COPY RIGHT

Title: COMPLEXITY REDUCTION IN INTER LAYER INTER PREDICTION IN SCALABLE HIGH EFFICIENCY VIDEO CODING

Paper Authors

C.SWETHA, E.RAMA KRISHNA

Sri Krishnadevaraya Engineering College, AP, India

.

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

COMPLEXITY REDUCTION IN INTER LAYER INTER PREDICTION IN SCALABLE HIGH EFFICIENCY VIDEO CODING

C.SWETHA[1], E.RAMA KRISHNA[2]

[1]PG Scholar, Dept of E.C.E, Sri Krishnadevaraya Engineering College, AP, India
[2]Assistant Professor, Dept of E.C.E, Sri Krishnadevaraya Engineering College, AP, India

Abstract—Memory bandwidth and on-chip memory requirements are critical issues in motion estimation (ME) implementations for video compression. The H.264/AVC scalable extension (SVC) provides variable frame rate and resolution video in a compressed digital sequence with interlayer prediction, which complicates the problems of limited memory bandwidth and onchip memory size. In this paper, an ME algorithm is proposed for the hardware encoder design of SVC that meets memory bandwidth and on-chip memory requirements. Clustered motion estimation and coding sequence reordering at macroblock and frame level processing are proposed. Compared with existing algorithms, the proposed algorithm has a 49.20% lower external memory bandwidth and reduces the on-chip memory requirement by 80.45% with video quality enhancements of up to 0.087, 0.090, 0.078, and 0.070 dB for four-layer (FullHD-HD-D1-CIF) spatial scalability, respectively.

Index Terms—H.264/AVC scalable extension (SVC), interlayer prediction, memory bandwidth, motion estimation (ME), on-chip memory, SVC.

1 INTRODUCTION

A video signal represented as a sequence of frames of pixels contains vast amount of redundant information that can be eliminated with video compression technology enhancing the total transmission and hence storage becomes more efficient. To facilitate interoperability between compression at the video producing source and decompression at the consumption end, several generations of video coding standards have been defined and adapted by the ITU-G and VCEG etc... Demand for high quality video is growing exponentially and with the advent of the new standards like H.264/AVC it has placed a significant increase in programming and computational power of the processors. In H.264/AVC, the

motion estimation part holds the key in capturing the vital motion vectors for the incoming video frames and hence takes very high processing at both encoder and the decoder. Wireless video transmission presents several problems to the design of a video coding system. First of all, some form of compression is needed for a bandwidth-limited system. Often, in a network environment for example, a certain amount of bandwidth is allocated to an individual user. Under these circumstances, a certain amount of "headroom" is allowed for each of the signal processing components based on user needs. The headroom for each of these components is usually not fixed, and is based on restricted channel capacity and

networking protocols needed to service the needs of its users. Given this, and the fact that video requires the highest bandwidth in a multimedia environment, the ability to vary the compression rate in response to varying available bandwidth is desirable. To achieve a certain bandwidth requirement, some combination of the following are required.

i)**Inter frame compression**: the idea behind interfame compression is that consecutive frames tend to have a high degree of temporal redundancy, and that the difference frame between the two would have a large number of pixel values near zero. So the result is a much lower energy frame than the originals, and thus more amenable to compression. Figure 1 shows the strategy for interframe coding. Because of the complexity and power increase in implementing motion estimation for interframe coding (requiring more than 50% of the total number of computations per frame), the cost value is high for interframe coding . Algorithms using interframe coding are often termed video coding algorithm.
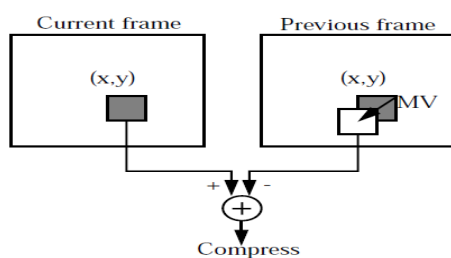
2 MOTION ESTIMATION AND COMPENSATION

