Algorithm and Hardware Architecture Design for Weighted Prediction in H.264/MPEG-4 AVC

Chi-Sun Tang, Chen-Han Tsai, Shao-Yi Chien, and Liang-Gee Chen
DSP/IC Design Lab and Media IC and System Lab
Graduate Institute of Electronics Engineering
National Taiwan University
Email: {chamot, phenom}@video.ee.ntu.edu.tw

***Abstract*— Weighted prediction (WP) is a tool to compensate the brightness difference in video sequences with brightness variations. In this paper, some weight parameter determination methods are surveyed, and a weighted prediction algorithm together with the hardware architecture design is proposed. The main idea of the algorithm is to limit the number of weight parameters transmitted by quantizing the parameter into levels and using only offset as the parameter. As a result, the extra parameters sent in each slice header is thus limited by the number of levels, and the parameter determination process requires much less computations. By further utilizing sub-sampling of brightness levels and estimated offset sum, a simplified architecture is also proposed. Simulation result shows that the later architecture achieves a coding gain of about 0.5dB over weighted prediction method in JM9.6 [1] and has a minimum overhead to hardware implementation.**

I. INTRODUCTION

In coding video sequences with change of ambient light, change of light source, fade-in/out, camera-iris adjustment, camera flash ... etc, the variation of light intensity globally or locally between the current frame and the reference frame would introduce inaccuracy to the motion estimation in video coding. This has a negative effect on the compression efficiency and bit-rate in particular applications like video conference and video capturing in mobile devices.

Weighted prediction (WP) is a tool to compensate the brightness difference so that the reference frame is more strongly correlated to the current frame after applying a set of multiplicative weight with an additive offset. When performing motion estimation in video with brightness variation, WP can assist in uncovering the best candidate even though the best candidate might not have a minimum cost of SAD (sum of absolute difference). More accurate motion estimation and smaller residue can be obtained at the expense of higher complexity in the encoder and extra headers transmitted.

H.264/AVC is the first international standard to include weighted prediction tool in the Main, Extended and High Profiles. The weighted prediction flow includes an estimation of a set of weight parameter with a multiplicative weight and an additive offset. There are two modes of weighted prediction in H.264, the explicit mode and the implicit mode. The determination of WP parameters is outside the scope of the H.264 specification.

In this paper, a scheme in determining the weight parameters which is suitable for hardware implementation is proposed. In section II, we would first review some weight parameters determination methods and later discuss the challenges in hardware implementation in section III. We will present our scheme and the execution flow together with hardware architecture in section IV. Simulation results and conclusion will be given in section V and VI respectively.

II. SURVEY OF WEIGHTED PREDICTION METHODS

The process of weighted prediction in hybrid video coding can be divided into three parts. Survey and analysis on the methods proposed by some papers is discussed below.

A. *Brightness Variation Detection*

In [2], the input is tested against the reference frames based on cross-entropy between them and decide whether or not brightness variation model will be applied. In [3], the algorithm calculate the histogram difference for variation detection. Weighted prediction will be applied only in frames which brightness variation is detected.

B. *Weight Parameter Determination*

In H.264 reference software JM9.6, only the scaling factor is determined by calculating the reference frame to current frame average brightness ratio. In [4], [3], [5], the optimal weight parameter is determined by evaluating the set of linear parameter (weight, offset) that could give the minimum MSE (minimum square error) between the current and WP compensated frame. Other two modes of linear transforms, scale only and offset only are discussed in [6].

C. *Refining the Motion Estimation*

The scaled reference frames are stored in another array which is used in motion estimation for more accurate result in H.264 reference software JM9.6 and [2]. This method is frame to frame based, and can not handle brightness variations that happen spatially across an image sequence efficiently.

