New Fast Binary Pyramid Motion Estimation for MPEG-2 and HDTV Encoding

Xudong Song *Member, IEEE*, Tihao Chiang, *Senior Member, IEEE*,
Xiaobing Lee, *Member, IEEE* and Ya-Qin Zhang, *Fellow, IEEE*

Abstract— **A novel Fast Binary Pyramid Motion Estimation (FBPME) algorithm is presented in this paper. The proposed FBPME scheme is based on binary multiresolution layers, XOR Boolean block matching and a *N-scale* tiling search scheme. Each video frame is converted into a pyramid structure of *K-1* binary layers with resolution decimation plus one integer layer at the lowest resolution. At the lowest resolution layer, the *N-scale* tiling search is performed to select initial motion vector candidates. Motion vector fields are gradually refined with the XOR Boolean block matching criterion and the *N-scale* tiling search schemes in higher binary layers.**

FBPME performs several thousands times faster than the conventional full search block matching scheme at the same PSNR performance and visual quality. It also dramatically reduces the bus bandwidth and on-chip memory requirement. Moreover, hardware complexity is low due to its binary nature.

Fully functional software MPEG-2 MP@ML and ATSC HDTV encoders based on the FBPME algorithm have been implemented. FBPME hardware architecture has been developed and is being incorporated into single-chip MPEG encoders. A wide range of video sequences at various resolutions has been tested. The proposed algorithm is also applicable to other digital video compression standards such as H.261, H.263 and MPEG4.

Keywords— **Binary Pyramid, Motion Estimation, MPEG, HDTV, Block Matching, Tiling Search**

I. Introduction

THE motion estimation (ME) and compensation are critical components for digital video compression systems such as High Definition Television (HDTV) and Standard Definition Television (SDTV) broadcasting equipment, video conferencing transmitters, and multimedia servers for Web applications. Block-based ME has been widely adopted by several important international standards such as Motion Picture Expert Group (MPEG), Advanced Television Standard Committee (ATSC), Digital Video Broadcasting (DVB) and International Telecommunications Union (ITU). The relevant standards include ISO/IEC MPEG-1, MPEG-2, MPEG-4, MPEG-7, H.261, H.263, DVB and ATSC specifications [1], [2], [3], [4], [5].

The optimal block-based ME is typically defined as the exhaustive search algorithm using Minimum Absolute Difference (MAD) block matching criterion. The technique evaluates motion vectors for all locations within the search window. Many fast ME algorithms with various sparse searching schemes have been developed to reduce computational complexity. Examples include the three-step search [6], the 2D logarithmic search [7], the conjugate directional search [8], the genetic search [9], [10], the unrestricted center-biased diamond search [11], the feature-based block motion estimation using integral projection [12], and sub-sampled motion field estimation with alternating pixel-decimation patterns [13]. These sparse searching schemes provided much reduced complexity at the expense of motion accuracy. In addition, fast search schemes are less robust in the sense that they are often trapped into local minimum.

The multi-resolution ME techniques [14], [15], [16], [17], [18], [19] were developed to search a smaller window from lower to higher resolution layers similar to the hierarchical motion projection search. A small search window at reduced resolution can cover the same area as the full search in the full resolution with reduced complexity. With an efficient hierarchical structure, decimation and filtering algorithms, the multi-resolution ME also provides greater robustness than sparse search algorithms.

To further reduce the computation, binary ME algorithms have been developed. Binary ME algorithms use an XOR Boolean block matching criterion instead of the integer MAD criterion for block matching. The binary representation provides much lower computational complexity and data bus throughput for the basic block matching module [20], [21], [22], [23], [24], [25], [26], [27]. The binary ME algorithms can be combined with any fast search schemes for further speed improvement.

In this paper, a new Fast Binary Pyramid Motion Estimation (FBPME) algorithm is presented, based on a binary pyramid structure, XOR Boolean block matching, and *N-scale* tiling search scheme. The block matching now uses only bit-wise XOR logic operations that are much simpler and faster to implement than integer MAD operations. The *N-scale* tiling scheme provides multi-path hierarchical projection searches from layer to layer to achieve more robust performance. FBPME is about 4000 times faster than the integer MAD full search ME, and 16 to 256 times faster than the HFM-ME algorithm [21], [22]. It also requires much less data bus bandwidth due to the binary nature of the data transfer between frame and reference buffers.

This paper is organized as follows. Section 2 discusses previous efforts in binary ME algorithms, and briefly summarizes the difference and advantages of the proposed

Tihao Chiang was with Sarnoff Corporation and is with Dept. of Electronics Engineering at National Chiao Tung University, R.O.C.

Ya-Qin Zhang was with Sarnoff Corporation in Princeton, NJ. He is now with Microsoft Research China in Beijing, PRC.

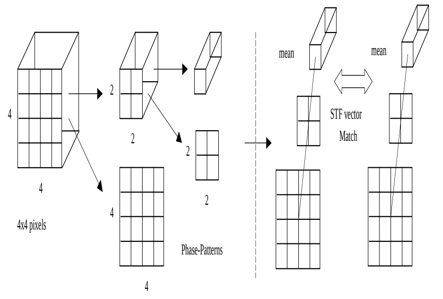

Fig. 1. Illustration of multi-resolution HFM block matching with STF vectors

FBPME. Section 3 describes the construction of the binary pyramid. Section 4 presents the XOR Boolean block matching and its parallel hardware architecture. Section 5 describes the *N-scale* tiling multi-path search for the FBPME algorithm. Section 6 derives the computation complexity and data bus bandwidth. Section 7 shows the testing results of FBPME used in MPEG-2 MP@ML and ATSC HDTV encoders for both SDTV and HDTV video inputs.

II. Binary Motion Estimation Algorithms

A binary based fast Hierarchical Feature Matching Motion Estimation scheme (HFM-ME) was proposed to reduce the matching computational complexity and bus throughput in [21], [22]. The HFM-ME algorithm performs ME with feature vector matching using integer means and binary phases. Zakhor and Lari [26] used one-bit edge information for global motion parameters to achieve camera stabilization with comparable performance to that of intensity-based integer ME. Natarajan, Bhaskaran, and Konstantinides [23] presented a fast binary motion estimation algorithm based on a simple one-bit transformation and conventional search schemes. This algorithm provides good performance for simple texture and slow motion sequences. For a complex texture scene sequence such as “Flowergarden”, it is reported in [23] that this binary ME is 2 dB below that for the full search MAD. Gharavi and Mills [20] presented a threshold method to convert the absolute difference of pixel matching into “0” or “1” bit representation. Thus the accumulator of the Mean Absolute Difference (MAD) becomes a simple counter for “1’s”. Feng, Lo, Mehrpour, and Karbowiak [27] used the mean of 8x8 pixels to threshold all the pixels to form an 8x8 bit-plane. The full search binary match is performed in the bit-planes to find several locations with mismatches less than a threshold. The full search MAD matching is then performed on these selected locations to obtain the best-matched position.

