coordinate of pixel in the encoding block 0. The formula (2) means that the weighted sum of the distance between $\mathbf{MV}_{\mathbf{awf}}$ and the remaining eight motion vectors is less than or equal to the weighted sum of the distance between any other motion vector (belonging to the nine vectors) and the remaining eight motion vectors. Through the filtering, we can acquire the consistency-improved internal motion vectors of non-edge areas.

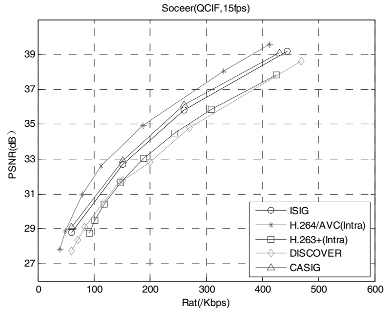
4. Experimental results

In order to assess the coding capability of the proposed Content-adaptive Side Information Generation Method (CASIG) scheme, we endeavour a comparison with the state-of-the-art DISCOVER [12] codec, which is feedback channel-based and operates in the DCT domain. The performance of the proposed codec is tested for a number of QCIF video sequences at 15 fps. In the experiments, GOP size is 2, DCT size is selected as 4×4 and the block size for motion estimation is selected as 8×8. The bit rate and PSNR are calculated for the luminance component of all frames by averaging them over the sequence. Key frames are intra-coded using H.264/AVC and quantization parameters are selected to match the average PSNR of WZ frames and key frames.

Tests were carried out for the Coastguard, Foreman and Soccer sequences for containing different motion levels from low to high in the given order. Fig 7 show the overall rate-distortion (RD) performance compared to the iterative side information generation method (ISIG)[5], H.264/AVC (intra), H.263+ (intra), proposed CASIG and DISCOVER codec. The performance of H.263+ or H.264/AVC is obtained by coding with H.263+ or H.264/AVC without exploiting temporal redundancy in order to have a similar encoder complexity. A significant PSNR gain up to 1.6 dB has been achieved by the proposed scheme against DISCOVER codec .For ISIG, the performance of CASIG is also improved to 0. 45dB. Also gap between CASIG and DISCOVER closes for the Coastguard sequence and increases for the Soccer sequence. As the results turn out, proposed technique improves the rate distortion performance more significantly for the Soccer sequence than Coastguard sequence. This is because of the high-speed motion in the sequence. Therefore, it is possible to conclude that the DISCOVER WZ codec can exploit both the temporal correlation and the spatial correlation in an efficient way while using a rather simple encoder and still be competitive when compared to the (more complex) H.264/AVC Intra encoder.



(a) (b)



(c)

Fig. 7 Comparison: (a) RD performances of different codecs for Coastguard sequence; (b) RD performances of different codecs for Foreman sequence; (c) RD performances of different codecs for Soccer sequence

5. Conclusion

This paper proposes an advanced SI creation framework and evaluates its RD performance in the TD-LDPC based WZ video codec with feedback channel. The proposed DVC codec improves the RD performance by proposing a novel SIG algorithm which considers both the temporal correlation and the spatial correlation. Experimental results demonstrate the superior performance of our method in comparison with the state-of-the-art methods typically for complex video sequences.

Reference

[1] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, "Distributed Video Coding," Proc. of the IEEE, Vol. 93, No. 1, pp. 71-83, Jan. 2005.

[2] Nikos Deligiannis, Adrian Munteanu, TomClerckx,Jan Cornelis,Peter Schelkens, "Overlapped Block Motion Estimation and Probabilistic Compensation with Application in Distributed Video Coding ," IEEE SIGNAL PROCESSING LETTERS,VOL.16,NO.9, pp. 743-746, Sep.2009.

[3] L. Natario, C. Brites, J. Ascenso and F. Pereira, "Extrapolating Side Information for Low-Delay Pixel Domain Distributed Video Coding," Int. Workshop on Very Low Bitrate Video Coding, Sep. 2005.

[4] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation With Spatial Motion Smoothing for Pixel Domain Distributed Video Coding," 5th EURASIP Conf. On Speech and Image Processing, July 2005.

[5] Badem M.B., Fernando, W.A.C., Martinez, J.L, Cuenca, P." An iterative side information refinement technique for transform domain Distributed Video Coding," Multimedia and Expo,ICME 2009, pp.177 – 180,2009.

[6] M.B. Badem, M. Mrak and W.A.C. Fernando, "Side information refinement using motion estimation in DC domain for transform-based distributed video coding," IET Electronics Letters, Vol. 44, No. 16, pp. 965 – 966, July 2008.

[7]Liu RongKe,Yue Zhi, Chen Changwen "Side Information Generation Based on Hierarchical Motion Estimation in Distributed Video Coding", Chinese Journal of Aeronautics, 22(2009),pp. 71-83.

[8] Otsu N, "Discriminant and least squares threshold selection," Proc. 4th IJCPR, pp. 592-596, 1978.

[9] N.Otsu, ''A threshold selection method from gray level histograms,'' IEEE Trans.Syst.ManCybern.SMC-9, pp. 62–66 1979.

[10]Astola J, Haavisto P, Neuvo Y "Vector median filters", Proceedings of the IEEE,1990, 78(4): 678- 689.

[11]Brownrigg, D.R.K "The weighted median filter", Communications of the ACM, 1984, 27(8): 807- 811.

[12] http://www.discoverdvc.or