Joint Prediction Algorithm and Architecture for Stereo Video Hybrid Coding Systems

Li-Fu Ding, Shao-Yi Chien, and Liang-Gee Chen, *Fellow, IEEE*

Abstract—3-D video will be the most prominent video technology in the next generation. Among the 3-D video technologies, stereo video systems are considered to be realized first in the near future. Stereo video systems require double bandwidth and more than twice the computational complexity relative to mono-video systems. Thus, an efficient coding scheme is necessary for transmitting stereo video. In this paper, a new structure of prediction core in stereo video coding systems is proposed from the algorithm level to the hardware architecture level. The joint prediction algorithm (JPA), which combines three prediction schemes, is proposed for high coding efficiency and low computational complexity. It makes the system outperform MPEG-4 temporal scalability and simple profile by 2–3 dB in rate–distortion performance. Besides, JPA also utilizes the characteristics of stereo video and successfully reduces about 80% computational complexity. Then, a new hardware architecture of the prediction core based on JPA and a modified hierarchical search block-matching algorithm is proposed. With a special data flow, no bubble cycles exist during the block-matching process. The proposed architecture also adopts the near-overlapped candidates reuse scheme to save the heavy burden of data access. Besides, both on-chip memory requirement and off-chip memory bandwidth can be reduced by the proposed new scheduling. Compared with the hardware requirement for the implementation of full search block-matching algorithm, only 11.5% on-chip SRAM and 3.3% processing elements are needed with a tiny PSNR drop, making it area-efficient while maintaining high stereo video quality and processing capability.

Index Terms—Hardware architecture, joint prediction algorithm (JPA), stereo video coding, 3-D video.

I. INTRODUCTION

STEREO video can provide the users with a sense of depth perception by showing two frames to each eye simultaneously. It can give users vivid information about the scene structure. With the technology of three-dimensional (3-D) TV getting more and more mature [1], stereo and multiview video coding are drawing more and more attention. In recent years, the MPEG 3-D auido/video (3DAV) group has worked toward the standardization for multiview video coding [2], which also

L.-F. Ding and L.-G. Chen are with the DSP/IC Design Laboratory, Graduate Institute of Electronics Engineering and Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan, R.O.C. (e-mail: lifu@video.ee.ntu.edu.tw; lgchen@video.ee.ntu.edu.tw).

S.-Y. Chien is with the Media IC and System Laboratory, Graduate Institute of Electronics Engineering, Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan, R.O.C. (e-mail: sychien@cc.ee.ntu.edu.tw).

advances the stereoscopic video applications. Although stereo video is attractive, the amount of video data and the computational complexity are doubled. A good coding system is first required to solve the problem of huge data with limited bandwidth. In a mono-video coding system, motion estimation (ME) requires the most computational complexity [3]. By comparison, computational loading is even heavier in stereo video coding systems due to additional ME and disparity estimation (DE). Therefore, an efficient prediction scheme is required to overcome these problems. Moreover, it is preferred that the proposed video encoding system can be easily integrated by the existing video standards.

Some stereo video coding systems have been proposed. Stereo video coding can be supported by temporal scalability tools of existing standards, such as the MPEG-2 multiview profile (MVP) [4], where a view is encoded as the base layer, and the other one is encoded as the enhancement layer. This approach does not have good coding efficiency [5]. The I3-D [6] is a famous approach, in which the texture information is collected in a synthetic view, and the depth information is recorded in a disparity map. It has good coding efficiency and compatibility with MPEG-4 standard. However, additional operations for extracting disparity maps and synthesizing stereo views are required in the encoder and the decoder, respectively, which are not the building blocks of conventional video coding systems. A mesh-based and block-based hybrid approach is proposed by Wang *et al.* [7]. However, it needs additional preprocessing for segmentation to prevent matching failure around the object boundary [8]. In addition, the computational complexity is very high.

DE is the core of stereo video coding systems. The DE algorithms of the previous systems can be roughly classified into three categories: pixel-based, mesh-based, and block-based. Pixel-based algorithms, such as dynamic programming [9], [10], and mesh-based algorithms [7] can generate more precise disparity or depth maps than block-based algorithms do. Because of the feature of noncrossing order of vectors, the mesh-based algorithms have good view-synthesis ability. The main disadvantage of pixel-based and mesh-based algorithms is that they cannot be compatible with the existing video coding standards [11]. An ultrahigh computational complexity is usually required for this approach. In addition, a segmentation step is usually needed for more accurate estimation in the mesh-based algorithms. On the other hand, the main advantage of block-based algorithms is that they have much better compatibility with the existing standards.

In this paper, a new stereo video coding system with joint prediction algorithm (JPA) and its architecture are proposed. For
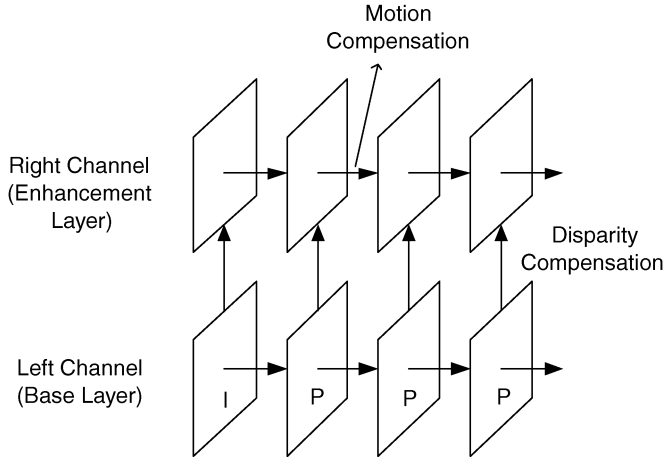
Fig. 1. Base-layer/enhancement-layer scheme of the proposed system. The base layer is encoded with MPEG-4 SP encoder.

better compatibility, the system is based on the hybrid coding scheme. Block-based algorithms are adopted for ME and DE, which are combined as the prediction core with JPA. To improve the coding efficiency and reduce the computational complexity, JPA is composed of three coding tools, joint block compensation, motion vector-disparity vector (MV–DV) prediction, and mode pre-decision. To meet the real-time constraint, the hardware architecture is designed based on the modified hierarchial search block-matching algorithm (HSBMA) and the joint block compensation scheme. A new scheduling and bandwidth-reduction scheme is proposed for improving the hardware utilization.

The remainder of this paper is organized as follows. Section II describes the proposed stereo video hybrid coding system with proposed JPA. Next, the analysis and algorithms of BMA are presented in Section III. Section IV describes the proposed prediction core architecture. Finally, Section V concludes this paper.

II. PROPOSED JPA

For the purpose of compatibility, the coding system adopts a base-layer/enhancement-layer scheme, as shown in Fig. 1. 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 Simple Profile (SP) encoder [12]. The block diagram of the proposed stereo video encoder is shown in Fig. 2. The main differences between the left channel and the right channel are disparity estimation, joint block generation, and other functional blocks such as mode predecision and MV-DV prediction, which will be introduced later. Note that reference frames from the left and right channels are both reconstructed. After encoding, the left compressed data, M and L, and the right compressed data of a small amount, N and R, are transmitted.

In the stereo video coding system, the prediction is the most important part. It is not only computationally intensive but also critical for the coding efficiency. Motion/disparity estimation/compensation are the key operations in the prediction core. Fig. 3 illustrates the prediction directions and the search windows (SWs) of two reference frames for ME and DE, respectively. For the current block in the right channel at

$t = 1$, in addition to ME with $SW_{r,\mathrm{ME}}$, there exists another way to find good prediction, that is, DE with $SW_{l,\mathrm{DE}}$. ME can remove the temporal redundancy. On the other hand, DE can remove the interview redundancy [13]. Therefore, the frames in the right channel have more than one choice to find their best matching blocks.

In order to improve the coding efficiency and reduce the computational complexity in the right channel, JPA is proposed and based on three prediction schemes. First, a block is compensated not only by the block of left or right reference frames but also by the combination of them according to different types of content of it. Second, the properties of stereo video are considered for the accurate motion vector prediction to reduce the computational complexity of ME. Third, the computational complexity of DE is reduced by the proposed mode predecision scheme. Based on these three schemes, in this section, the details of JPA are shown in the subsections.

A. Joint Block Compensation

1) Joint Block: In ME and DE steps of the right channel, the current block has two reference frames, as shown in Fig. 3. The gray region is the SW of a reference frame. Note that the SW of the left reference frame for DE is not a square because cameras are assumed to be parallel-structured, so the candidate blocks are only on the belt of the region [14]. There are mainly two types of compensated blocks in the right channel. They are the motion-compensated (MC) block and the disparity-compensated (DC) blocks which are illustrated as $B'_{r,\mathrm{ME}}$ and $B'_{l,\mathrm{DE}}$ in Fig. 3, respectively. The MC block often occurs in the background because of its zero or slow motion. Since DE is usually not able to predict well in the case of occlusions between the left and right frames, these blocks will also be compensated by this type of blocks. On the other hand, DC block often occurs in the moving objects because of their deformation during motion. In this case, DC blocks usually have better prediction capability.

However, when a frame is divided into several blocks, there is a high probability that a block may contain more than one type of video objects. For example, a block may contain both moving objects and background in it. In this case, neither the MC block nor the DC block can predict well. As a result, a new type of compensated blocks, the joint block, is proposed. Fig. 4 shows the illustration of the proposed joint block generation and compensation. After ME and DE, the best matching blocks in the two SWs are derived. Then, the joint block generation step starts. They are the linear combination of the MC and the DC blocks. By the specified weighting parameters, several different joint block candidates are generated.