A. Hierarchical Feature Matching Motion Estimation

Lee and Zhang explored the basic binary block matching concepts as illustrated in the Figure 1. The 4x4 integer array of MAD block matching can be replaced by the sign truncated feature (STF) vectors matching with 8 bits of

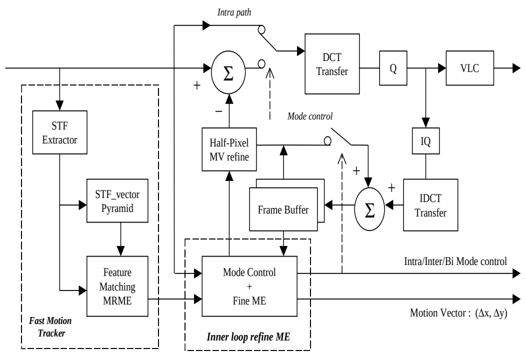

Fig. 2. HFM fast motion tracker and refinement ME of the inner encoder loop

mean, 2x2 and 4x4 bits arrays of phases. As shown in Figure 2, the HFM-ME algorithm is implemented as a fast motion tracker to provide coarse motion vectors and a refined motion vector with a small (3x3) pixels search window. A half-pixel mode decision module is used to find the final motion vector with half-pixel accuracy. The HFM-ME is executed as follows:

1. Build multi-resolution STF feature vectors, frame-by-frame
2. Perform 16x16 XOR Boolean matching and 4x4 integer array of MAD matching at each search point
3. Use the relative full search scheme on the binary bit-planes
4. Use the inner encoding loop MAD ME on reconstructed data within 3x3 pixels of the refined window.

In step 1, a 4x4 sub-block of luminance integers is represented as a feature vector of one integer of the mean, 2x2 bits array and 4x4 bits array of binary phase pattern. They can be simplified as one mean with 4x4 bits array of phases. Then a 16x16 macro-block is expressed as 4x4 sets of the mixed resolution feature vectors. As the block matching, the 4x4 integer matching of MAD and 16x16 binary matching of XOR Boolean operations are performed. In the XOR Boolean block matching, a “0” stands for a good matched pixel, and a “1” for a mismatched pixel. A Look Up Table (LUT) of the counted “0’s” is developed for software emulation of the fast numeric mapping of the XOR matching results. The hardware logic implementation should be much simpler and faster.

In step 2, the full-search scheme is applied to the 4x4 array of STF feature vector matching, vector-by-vector. As a straightforward implementation with a ± 1 pixel search step, each mean matching only needs 1/16 of integer computations and data retrieval as in the 16x16 array of integer MAD matching. The 4x4-integer array of matches can be performed with a ± 4 pixel search step. Around each step at location (x, y), 16x16 array of binary phase matching is performed with ± 1 pixel step skew of (dx, dy).

In step 3, the relative full search is performed around each checkpoint without phase skew as in step 2. The 4x4 array of STF feature of the reference frame is matched

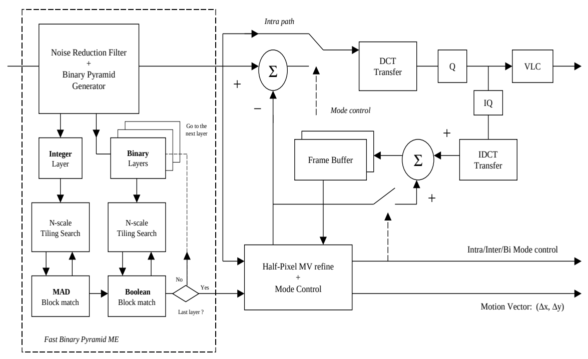

Fig. 3. Illustration of fast binary pyramid motion estimation

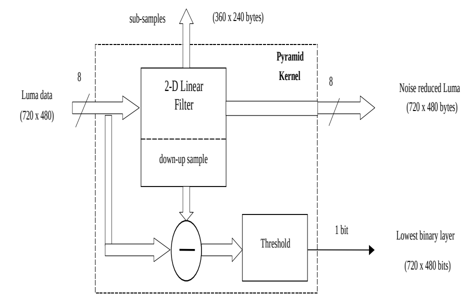

Fig. 4. Basic pyramid filtering and binary bit-plane construction

with the shifted integer means and binary phases of current blocks rather than the skewed reference blocks, in order to eliminate the phase skew errors without retrieving the shifted means of the reference block from the external memories. The STF vectors of (dx,dy) shifted current block can be derived from the input data in the on-chip memories. Therefore, the final motion vector (x+dx, y+dy) is shifted from the input block location (i, j) to (i-dx, j-dy).

In step 4, a refined full search ME module is done in the inner encoding loop to offset the (dx,dy) shift as shown in the Figure 2. Experiment results show that HFM-ME with relative full-search provides almost the same PSNR performance as the full search, but with much less computation complexity and bus bandwidth.

The disadvantages of the HFM-ME algorithm are: (a) difficulty in combination with sparse fast search; (b) inaccuracy of phase skewing; (c) limited number of layers due to the fixed MB size; (d) the need for the inner encoding loop ME module, and (e) need for fusion of integer matching and binary matching results.

B. Fast Binary Pyramid Motion Estimation

In this paper we propose a new FBPME technique as shown in Figure 3, which has low computational complexity, reduces data bandwidth, and requires simple hardware implementation. FBPME consists of the binary pyramid construction, XOR block matching logic and *N-scale* tiling search scheme. FBPME eliminates the phase skew of the HFM-ME in the binary pyramid construction of the FBPME. The XOR matching is implemented with simple parallel hardware architecture. *N-scale* tiling search is used to refine the motion vector candidates obtained from the lowest resolution layer.

The tiling search scheme combines the advantages of the hierarchical motion projection search and multi-path search similar to the Viterbi algorithm. The multi-path tiling search can greatly reduce the computations with multiple smaller search windows to preserve the robust performance of full search block matching, eliminating the need for large windows using binary full search and the inner-loop refinement in the HFM-ME algorithm.