Fig 1 Motion Estimation and Compensation.

ii) **Intra frame compression**: this implies spatial redundancy reduction, and is applied on a frame by frame basis. For situations where bandwidth is limited, this method allows for great flexibility in changing the compression to achieve a certain bandwidth. The key component in intra frame compression is the quantization, which is applied after an image transform. Because of the spatial correlation present after performing a transform (DCT or wavelet for example), quantization can be applied by distributing the bits based on visual importance of a spatially correlated image. This method of compression has the added advantage, that the compression can be easily varied based on available bandwidth on a frame by frame basis.

iii) **Frame rate reduction**: Another form of compression is reducing the frame rate of coded images. This results in a linear (1/N factor) reduction in the bitrate, where N is the current frame rate divided by the reduced frame rate. The resulting decoded frames at the decoder is also reduced by 1/N.

iv) **Frame resolution reduction**: The final form of compression is reducing the frame resolution. This results in a quadratic (1/N2) reduction in the bitrate, assuming uniform reduction in the horizontal and vertical directions. The encoder and decoder must have the ability to process variable resolution frames, thus making the design more complicated.

3 H.264/AVC SCALABLE EXTENSION

An important feature of SVC is the scalability of a single bit stream. Temporal scalability provides hierarchical coding structures with B or P frames that apply different frame rates according to

requirements. The hierarchical coding structures include hierarchical B frames and low-delay coding structure classes, as shown in Fig. 1. The frames are predicted from only the previous frames of a given layer. The B frames can be bidirectionally predicted, but P frames can only be predicted from previously encoded frames. The hierarchical prediction structure provides temporal scalability, and when compared to classical IBBP coding, also improves coding efficiency. The hierarchical structures result in coding delay, which can be controlled by restricted motion-compensated prediction from future frames. To reduce the coding delay, SVC provides a hierarchical low-delay coding structure which has the same degree of temporal scalability. In SVC, a nondyadic hierarchical prediction structure is provided, which affects the nondyadic frame rate in temporal scalability. The hierarchical prediction structure can also be the multiple reference picture concept of H.264/AVC. To increase the coding efficiency to meets user requirements, the group of pictures (GOP) size or the prediction structure can be varied. Spatial scalability means that the base layer (BL) has the lowest resolution and the ELs have a higher resolution. For SVC decoders, the BL can be decoded individually, but ELs need BL information to be decoded. The coding of each layer is similar to that in the H.264/AVC standard. In order to improve the EL coding efficiency, interlayer prediction techniques are adopted to improve rate distortion performance at the cost of increased encoding computational complexity. There are three interlayer

prediction schemes, namely interlayer intraprediction, interlayer motion prediction, and interlayer residual prediction. Based on these techniques, information from the BL, such as motion vectors and residuals, are applied to the encoding procedure of the ELs. The BL is important for the quality of an SVC bitstream because its quality influences that of ELs.Quality scalability is similar to spatial scalability, but they differ in their quantization step sizes for the BL and ELs. In other words, quality scalability is a special case of spatial scalability with identified resolution for the BL or ELs. A smaller quantization step size is used to refine texture information. Quality scalability is used to apply different bitrates for each layer to fulfill different requirements.

4 ON-CHIP MEMORY AND BANDWIDTH

Requirements of SVC Encoding Predicted motion vectors (PMV) and co-location as thesearch center are methods used to create a search window. The motion vector cover rates for the two methods. In this paper, the golden motion vectors were searched in the range ±128 using co-location. Neighbor MBs that hold movement behavior and PMV refine the center of the search window, which reduces the search range for finding the best motion vector. In other words, PMV performance is better than co-location with the same search range. In the ME procedure, each MB must reload self reference data. With the PMV method, the reference data is often not reduplicated between neighbor MBs, increasing memory BW that is defects for hardware implementations. Co-location

