DC-12

720×480 30fps Efficient Prediction Core Chip for Stereo Video Hybrid Coding System

Li-Fu Ding, Shao-Yi Chien, and Liang-Gee Chen
DSP/IC Design Lab, Graduate Institute of Electronics Engineering and
Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
Email: {lifu,shaoyi,lgchen}@video.ee.ntu.edu.tw

***Abstract*— The chip design of prediction core in the stereo video hybrid coding system is implemented with 0.18 μm 1P6M technology by TSMC. The die size is 4.53 mm^2. This IC can can achieve real-time requirement under the operating frequency of 81 MHz for 30 D1 frames per second (fps) in the left and the right channel simultaneously, with ME/DE search range of [-64, +63] in horizontal direction and [-32, +31]/[-16, +15] in vertical direction. Compared with the hardware requirement for implementation of full search block matching algorithm (FSBMA), only 11.5% on-chip SRAM and 1/30 amount of PEs are needed. It shows that the hardware cost is quite small.**

I. Introduction and Motivation

Stereo video can make users have 3D scene perception by showing two frames to different eyes simultaneously. With the technologies of 3D-TV getting more and more mature [4], stereo and multi-view video coding draw more and more attention. In recent years, MPEG 3D auido/video (3DAV) Group has worked toward the standardization for multi-view video coding [5], which also advances the stereoscopic video applications. Although stereo video is attractive, the amount of video data and the computational complexity is doubled. A good coding system is required to solve the problem of huge data with limited bandwidth. Besides, in a mono video coding system, the prediction in temporal domain, motion estimation (ME), requires the most computational complexity [1]. By comparison, computational load of prediction is heavier in stereo video coding systems due to additional ME and disparity estimation (DE), which is the prediction in spatial domain, in the additional channel. For real-time applications, a hardware accelerator solution is urgently required due to the heavy computational complexity.

Under real-time constraint, it is preferred that ME and DE unit, which is called "prediction core" in the system, should be realized by a hardware accelerator due to its heavy computational load. Among all the block matching algorithms, full search block matching algorithm (FSBMA) is the most popular [3]. Many FSBMA architectures were developed based on the regular data flow [7]. However, for D1 30 fps stereo video contents, hardware processing parallelism needed is more than 4096 (that is, 4096 absolute difference operations execute simultaneously) to achieve real-time requirement. The hardware cost is quite large. Hierarchical search block matching algorithm (HSBMA) has been regarded as powerful computational configuration in BMA [3]. It can effectively reduce not only computational load but also on-chip memory size. However, due to its irregular data access in fine levels, data of search windows (SWs) are hard to be reused effectively. It causes much more data access load than FSBMA. Therefore, an algorithm and hardware architecture for stereo video coding system are designed for the compromise between area cost, stereo video quality, and processing capability. This IC is implemented based on the developed algorithm and architecture.

The stereo video coding system is first described in the next section. Then the hardware architecture and the chip implementation are shown in Section III and IV, respectively. Finally, in Section V gives the conclusion.