In [3], [5], it is adaptive in whether to apply weighted prediction to certain macroblocks when there is variation that happens spatially. Although this model could perform well in software environment, the brightness variation detection step proposed is difficult to apply to macroblock-pipeline based hardware architectures. For spatial brightness variations having many brightness difference levels across the sequence,

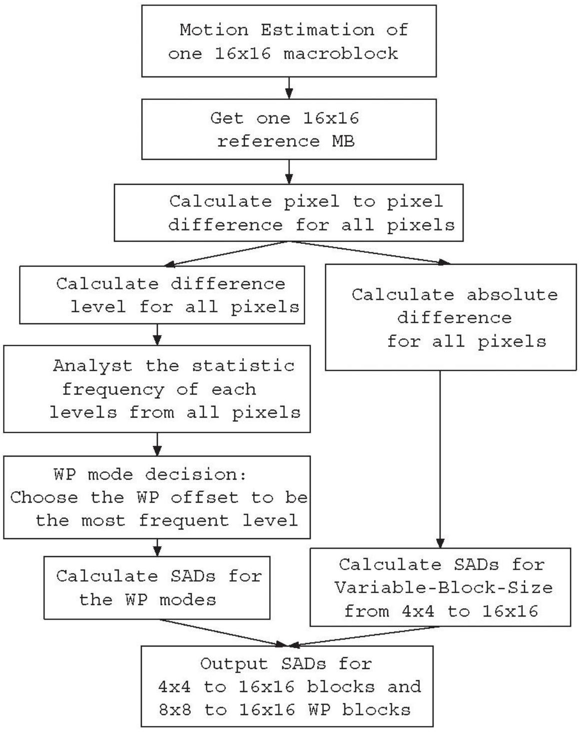


Fig. 1. Flow diagram of the proposed algorithm.

adaptively choosing to apply a single set of weight parameter or not applying to individual macroblocks may not be enough. A larger set of weighted parameters will be needed to model the above situation. The high computational complexity makes it expensive to implement in real-time hardware design.

III. CHALLENGE IN HARDWARE IMPLEMENTATION

After the above analysis, we could find that the difficulties in finding the best weight prediction compensated candidate and the optimal WP parameter in video coding would be the large computations in finding the best candidate in motion estimation and simultaneously considering the brightness variation with weighted prediction. For hardware implementation, motion estimation is already the most computational intensive part and iterative algorithms would be impractical in real-time applications.

On the other hand, the extra bits in transmitting the weight parameters could diminish the gain from smaller residue and accurate prediction if the levels and numbers of weighted parameters transmitted are not carefully controlled. For variations that appear non-uniformly across the whole frame, a large number weight parameters are required and thus the effectiveness in applying WP tool will be reduced.

IV. PROPOSED IMPLEMENTATION SCHEME

The goal of our design is to find a method for determination of a coarse WP parameter which can obtain moderate gain in handling video with brightness variation. The hardware implementation of the algorithm should be cost effective in view of smaller architectural size. Integration of the WP hardware with existing designs of motion estimation module could further increase the speed and lower the overall size of the design. It should also be capable to handle variable block sizes in motion estimation as well.

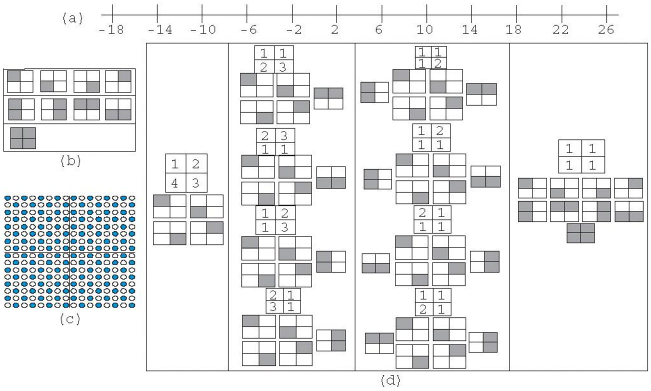


Fig. 2. (a) Quantization ranges for offset parameters, (b) Nine modes of weighted prediction for a 16x16 MB, (c) Sub-sampling points, (d) The output patterns of SAD-WP calculated for various number of different WP parameters.