According to the criterion of sum of absolute difference (SAD), the best type of compensated block is selected. For each macroblock (MB) of the current frame, the distortion of the three types of blocks are computed by

$$D_{\mathrm{motion}} = \min\left\{ \sum_{t\in B_r, t'\in B'_{r,\mathrm{ME}}} |I_r(t) - I_{r-1}(t')| \big|_{B'_{r,\mathrm{ME}}\in SW_{r,\mathrm{ME}}(B_r)} \right\}$$
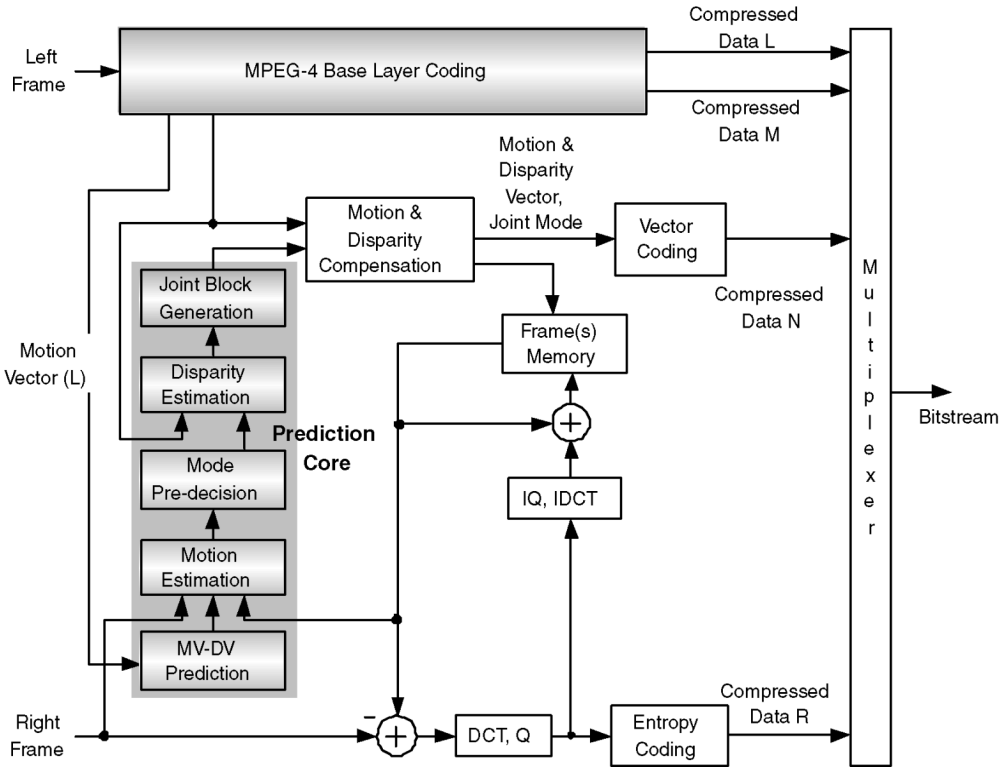
(1)

Fig. 2.   Block diagram of the proposed stereo video encoder. The system is based on the hybrid coding scheme.
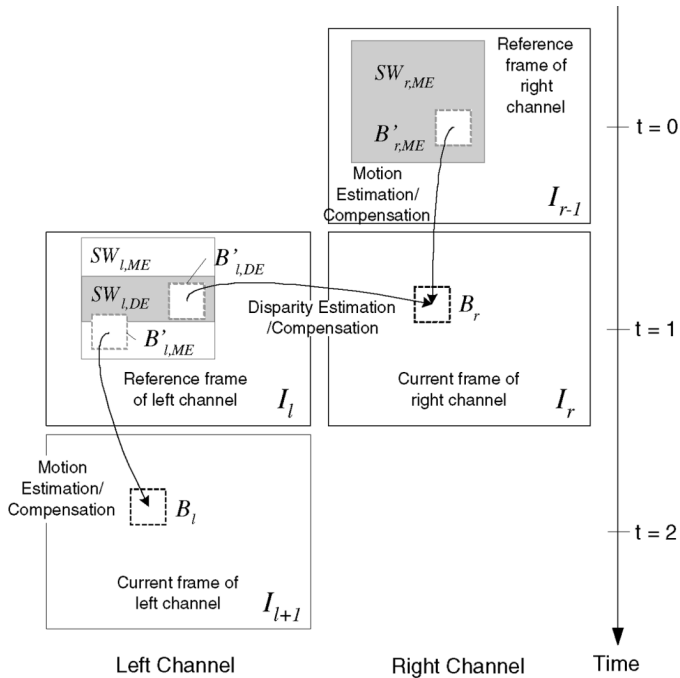

Fig. 3.   Illustration of the prediction directions and the SWs of two reference frames for ME and DE, respectively. The gray regions are the SWs for B_r.
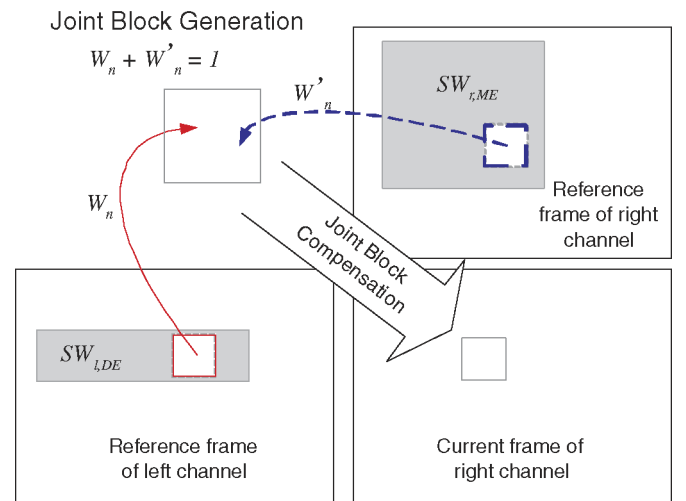

Fig. 4.   Proposed joint block generation and compensation. The joint block is the weighted sum of the MC and the DC blocks.

$$D_{\text{disparity}} =$$
$$\min\left\{ \sum_{t\in B_r, t'\in B'_{l,\text{DE}}} |I_r(t) - I_l(t')|\big\|_{B'_{l,\text{DE}}\in \text{SW}_{l,\text{DE}}(B_r)} \right\} \quad (2)$$

$$D_{j_n} =$$
$$\min\left\{ \sum_{t\in B_r, t'\in B'_{l,\text{DE}}, t''\in B'_{r,\text{ME}}} |I_r(t) - [W_n \cdot I_l(t') \right.$$
$$\left. + W'_n \cdot I_{r-1}(t'')]|\big\|_{W_n+W'_n=1} \right\} \quad (3)$$

where D_{motion} and $D_{\text{disparity}}$ are the minimum SADs of MC and DC blocks, respectively. B_r is the current block in the

TABLE I
STATISTICS OF THE PROBABILITY OF VARIOUS JOINT BLOCK PATTERNS

W_n	Percentage	Pattern combination	Percentage
0.000	23.1%	Pattern1	2.1%
0.125	2.8%	Pattern2	1.5%
0.250	2.4%	Pattern3	0.1%
0.375	14.5%	Pattern4	0.3%
0.500	18.8%	Pattern5	1.6%
0.625	11.5%	Pattern6	0.2%
0.750	8.5%	Pattern7	0.7%
0.875	5.6%	Pattern8	1.5%
1.000	4.8%		
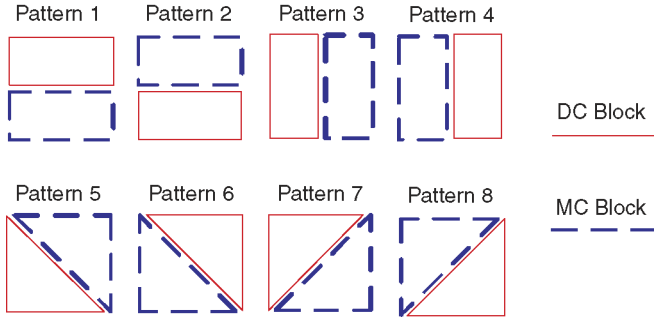
Fig. 5. Selection of joint block patterns. Eight patterns are chosen for experimental analysis.

Fig. 6. Rate–distortion performance of various numbers of joint block modes. Curve I contains both weighting modes and pattern combination, whereas curves II and III contain only weighting modes.

right channel. $B'_{r,\mathrm{ME}}$ is the reference block in the right channel. $B'_{l,\mathrm{DE}}$ is the reference block in the left channel. $SW_{r,\mathrm{ME}}(B_r)$ and $\mathrm{SW}_{l,\mathrm{DE}}(B_r)$ are the SWs in the right and left reference frames of the block B_r, respectively. The proposed joint block is then generated as the weighted sum of the two blocks. W_n and W'_n are complementary weighting functions that describe the weighting parameters. In (3), D_{j_n} is derived. Finally, the mode decision is described as

$$\mathrm{Mode} = \arg \min_{\mathrm{mode}} \{D_{\mathrm{motion}}, D_{\mathrm{disparity}}, D_{j_1}, \ldots, D_{j_n}\}. \quad (4)$$

In our stereo video encoder, the modes are compressed by the arithmetic coding process. Note that the proposed algorithm is applied on the luminance domain. When performing the joint block compensation, the chrominance data are compensated by the same set of vectors, as in the existing hybrid coding standards [12], [15].

2) Selection of Joint Block Weighting Parameters and Patterns: There are infinite weighting parameters or patterns that can be used for joint block generation. In the previous works, the MB is only compensated by the MC and the DC blocks with fixed weighting parameters [16], [17]. In our experiment, seventeen modes are considered. As shown in Table I, nine modes of joint blocks are generated by the weighted sum of MC and DC blocks. Five stereo video sequences are tested and averaged. For example, The value of W_n, 0.625, means the pixels in the joint block are the sum of the pixels of the DC block multiplied by 0.625 and the pixels of the MC block multiplied by 0.375. Note that the zero W_n makes the joint block mode as the same as using ME. $W_n = 1$ means the joint block use DE only for prediction. On the other hand, there are also eight kinds of joint blocks generated by the combination of complementary shapes,
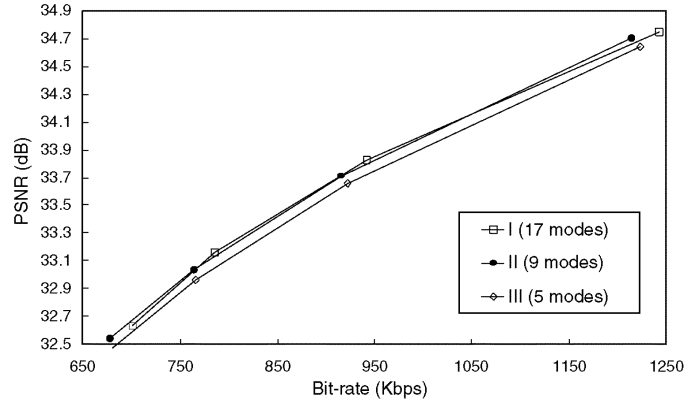
as shown in Fig. 5. In Table I, it shows that, after mode decision, 92% of blocks choose the joint block modes formed by the weighting parameters. Only 8% blocks choose the combination of complementary shapes to form the joint blocks. The reason is that the edges of patterns used are very sharp and straight, while the shape of an object is not so in general cases. Therefore, the distortion of these kinds of joint blocks is often large.