The FBPME is performed according to the following steps:

1. Build the binary pyramid structure
2. Perform the *N-scale* tiling search with integer MAD block matching at the top integer layer
3. Perform the *N-scale* tiling search with XOR Boolean block matching at binary layers
4. Perform full search XOR Boolean matching with a ± 3 pixel refinement window at original resolution.

In step 1, the Binary Pyramid Generator performs k layer decimation filtering to build a conventional pyramid structure with k integer layers. Each integer layer (except the bottom layer) is up-sampled to approximate its lower layer data. The difference between the integer layer and its approximation is compared to a threshold to derive the binary layer. The resultant binary pyramid structure consists of one top integer layer and k-1 binary layers. This new binary pyramid structure offers more precise representation for the residual information of the pyramid layers without the phase skew errors caused by the simple mean threshold methods.

In step 2, *N-scale* ($N = 4$) tiling search is performed for the top integer layer to locate four initial motion vector candidates corresponding to the best MAD block matching with different shapes of tiling super-blocks. These four initial motion vector candidates will be projected into the next binary layer for multi-path motion vector refinements. Each candidate is related to one of the four shapes of tiling such as 8x8, 4x8, 8x4, and 4x4 blocks, as shown in Figure 13. This tiling search with various shaped super-blocks covers greater geometrical region of multiple macroblocks in the original resolution layer to prevent the FBPME from being trapped into local minima. The FBPME scheme, therefore, can achieve more reliable and accurate motion estimation with similar or smaller search windows.

In step 3, the XOR Boolean block match is combined with the *N-scale* tiling search at higher resolution binary layers. Centering on each motion vector projection location, the tiling search is performed in a 3x3 pixel window. The four motion vector candidates with the best XOR block match will be projected into the next binary layer. The same process will be repeated, layer-by-layer, until the last binary layer. The XOR Boolean block match can be implemented with look-up-table methods using soft-

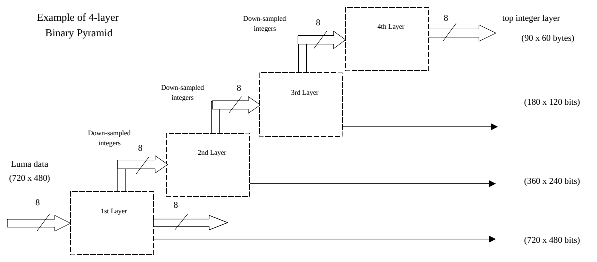

Fig. 5. Illustration of 4-layer binary pyramid construction

ware emulation or hardware logic as indicated in Section 4.

In step 4, at the original resolution binary layer, the binary motion vector refinement is performed with the XOR Boolean block match of 16x16 bits and full search around the best block match of the four projected candidates from the previous layer. The search window is ±3 pixels.

III. Construction Of the Binary Pyramid

The binary pyramid is constructed as follows:

1. Initialization: Let k be the number of pyramid level. Let $l = 0$ and level 0 be the original image, i.e., $\mathbf{X_l}$=$\mathbf{X_0}$.
2. Generate level l+1: An image $\mathbf{X_l}$ is filtered by a low-pass filter $\mathbf{H}$, the output of a low-pass filter $\mathbf{H}$ is $\mathbf{H}(\mathbf{X_l})$. $\mathbf{H}(\mathbf{X_l})$ is decimated by a factor of M in each dimension which produces a reduced image $\mathbf{X_{l+1}} = \mathbf{D}_\mathrm{M}(\mathbf{H}(\mathbf{X_l}))$. The reduced image $\mathbf{X_{l+1}}$ is interpolated by an interpolating filter $\mathbf{I_M}$ giving the expanded image $\mathbf{I_M}(\mathbf{X_{l+1}})$.
3. Form the Binary Pyramid $\mathbf{B_l}$:

$$\mathbf{B_l} = \begin{cases} 1 & \text{if } (\mathbf{X_l} - \mathbf{I_M}(\mathbf{X_{l+1}})) \geq \mathbf{T} \\ 0 & \text{otherwise} \end{cases}$$

   Where T is a threshold.
4. Termination: Let $l = l + 1$ and if $l \neq k - 1$, go to step 2; otherwise, $\mathbf{B_l} = \mathbf{X}_(\mathbf{k-1})$, stop.

Figure 4 illustrates the basic functions to construct the binary pyramid. The 2-D filter and down-sampling operations produce the conventional pyramid data structures. The difference between the input data and the up-sampling approximations is compared with a threshold to generate the binary plane with the same spatial resolution as the input data. This process is repeated on the down-sampled data, layer-by-layer, until reaching the desired binary level. An example of 4-layer binary pyramid construction is shown in the Figure 5.

An example to illustrate the binary pyramid construction process follows. A 4x4 sub-block of input data is filtered and decimated by a factor of 2 as shown in Equation (1).

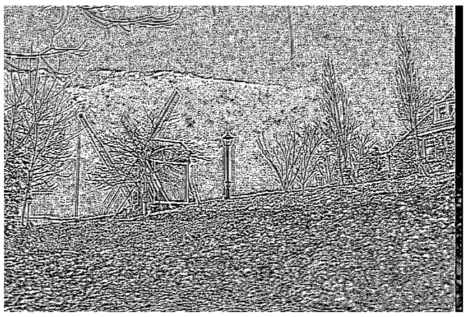

Fig. 6. The image of level 1 of the binary pyramid generated from the 2nd frame in the "Flowergarden" sequence.

$$Subsample \begin{pmatrix} 74 & 59 & 100 & 158 \\ 74 & 69 & 59 & 80 \\ 87 & 86 & 65 & 69 \\ 100 & 118 & 72 & 60 \end{pmatrix} = \begin{pmatrix} 70 & 94 \\ 87 & 72 \end{pmatrix} \tag{1}$$

The 2x2 decimated data is up-sampled as shown in Equation (2).

$$Upsample \begin{pmatrix} 70 & 94 \\ 87 & 72 \end{pmatrix} = \begin{pmatrix} 70 & 82 & 94 & 94 \\ 78 & 80 & 83 & 83 \\ 87 & 79 & 72 & 72 \\ 87 & 79 & 72 & 72 \end{pmatrix} \tag{2}$$

Finally, the difference between Equation (1) and (2) is compared with a threshold of zero to give a binary plane as shown in Equation (3).