reuses partial search windows between neighbor MBs. And, common search ranges are not loaded into on-chip memory from external memory, decreasing the BW requirement. Therefore, the co-location method is suitable for hardware implementation. For temporal scalability, SVC adopts the hierarchical B frame encoding structure, as illustrated in Fig. 3. Two search ranges of ME are thus applied to the frames of the lower temporal layer, which doubles the ME operations and n-chip memory requirements. In MB ME processing, the search ranges of the reference frames of list0 and list1 are loaded into on-chip memory from external memory. The current MB encoding, respectively, performs ME on the search ranges SR0 and SR1 in the lower layer, reference frames F1 and F3, i.e., SR0 and SR1 data is kept in onchip memory during the current MB encoding. If the search range is ±128, the process requires an on-chip memory size of $2\times(128 + 16 + 128)2$ pixels for the higher temporal layer encoding. Since a pixel represents one byte, 1 46 968 bytes of storage are required. An external memory BW of 34545 MBps (megabytes per second) is required for SVC encoding for a 30 frames/s clip with full HD resolution.

5 PROPOSED SVC ME ALGORITHM

The on-chip memory size and the memory BW dominate the ME design of the SVC encoder. The CME is presented in this section. It predicts a common search center for several MBs. To adapt to different hardware resources, this paper proposes sub-block CME (SCME) and sub-block CME with small (SCMES) and large (SCMEL)

search ranges. In addition, this paper proposes CSR to extend the reuse concept to framelevel ME for the SVC hierarchal B frame encoding. The proposed algorithms efficiently reduce on-chip memory size and memory BW requirements.

A. Clustered Motion Estimation (CME)

For the variable MB partitions of SVC (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4), a total of 41 candidate MVs are generated during the ME of an MB. The PMV of the 16×16 pixel MB partition is an important index of the motion of an MB. Table II lists the percentages of the MVs of all the partition sizes located within the 3-pixel distance range of a 16×16 block PMV in the horizontal and vertical directions. The table shows that most MVs are covered by the PMV of the 16×16 partition size even in the small range. Each MB has 41 search window centers. Different search ranges must be loaded from external memory to on-chip memory for each motion search. To reduce memory BW, common hardware designs use the co-location point as a search center.
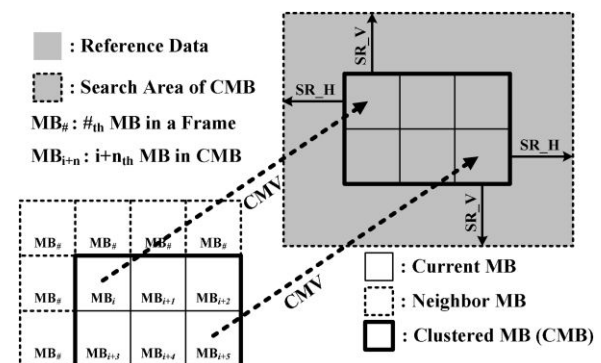

Fig. 2. Concept of proposed CME.

However, the video quality loss using the method is more significant than that using PMV as a search center in a given search range. In order to reduce the search range,

decrease computing power, and increase video quality, this paper proposes the CME, which adopts the PMVs of neighbor MBs to generate a clustered motion vector (CMV). The CMV is the common search center of the clustered MBs (CMBs). Each block size of an MB searched best motion vector in the same search window. Since the CMV was selected using a 16 × 16 block PMV, a smaller search range with CMV still provides a good-quality MV. Fig. 2 shows a 3 × 2 CME example in SVC encoding. In the figure, the CMB is composed of six MBs. The motion vectors of neighbor MBs is selected using the CMV as

$$CMV_{m \times n} = \frac{\sum_{k=0}^{m \times n} PMV_k(16 \times 16 \text{ block})}{m \times n} \quad (1)$$