A. *Weight Parameter Determination*

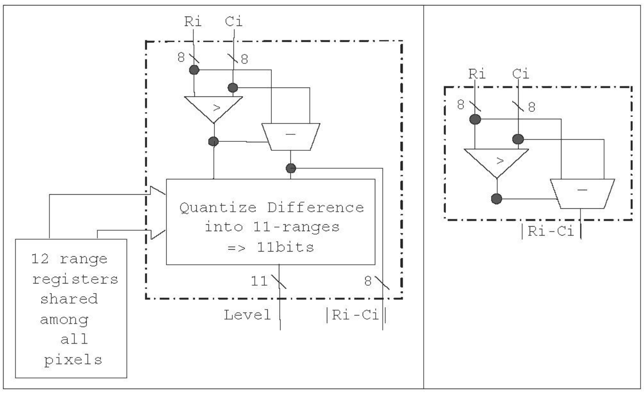
Figure 1 shows the flow of the proposed weighted prediction scheme. This approach is based on the full search block-matching algorithm (FSBMA) for motion estimation of MBs with variable block sizes as presented in [7].

In the process of motion estimation, a 16x16 current block and a 16x16 reference block is loaded in a 256-Byte SRAM module and a 256 shift-register file respectively. By shifting the shift-register file array, different 16x16 MBs is loaded in each cycle for motion estimation.

The quantization levels for offset are showed in figure 2(a), and they can be dynamically updated. These range levels are chosen for applications with moderate (but not abrupt) brightness variations between frames.

On one path of the flow in figure 1, the absolute difference values of all pixels are added together in a multi-level 2-D adder tree that will give the SADs for variable block sizes of 4x4, 4x8, 8x4, 8x8, 8x16, 16x8 and 16x16 simultaneously[7]. On the other path, the difference levels from all pixels are analyzed for the four 8x8 MB-partitions and the most frequent level is chosen for each of the four WP offsets if its frequency is higher than a preset threshold.

There is a total of 9 modes of weighted prediction which consists of four 8x8, two 8x16, two 16x8 and one 16x16 modes as shown in figure 2(b). After obtaining the four 8x8 WP offsets for the 16x16MB, we can calculate the 9 modes of SAD-WP simultaneously in another multi-level 2-D adder tree or by an approximation of SAD-WP that will be described later. Further mode decision and rate-optimization strategy is possible using the 41 SADs and 9 SAD-WP obtained, and this process is not discussed in this paper.



(a) PE for one Pixel(i)

(b) PE for one Pixel(i) of decimated pixels

Fig. 3. Proccessing unit for one pixel; (a) calculate both absolute difference and difference level; (b) calculate only the absolute difference.

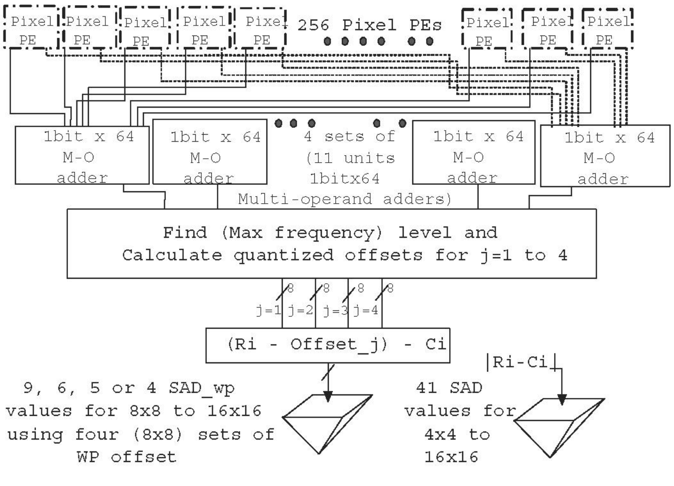


Fig. 4. Circuit diagram of the first architecture.

This algorithm can be further refined by using the fact that brightness difference levels usually have a low spatial frequency. Sub-sampling on a 2 to 1 pixel basis will be enough for sampling brightness difference levels in finding the WP offset for each block. Figure 2(c) shows the sub-sampling space for each 16x16 MB. Sub-sampling could result in a dramatic decrease in area with hardware implementation.

B. Hardware Architecture

Two different architectures are proposed in the implementation of the above mentioned algorithm. Their design will be discussed in the following.