Fig. 6 shows the rate–distortion performance of various number of joint block modes. Five, nine, and seventeen modes are taken into consideration. Curve I contains nine weighting modes and eight pattern combination modes listed in Table I. Curves II and III contain only weighting modes without pattern combination modes. The performance of Curves I and II are similar. The reason is that, although joint block compensation with more modes has better prediction ability and significantly reduces bit rate for encoding residue, it has the penalty for encoding mode information. As a result, we decide to choose the weighting parameters to generate all the joint blocks in the stereo video coding system.

B. MV-DV Prediction

In general stereo video coding systems, ME and DE are the key operations. However, compared with mono-video systems, additional ME and DE of the right channel greatly increase the computational burden. Therefore, the MV–DV prediction scheme is proposed.

1) Correlation Between DVs and MVs: The correlation is shown in Fig. 7. It can be described as [18]

$$\mathrm{DV}_{k-1} + \mathrm{MV}_R = \mathrm{MV}_L + \mathrm{DV}_k. \quad (5)$$

If parallel-setup cameras are unchanged, DVs of an object in different time slots are almost the same. Therefore

$$\mathrm{DV}_{k-1} \approx \mathrm{DV}_k \Rightarrow \mathrm{MV}_R \approx \mathrm{MV}_L. \quad (6)$$

According to the correlation, MV_L is used as the predictor of MV_R. Because of the parallel-setup camera structure, there is a global horizontal displacement between the left and right channels, which is called the global disparity. In order to find the
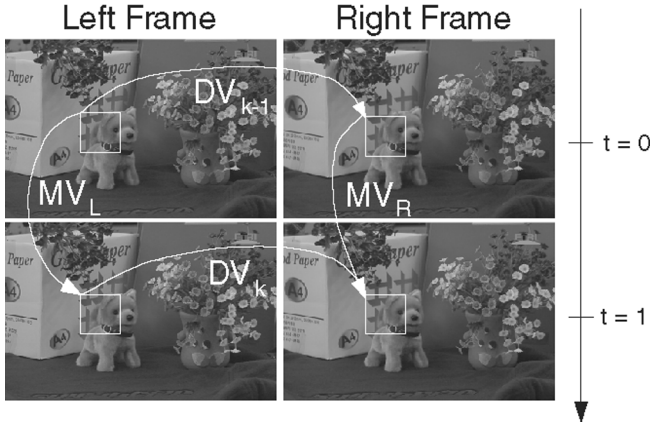
Fig. 7.   Relation between DVs and MVs. If parallel-setup cameras are unchanged, DVs of an object in different time slots are almost the same.
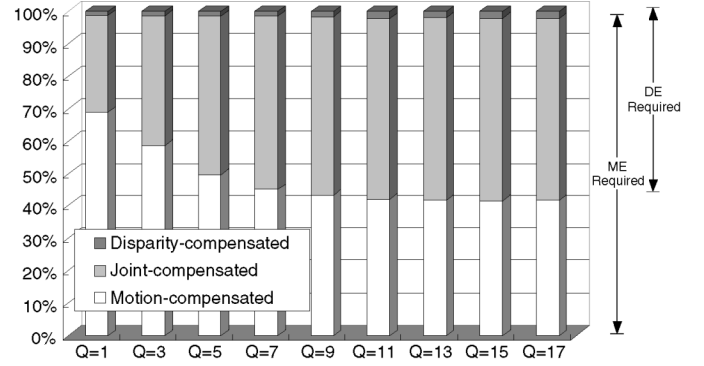

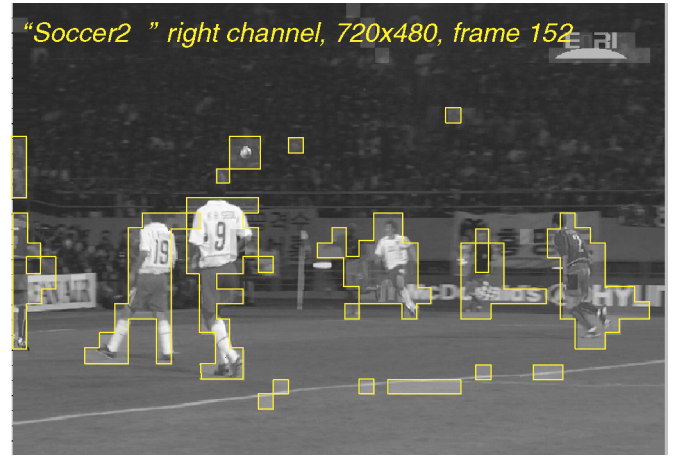Fig. 8.   Percentage of block prediction types of sequence "Race2."

Fig. 9.   Subjective view of statistics of compensated block types. The highlighted blocks are DV-predicted. It shows that moving objects, such as soccer players or the ball, are usually DV-predicted.

predictors, the global disparity should be derived first because of the relation between MVs and DVs introduced above. Here, we use a simple way to find the global disparity rather than the complex global motion estimation (GME) scheme [19], that is, DE is performed with the statistic process in the first P-frame in the right channel. Since the background usually occupies the largest area, the disparity that occurs most frequently is set as the initial global disparity. The global disparity is dynamically updated due to unexpected conditions such as scene change and moving background. The details will be introduced in the next subsection. The background detection is determined by

$$F_{\text{diff}}(N) = \sum_{t \in B_{N,r}, t' \in B_{N,r-1}} |I_r(t) - I_{r-1}(t')| \quad (7)$$

$$\text{Background}(N) = \begin{cases} \text{true}, & \text{if } MV_N(x,y) = (0,0) \text{ or} \\ & \quad F_{\text{diff}}(N) < \text{Threshold} \\ \text{false}, & \text{otherwise} \end{cases} \quad (8)$$

where $B_{N,r}$ and $B_{N,r-1}$ are the Nth blocks in I_r and I_{r-1}, respectively. $\text{Background}(N)$ is the state of the Nth block in the right frame. $MV_N(x,y)$ is the MV of the Nth block. The Threshold of $F_{\text{diff}}(N)$ shown here is empirically chosen for the additional criterion. According to the background information, the disparity vectors of these background blocks are counted in the statistical analysis. Then, the global disparity vector GD is derived as

$$\text{GD} = \arg \max_{\text{DV}} \{\text{Num}(\text{DV})\} \quad (9)$$

where $\text{Num}(\text{DV})$ is the histogram of DV. For the first P-frame in the left channel, a background detection scheme is used to find GD. Before ME in the right channel, the corresponding block in the left frame of the current block in the right frame can be found by use of GD. Then, the MV of the corresponding block is used as the predictor of the current block. MV_R is derived within a small SW to reduce computation. However, the DVs of background are usually smaller than those of foreground. If the SAD is larger than a empirically chosen threshold, the block is viewed as a foreground block. Its search range extends

adaptively to find a better MV. Next, a more precise GD is fed back to the system. To avoid error propagation, GD can be updated after every M frames, where M is a flexible parameter. In the proposed stereo video coding system, the MVs in the left channel and the DVs between two channels are coded separately by checking the vector differences in the self-defined lookup tables (LUTs). The MVs in the right channel are then predicted by these two vectors in the above-mentioned way.

C. Mode Predecision

In addition to the reduction of ME complexity, a new method is proposed for computational complexity reduction of DE. In our experiments shown in Fig. 8, 40%–70% blocks in the right frame are motion-compensated, 25%–60% blocks are joint-compensated, whereas only about 5% blocks are disparity-compensated. From the above analysis, over 95% blocks must perform ME, while only 30%–60% blocks must perform DE. Thus, an unnecessary DE could be skipped to reduce computational complexity. From our analysis, the MV-predicted blocks often have zero motion, such as blocks in the background, or have slow motion caused by moving cameras. An example is shown in Fig. 9. The highlighted blocks are DV-predicted. It shows that moving objects, such as soccer players or the ball, are usually DV-predicted while the other blocks composed of the background are usually MV-predicted.

TABLE II
HIT RATE AND PSNR DROP OF MODE PREDECISION

Sequences	Hit rate	PSNR drop (dB)
Soccer2	89.72%	0.025
Puppy	94.30%	0.018
Golf	90.15%	0.023
Flamenco	85.36%	0.046
Race2	77.32%	0.096

By use of these properties, mode predecision scheme is applied after the minimum SAD of ME $\mathrm{SAD_{ME}}$

$$\mathrm{Skip} = \begin{cases} \text{true,} & \text{if } F_{\mathrm{diff}} < \mathrm{Threshold}_1 \text{ and} \\ & \quad \mathrm{SAD_{ME}} < \mathrm{Threshold}_2 \\ \text{false,} & \text{otherwise.} \end{cases} \quad (10)$$

If Skip is true, the block is usually MV-predicted. Then DE is skipped, and the computational complexity is reduced. The $\mathrm{Threshold}_1$ and $\mathrm{Threshold}_2$ shown above are similar with the Threshold in (8). They are sequence dependent and closely related with rate–distortion condition. For example, in our simulation, we set 600 as $\mathrm{Threshold}_1$ and 1200 as $\mathrm{Threshold}_2$ in the sequence "Soccer2." Table II shows the hit rate and PSNR drop of five test sequences. The average hit rate is about 87% and the PSNR drop is approximately 0.02 dB.

D. Experimental Analysis and Comparison

1) Improvement of Coding Efficiency: The proposed system is compared with MPEG-4 SP [20] and the temporal scalability profile (TSP) encoder [21]. Rate–distortion performance of only right channels (enhancement layer) are compared because the left channels are all encoded by MPEG-4 SP. The performance of the left channels are similar to the right channels also encoded by MPEG-4 SP because of the similar video contents of two channels. Stereo video sequence "Race2" (320×240, 30 fps) and "Soccer2" (720×480, 30 fps) are taken as test sequences.

