COMPUTATION-FREE MOTION ESTIMATION WITH INTER-VIEW MODE DECISION FOR MULTIVIEW VIDEO CODING

Li-Fu Ding, Pei-Kuei Tsung, Shao-Yi Chien, Wei-Yin Chen, and Liang-Gee Chen

DSP/IC Design Lab, Graduate Institute of Electronics Engineering,
Department of Electrical Engineering, National Taiwan University

ABSTRACT

Multiview video coding (MVC) plays a critical role in 3D-TV systems. However, MVC require much more memory bandwidth and computational complexity relative to mono-view video coding. Therefore, an efficient prediction scheme is necessary for computation reduction. In this paper, a new fast prediction algorithm, computation-free motion estimation (ME) with inter-view mode decision, is proposed. By utilizing disparity estimation to find corresponding blocks between different views, the coding information, such as modes and motion vectors, can be effectively reused from the coded view channel. Therefore, the computation for ME can be almost totally neglected in most view channels. Besides, the proposed algorithm also provides the computation scalability. Experimental results show that compared with the full search block matching algorithm (FSBMA), the proposed algorithm not only save 61–72% computation with near-FSBMA quality, but save almost 100% computation with a little quality drop of 0.23–0.36 dB. Thus, computation-free ME is achieved.

Index Terms— Multiview video coding, motion estimation, disparity estimation, 3-D video, H.264/AVC.

1. INTRODUCTION

Multiview video can provide users with a sense of complete scene perception by transmitting several views to the receiver simultaneously. It can give users a vivid information about the scene structure. With the technology of 3D-TV and free viewpoint TV (FTV) getting more and more mature [1][2], multiview video coding (MVC) draws more and more attention. In recent years, MPEG 3D auido/video (3DAV) group has worked toward the standardization for MVC [3], which also advances the multiview video applications. From the discussion of MPEG 3DAV meeting, the developed coding scheme for multiview video settings mainly uses H.264/AVC to exploit temporal and inter-view dependencies [4]. Therefore, many coding aspects of MVC in the related research area are based on the hybrid coding scheme and highly related to the H.264/AVC.

Although MVC is an emerging technology, huge amount of video data and high computational complexity make it difficult to be realized. Different from mono-view video coding, disparity estimation (DE) is utilized to exploit inter-view dependencies in MVC. According to our instruction analysis, the prediction part which consists of ME and DE occupies 95% computational complexity in a MVC system. Consequently, the prediction part becomes the most computation-consuming part in a MVC system. Ultra-high computational complexity is a critical design challenge for MVC.

In a MVC system, ME removes the temporal redundancy while DE removes the inter-view redundancy. In addition to temporal and inter-view redundancies, another kind of redundancy called "computational redundancy" exists because of the close relation between motion vectors (MVs) and disparity vectors (DVs) in neighbor frames. According to the correlation, a fast prediction algorithm have been proposed to save the computation of ME for stereo video coding in our previous work [5]. However, the coding structures in MVC are more complicated than those in stereo video coding. In this paper, an new fast prediction algorithm, computation-free ME with inter-view mode decision, is proposed for MVC. Based on the fact that the video content is highly related between view channels, the proposed algorithm greatly reduces computational complexity while maintains video quality. The rest of the paper is organized as follows. The computation-free ME with inter-view mode decision is presented in Section 2. Section 3 shows the simulation results. Finally, Section 4 concludes this paper.

2. PROPOSED COMPUTATION-FREE ME WITH INTER-VIEW MODE DECISION

The computation-free ME algorithm with inter-view mode decision is proposed to retrieve the correlation between view channels. The proposed algorithm is based on the concept that MB prediction modes and their corresponding MVs can be predicted with the aid of DE and the coding information of neighboring views. Therefore, almost all the computational complexity of ME can be reduced. In this section, the system architecture of the multiview hybrid coding system is introduced first. Then the details of the proposed algorithm are presented.