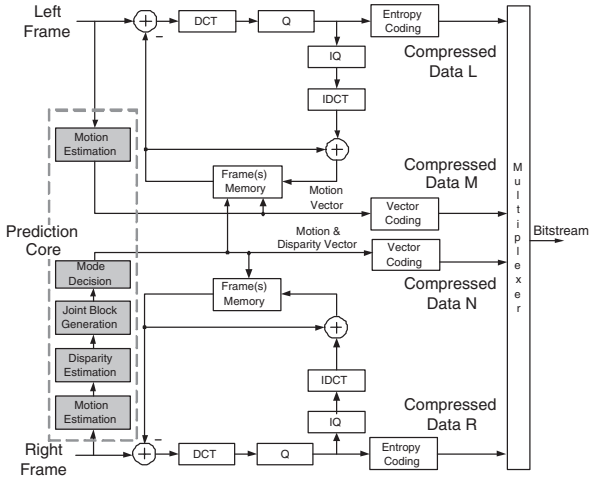

Fig. 1. Block diagram of the proposed stereo video encoder

II. Stereo Video Coding System

For the purpose of compatibility, the coding system adopts a base-layer-enhancement-layer scheme. The left view is set as the base layer, and the right view is set as the enhancement layer. The base layer is encoded with MPEG-4 video encoder. The implementation of the proposed stereo video coding system is based on two hardware-oriented concepts. First, in order to greatly reduce the area of processing elements (PEs) and on-chip memory, we adopt HSBMA to perform ME and DE. Second, in the compensation step, a block is not only compensated by the block of left or right reference frames, but also the combination of them for different types of content in the current block. Based on these concepts, the system block diagram of the proposed stereo video encoder is shown in Fig. 1. The proposed prediction core mainly includes ME, DE, joint block generation, and mode decision, which is the most computation-consuming part in the system. The main differences between the encoding flows of the left channel and the right channel are DE, joint block generation, and mode decision, which are introduced later. After encoding, the compressed data of the left channel, M and L, and the compressed data of the right channel, which is of a small amount, N and R, are transmitted.

In ME and DE steps of the right channel, the current block has two reference frames, as shown in Fig. 2. Grey region is the search range of a reference frame. There are three types of compensated blocks for right channel in the proposed stereo video encoder. They are motion-compensated block, disparity-compensated block, and proposed joint

0-7803-9162-4/05/$20.00 ©2005 IEEE

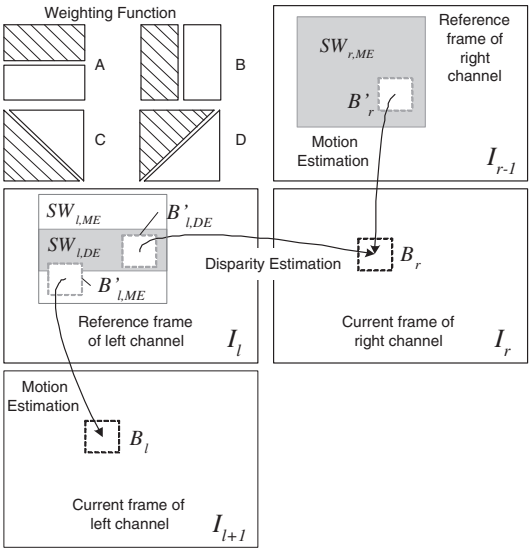

Fig. 2. The illustration of the prediction directions and the search range of two reference frames.

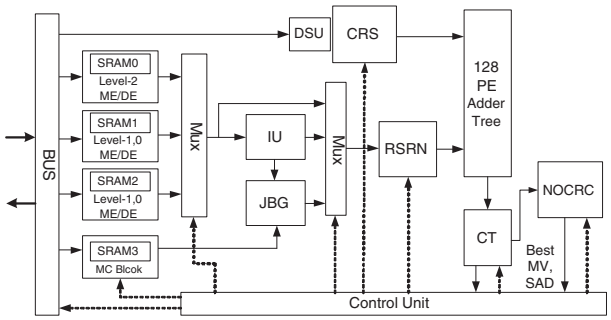

Fig. 3. Architecture of the prediction core chip for stereo video hybrid coding system.

block. The joint block can be predicted by the weighted sum or linear combination of different types of blocks. According to the criterion of sum of absolute difference (SAD), the best type of compensated block is selected by the mode decision. PSNR gain is 2−3 dB with joint block compensation [2]. The proposed scheme greatly improves the coding efficiency of stereo video systems.