Fig. 10 shows the comparison between the proposed algorithm, MPEG-4 TSP, and MPEG-4 SP. The proposed joint prediction scheme is 3 and 2 dB better than MPEG-4 SP and TSP, respectively. It shows that the joint block compensation scheme successfully reduces more redundancy. Fig. 11 shows the performance of different coding tools. Without the joint prediction scheme (curve 1), the PSNR degradation is serious, as in the MPEG-4 SP. After the DE operation is turned on (curve 2), the coding efficiency is improved, just like the effect of multiple reference frames. When joint block compensation scheme is applied (curve 3), there is a 3-dB gain on coding efficiency. Besides, after applying MV-DV prediction scheme and mode predecision scheme (curve 4), not only the video quality is maintained but also most of the computational complexity can be reduced, which will be introduced in the next subsection.

2) Reduction of Computational Complexity: Table III shows the reduction of search points. Note that every search point contains 256 substraction and addition operations. In our experiments, the search ranges of ME and DE are $[\pm 32, \pm 16]$ and $[\pm 32, \pm 8]$ for 320×240 sequences, and are $[\pm 64, \pm 32]$ and $[\pm 64, \pm 16]$ for 720×480 sequences, respectively. The proposed algorithm reduces about 80% computational complexity with negligible quality degradation. From Fig. 12, we can see
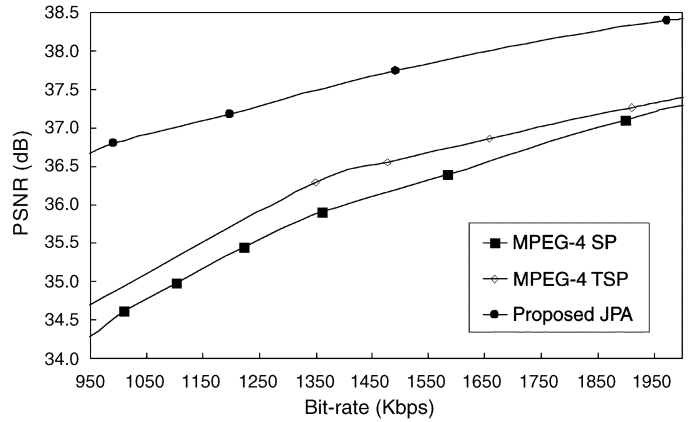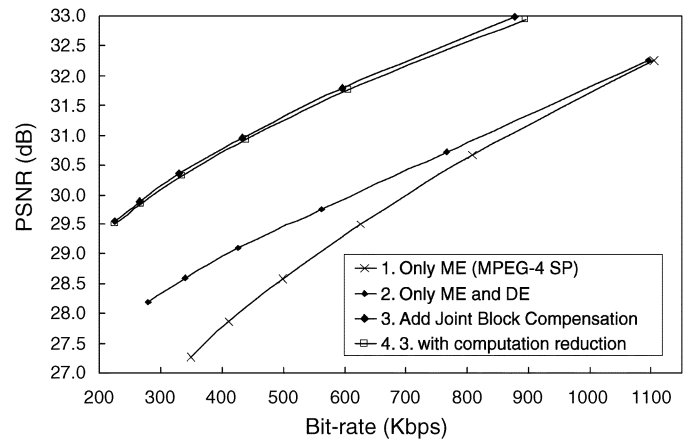

Fig. 10. Rate–distortion curve of sequence "Soccer2."

Fig. 11. Rate–distortion curve of sequence "Race2."

that if the search range is reduced from ± 16 to ± 2, the PSNR degradation is only about 0.1 dB, while both the computational complexity of ME and DE are greatly reduced.

III. BLOCK MATCHING ALGORITHM FOR ME/DE

ME is a key unit in hybrid video coding systems. In the proposed stereo video coding system, additional ME and DE are required when encoding frames in the right channel. It increases the design challenges in large on-chip memory, memory bandwidth, and computational complexity. Due to the high hardware cost of FSBMA architecture, it is not suitable for the hardware architecture design of the prediction core. There are several kinds of fast motion estimation algorithms, such as three-step search [22], four-step search [23], diamond search [24], hexagon-based search [25], and hierarchical search [26]. Among those fast algorithms, hierarchical search block-matching algorithm (HSBMA) can reduce not only the computational complexity but also the requirement of on-chip memory. Therefore, HSBMA is adopted with further improvement.

A. Modified Hierarchical ME/DE Block Matching Algorithm

Compared with FSBMA, conventional HSBMA suffers from the problem of quality degradation [27]. Then, error propagation will make the MVs not the best. To prevent this situation, the

TABLE III
SEARCH POINTS REDUCTION

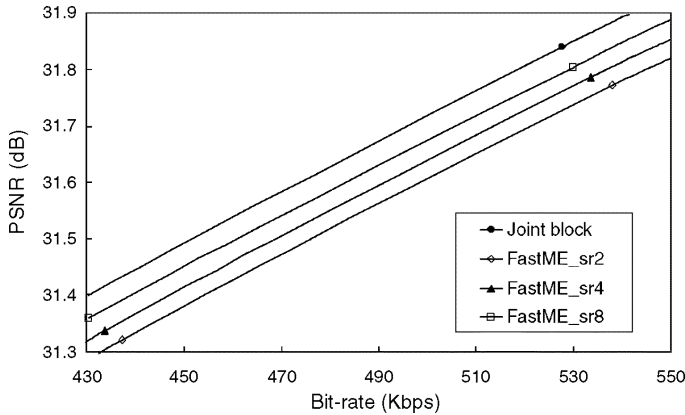| Sequences	Race2	Flamenco	Golf	Soccer2	Puppy
None	2560	2560	2560	10240	10240
MV-DV prediction	1088	1088	1088	4352	4352
Mode pre-decision	1574	1160	1850	4204	3004
Combination of 2 schemes	609	400	706	1787	1277
Search point reduction ratio	76.20%	84.37%	72.41%	82.54%	87.52%

Fig. 12.   Rate–distortion curve of fast algorithm with various search ranges of sequence "Race2."

multiple-candidates scheme is adopted, that is, several motion vector candidates are chosen after performing the coarser level block-matching process (BMP). Take a three-level HSBMA, for example: level 2 is defined as the coarsest level, and level 0 is defined as the finest level. Three motion vectors with smaller SADs are first chosen in level-2 BMP. Then, three level-1 BMPs begins. After that, only three rather than nine better MVs are chosen in these three level-1 BMPs. Then, three level-0 BMPs starts. Finally, the best MV is chosen in these three SWs of level-0 BMPs.

Usually, more candidates chosen in the coarser levels and larger SWs in the refinement levels can provide better video quality. However, these are accompanied with some side effects. The computational complexity and the system memory bandwidth increase rapidly with more candidates chosen and larger SWs. To find the suitable combination of the number of candidates chosen and the search range, we did an experimental analysis focused on the rate–distortion and system bandwidth. The D1 (720×480) size sequences are tested, and the search ranges are $[-64, +63]$ in the horizontal direction and $[-32, +31]$ in the vertical direction. Fig. 13 shows the rate–distortion performance of various HSBMAs with different levels, different number of candidates, and different refinement ranges. For example, L_3_5_2 means three levels, five candidates refinement, and 5×5 refinement range are selected. Four video sequences are tested. Compared with FSBMA, we observed that the video quality is acceptable. However, the bandwidth requirement of data access is very much different between different cases. Table IV shows the bandwidth requirement for various specifications. The reference frames are in the off-chip frame buffer. The bottom SW represents the SW of level 2 that is required to be loaded from the off-chip frame buffer. On the other

hand, the top SW represents the SW of level 0. Take the algorithm "HS2" for example, a current block contains 256 pixels, thus it spends 256 bytes of memory bandwidth per block. The memory bandwidth of loading a bottom SW is reduced by utilizing level-C data reuse scheme [28] whatever the adopted algorithm is. However, when loading the mid or top SW, the regular data reuse scheme cannot be applied. In this case, $(16 + 32) \times (16 + 32) \times 5 = 11520$ B of memory bandwidth is required for the Top SW of every block. It shows that the bandwidth requirement of HSBMA is much higher than that of FSBMA. The main reason is that finer level cannot be applied with effective data reuse scheme, such as level-C data reuse scheme of FSBMA. The situation is more serious when the number of candidates chosen is five. From this table, HS7 is a suitable choice.

B. Near-Overlapped Candidates Reuse Scheme (NOCRS)

Still, the problem of bandwidth requirement is not fully solved. SW data cannot be reused effectively in the refinement levels, level-1 and level-0 BMPs. It will cause serious overhead on bus bandwidth. However, the experimental analysis shows that MVs of the best three candidates are usually very close. It means that the SWs of them in the next refinement level are partially overlapped. Therefore, we propose the near-overlapped candidates reuse scheme (NOCRS) to reduce the bandwidth requirement. After finding the best three candidates during level-2 and level-1 BMPs, three MVs will be checked by calculating their differences mutually. If the differences are smaller than a threshold, the overlapping condition is satisfied. The threshold is statistically analyzed according to the designer's specifications. We set four and two as the thresholds for motion vector differences in the x- and y-directions, respectively, in the simulation. The union of two or even three SWs is loaded only once from the off-chip frame buffer, as shown in Fig. 14.

In summary, Fig. 15 shows the flow chart of HSBMA with NOCRS. NOCRS not only reduces off-chip memory bandwidth successfully but also avoids unnecessary computation on duplicated search candidates in two separate SWs. Table V shows the system bandwidth requirement of three ME/DE algorithms. After NOCRS is applied, 35.5% system bandwidth can be saved. Although FSBMA still requires less system bandwidth by regular data reuse scheme, for example, level-C data reuse scheme, the proposed HSBMA with NOCRS has much less on-chip memory requirement and computational complexity.