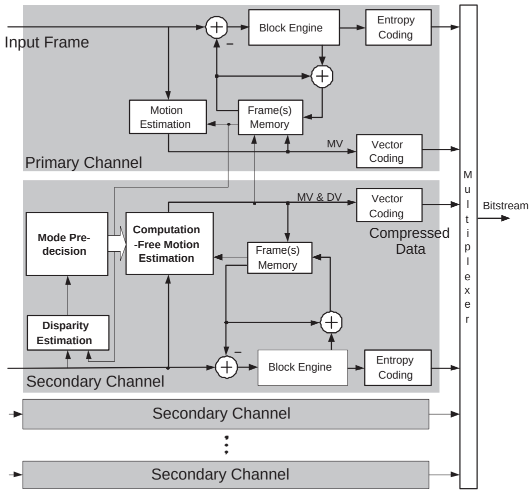


Fig. 1. Block diagram of the proposed multiview video encoder.

2.1. System Architecture

The block diagram of the proposed multiview video encoder is shown in Fig. 1. There are two kinds of view channels, the primary channel and the secondary channel. The main differences between the primary and secondary channel are DE and mode pre-decision part, which is introduced later. There is no DE operation in the primary channel. The number of primary and secondary channels depends on the coding structure. The block engine includes quantization, transform, and deblocking filter, etc.. After encoding, the compressed bitstream of each channel is transmitted. In addition, in the proposed algorithm, the prediction modes are clasified into four categories, that is, INTER_ME, INTER_DE, INTRA, and SKIP modes. In the subsection below, the mode pre-decision scheme is proposed to predict these modes.

2.2. INTER_DE/SKIP/INTRA Mode Pre-decision

In the proposed encoder, the best prediction mode is decided by the Lagrangian mode decision, which contains the distortion part and the rate part. According to the criterion of sum of absolute difference (SAD), the DV which results in the minimum distortion is computed by

$$DV = \arg\min_{DV}\{\sum_{t\in B_s, t'\in B'_p} |I_s(t) - I_p(t')| \,|_{B'_p\in SW_p(B_s)}\}. \quad (1)$$

B_s represents the current MB in a secondary channel. B'_p represents the search candidates located in the SW in the primary channel. Note that the cost for the rate part is not considered here for finding the best matching disparity-compensated (DC) block in the primary channel. However, the best DC block for coding is still searched by the Lagrangian mode decision. The cost, $Cost_{DE}$, which contains both the distortion and rate part, is stored. Then the threshold of early termination for ME in the secondary channel is chosen as following,

$$TH_{Cost_{ME}} = \arg\max_{TH_{Cost_{ME}}} \{R_p(MB_1), R_p(MB_2), ..., R_p(MB_4)\}. \quad (2)$$

$R_p(MB)$ means the overlap region of the MB which is overlapped by the DC block. After $TH_{Cost_{ME}}$ is decided, the early termination is activated and described by the following equation,

$$Skip_{ME_s} = \begin{cases} true, & \text{if } Cost_{DE} < \alpha \times TH_{Cost_{ME}}\,, \\ false, & \text{otherwise.} \end{cases} \quad (3)$$

Therefore, INTER_DE mode pre-decision is achieved with early termination of ME. It usually exists color mismatch between view channels due to the mismatch of sensors and preprocessing units in multiple cameras. The mismatch affects the performance of the scheme. Therefore, the purpose of the adjustable parameter α is to solve this problem.

SKIP mode is a useful and simple coding tool in H.264/ AVC. MV predictor is adopted for the current MB to generate a compensated block in this mode. Therefore, the computation of ME can be entirely saved if SKIP mode can be predecided. SKIP mode is powerful as well in MVC. According to our analysis, SKIP mode is chosen by 50%–90% MBs in the conditions of median and low bit-rates. In the same way, INTRA mode is chosen by most MBs in a frame in the condition of high bit-rate. Therefore, many computation operations for ME are wasted and required to be saved. The inter-view correlation can be utilized to save the unnecessary ME computation.