1) About the First Architecture Implemented: The first architecture is shown in figure 4. In processing each 16x16 motion estimation block, calculation of absolute difference and the quantization of the absolute difference into offset levels for each pixel are performed in the same PE unit. The architecture of a PE which performs offset level calculation and a PE which do not are shown in figure 3. The pixel differences are compared with the 12 range registers for quantization into 11 levels. Range registers are shared among all pixels and the register values are defined as in figure 2(a). One advantage of using registers to hold the values is the flexibility to adjust the quantization levels when handling multiple reference frames, although it is not studied in this paper.

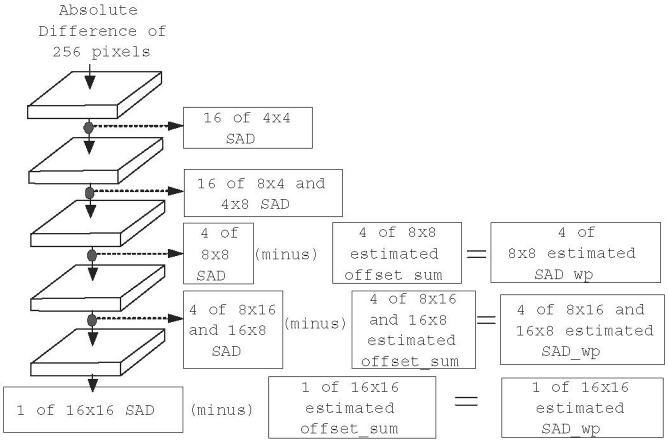


Fig. 5. SAD and SAD-WP values obtainable at various levels of the 2-D adder tree in the second architecture.

After the processing by the 256 PE array, the 11 offset levels from 256 pixels are feed into multi-operand adders. Multi-operand adders are divided into 4 sets of 8x8 size. Each set contains 11 units of 1bit x 64inputs (or 1bit x 32inputs for sub-sampled case) multi-operand adders. These 4 sets are used to calculate the offset level of the four 8x8 MB-partition. The most frequent level will be taken as the offset value for the 8x8 block if the frequency is higher than a threshold of 32 (or 16 in sub-sampled case). Otherwise an offset of 0 (no WP) will be applied. This is illustrated in figure 2(d). Nine different SAD-WP is produced only when the four offset values of each 8x8 MB-partition is the same. For 16x16 MB having 2 different 8x8 offset values, 6 SAD-WP is produced, 5 and 4 SAD-WP is produced for 3 and 4 different 8x8 offset values respectively.

After obtaining the absolute differences for each pixel, the 41 SAD values are calculated in one of the multi-level adder trees. Meanwhile the SAD-WP are calculated in another 2D adder-tree.

2) About the Second Architecture Implemented: The second architecture is designed for smaller area and faster implementation with a minor degrade in quality (weighted prediction accuracy). Instead of using a separate 2-D multi-level adder tree for calculating the SAD-WP for weighted prediction, only one adder tree is used to calculate both SAD and SAD-WP simultaneously. The estimated SAD-WP is calculated using the side information of number of positive and negative results in pixel subtraction in the pixel PE as shown in figure 3 and the calculation is given by

$$OffsetSum = (no.positive - no.negative) \times Offset$$
$$SADwp = SAD - OffsetSum$$

The SAD and SAD-WP values are obtained in one adder tree as shown in figure 5.

Fig. 6. The test sequence 'rabbit' used in the simulation.

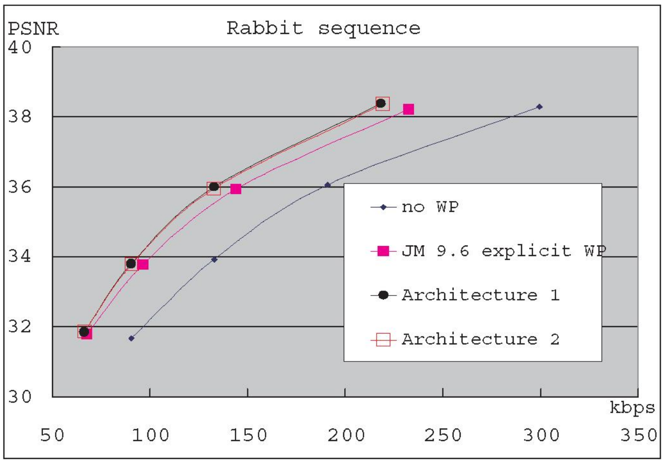


Fig. 7. Simulation result for the proposed architectures.