III. Hardware Architecture of Prediction Core Chip

The hardware architecture designed for the hardware-oriented algorithm is shown in Fig. 3. There are eight main units: control unit, reference shift register network (RSRN), current register set (CRS), 128-PE adder tree, compare tree (CT), NOCR checker (NOCRC), interpolation unit (IU), and joint block generator (JBG). RSRN is composed of a reconfigurable shift register array. After data loading of search window is finished, RSRN starts to fetch data from on-chip memory. Meanwhile, PE-adder tree accumulates the distortion (absolute difference). Except for several cycles in the beginning for data preparing, SADs of eight candidate blocks in level-2 block matching process (BMP), two candidate blocks in level-1 BMP, or half candidate blocks in level-0 BMP can be derived in every cycle. Then compare tree can compare these SADs in one cycle. The best three candidate MVs are chosen for refinement process. NOCRC checks the degree of overlapping and outputs the post-processed MVs to address generator (AG) in the control unit, which decides when

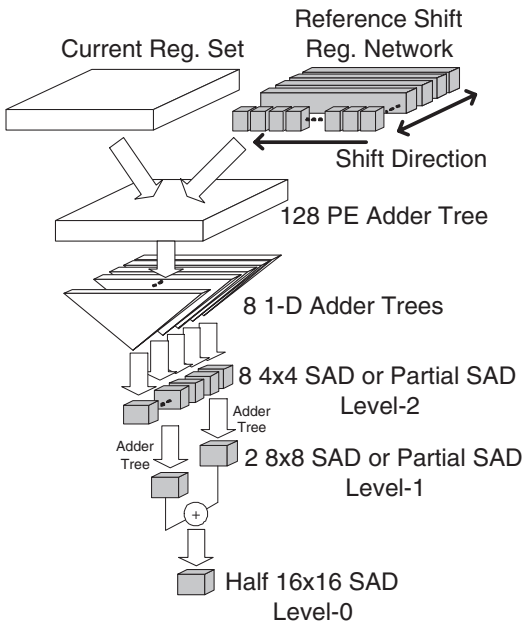

Fig. 4. Architecture of RSRN connected by 128 PE adder tree.

block matching process of the next level begins. IU generates sub-pixels in half pixel refinement process. JBG generates joint block for mode decision for improving coding efficiency of stereo video. The detail architectures of RSRN, JBG, and NOCRC are described in the following paragraphs. Furthermore, data reuse scheme and memory organization are also shown. Besides, a new scheduling is proposed to reduce the demand of on-chip memory and bandwidth for data access.

To achieve the design goal of hierarchical block-matching operation with only one hardware resource, RSRN is composed of a set of reconfigurable shift register array, which consists of 128 8-bits registers. As shown in Fig. 4, the outputs of these registers are connected to 128 processing elements (PEs) in the adder tree, which can calculate sum of absolute differences of 128 pixels and accumulate them in one cycle. It has high reconfigurability and can change the connection configuration by the control unit. One column of pixels are fetched from on-chip memory to RSRN every cycle. Because of its reconfigurable feature, data in RSRN can shift left-ward, down-ward, and right-ward, there are no bubble cycles when the search position is changed in the vertical position.

After SADs are generated from 128-PE adder tree, compare tree compares eight 4×4 SADs in level-2 BMP or two 8×8 SADs in level-1 BMP. Both in level-2 and level-1, the best three MVs are chosen, and then inputted to NOCR checker. MV differences are calculated mutually, and then the overlapping condition is decided. This architecture is simple, but it can effectively reduce over 30% unnecessary data access from off-chip. Furthermore, it also reduces unnecessary computation and saves processing cycles.

When ME of the right channel is finished, the best candidate block must be hold for joint block generation step. As motion compensation process, the best candidate block is loaded into on-chip SRAM "MC block," as shown in Fig 3. After DE of the right channel is also finished, mode decision for joint block starts. Figure 5 shows one of the sixteen processing elements in JBG. Only adders are used to generate weighted sum in joint pel generation unit. In our chip design, there are sixteen PEs to generation eight kinds of joint blocks at the

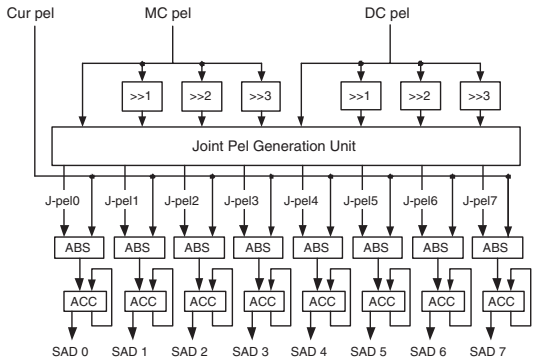

Fig. 5. Architecture of one PE of the joint block generator.