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 74 & 59 & 100 & 158 \\ 74 & 69 & 59 & 80 \\ 87 & 86 & 65 & 69 \\ 100 & 118 & 72 & 60 \end{pmatrix} - \begin{pmatrix} 70 & 82 & 94 & 94 \\ 78 & 80 & 83 & 83 \\ 87 & 79 & 72 & 72 \\ 87 & 79 & 72 & 72 \end{pmatrix} \tag{3}$$

Figure 6 shows the first binary layer image of the 2nd frame in the "Flowergarden" sequence. Figures 7-8 show the 2nd and 3rd binary layer images, respectively. Figure 9 shows the top layer integer image, in 8 bits per pixel, as the 4-th layer of conventional sub-samples. The binary layer images have not only preserved the edge information but also emphasized the complex details from relatively flat areas, such as the cloud area in the "Flowergarden" sequence.

IV. XOR Boolean Matching Criterion and Parallel Architecture

Block matching algorithms are widely used in motion compensated video-coding applications. In block matching

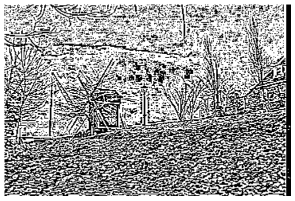

Fig. 7. The image of level 2 of the binary pyramid generated from the 2nd frame in the "Flowergarden" sequence.

Fig. 8. The image of level 3 of the binary pyramid generated from the 2nd frame in the "Flowergarden" sequence.

ME, a frame F_c is divided into blocks with size of $G \times G$. For each block in the current frame F_c, a block from the reference frame F_r (a previous or future frame) with corresponding displacement (x, y) is selected by satisfying the block matching criteria. This displacement (x, y) in the reference block shifted from the current block location is defined as the motion vector.

A. Block Matching Criteria

The Mean Absolute Difference (MAD) is often used as the block-matching criterion. The MAD calculation involves one subtraction, one absolute value operation, and one accumulation per pixel.

$$MAD(x, y) = \frac{1}{G^2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |I_c(i, j) - I_r(i + x, j + y)|$$

where $I_c(i, j)$ is the value of pixel intensity at location (i, j) within the current frame. $I_r(i + x, j + y)$ is the value of pixel intensity at location (i+x, j+y) within the reference frame.

In MAD block matching, each pixel is represented as an 8-bit integer. In binary pyramid, a byte is reduced to a bit so only bit-wise operations are needed. The absolute difference computation then becomes a simple logic XOR operation, such as:

$$\begin{aligned} A - B &= A \oplus B \textit{ with a borrow bit } Cy = \overline{A}B \\ |A - B| &= A \oplus B \textit{ without any borrow bit} \end{aligned}$$

where $\oplus$ denotes the exclusive-or (XOR) operation.

The XOR binary block match is the simplest case of the MAD matching criterion with the minimum hardware complexity. The XOR match criterion is expressed as the following:

$$XOR(x, y) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \phi(I_c(i, j) \oplus I_r(i + x, j + y))$$

Fig. 9. The image of level 4 of the binary pyramid generated from

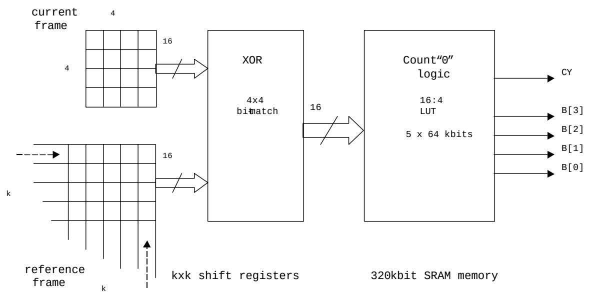

Fig. 10. Block diagram of the 4x4 kernel operator of 4×4 arrays of XOR Boolean block matching

$$\phi(I_c(i, j) \oplus I_r(i + x, j + y)) = \begin{cases} 0 & \text{if } I_c(i, j) = I_r(i + x, j + y) \\ 1 & \text{otherwise} \end{cases}$$

It shows that the binary block matching becomes simply counting the number of "0"s in the $G \times G$ match pattern between the current and the reference binary blocks. The more the "0"s, the better the two binary blocks are matched.

B. The Parallel XOR Matching Architecture

The binary XOR matching operations can be easily built into NxN arrays of parallel logic architecture. In a simple Field Programmable Gate Array (FPGA) design, the kernel operator can be 4x4 arrays and 4x4 such kernel operators can perform the 16x16 of Boolean XOR match of a macro-block in a single clock cycle.

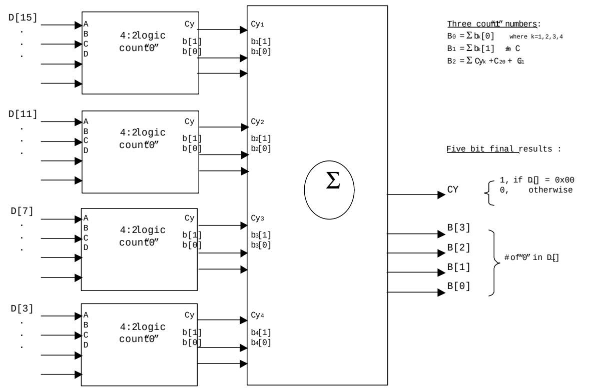

Fig. 11. Partitions of combinatorial logic for the 16:4 count "0" logic

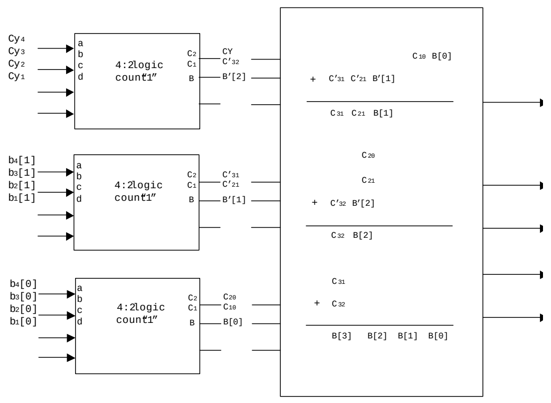

Fig. 12. Partition of the summation logic into three count and a complex carry logic