IV. PREDICTION CORE ARCHITECTURE

The overall architecture of the prediction core is shown in Fig. 16. There are nine main units: control unit, reference shift register network (RSRN), current register set (CRS), current
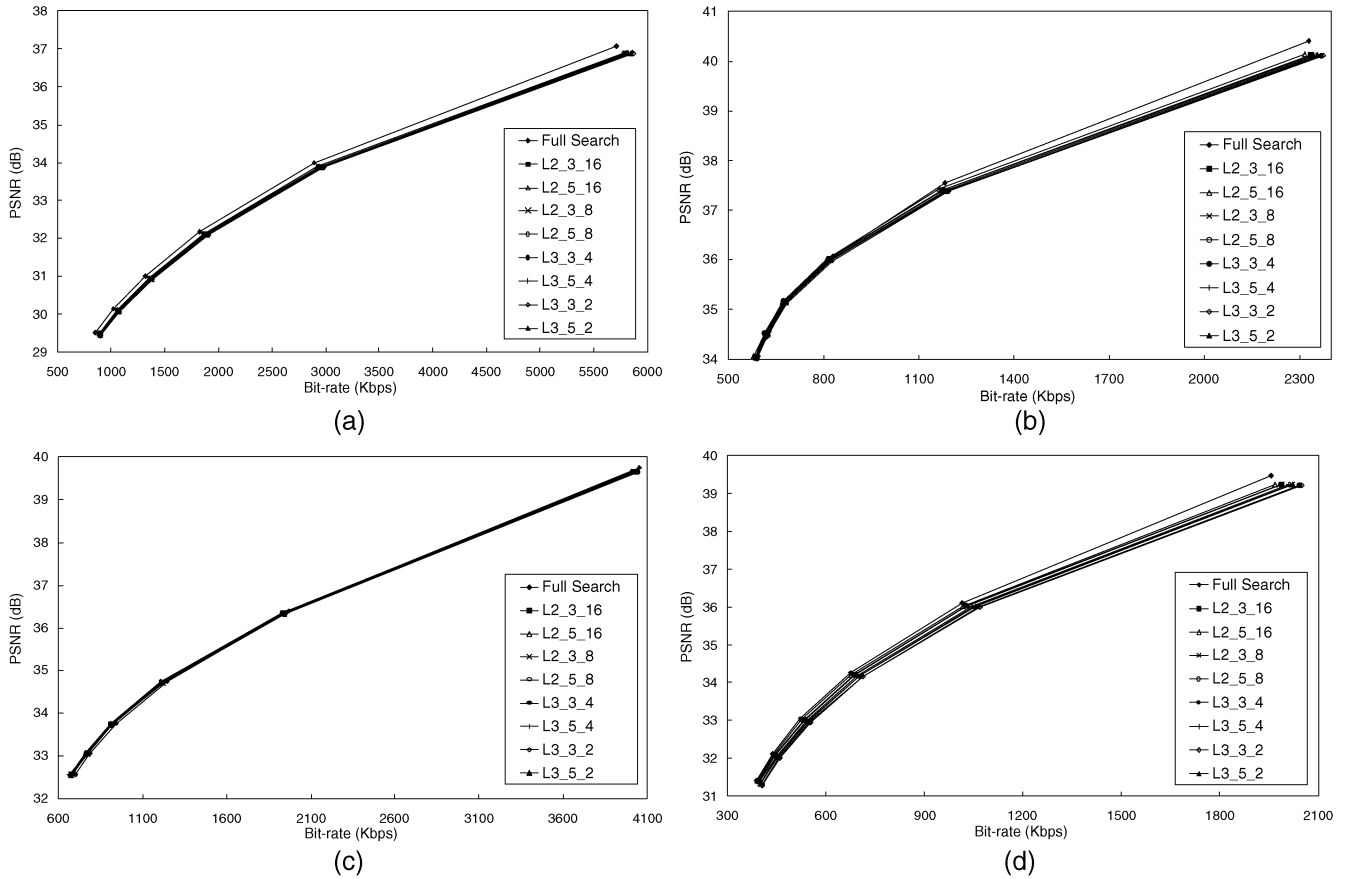
Fig. 13. Rate–distortion performance of four test sequences (a) "Wendy," (b) "Angel," (c) "Toshiba," and (d) "Taxi." Various levels, number of candidates, and refinement ranges are tested. For example, L_3_5_2 means three levels, five candidates refinement, and 4 × 4 refinement range are selected.

TABLE IV
OFF-CHIP MEMORY BANDWIDTH REQUIREMENT OF VARIOUS HSBMA

Algorithm	FS	HS1[a]	HS2[b]	HS3[c]	HS4[d]	HS5[e]	HS6[f]	HS7[g]	HS8[h]
Current block (Byte/block)	256	256	256	256	256	256	256	256	256
Bottom SW (Byte/block row)	6400	4000	4000	4000	4000	4000	4000	4000	4000
Mid SW (Byte/block)	0	0	0	0	0	768	1280	1200	2000
Top SW (Byte/block)	0	6912	11520	3072	5120	1728	2880	1200	2000
Total bandwidth (MByte/sec)	64.8	252.5	458.3	132	211	109.7	174	106	167.8

[a] 2 levels, 3 candidates refinement, [-16, +16] refine search range
[b] 2 levels, 5 candidates refinement, [-16, +16] refinement range
[c] 2 levels, 3 candidates refinement, [-8, +8] refinement range
[d] 2 levels, 5 candidates refinement, [-8, +8] refinement range
[e] 3 levels, 3 candidates refinement, [-4, +4] refinement range
[f] 3 levels, 5 candidates refinement, [-4, +4] refinement range
[g] 3 levels, 3 candidates refinement, [-2, +2] refinement range
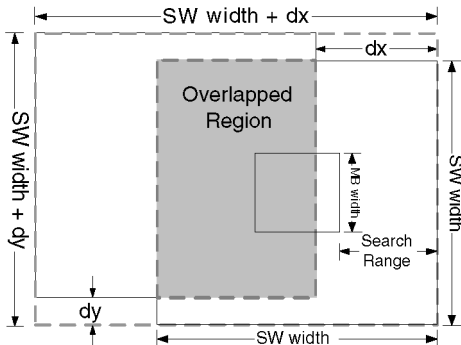[h] 3 levels, 5 candidates refinement, [-2, +2] refinement range

Fig. 14. Union of overlapped SWs. The SWs in the next refinement level are partially overlapped.

MUX network (CMN), 128-PE adder tree, comparison tree (CT), NOCR checker (NOCRC), interpolation unit (IU), and joint block generator (JBG). The CMN outputs three kinds of current blocks for three-level BMPs. The RSRN is composed of a reconfigurable shift register array. After data loading of SW is finished, RSRN starts to fetch data from the on-chip memory. Meanwhile, 128-PE adder tree generates SADs. Then, CT compares these SADs in one cycle. The best three candidate MVs are chosen for refinement. NOCRC checks the degree of overlapping and outputs the postprocessed MVs to the address generator (AG) in the control unit, which decides when to begin the next level BMP. IU generates subpixels in the half pixel refinement process. JBG generates joint blocks
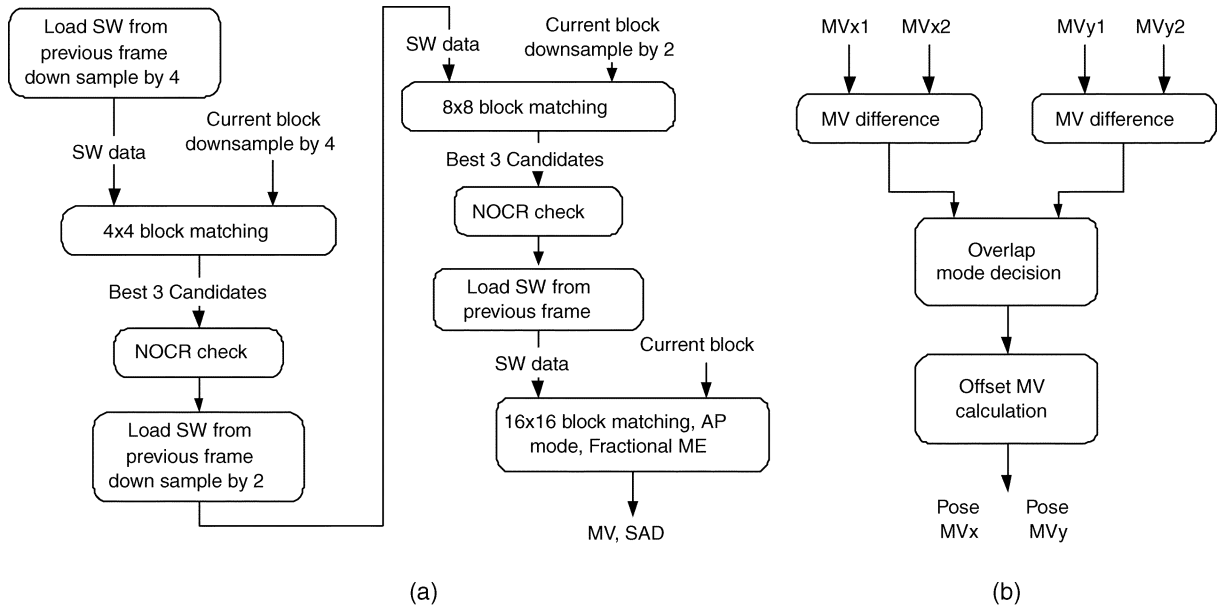
Fig. 15. Data flow of the proposed ME/DE algorithm. (a) Flow of proposed HSBMA with NOCRS. (b) Flow of NOCRS.

TABLE V
SYSTEM BANDWIDTH REQUIREMENT OF THREE BMAS

Data loaded from off-chip frame buffer	FSBMA	HSBMA without NOCRS	HSBMA with NOCRS
Current frame	9.9	9.9	9.9
SW for 4×4 BMP	0	3.4	3.4
SW for 8×8 BMP	0	46.4	29.5
SW for 16×16 BMP	55	46.4	31
Reconstruct frame	9.9	9.9	9.9
DS Reconstruct frame	0	7.4	7.4
Total bandwidth (MByte/sec)	74.8	123.4	91.1

for mode decision for improving the coding efficiency of the stereo video. The detailed architecture will be described in the following. Furthermore, the data reuse scheme and memory organization are also shown. In addition, a new scheduling is proposed to reduce the demand of on-chip memory and off-chip memory bandwidth [29].