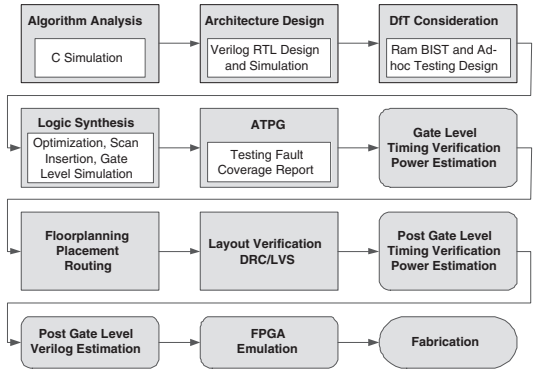

Fig. 6. Design flow and methodology of the prediction core chip.

same time. After 16 cycles, eight 16×16 SADs of joint blocks are generated. Then they are outputted to the compare tree to choose the best SAD, and the best mode is derived as well.

IV. Chip Implementation

The prediction core chip is implemented on a 4.53 mm^2 die with 0.18 μm 1P6M technology by TSMC. The detailed implementation flow is described from now on.

A. Design Flow and Methodology

Figure 6 shows the design flow and methodology for the prediction core chip. First, the platform of stereo video coding system is built up. The stereo video encoder and decoder is implemented by software C and followed by the algorithm analysis of the prediction core. Then, architecture design starts. It is realized with Verilog HDL. After a large amount of Verilog-XL simulations are all correct, RAM BITS and Ad-hoc testing design are applied for DfT consideration. Then, the RTL design is synthesized by Synopsys Design Compiler with the 0.18 μm Artisan cell library. Meanwhile, scan chain insertion is applied. Then, Verilog-XL is used to perform the gate-level simulation to make sure that the target specification is met and followed by the power estimation by the Synopsys Power Compiler. After front-end is finished, back-end of the chip design begins. We use Synopsys Astro as the backend tool that performs timing-driven automatic place-and-route. The layout verification including DRC and LVS are then performed. Finally, power estimation and post gate-level simulation are done and followed by FPGA emulation. Figure 7 shows the layout view. The detailed features of the chip is shown in Table I.

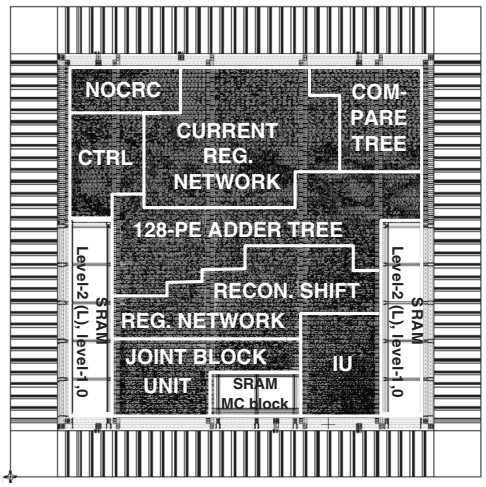

Fig. 7. Chip layout of the proposed prediction engine architecture.

TABLE I
Chip specifications.