Figure 2 shows the decision and data flow of the proposed algorithm. Solid lines and grey blocks represent the decision flow, and dotted lines represent the data flow. First, $Cost_{DE}$ and the cost of SKIP/INTRA mode, $Cost_{SKIP}$/$Cost_{INTRA}$, are computed and compared. If $Cost_{SKIP}$ or $Cost_{INTRA}$ is smaller, the mode of the MB which is overlapped mostly by DC block, called "twin-MB" (MB), is checked. If TMB is also predicted by SKIP or INTRA mode, its cost, $Cost_{TMB}$, is compared with $Cost_{SKIP}$ or $Cost_{INTRA}$ of the current MB. Finally, the current MB is predicted by SKIP or INTRA mode if $Cost_{SKIP}$/ $Cost_{INTRA}$ is still smaller than $Cost_{TMB}$. On the other hand, if $Cost_{DE}$ is smaller in the beginning, the INTER_DE mode pre-decision scheme is applied.

2.3. INTER_ME Mode Pre-decision

To further reduce the computational complexity of ME in secondary channels, INTER_ME mode pre-decision scheme is proposed. After the decision flow of INTER_DE/SKIP/ INTRA mode pre-decision is executed, if the current MB mode is not classified to INTER_DE, SKIP, or INTRA mode, INTER_ME mode pre-decision starts. There are seven MB types according to their block size such as 16×16, 8×8, 4×4, etc.. It occupies over 90% computational complexity in video

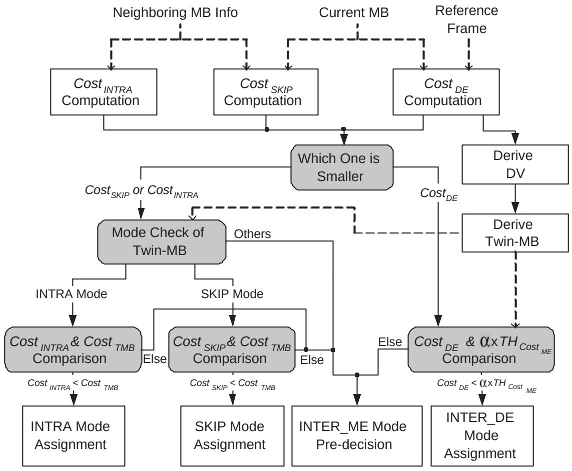


Fig. 2. Decision and data flow of the proposed algorithm.

coding. Thus, INTER_ME mode pre-decision scheme is proposed to predict the mode partition and the corresponding MVs of the current MB.

The illustration of INTER_ME mode pre-decision scheme is shown in Fig. 3. The location of DC block has been derived from DE in the process of INTER_DE/SKIP/INTRA mode pre-decision and is shown as the grey area in Fig. 3. Then the grey area is split into sixteen 4×4 sub-blocks. Each sub-block covers a 4×4 area in the reference frame, and the coding information of the reference frame can be also extracted. Then each sub-block is assigned a MV which is the same as the MV in the 4×4 area. Note that if the 4×4 area contains more than one MV, the MV of the coded sub-MB with the largest overlapping area which is covered by the 4×4 grey sub-block is assigned. To prevent the prediction error propagation, there is not any early termination and fast prediction scheme applied in the primary channel. It means all kinds of cost must be calculated in the primary channel. No matter what kind of mode are chosen, the best inter prediction mode and its corresponding MVs are stored. Therefore, if the covered MB in the reference frame is predicted by INTRA or SKIP mode, the MB partition and its corresponding MVs can be still adopted for INTER_ME mode pre-decision for the secondary channels.

After each sub-block is assigned a MV, the process of partition grouping begins. The MB partitioning defined in H.264/AVC follows a hierarchical and symmetric rule. Thus, the partition grouping also follows a hierarchical and symmetric manner. It is a bottom-up grouping process from 4×4 sub-block to 16×16 block. The neighboring four 4×4 sub-blocks with the same label are merged to a 8×8 sub-block, and so forth. On the other hand, when the neighboring four 4×4 sub-blocks belong to different labels, the costs of more possible MB partition are calculated according to a decision rule. From this manner, all MVs in a current MB can be predicted.