A. RSRN

To achieve the design goal of hierarchical BMP with only one hardware resource, RSRN is composed of a reconfigurable shift register array, which consists of 128 8-bit registers, as shown in Fig. 17. It has high reconfigurability and can reconfigure to shift downward, leftward, and rightward. After the SW are loaded from off-chip to the on-chip memory, one column of SW pixels are fetched to RSRN every cycle. There are no bubble cycles when the search position is changed in the vertical position. An example of the detailed data flow is shown in Fig. 18, which is the data flow of level-1 BMP with the maximum search range [−6, +6] in our design, and the SW is 20 words × 20 bits. When BMP starts, RSRN fetches one column of SW pixels at each cycle. At cycle 7, all the candidate block data of search positions (−6, −6) and (−6, −5) are stored in RSRN. Thus SAD0 and SAD1 are generated at cycle 7. Then, two SADs are generated in each cycle. At cycle 12, two additional pixels must be prefetched in additional registers to avoid bubble cycles during the reconfiguration step. At cycle 19, these additional

sixteen pixels are ready to input to RSRN. Then the RSRN shifts downward, and two SADs of search position (6, −4) and (6, −3) are generated without any bubble cycles. At cycle 21, RSRN changes the connection configuration again and shifts leftward. In this way, all the bubble cycles can be avoided besides the initial cycles, so the utilization is near 100%. Level-2 and level-0 BMP have flows similar to level-1 BMP.

B. 128-PE Adder Tree

Fig. 19 illustrates 128-PE adder tree. The pixels of the current block and the reference candidates are fetched into 128-PE adder tree every cycle. Except for several cycles in the beginning for data preparing, SADs of eight candidate blocks in level-2 BMP, two candidate blocks in level-1 BMP, or half candidate blocks in level-0 BMP can be derived. Then the CT compares the SADs and chooses the proper candidates for the next level BMP.

C. NOCRC

The architecture of NOCRC is shown in Fig. 20. The best three MVs chosen by CT are inputted to the NOCRC after level-2 and level-1 BMPs. The MV differences are calculated mutually, and then the overapping condition is decided. For example, if three outputs of threshold units are all logic 1, it means SW of next level should be loaded only once rather than three times. It can effectively reduce over 35% unnecessary data access from off-chip. Furthermore, it also reduces unnecessary computation and saves processing cycles.

D. JBG

When the ME of the right channel is finished, the best candidate block must be stored for the joint block generation step. The best candidate block is loaded into on-chip RAM_MC, as shown in Fig 16. After DE of the right channel is finished, mode decision for the joint block starts. Fig. 21 shows one of the 16 JPG units in JBG. Only adders are used to generate weighted
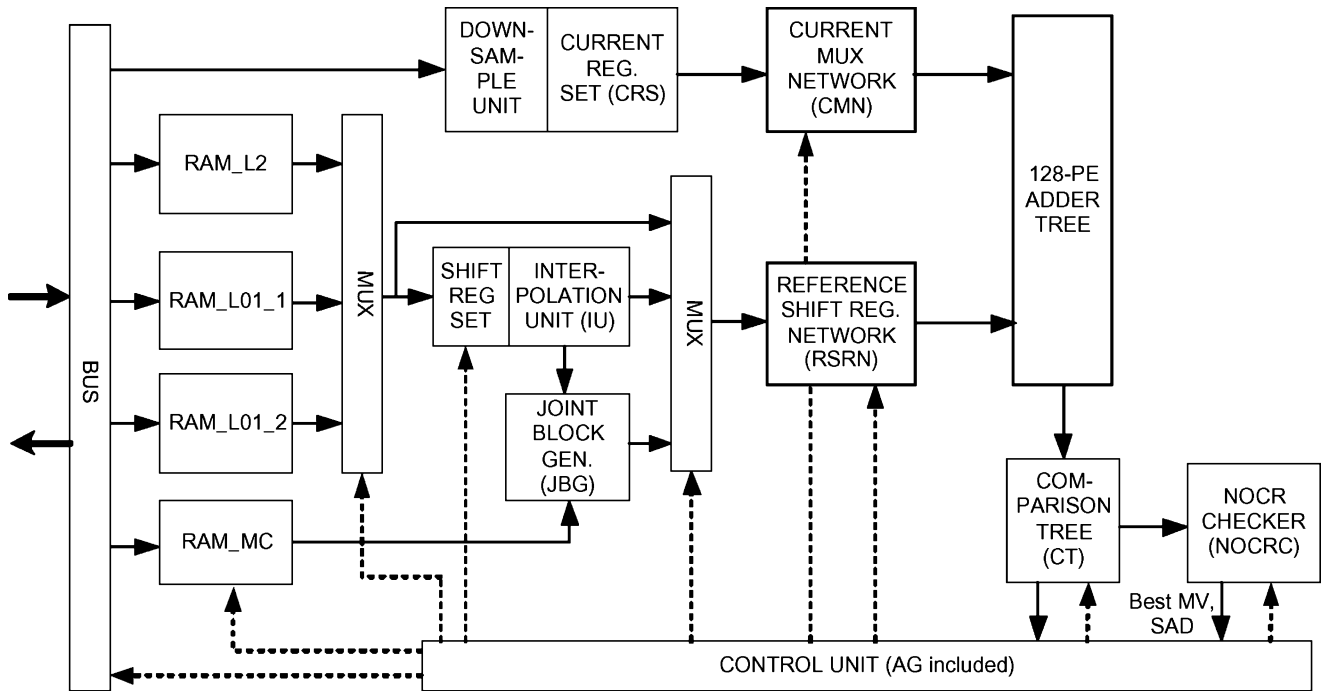
Fig. 16. Architecture of the prediction core. The architecture performs the prediction task in both channels.
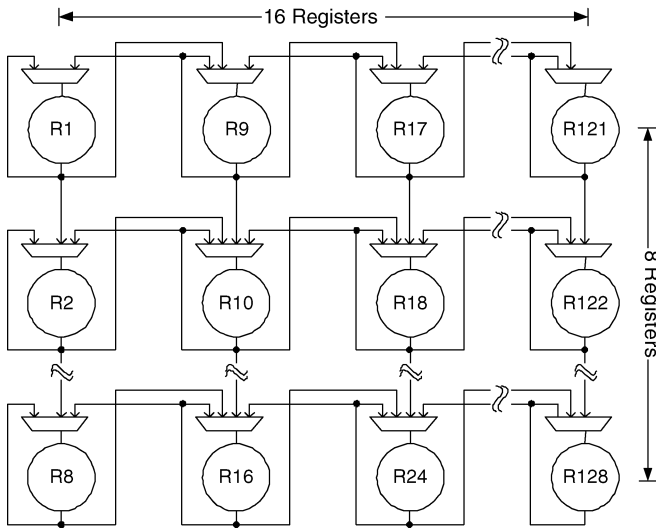

Fig. 17. RSRN. It is composed of a set of reconfigurable shift register arrays so that the pixels can shift downward, leftward, and rightward.
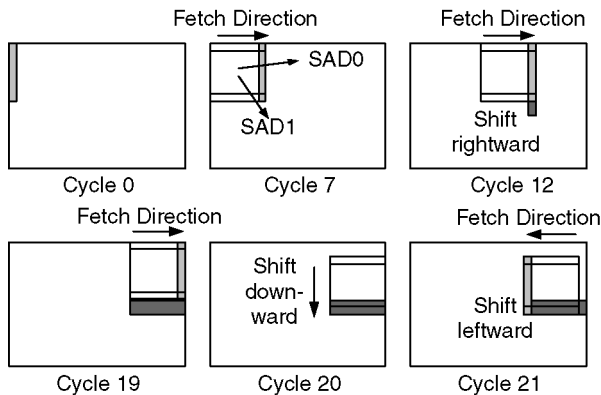

Fig. 18. Data flow of level-1 BMP. Level-2 and level-0 BMP have flows similar to level-1 BMP.
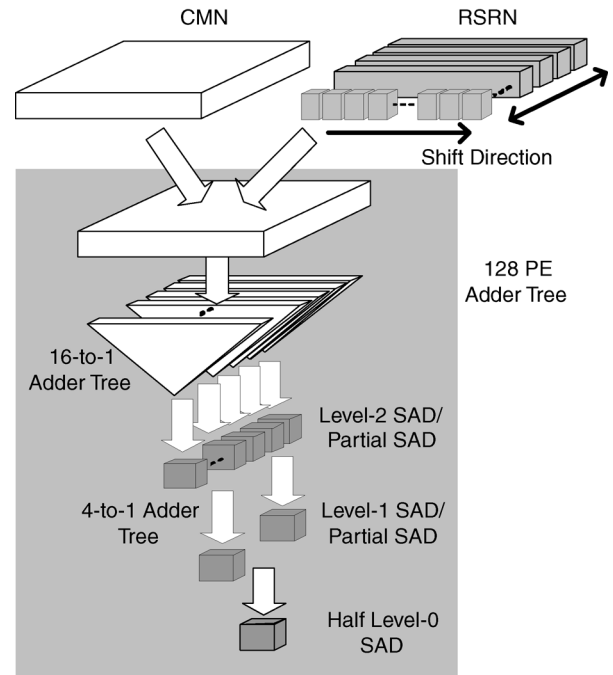

Fig. 19. 128-PE adder tree. The pixels of the current block and the reference candidates are fetched into 128-PE adder tree every cycle.

sum in the joint pel generation. Every JBG unit generates one column of all the joint block in one cycle. After 16 cycles, eight SADs of joint block candidates are generated. Then, they are input to the CT to choose the best SAD, and the best mode is derived as well.