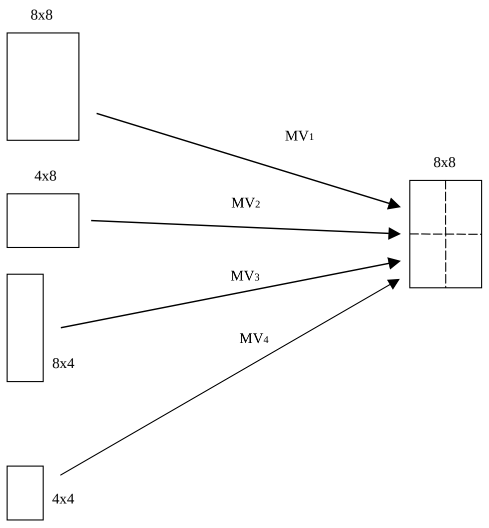

Fig. 13. Illustration of 4-scale tilings search scheme

The Figure 10 shows a block diagram of the 4x4 operator. It consists of a 4x4 register file, a $k \times k$ shi ister arrays, a 4x4 XOR match operator, and a 16: Up Table (LUT) memory. The 4x4 register file hold binary sub-block from the current frame F_c and the shift register arrays hold the bits of the search windo the reference frame F_r. The 4x4 XOR operator pe the 4×4 Boolean match, and the 16:4 LUT conve number of "0" in the 16 bit match result into the numeric data. All the registers, shift registers, an memory cells can be implemented by Static Rando cess Memory (SRAM) technologies. The 16:4 LUT m utilizes 320kbits SRAM cells, as $(4+1) \times 2^{16}$ bits an of them saved the same value. Using two of 8:3 LUT and a 4bit adder can greatly reduce the SRAM size with compromising the speed. A 4bit counter and 16bit shift register can easily count the number "0" of the matched bits in 16 clock cycles with much less logic cells.

The "Hard-Wire" Application-Specific Integrated Circuits (ASIC) solutions can implement the same 16:4 logic of counting "0" with much less logic gates and the fast performance. The Figure 11 illustrates the basic functions of the 16:4 mapping logic with the "Hard-Wire" combinatorial logic implementations. The 16:4 logic can be partitioned into 4 of 4:2 count "0" logic circuits and a summation logic to add up the 4 of 2bit numerical numbers.

The 4:2 count "0" logic converts the 4 input bits $[A, B, C, D]$ to 2 output bits $b[1], b[0]$, and Cy (carry bit). When the Cy = 1, the $b[1]$ and $b[2]$ will be equal to 0, because there are only 4 input bits. The 4:2 count "0" logic is expressed as:

$$\begin{aligned} Cy &= \overline{A}\ \overline{B}\ \overline{C}\ \overline{D} \\ b[0] &= A \oplus B \oplus C \oplus D \\ b[1] &= \overline{Cy}\ \overline{ABC}\ \overline{ABD}\ \overline{ACD}\ \overline{BCD} \end{aligned}$$

where the $b[0]$ shows there are 1 or 3 "0"s and the $b[1]$ with 2 or 3 "0" bits in the 4 input bits, respectively. The $b[1]$ can be illustrated as there are no 1 "0", nor 4 "0"s, nor 4 "1"s.

The summation logic can be further partitioned into three count "1" logic circuits, as illustrated in the Figure 12. At first, the B[0] count "1" logic adds all the least significant bits {b1[0], b2[0], b3[0], b4[0]} from the outputs of four count "0" circuits. Then, the B[1] logic adds the second bits {b1[1], b2[1], b3[1], b4[1]} and the carry bit from B[0]. Finally, the B[2] logic adds all the carry bits {Cy1, Cy2, Cy3, Cy4} plus carry bits from B[0] and B[1], respectively. They are expressed as the follows:

$$\begin{aligned} B[0] &= b_1[0] \oplus b_2[0] \oplus b_3[0] \oplus b_4[0] \\ B[1] &= b_1[1] \oplus b_2[1] \oplus b_3[1] \oplus b_4[1] \oplus C_{10} \\ B[2] &= Cy1 \oplus Cy2 \oplus Cy3 \oplus Cy4 \oplus C_{20} \oplus C_{21} \\ B[3] &= C_{31} \oplus C_{32} \\ CY &= Cy1\ Cy2\ Cy3\ Cy4 \end{aligned}$$

where the C_{10} and C_{20} are the carry bits from the B[0] logic. The bits C_{21}, C_{31} and C_{32} are carry bits from the B[1] logic and B[2] logic circuits, respectively. When CY=1, the set {B[3],B[2].B[1],B[0]} equals {0,0,0,0} because this 16:4 count "0" logic has only 16 input bits.

The carry bits from the B[0] logic are expressed as:

$$\begin{aligned} C_{20} &= b_1[0]\ b_2[0]\ b_3[0]\ b_4[0] \\ C_{10} &= \overline{C_{20}}\ \overline{\overline{b_1[0]}\ \overline{b_2[0]}\ \overline{b_3[0]}}\ \overline{\overline{b_1[0]}\ \overline{b_2[0]}\ \overline{b_4[0]}} \\ &\quad \overline{\overline{b_1[0]}\ \overline{b_3[0]}\ \overline{b_4[0]}}\ \overline{\overline{b_2[0]}\ \overline{b_3[0]}\ \overline{b_4[0]}} \end{aligned}$$

where the C_{10} can be interpreted as that there exist no four "0"s nor one "1" nor four "1"s in the least significant bits {b1[0], b2[0], b3[0], b4[0]}.

To simplify the logic, three of 4:2 count "1" circuits are used. The B'[1] and B'[2] logic are expressed as:

$$\begin{aligned} B'[1] &= b_1[1] \oplus b_2[1] \oplus b_3[1] \oplus b_4[1] \\ C'_{31} &= b_1[1]\ b_2[1]\ b_3[1]\ b_4[1] \end{aligned}$$

$$
\begin{aligned}
C'_{21} &= \overline{\overline{b_1[1]}\;\overline{b_2[1]}\;\overline{b_3[1]}}\;\overline{\overline{b_1[1]}\;\overline{b_2[1]}\;\overline{b_4[1]}}\;\overline{\overline{b_1[1]}\;\overline{b_3[1]}}\\
&\quad \overline{\overline{b_2[1]}\;\overline{b_3[1]}\;\overline{b_4[1]}}\;\; C'_{31}\\
B'[2] &= Cy1 \oplus Cy2 \oplus Cy3 \oplus Cy4\\
C'_{32} &= \overline{\overline{Cy1}\;\overline{Cy2}\;\overline{Cy3}}\;\overline{\overline{Cy1}\;\overline{Cy2}\;\overline{Cy4}}\;\overline{\overline{Cy1}\;\overline{Cy3}\;\overline{C}}\\
&\quad \overline{\overline{Cy2}\;\overline{Cy3}\;\overline{Cy4}}\;\overline{CY}
\end{aligned}
$$