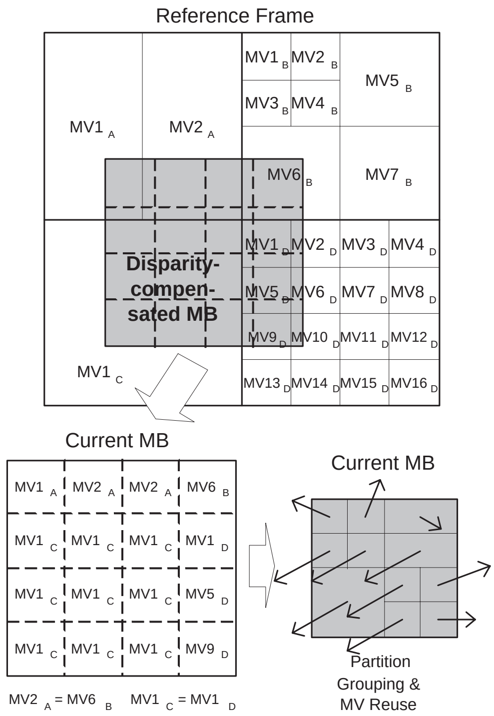
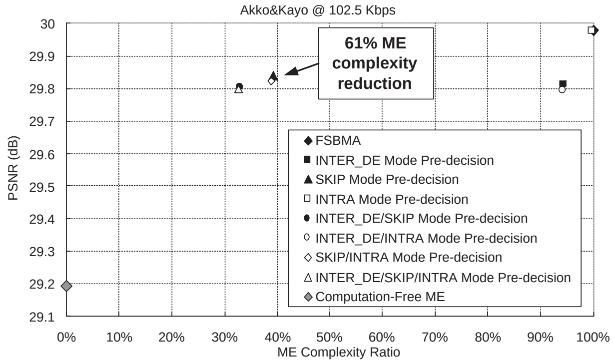


Fig. 3. The illustration of INTER_ME mode pre-decision.

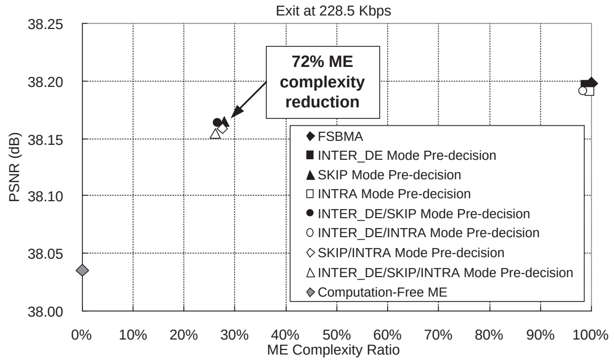
3. SIMULATION RESULTS

The proposed method is implemented by modifying the MVC -configuration in JSVM4.5 released from HHI [6]. It is compared with coding structures in which the full search block matching algorithm (FSBMA) is applied to both ME and DE. Rate-distortion performance of only secondary channels are compared because the ME parts in the primary channels in both cases are implemented with FSBMA. Sequences "Exit," "Ballroom," "Rena," "Akko&Kayo," with size 640×480 are tested. Two, three, and five view channels of these sequences are chosen for simulation. Note that the search ranges of DE and ME are both in [-64, +63] in the horizontal direction, while the search range of DE is [-16, +15] in the vertical direction.