Technology	TSMC 1P6M 0.18um
Chip size	2.13mm×2.13mm
Package	128 CQFP
On-chip memory	21,248 bits
Logic gate count	137,838
Maximum frequency	100 MHz
Power supply	1.8V
Power consumption	95.85 mW @ 100 MHz
Search range	ME: horizontal [-64, +63], vertical [-32, +31] DE: horizontal [-64, +63], vertical [-16, +15]
Processing capability	30 D1(720x480) frames/sec in both channels including 2 ME and 1 DE operations

B. Prototype of Real-time Stereo Video System

In addition to chip implementation, the architecture also passed FPGA emulation and is integrated to the prototype of real-time stereo video system, as shown in Fig. 8. First, stereo video is captured by a stereo video camera and followed by pre-processing step. Then, the encoding process is performed with HW/SW co-design. Annapolis Firebird FPGA card is used for hardware acceleration of the prediction core. It also proves the correctness of chip architecture. The encoded bitstream is transmitted to the client PC, and then it is decoded by the developed stereo video decoder software. After the post-processing step, the stereo video is displayed on the 3D LCD displayer on real-time.

C. Test Considerations

In this design, three design for testability (DfT) techniques are applied. They are ad-hoc testing, RAM Build-In-Self-Test (BIST), and scan chain insertion.

By multiplexing the input and output data in specific modules, the errors can be shrunk to a particular area. In the ad-hoc mode, we make input signals directly connected to inputs of certain module and observe the output signals from the output port. By doing this, the module can be fully controlled and tested to see whether it is functionally work or not. In the ad-hoc modes, we can control and observe RSRN, CT, NOCRC, and 128 PE Adder Tree modules. The data output such as SAD, MVs, DVs in each cycle can be easily observed.

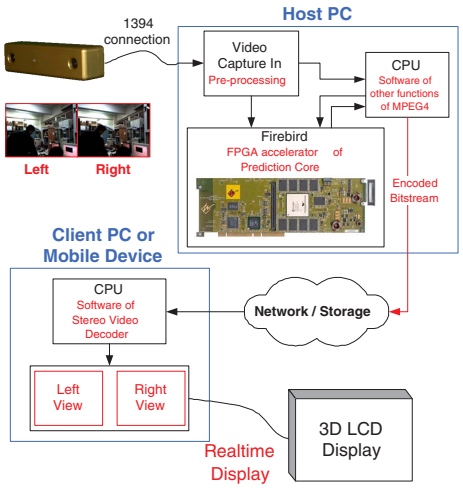

Fig. 8. Prototype of stereo video system.

TABLE II
COMPARISON BETWEEN PREVIOUS WORK AND PROPOSED ARCHITECTURE.

Architecture	Samsung [7]	Proposed
Gate count	140 k	137k
On-chip SRAM	2.5 kBytes	2.6 kBytes
System bandwidth	125 Mbytes/s	91.1 Mbytes/s
Normalized freq.	135MHz	81MHz
Processing capability	ME, D1, 30fps	2ME, 1DE, JBC, D1, 30fps

Because there are 12 embedded SRAM modules in our design, the testability of the embedded SRAM modules is quite important. By sharing the BIST controller for each memory module, the overhead of the design caused by BIST circuits can be minimized. Twelve SRAM modules are divided into three groups in "BIST mode". At "BIST mode", these groups of SRAM modules are enabled at different time by the controlling of the signal MemGroupSel. Hence, the chip will not be damaged during testing. Through this algorithm, the number of the test patterns can be minimized with 100% fault coverage.

By the scan test, a fault can almost always detectable if the fault coverage is high enough. There are 6610 registers in our design. In order to reduce the problem of global routing and testing time, seven scan chains are used. Each scan chain connects 945 or 946 the registers in a line. The use of multiple scan chains greatly reduce the time for testing and the task of Automatic Test Pattern Generation (ATPG). According to the report of Tetra-MAX, the number of faults are 466484 and the test coverage is 98.63%.

D. Comparison

So far there is no other architecture of prediction core of stereo video system. Our architecture can also handle motion estimation. Therefore, it is compared with the previous HSBMA architecture [6]. Table II shows the comparison. Although the hardware cost such as logic gate count and SRAM are similar, much less system bandwidth and normalized operating frequency are needed for our architecture. Besides, this chip has more functionalities such as DE, joint block compensation for the stereo video prediction. Because of its various functionalities, it can be easily integrated into mono or stereo video coding systems.