V. SIMULATION RESULTS

Simulation is performed in the modified H.264 reference software JM9.6 that support multiple reference-picture indices. The objective is to enable MBs in the same picture to use different weighting parameters even when predicted from the same reference picture [8]. The GOP structure is IBBPBB and the JM quantization parameters QP tested are 28, 32, 36 and 40.

In order to simulate the performance in particular applications like video conferencing and video capturing in mobile devices, a self recorded short video sequence "rabbit" fig 6, is used[9]. The sequence has a 320x240 resolution and some translational motion of a character moving before a background photo. The lighting is adjusted by a dimmer lamp and the video shows continuous variations of brightness temporally and spatially.

The simulation results are shown in the rate-distortion curve in figure 7. The proposed architectures of the first and second ones are simulated and compared. The performance of the JM 9.6 explicit-WP and JM 9.6 without-WP is also shown on the same figure. At 200kbps, the first and second architectures performs 0.57dB and 0.54dB higher than the weighted prediction method as in JM9.6 in explicit mode respectively. In comparison with JM9.6 with weighted prediction disabled, they performed 1.95dB and 1.90dB better for the first and second architectures. The result shows that the proposed algorithm performs better than the algorithm used in the reference software. The design using the second architecture performs nearly the same as the first with only about 0.04dB lower PSNR at 200kbps. This shows that the second simplified hardware architecture is justifiable despite the lower PSNR.

VI. CONCLUSION AND FUTURE WORK

This paper presents a scheme in the determining of the weight parameters which is suitable for hardware implementation. The proposed algorithm is capable to handle moderate brightness variations in a video sequence. The association of more than one weight parameters for a particular reference picture is used in this algorithm. Hardware architectures are presented base on the above algorithm and the performance is compared.

The simulation result reveals that the proposed architectures perform about 0.5 dB better than the weighted prediction method as in JM9.6. It also shows that the second architecture with sub-sampling of brightness level and approximated SAD-WP would be the more efficient hardware to be implemented, although it has a minor 0.04dB PSNR lower than the first architecture.

Further investigation will be performed on the using of different level boundaries by loading different values into range registers. Efficient architecture using a scale and offset will also be evaluated. An implementation of the architecture with the motion estimation module will be performed in an ASIC environment in the future.

REFERENCES

[1] "JVT reference software version JM9.6." [Online]. Available: http://http://iphome.hhi.de/suehring/tml/doc/index.htm

[2] S. H. Kim and R.-H. Park, "Fast local motion-compensation algorithm for video sequences with brightness variations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 289 – 299, Apr. 2003.

[3] Y. Shen, D. Zhang, C. Huang, and J. Li, "Adaptive weighted prediction in video coding," in *IEEE International Conference on Multimedia and Expo (ICME'04)*, vol. 1, June 2004, pp. 427–430.

[4] H. Kato and Y. Nakajima, "Weighting factor determination algorithm for h.264/mpeg-4 avc weighted prediction," in *2004 IEEE 6th Workshop on Multimedia Signal Processing*, Sept. 2004, pp. 27–30.

[5] K. Kamikura, H. Watanabe, H. Jozawa, H. Kotera, and S. Ichinose, "Global brightness-variation compensation for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 988–1000, Dec. 1998.

[6] D. Liu, Y. He, S. Li, Q. Huang, and W. Gao, "Linear transform based motion compensated prediction for luminance intensity changes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, pp. 304–307, May 2005.

[7] S.-Y. Chien, Y.-W. Huang, C.-Y. Chen, H. H. Chen, and L.-G. Chen, "Hardware architecture design of video compression for multimedia communication systems," *IEEE Commun. Mag.*, vol. 43, no. 8, pp. 122–131, Aug. 2005.

[8] J. Boyce, "Weighted prediction in the h.264/mpeg avc video coding standard," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS'04)*, May 2004, pp. III–789–792.

[9] "Rabbit YUV 4:2:0 sequence," DSP / IC Design LAB, GIEE NTU. [Online]. Available: http://video.ee.ntu.edu.tw/ chamot/video/