Consequently, the carry bits from B[1] and B[2] c
are as follows:

$$
\begin{aligned}
C_{21} &= C'_{21} \oplus (C_{10}B'[1])\\
C_{31} &= C'_{31} \oplus (C_{10}B'[1]C'_{21})\\
C_{32} &= C'_{32} \oplus (C_{20}C_{21} + C_{20}B'[2] + C'_{21}B'[2]
\end{aligned}
$$

The special "Hard-Wire" ASIC implementation
parallel logic architecture for the Boolean XOR
matching can be much fastser than the FPGA LU
signs and the software emulations of the LUT algo
with the general purpose DSP and CPU processors.

V. Binary Pyramid Motion Estimation W *N-Scale* Tilings

Let F_m denote the m-th image frame in a video seq
The binary pyramid F_m is generated using the alg
described in Section 3. Each pyramid layer is parti
into non-overlapping blocks of size $\mathbf{s}$. The resulting
tion or *tiling* at layer l is represented by $F_m^{l,\mathbf{s}}$. The in
value of the pixel with coordinates $\mathbf{x} = [x_1, x_2]^T$ in
age frame m at layer l is denoted by $F_m^l(\mathbf{x})$ where
and T denote the row index, the column index, and "transpose", respectively. For a given $\mathbf{s} = [s_1, s_2]^T$, the block of pixels with upper left corner at image position $\mathbf{x}$ at layer l is referred to as $B_m^l(\mathbf{x}, \mathbf{s}) = \{F_m^l(\mathbf{q}) \in F_m^l | \mathbf{x} \le q < \mathbf{x} + \mathbf{s}\}$. The sum of the absolute differences between pixels from a block $B_m^l(\mathbf{x}, \mathbf{s})$ from frame m and corresponding pixels in block $B_{m-1}^l(\mathbf{x}, \mathbf{s})$ from frame $m-1$ can be represented as:

$$
\begin{aligned}
&B_m^l(\mathbf{x}, \mathbf{s}) \ominus B_{m-1}^l(\mathbf{x}, \mathbf{s})\\
=&\; \textstyle\sum_{\mathbf{0} \le \mathbf{d} < \mathbf{s}} |F_m^l(\mathbf{x} + \mathbf{d}) - F_{m-1}^l(\mathbf{x} + \mathbf{d})|
\end{aligned}
$$

The sum of *XOR Boolean* operations between pixels from a block $B_m^l(\mathbf{x}, \mathbf{s})$ from frame m and corresponding pixels in block $B_{m-1}^l(\mathbf{x}, \mathbf{s})$ from frame $m-1$ can be denoted as:

$$
\begin{aligned}
&B_m^l(\mathbf{x}, \mathbf{s}) \oplus B_{m-1}^l(\mathbf{x}, \mathbf{s})\\
=&\; \textstyle\sum_{\mathbf{0} \le \mathbf{d} < \mathbf{s}} \phi(F_m^l(\mathbf{x} + \mathbf{d}) \oplus F_{m-1}^l(\mathbf{x} + \mathbf{d}))
\end{aligned}
$$

$$
\begin{aligned}
&\phi(F_m^l(\mathbf{x} + \mathbf{d}) \oplus F_{m-1}^l(\mathbf{x} + \mathbf{d}))\\
=&\; \begin{cases} 0 & \text{if } F_m^l(\mathbf{x} + \mathbf{d}) = F_{m-1}^l(\mathbf{x} + \mathbf{d})\\ 1 & \text{otherwise} \end{cases}
\end{aligned}
$$

Let $F^{0,\mathbf{s}}$ be a tiling defined on the full resolution image with cardinality κ. Let $F^{0,\mathbf{s/2}}$ be a tiling defined on level 1. For the lower resolution image, *N-scale* tilings are considered. They are $\{F^{l,\mathbf{s1}}\}_{l=2}^{k-1}$, $\{F^{l,\mathbf{s2}}\}_{l=2}^{k-1}$, $\cdots$, and $\{F^{l,\mathbf{sN}}\}_{l=2}^{k-1}$.

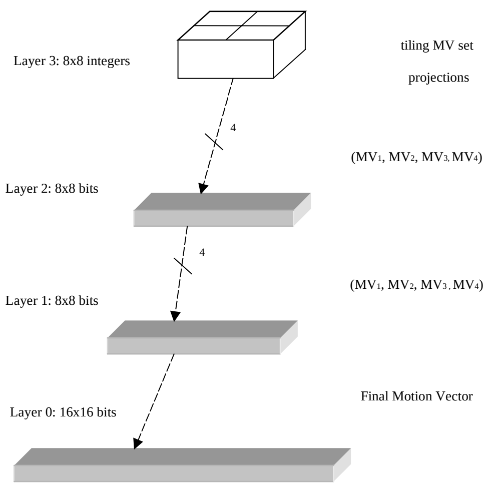

Fig. 14. Tiling set of motion vector projection into multi-layer pyramid

1. Initialization: Let k be the number of binary pyramid level and $l = k-1$. At the Level $k-1$, the motion vector fields $\mathbf{v}_{s1}^{k-1}$, $\mathbf{v}_{s2}^{k-1}$, $\cdots$, and $\mathbf{v}_{sN}^{k-1}$ are defined as follows:

   $$\mathbf{v}_{sj}^{k-1} = \arg \min_{v \in \Omega^{k-1}} B_m^{k-1}(\mathbf{x}, \mathbf{s}j) \ominus B_{m-1}^{k-1}(\mathbf{x} + \mathbf{v}, \mathbf{s}j)$$

   where

   $$
   \begin{aligned}
   1 &\le j \le N\\
   \Omega^{k-1} &= \{\mathbf{v} : -\mathbf{d}^{k-1} \le \mathbf{v} \le \mathbf{d}^{k-1}\}
   \end{aligned}
   $$