Table III shows the system bandwidth requirement of three ME/DE algorithms. 35.5% system bandwidth can be saved after NOCRS is applied. Although FSBMA requires less system bandwidth by regular data reuse scheme, for example, level-C data reuse scheme [7], Table IV shows the proposed HSBMA with NOCRS has much less on-chip-memory requirement and computational load. Only 11.5% on-chip SRAM and 1/30 amount of PEs are needed in this chip. At the same time, the proposed algorithm can still maintains good objective and subjective quality.

TABLE III
SYSTEM BANDWIDTH REQUIREMENT OF THREE BMA.

Data loaded from off-chip frame buffer	FSBMA	HSBMA without NOCRS	HSBMA with NOCRS
Current frame	9.9	9.9	9.9
SW for 4x4 BMP	0	3.4	3.4
SW for 8x8 BMP	0	46.4	29.5
SW for 16x16 BMP	55	46.4	31
Reconstructed frame	9.9	9.9	9.9
DS reconstructed frame	0	7.4	7.4
Total (Mega-Bytes / sec)	74.8	123.4	91.1

TABLE IV
COMPARISON BETWEEN PROPOSED ALGORITHM AND FSBMA.

Criterion	FSBMA	HSBMA with NOCRS
On-chip memory	180k bits	20.75k bits
Nomalized search points / MB	8192	100 - 234
Parallelism of PEs for real-time	>4096	128
Average quality drop	0	<0.2 dB

V. CONCLUSION

A prototype chip is currently under fabrication with 0.18um 1P6M technology by TSMC. This chip can achieve real-time requirement for D1 (720×480) stereo video system at 81MHz in general video case. Due to the irregular (cycle-variant) property of hierarchical ME/DE block matching algorithm, the maximum working frequency is designed at 100MHz to handle the worst case in real time. Note that the chip can perform two ME operations of the left and right channel and one DE operations of the right channel in 1/30 second. Compared with the hardware requirement for implementation of FSBMA, only 11.5% on-chip SRAM and 1/30 amount of PEs are needed in this chip. It is quite area-efficient. Algorithm analysis also shows that this chip maintains good video quality.

REFERENCES

[1] H.-C. Chang, L.-G. Chen, M.-Y. Hsu, and Y.-C. Chang, "Performance analysis and architecture evaluation of mpeg-4 video codec system," in *Proceedings of 2000 IEEE International Symposium on Circuits and Systems (ISCAS 2000)*, 2000.

[2] L.-F. Ding, S.-Y. Chien, Y.-W. Huang, Y.-L. Chang, and L.-G. Chen, "Stereo video coding system with hybrid coding based on joint prediction scheme," in *Proceedings of 2005 IEEE International Symposium on Circuits and Systems (ISCAS 2005)*, 2005.

[3] Y.-W. Huang, "Algorithm and architecture design for motion estimation, h.264/avc standard, and intelligent video signal processing," Ph.D. dissertation, Nation Taiwain University, Taipei, Dec. 2004.

[4] F. Isgrò, E. Trucco, P. Kauff, and O. Schreer, "Three-dimensional image processing in the future of immersive media," *IEEE Transations on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 388–303, Mar. 2003.

[5] *Requirements on multi-view video coding*, MPEG-4 Std. ISO/IEC JTC1/SC29/WG11 N6501, 2004.

[6] B.-C. Song and K.-W. Chun, "Multi-resolution block matching algorithm and its vlsi architecture for fast motion estimation in an mpeg-2 video encoder," *IEEE Transations on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 1119–1137, 2004.

[7] J. C. Tuan, T. S. Chang, and C. W. Jen, "On the data reuse and memory bandwidth analysis for full-search block-matching vlsi architecture," *IEEE Transations on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 61–72, Jan. 2002.