E. Memory Organization and Data Reuse Scheme

Fig. 16 shows that SRAM_L2 stores the level-2 SW data for the left and right channels. RAM_L01_1 and RAM_L01_2 store
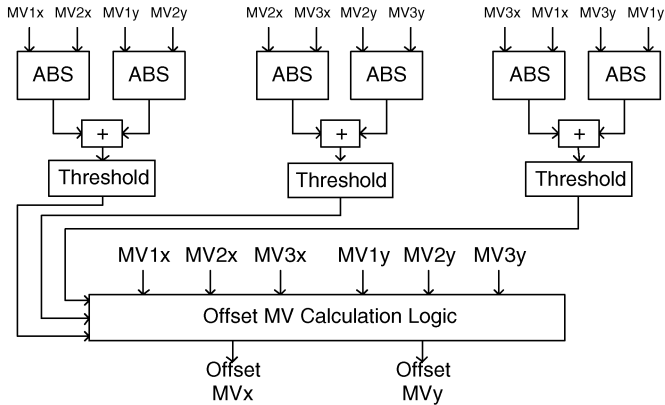
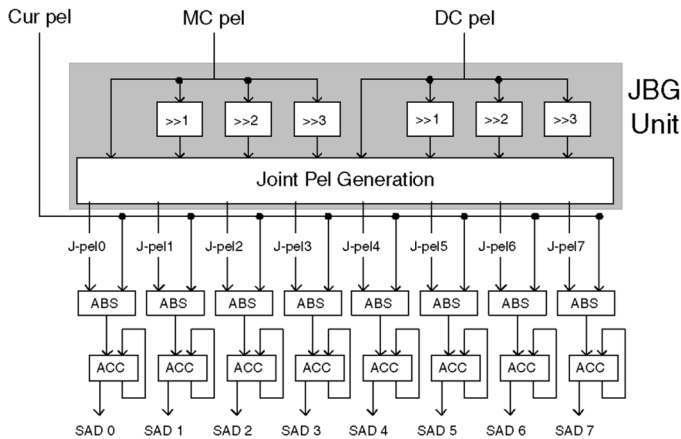Fig. 20. NOCRC. The MV differences are calculated mutually, and then the overlapping condition is decided.

Fig. 21. One of the 16 JBG units. Every JBG unit generates one column of all of the joint blocks in one cycle.

the refinement SW data for level-1 and level-0 BMPs. Since there might be 1 to 3 SWs for block matching, RAM_L01_1 and RAM_L01_2 are accessed with ping-pong mode to store SW data from the off-chip buffer. RAM_MC buffers the best ME candidate block in the right channel, which is used for joint block generation. In the proposed architecture, only 20.75 Kbit on-chip SRAM are required, which is only an 11.5% requirement compared with FSBMA.

Because of the regular data access of level-2 BMP, the level-C reuse scheme is applied for the SW loading in level-2 BMP. The disadvantage of conventional HSBMA [23] is that the SW required for refinement level (level-1, level-0) cannot be reused due to its irregular flow. It increases the data access burdens. However, the proposed NOCRS effectively solves this problem. In other words, data reuse scheme is also applied in level-1 and level-0 BMP to save bandwidth.

F. Proposed Scheduling for Stereo Video Coding System

The proposed scheduling is modified from our prior stereo video system [29] for hardware implementation consideration. The original frame-based scheduling of the prediction engine is shown in the upper part of Fig. 22. It shows that ME in the left channel cannot start until MVs and DVs of all the blocks of a frame in the right channel are derived. However, the SW for DE

in Fig. 3 is enclosed by the search window for ME in the left frame. In the original scheduling, $SW_{l,DE}$ is loaded twice from the off-chip frame buffer. Therefore, this wastes bus bandwidth.

This problem can be solved by the new scheduling, as shown in the lower part of Fig. 22. Before DE of B_r, $SW_{l,ME}(B_l)$ is loaded from off-chip instead of $SW_{l,DE}(B_r)$. After DE and joint block mode decision are done, ME of B_l in the left channel starts. No loading process is needed for ME of B_l. On-chip memory for $SW_{l,DE}(B_r)$ is shared with that for $SW_{l,ME}(B_l)$. The proposed scheduling reduces both the requirements of off-chip memory bandwidth and loading cycles. Moreover, 23% on-chip memory can be saved.

G. Chip Implementation

The proposed prediction core architecture was verified by the VLSI implementation. The processing capability is listed as follows: 720×480 frame size and 30 fps both in the left and right channels. In the ME case, the search range is $[-64, +63]$ in the horizontal direction and and $[-32, +31]$ in the vertical direction. While in the DE case, the search range is $[-64, +63]$ in the horizontal direction and $[-16, +15]$ in the vertical direction. The die photograph is shown in Fig. 23. There are three groups of on-chip single-port SRAM on the chip. The core size is 2.13×2.13 mm^2. The detailed chip features are shown in Table VI.

H. Comparisons

So far, there is no other architecture of prediction core for stereo video systems. The proposed architecture can also perform ME. Table VII shows the comparison of FSBMA and the proposed HSBMA. A search point means one SAD calculation process of a MB. Since the computational complexity of search points in different levels of hierarchical search algorithms is not the same, they are normalized first. Compared with FSBMA, only 11% of on-chip memory is required, and the computational complexity is also greatly reduced. The number of PEs of HSBMA architecture is 3.3% of FSBMA architecture. Fig. 24 shows the rate–distortion performance of FSBMA and proposed HSBMA with NOCRS. Three video sequences are tested by simulation with hardware description language (HDL). Compared with FSBMA, the proposed algorithm maintains good objective video quality.

Table VIII shows the comparison with the previous HSBMA architecture [26]. Although the hardware cost such as logic gate count and on-chip memory are similar, the proposed architecture provides more functionalities, such as DE and joint block compensation for the stereo video prediction, with less PE and system bandwidth requirement. The improvement results from the proposed NOCRS, scheduling, and efficient reconfigurability of the architecture. In addition, it can be easily integrated into mono or stereo video coding systems because of its various functionalities.

V. CONCLUSION

This paper presents an efficient structure of the prediction core in stereo video coding systems from the algorithm level to the hardware architecture level. A stereo video hybrid coding system with the joint prediction scheme is designed for the
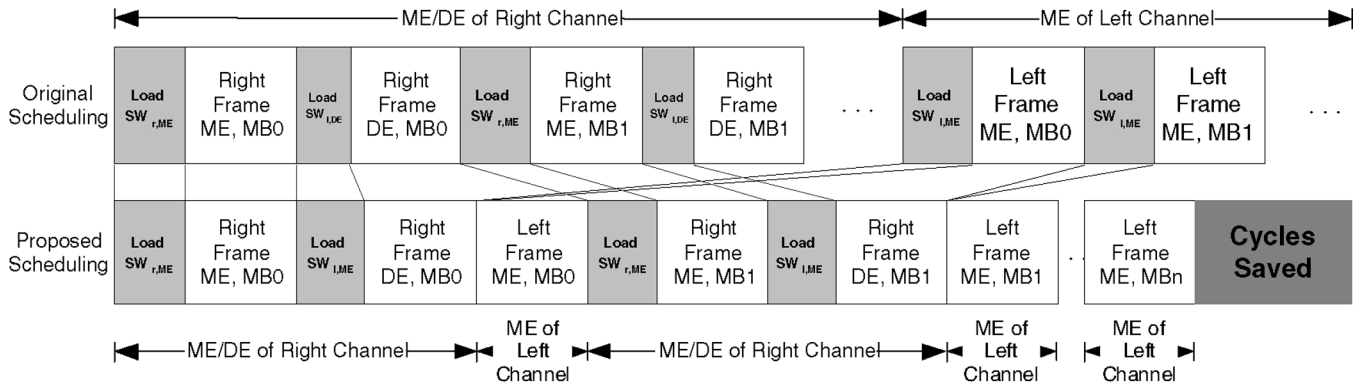
Fig. 22. Proposed scheduling of the stereo video system. No loading process is needed for ME in the left channel.
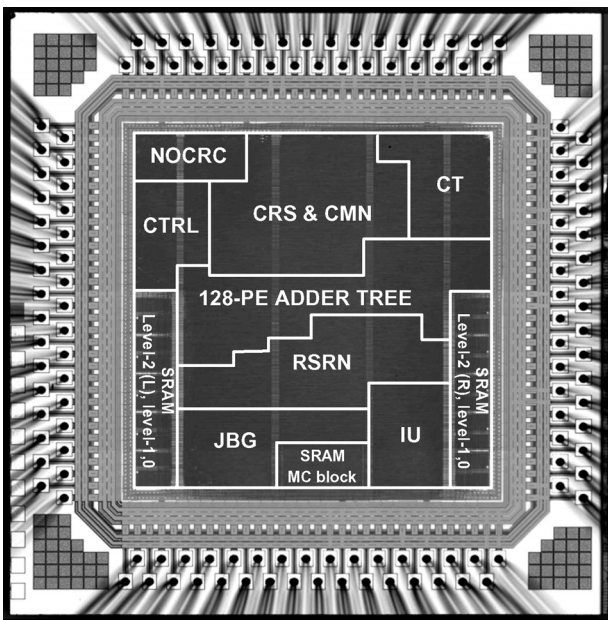

Fig. 23. Die photograph of the prediction core design.

TABLE VI
CHIP SPECIFICATIONS

Technology	TSMC 1P6M 0.18um
Package	128 CQFP
Core size	2.13×2.13mm^2
Logic gate count	137,838 (2-input NAND gate)
On-chip memory	20.75 Kbits
Maximum frequency	100 MHz
Power supply	1.8 V
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 left and right channels simultaneously, including 2 ME and 1 DE operations

TABLE VII
ALGORITHM COMPARISON

Algorithm	FSBMA	HSBMA with NOCRS
On-chip memory	180 Kbits	20.75 Kbits
search points/MB[a]	8192	100 - 234
PE requirement	> 4096	128
Quality drop	0	< 0.2 dB
Bandwidth	74.8 MB/s	91.1 MB/s

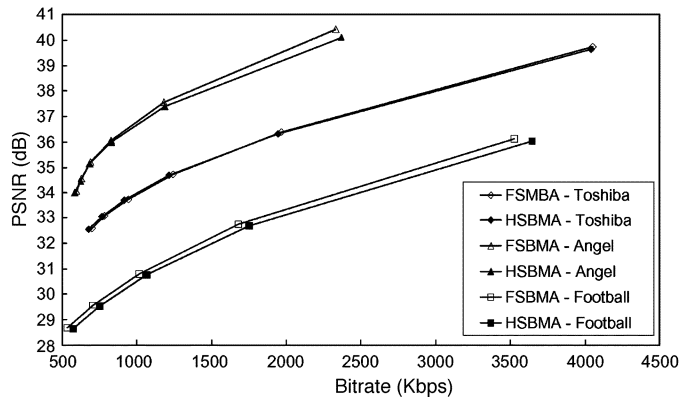[a]The search point is normalized because of different computational complexity in various BMPs.