2. Projection and Refinement: From the lower resolution $l+1$ to the higher resolution l, the motion vector fields are projected according to the following:

   $$\mathbf{u}_{sj}^l = 2\mathbf{v}_{sj}^{l+1}$$

   where

   $$1 \le j \le N$$

   The motion vector fields at the resolution l are refined as follows:

   $$
   \begin{aligned}
   \mathbf{v}_{sj}^l &= \arg \min_{\mathbf{v} \in \{\mathbf{w}_{j1}^l, \mathbf{w}_{j2}^l, \cdots, \mathbf{w}_{jN}^l\}}\\
   &\quad \left(B_m^l(\mathbf{x}, \mathbf{s}j) \oplus B_{m-1}^l(\mathbf{x} + \mathbf{v}, \mathbf{s}j)\right)
   \end{aligned}
   $$

   where

   $$\mathbf{w}_{ji}^l = \arg \min_{v \in \Omega^l} B_m^l(\mathbf{x} + \mathbf{u}_{si}^l, \mathbf{s}j) \oplus B_{m-1}^l(\mathbf{x} + \mathbf{u}_{si}^l + \mathbf{v}, \mathbf{s}j)$$

   $$1 \le j \le N$$

   $$1 \le i \le N$$

   $$\Omega^l = \{\mathbf{v} : -\mathbf{d}^l \le \mathbf{v} \le \mathbf{d}^l\}$$

3. Let $l = l - 1$ and if $l \neq 1$, go to step 2; otherwise, go step 4, continue.
4. Level 1:

$$\mathbf{u}^1_{sj} = 2\mathbf{v}^2_{sj}$$

where

$$1 \leq j \leq N$$

$$\mathbf{v}^1_{s/2} = \arg \min_{\mathbf{v} \in \{\mathbf{w}^l_1, \mathbf{w}^l_2, \cdots, \mathbf{w}^l_N\}} \left(B^1_m(\mathbf{x}, \mathbf{s/2}) \oplus B^1_{m-1}(\mathbf{x} + \mathbf{v}, \mathbf{s/2})\right)$$

where

$$\mathbf{w}^1_i = \arg \min_{v \in \Omega^1} \left(B^1_m(\mathbf{x} + \mathbf{u}^1_{si}, \mathbf{s/2}) \oplus B^1_{m-1}(\mathbf{x} + \mathbf{u}^1_{si} + \mathbf{v}, \mathbf{s/2})\right)$$

$$1 \leq i \leq N$$

$$\Omega^1 = \{\mathbf{v} : -\mathbf{d}^1 \leq \mathbf{v} \leq \mathbf{d}^1\}$$

5. Termination: Level 0:

$$\mathbf{u}^0 = 2\mathbf{v}^1_{s/2}$$

$$\mathbf{v}^0 = \arg \min_{v \in \Omega^0} B^0_m(\mathbf{x} + \mathbf{u}^0, \mathbf{s}) \oplus B^0_{m-1}(\mathbf{x} + \mathbf{u}^0 + \mathbf{v}, \mathbf{s})$$

where

$$\Omega^0 = \{\mathbf{v} : -\mathbf{d}^0 \leq \mathbf{v} \leq \mathbf{d}^0\}$$

Figure 13 shows a *4-scale* tiling example. A binary pyramid layer can be tiled into 8×8, 8×4, 4×8, 4×4 super-blocks with various shapes, with respect to the scale $N = 4$. There are 3 advantages in the *N-scale* tiling search in the binary pyramid structure: (a) Large super-blocks can be used for more hierarchical layer ME; (b) Natural scenes frequently contain motion at different scales; various shapes can be better fitted with motion objects; (c) Viterbi-like multi-path hierarchical projection search can prevent the ME from being trapped in local minima. The tiling multi-path search is much more important for binary based ME, as illustrated in Figure 14.

Figure 14 shows the fast binary pyramid motion estimation using *4-scale* tiling. Motion estimation is first performed at level 3 using four different tiling block sizes as shown in Figure 13 in the conventional block matching. Each detected motion vector from each scale is propagated to the next lower level and is refined using XOR matching criterion at four different scales. This process repeats once until level 1 is reached. At level 1, the four motion vectors projected from level 2 are refined using XOR matching criterion at the 8×8 tiling block. The best motion vector from these four motion vectors, based on the most "0's" from XOR matching, is projected to level 0. At level 0, the final motion vector is obtained refining the motion vector projected from level 1 using XOR matching criterion for the 16×16 tiling block.

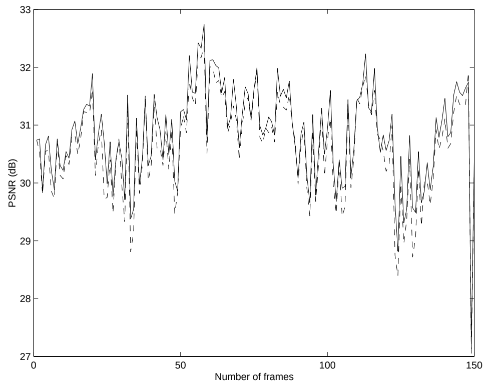

Fig. 15. PSNR versus frame number for an MPEG-2 encoder. Test sequence: Flowergarden. Solid line represents Full Search. Dashed line represents FBPME.

VI. Complexity and Bus Bandwidth

A. *Complexity Of the FBPME*

The width and the height of the image sequence is W and H, respectively. The search window has $\pm M$ pixels. In the conventional full search block matching algorithm, it requires one subtraction, one addition, and one absolute value operation for a single pixel matching. The computational complexity (operations per frame) is approximated as:

$$C_{FULL} \simeq W \times H \times 4 \times M^2 \times 3 = 12WHM^2$$

In FBPME, we use $k = 4$ and $N = 4$. For the FBPME the same full search scheme is used. The tiling block size at level 0 is 16×16. The tiling block size at level 1 is 8×8. The same tiling block sizes of 8×8, 8×4, and 4×8, 4×4 are used at level 2, and level 3. The effective search range at level 0, level 1, and level 2 is set to ± 3 pixels. The effective search range at level 3 is set to $\pm M/8$ pixels.