Figure 4 shows the complexity-quality plots under different conditions of bit-rates. The computational complexity of ME in the secondary channel is considered. Note that in the simulation of ME complexity ratio, "search point per MB" is adopted as the unit of computation to make the results independent of different platforms. The computational complexity and coding performance of FSBMA is regarded as a criterion for comparison. Because there are four pre-decision schemes, it results in sixteen configurations for computation reduction for ME. By modifying different mode configurations, the coding performance varies case by case. Therefore, a computation-scalable ME can be achieved to fulfill various video quality under the conditions of different computational



(a)



(b)

Fig. 4. Complexity-quality plots of various test sequences with various bit-rates. (a) "Akko&Kayo" at 102.5 Kbps. (b) "Exit" at 228.5 Kbps.

budgets. After removing some unnecessary combinations of mode pre-decision schemes. There are nine computation-scalable configurations shown in Fig. 4. Take "Exit" in Fig. 4 (b) as an example, if only SKIP mode pre-decision scheme is turned on, 72% computational complexity can be saved with only 0.03 dB quality loss. Therefore, near-FSBMA quality is maintained.

In addition to the computation-scalability of the proposed algorithm, computation-free ME can be realized in secondary channels. Table 1 shows the comparison of rate-distortion performance between the proposed algorithm and FSBMA. Because the number of search candidates for refinement is negligible, almost 100% ME computation in secondary channels can be saved. According to the proposed algorithm, FSBMA is only performed for ME in the primary channel. The prediction modes and MVs in the primary channel become good reference for the proposed mode pre-decision schemes. Therefore, computation-free ME is achieved in most view channels. Even the penalty is the quality degradation of 0.23–0.36 dB in PSNR, the proposed algorithm still provides better coding gain than simulcast coding, that is, every channel is encoded independently, in most case. Note that because the

Table 1. Statistics of average quality drop and complexity reduction of computation-free ME

Sequences	Quality Drop	Complexity Reduction	Comparison with Simulcast	Complexity Ralative to Simulcast
Akko&Kayo	0.36 dB	99.96%	+0.55 dB	62.5%
Ballroom	0.23 dB	99.99%	+0.01 dB	50%
Exit	0.25 dB	99.99%	–0.12 dB	40%
Rena	0.32 dB	99.97%	+0.74 dB	40%

computational complexity of DE is 25% of that of ME, the proposed algorithm requires less computation than simulcast coding. The degree of computation reduction depends on the view numbers. Therefore, it means that the "computational redundancy" can be effectively exploited by the proposed algorithm.

4. CONCLUSION

This paper presents an efficient fast mode decision algorithm for the prediction part in MVC. The computation-free computation ME with inter-view mode decision is proposed to overcome the design challenge of ultra-high computational complexity in MVC. Based the concept of high inter-view correlation between views and the feature of mode distribution different from that in mono-view video coding, the proposed algorithm effectively reduces near 100% computational complexity for ME in most view channels. Thus computation-free ME is achieved with little quality degradation. Moreover, the configurations of INTER_DE/SKIP/INTRA mode pre-decision and INTER_ME mode pre-decision can be determined to be turned-on or not to fulfill the requirements of desired video quality or computational resource allocation.

5. REFERENCES

[1] F. Isgrò, E. Trucco, P. Kauff, and O. Schreer, "Three-dimensional image processing in the future of immersive media," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 388–303, Mar. 2003.

[2] T. Fujii and M. Tanimoto, "Free-viewpoint TV system based on ray-space representation," *Proceedings of SPIE*, vol. 4864, pp. 175–189, Mar. 2002.

[3] MPEG-4 Video Group, *Requirements on multi-view video coding*, Number ISO/IEC JTC1/SC29/WG11 N6501. MPEG-4, 2004.

[4] MPEG-4 Video Group, *Description of Core Experiments in MVC*, Number ISO/IEC JTC1/SC29/WG11 W8019. MPEG-4, Apr. 2006.

[5] L.-F. Ding, S.-Y. Chien, and L.-G. Chen, "Joint prediction algorithm and architecture for stereo video hybrid coding systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 11, pp. 1324–1337, Nov. 2006.

[6] MPEG-4 Video Group, *AHG on Multiview Video Coding*, Number ISO/IEC JTC1/SC29/WG11 N7829. MPEG-4, Jan. 2006.