Fig. 24. Comparison of rate–distortion between proposed HSBMA and FSBMA.

TABLE VIII
ARCHITECTURE COMPARISON

Architecture	[26]	This work
Area	140K	137K
No. of PEs	256	128
Memory	2.5 KBytes	2.6 KBytes
Bandwidth[a]	125 MBytes/sec [b]	91.1 MBytes/sec
Frequency	54 MHz	81MHz
Function	ME	2 ME, 1 DE, joint block compensation

[a]Only mono-channel ME in a P-frame is considered
[b]The system bandwidth requirement is estimated

purpose of overcoming the design challenges of poor coding efficiency and high computational complexity. Compared with MPEG-4 TSP and SP, the coding efficiency is improved by 2–3 dB. In addition, 80% of the computational complexity is reduced. Moreover, a hardware-oriented stereo video prediction algorithm and the associated hardware architecture for the prediction core are also presented. Compared with FSBMA,

the proposed HSBMA greatly reduces hardware cost while still maintaining good video quality. With NOCRS, the problem of critical memory bandwidth requirement can be solved by checking the overlap degree of the three SWs in the refinement levels. The hardware architecture is codesigned with the proposed algorithm with a set of reconfigurable shift register

array and its related circuits, which can be configured for all of the scan directions for three-level BMPs. Moreover, the proposed scheduling not only reduces cycles for loading data from off-chip frame buffer but also eliminates on-chip memory for level-2 of DE. It shows that the architecture is area-efficient and has both good processing capability and functionality.

There are still some extensions in the proposed stereo video hybrid coding system. There might be another suitable pattern combination of the joint block, such as gradual blending modes and H.264 block modes, for better coding efficiency. Mode decision can be further optimized by the Lagrange multiplier. The rate–control algorithms for stereo video coding can also be further explored. On the other hand, MV-DV prediction and mode predecision can be implemented in the architecture for higher area efficiency. They are challenged research topics and also belongs to our future work.

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] MPEG-4 Video Group, Requirements on Multi-View Video Coding , ISO/IEC JTC1/SC29/WG11 N6501. MPEG-4, 2004.

[3] 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 *Proc. IEEE Symp. Circuits*, May 2000, vol. 2, pp. 449–452.

[4] MPEG-2 Group, Proposed Draft Amendment No. 3 to 13818-2 (Multi-View Profile) , ISO/IEC JTC 1/SC 29/WG11 N1088. MPEG-2, 1995.

[5] S.-Y. Chien, S.-H. Yu, L.-F. Ding, Y.-N. Huang, and L.-G. Chen, "Efficient stereo video coding system for immersive teleconference with two-stage hybrid disparity estimation algorithm," in *Proc. IEEE Int. Image Process.*, 2003.

[6] J.-R. Ohm and K. Müller, "Incomplete 3-D multiview representation of video objects," *IEEE Trans. Circuits Syst. Video Technol*, vol. 9, no. 2, pp. 389–400, Mar. 1999.

[7] R.-S. Wang and Y. Wang, "Multiview video sequence analysis, compression, and virtual viewpoint synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 397–410, Apr. 2000.

[8] S.-Y. Chien, S.-H Yu, L.-F. Ding, Y.-N. Huang, and L.-G. Chen, "Fast disparity estimation algorithm for mesh-based stereo image/video compression with two-stage hybrid approach," in *Proc. SPIE Visual Commun. Image Process.*, Jun. 2003, vol. 5150, pp. 1521–1530.

[9] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 2, pp. 139–154, Mar. 1985.

[10] N. Grammalidis and M. G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 3, pp. 328–344, Jun. 1998.

[11] S.-Y. Chien, "Video segmentation: Algorithms, hardware architectures, and applications," Ph.D. dissertation, Dept. Elect. Eng., Nation Taiwan Univ., Taipei, Taiwan, R.O.C., May 2003.

[12] *Generic Coding of Audio-Visual Objects: Part 2-Visual 14496-2*, ISO/IEC JTC1/SC29/WG11 N2502a, FDIS, MPEG-4 Video Group, 1998.

[13] W. Yang, K.-N. Ngan, and J. Cai, "MPEG-4 based stereoscopic and multiview video coding," in *Proc. Int. Symp. Intell. Multimedia, Video Speech Process.*, 2004, pp. 61–64.

[14] Y. Wang, J. Ostermann, and Y.-Q. Zhang*, Video Processing and Communication.* Upper Saddle River, NJ: Prentice-Hall, 2001.

[15] *Advanced Video Coding for Generic Audiovisual Services*, ISO/IEC MPEG and ITU-T VCEG, Joint Video Team (JVT) , May 2003.

[16] J. N. Ellinas and M. S. Sangriotis, "Stereo video coding based on interpolated motion and disparity estimation," in *Proc. 3rd Int. Symp. Image Signal Process. Anal.*, 2003, pp. 301–306.

[17] Y. Luo, Z. Zhang, and P. An, "Stereo video coding based on frame estimation and interpolation," *IEEE Trans. Broadcasting.*, vol. 49, no. 1, pp. 14–21, Mar. 2003.

[18] G. Heising, "Efficient and robust motion estimation in grid-based hybrid video coding schemes," in *Proc. Int. Conf. Image Process.*, 2002, pp. 687–700.

[19] H.-Z. Jia, W. Gao, and Y. Lu, "Stereoscopic video coding based on global displacement compensated prediction," in *Proc. Int. Conf. Inf. Commun. Security*, 2003, pp. 61–63.

[20] MPEG-4 Group, The MPEG-4 Video Standard Verification Model version 18.0 , ISO/IEC JTC 1/SC 29/WG11 N3908, 2001.

[21] S. Cho, K. Yun, B. Bae, Y. Hahm, C. Ahn, Y. Kim, K. Sohn, and Y. h. Kim, Report for EE3 in MPEG 3DAV , ISO/IEC JTC1/SC29/WG11 M9186, Dec. 2002.

[22] T. Koga, K. Linuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion-compensated interframe coding for video conferencing," in *Proc. NTC*, Nov. 1981, pp. C9.6.1–9.6.5.

[23] L. M. Po and W. C. Ma, "A new center-biased search algorithm for block motion estimation," in *Proc. ICIP*, Oct. 1995, pp. 410–413.

[24] S. Zhu and K. K. Ma, "A new diamond search algorithm for fast block matching motion estimation," *Inf., Commun. Signal Process.*, pp. 9–12, Sep. 1997.

[25] C. Zhu, X. Lin, and L. P. Chau, "Hexagon-based search pattern for fast block motion estimation," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 12, no. 5, pp. 349–355, May 2002.

[26] 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 Trans. Circuits Syst. Video Technol.*, vol. 12, no. 9, pp. 1119–1137, Sep. 2004.

[27] W. M. Chao, C. W. Hsu, Y. C. Chang, and L. G. Chen, "A novel hybrid motion estimator supporting diamond search and fast full ssearch," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2002, vol. 2, pp. 492–495.

[28] 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 Trans. Circuits Syst. Video Technol.*, vol. 12, no. 1, pp. 61–72, Jan. 2002.

[29] L.-F. Ding, S.-Y. Chien, and L.-G. Chen, "Algorithm and architecture of prediction core in stereo video hybrid coding system," in *Proc. IEEE Workshop Signal Process. Syst.*, Nov. 2005, pp. 538–543.

Li-Fu Ding was born in Keelung, Taiwan, R.O.C., in 1981. He received the B.S. degree in electrical engineering and the M.S. degree in electronics engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 2003 and 2005, respectively, where he is working toward the Ph.D. degree in electronics engineering.

His major research interests include stereo and multiview video coding, motion estimation algorithms, and associated VLSI architectures.

Shao-Yi Chien (S'00–M'04) was born in Taipei, Taiwan, R.O.C., in 1977. He received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, in 1999 and 2003, respectively.

During 2003 to 2004, he was a Member of the Research Staff with the Quanta Research Institute, Tao Yuan Shien, Taiwan, R.O.C. In 2004, he joined the Graduate Institute of Electronics Engineering and Department of Electrical Engineering, NTU, as an Assistant Professor. His research interests include video segmentation algorithm, intelligent video coding technology, image processing, computer graphics, and associated VLSI architectures.

Liang-Gee Chen (S'84–M'86–SM'94–F'01) was born in Yun-Lin, Taiwan, T.O.C., in 1956. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Cheng Kung University, Tainan, Taiwan, R.O.C., in 1979, 1981, and 1986, respectively.

He was an Instructor (1981–1986) and an Associate Professor (1986–1988) with the Department of Electrical Engineering, National Cheng Kung University. In the military service during 1987 and 1988, he was an Associate Professor in the Institute of Resource Management, Defense Management College. In 1988, he joined the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, R.O.C. During 1993 to 1994, he was a Visiting Consultant with the DSP Research Department, AT&T Bell Laboratories, Murray Hill, NJ. In 1997, he was a Visiting Scholar with the Department of Electrical Engineering, University of Washington, Seattle. Currently, he is a Professor with NTU. Since 2004, he has also been the Executive Vice President and the General Director of Electronics Research and Service Organization (ERSO) in the Industrial Technology Research Institute (ITRI). His current research interests are DSP architecture design, video processor design, and video coding system.

Dr. Chen is a member of Phi Tan Phi. He was the General Chairman of the 7th VLSI Design CAD Symposium. He is also the General Chairman of the 1999 IEEE Workshop on Signal Processing Systems: Design and Implementation. He has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY since June 1996 and an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS since January 1999. He has been an Associate Editor of the *Journal of Circuits, Systems, and Signal Processing* since 1999. He served as the Guest Editor of *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology* in November 2001. He is also an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS. Since 2002, he has also been an Associate Editor of PROCEEDINGS OF THE IEEE. He received the Best Paper Award from ROC Computer Society in 1990 and 1994. From 1991 to 1999, he received Long-Term (Acer) Paper Awards annually. In 1992, he received the Best Paper Award of the 1992 Asia-Pacific Conference on Circuits and Systems in VLSI design track. In 1993, he received the Annual Paper Award of Chinese Engineer Society. In 1996, he received the Out-standing Research Award from NSC, and the Dragon Excellence Award for Acer. He was an IEEE Circuits and Systems Distinguished Lecturer from 2001 to 2002.