where m and n are the MB numbers of the horizontal and vertical directions of a CMB, respectively. PMVk represents the 16 × 16 block PMV of the kth MB in the CMB. CMVmxn is the mean of the 16 × 16 block PMV of the neighbor MBs. The same CMVs were applied to the MBs in a CMB for most reference data reuse between the MBs for external memory prefetching. In our simulation, the CMV was always located around the best MV after the ME procedure; therefore, the reduced vertical and horizontal search ranges (SR−V and SR−H) are sufficient to provide a high-quality ME. The size of the CMB depends on the image resolution. The higher spatial layers have higher resolution and use more MBs in a CMB. The on-chip memory size is the CMB data plus the search area of the CMB. A 3 × 2 CMB with a ±16 pixel search range requires a total of 5120 pixel data in the on-

chip memory. For a list0 or list1 ME encoding of CMB, the data access is 10 240 bytes.For different clusters, the reused data rates, PSNR values, and BWs are different. In this paper, 2 × 2, 3 × 2, 4 × 2, 5 × 2, 6 × 2, and 7 × 2 clusters were simulated with two layers (QCIF and CIF) and a ±16 search range for both layers. The ME algorithm, where the search center was selected using the PMV, adopted the traditional full search. The simulation results are shown in Table III. According to the results, this paper adopts the 5 × 2 cluster and a ±16 search range for each layer. H.264/AVC SVC provides spatial scalability, with the BL with the smallest resolution and ELs with higher resolution. The interlayer prediction methods are applied for interprediction from the BL to the EL. Based on the characteristic of H.264/AVC scalable extension, base layer quality influences sensitively enhancement layer quality. In base layer, one MB was CMB to improve video quality.

B. Sub-Block CME (SCME)

On-chip memory size and external memory BW are important issues for SVC encoder design since they influence video quality. More memory size and BW lead to better video quality. For high-end hardware resources, SCME provides better video quality than that provided by CME. In general, each block size has its own search window. To reduce on-chip memory size and BW, CME adopts the same search window for all block sizes. Different from CME, SCME provides five search windows to each MB at the BL. The block size is classified into two types. Type I includes 16 × 16, 16 × 8, and 8 × 16, and type II

includes 8 × 8, 4 × 8, 8 × 4, and 4 × 4. Since an MB is divided into four 8 × 8 blocks, one MB contains five search windows at the BL, one for type I and four for type II. Type I shares one of the search windows and type II shares the other search windows,where Pn is an 8 × 8 block number and SWn is the search window of Pn for ME.

C. Sub-Block CME With Different Search Range between the Base and Enhancement Layers (SCMEL)

In CME and SCME, the BL and ELs use a ±16 search range. Generally, increasing the search range improves video quality. However, enhancement layer have large amount of MBs that heavily increase computing power, on-chip memory size, and external memory bandwidth. The BL has the lowest resolution and fewer MBs, which influence the video quality. This paper addressed SCMEL based on SCME, whose search Fig. 5. SCME search window in base layer for (a) 16 × 16, 16 × 8, and 8 x 16 blocks (type I) and (b) 8 × 8, 4 × 8, 8 × 4, and 4 × 4 blocks (type II).
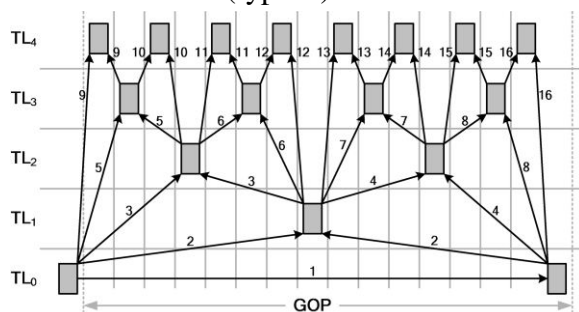

Fig. 3. Hierarchical B frame structure for five-temporal-layer SVC encoding.

range at base layer is larger than that of enhancement layer, with base layer at ±24 and enhancement layer at ±16. A smaller search range reduces on-chip memory size and external memory BW. This paper

addressed SCMES based on SCME. Assuming that a larger block size has larger movement and a smaller block size has smaller movement, type II (8 × 8,8 × 4, 4 × 8, and 4 × 4) block sizes have half the search range of type I (16 × 16, 16 × 8, and 8 × 16) block sizes, type I at ±16, type II at ±8, and enhancement layer at ±16.

D. Coding Sequence Reordering (CSR)

For frame-level data reuse, this paper proposes CSR for the SVC hierarchical B frame structure. Most hardware video encoders use MB-by-MB encoding. This allows better data reuse for a small on-chip memory design. A pipelined architecture always was applied in MB-level encoding to enhance performance. In Fig. 3, the numbers represent the corresponded encoding order from the BL (TL0) to the fifth layer (TL4) of temporal scalability. Each MB can be encoded by two processing elements, which calculate list0 and list1 ME concurrently. However, the different locations of the search windows for list0 and list1 result in double reference data access requirements. CSR is thus merged with CME.

6. CONCLUSION

External memory BW is a bottleneck for embedded multimedia systems. The on-chip memory size affects hardware cost. These parameters are thus important for realizing H.264/AVC SVC encoder hardware. This paper proposed four ME algorithms that reduce the external memory BW and on-chip memory size requirements while preserving high video quality. For different hardware resources, different proposed algorithms were adopted. The proposed CC algorithm effectively reduced external

memory bandwidth by 49.20% and the on-chip memory size requirement by 80.45%. Compared to previous researches, the PSNR of the proposed ME algorithm improved by 0.0813 dB on average. The proposed ME algorithms provide high encoded video quality with a small on-chip memory size requirement and low external memory BW.

REFERENCES

[1] ITU-T and ISO/IEC JTC1, JSVM-11 Software, JVT-X203, Oct. 2007.

[2] ITU-T and ISO/IEC JTC1, Joint Scalable Video Model JSVM-11, JVTX202, Oct. 2007.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[4] Z. Liu et al., "A 1.41W H.264/AVC real-time encoder SoC for HDTV1080P," in Proc. VLSI Circuits Symp., Jun. 2007, pp. 12–13.

[5] Y. H. Chen, T. D. Chuang, Y. J. Chen, C. T. Li, C. J. Hsu, S. Y. Chien, et al., "An H.264/AVC scalable extension and high profile HDTV 1080p encoder chip," in Proc. IEEE Symp. VLSI Circuits, Jun. 2008, pp. 104–105.

[6] X. Wu, W. Xu, N. Zhu, and Z. Yang, "A fast motion estimation algorithm for H.264," in Proc. IEEE Int. Conf. Signal Acquisition Process., Feb. 2010, pp. 112–116.

[7] A. Kundu, "Modified block matching algorithm for fast block motion estimation," in Proc. IEEE Int. Conf. Signal Image Process., Dec. 2010, pp. 260–264.

[8] H. T. Lin and J. S. Chiang, "A new diamond-arc-hexagon search," in Proc. IEEE Int. Symp. Signal Process. Inf. Technol., Aug. 2006, pp. 811–816.

[9] J. Byun, J. Choi, and J. Kim, "A fast multi-reference frame motion estimation algorithm," IEEE Trans. Consum. Electron., vol. 56, no. 3, pp. 1911–1917, Aug. 2010.

[10] Z. Chen, Q. Liu, T. Ikenaga, and S. Goto, "A motion vector difference based self-incremental adaptive search range algorithm for variable block size motion estimation," in Proc. IEEE Image Process., Oct. 2008, pp. 1988–1991.

[11] Y. H. Chen, T. D. Chuang, Y. H. Chen, C. H. Tsai, and L. G. Chen, "Frame-parallel design strategy for high definition B-frame H.264/AVC encoder," in Proc. IEEE Circuits Syst., May 2008, pp. 29–32.