At level 3, the computational complexity is approximated as

$$C_3 \simeq \frac{WH}{64} \times \frac{M^2}{64} \times 4 \times 3 = \frac{3WHM^2}{1024}$$

At level 2 and level 1, each 8×8 block can be represented by four 16-bit vectors. A 16-bit XOR Boolean operator can be implemented using one 16-bit exclusive-or arrays, a dual-port look-up table (LUT) with 65536 entries. The 8×8 block needs four additions for the 64 pixels for the basic matching. Likewise, the 8×4 block needs two additions for the 32 pixels basic matching. Then, the computational complexity at level 2 and level 1 is approximated as

$$C_2 \simeq \frac{WH}{16} \times \left(\frac{4}{64} + \frac{2}{32} \times 2\right) \times 49 = \frac{147WH}{256}$$

At level 1, The 8×8 block then needs four additions for the 64 pixels for the basic matching. The computational complexity at level 1 is approximated

$$C_1 \simeq \frac{WH}{4} \times \frac{4}{64} \times 49 = \frac{49WH}{64}$$

At level 0, the 16×16 block needs sixteen additions for the 256 pixels for the basic matching. The computational complexity is approximated as

$$C_0 \simeq WH \times \frac{32}{256} \times 49 = \frac{49WH}{16}$$

Therefore, the computational complexity of the FBPME can be approximated as

$$C_{FBPME} = C_3 + C_2 + C_1 + C_0 = \frac{3WHM^2}{1024} + \frac{1127WH}{256}$$

The ratio between C_{FBPME} and C_{FULL} is below.

$$\frac{C_{FBPME}}{C_{FULL}} = \frac{\frac{3WHM^2}{1024} + \frac{1127WH}{256}}{12WHM^2} = \frac{1}{4096} + \frac{1127}{3072M^2} \quad (4)$$

It is seen from Equation (4) that the computational complexity of FBPME is very low compared with the full search.

B. *Bus Bandwidth*

Bus throughput and on-chip memory requirement are often bottlenecks for cost-effective real-time MPEG encoder implementation. This subsection estimates the required data throughput for the proposed FBPME scheme.

The frame rate, the width and height of the image sequence is F_r, W, and H, respectively. The size of the image block is $G \times G$. A picture frame contains $\frac{H}{G}$ pictures slices, and there are $\frac{W}{G}$ blocks in each slice. The search window has $\pm M$ pixels. In a block matching motion estimation, search areas of adjacent blocks overlap quite a bit. This overlapped area data can be stored inside the on-chip memory buffer to reduce external memory bandwidth. We assume an on-chip memory buffer '**D**' whose size equals to the search area, $(G+2M) \times (G+2M)$ bytes. The new loading data size for buffer **D** is $G \times (G+2M)$ bytes when the next block is on the same picture slice. We need to load the complete buffer at the beginning of a slice while processing one picture slice. Thus, the total external memory bandwidth per slice is approximately $((G+2M)^2 + (\frac{W}{G} - 1) \times G \times (2M+G))$ bytes if boundary block cases are neglected. So, the bus bandwidth (bytes/s) of the Full search is approximated as

$$\begin{aligned} &MB_{full} \\ \simeq\ & \tfrac{H}{G}((G+2M)^2 + (\tfrac{W}{G} - 1) \times G \times (2M+G)) \times F_r \end{aligned}$$

A derivation of the memory bandwidth requirement for FBPME is given in the following.

At layer 3, the bus bandwidth (bytes) is approximated as

$$\begin{aligned} &MB_3 \\ \simeq\ & \tfrac{H}{64} \times ((8+M/4)^2 + (\tfrac{W}{64} - 1) \times 8 \times (M/4+8)) \end{aligned}$$

At layer 2, the bus bandwidth (bytes) is approximated as

$$\begin{aligned} MB_2 \simeq \frac{1}{8} \times (&\frac{HW}{16 \times 32} \times (8+6) \times (4+6) \times 2 \\ &+ \tfrac{HW}{32 \times 32} \times (8+6)^2 + \tfrac{HW}{16 \times 16} \times (4+6)^2) \end{aligned} \quad (5)$$

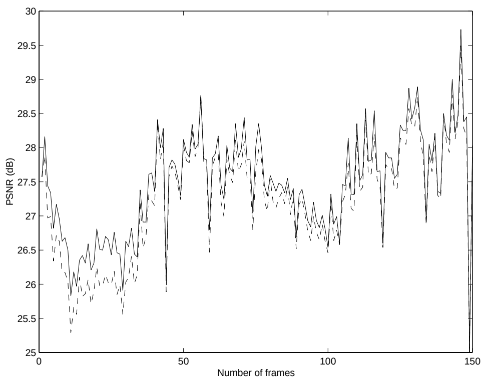

Fig. 16. PSNR versus frame number for an MPEG-2 encoder. Test sequence: Mobi. Solid line represents Full Search. Dashed line represents FBPME.

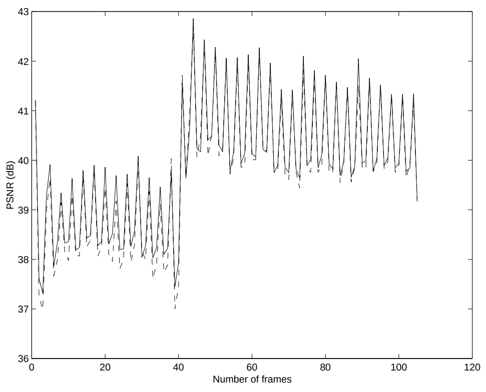

Fig. 17. PSNR versus frame number for an MPEG-2 encoder. Test sequence: Jeep. Solid line represents Full Search. Dashed line represents FBPME.

At layer 1, the bus bandwidth (bytes) is approximated as

$$MB_1 \simeq \tfrac{H}{16} \times \tfrac{W}{16} \times (8+6)^2 \times 4 \times \tfrac{1}{8}$$

At layer 0, the bus bandwidth (bytes) is approximated as

$$MB_0 \simeq \frac{H}{16} \times \frac{W}{16} \times (16+6)^2 \times \frac{1}{8}$$

Therefore, the bus bandwidth (bytes/s) of the FBPME is approximated as

$$MB_{FBPME} \simeq F_r \times (MB_0 + MB_1 + MB_2 + MB_3)$$

Table I lists bus bandwidth requirements in Mbytes/s for the FBPME and Full Search where BS represents Bus Bandwidth. It is showing that the bus bandwidth requirement of the proposed algorithms is much smaller than that of full search.

VII. Experimental Results

The proposed fast binary pyramid motion estimation algorithm was implemented in the MPEG-2 framework. The SDTV MPEG sequences "Flowergarden", " Mobi" of size 720×480 consisting of 150 frames each and the HDTV

TABLE I

Bus bandwidth in Mbytes/s for the two algorithms. $F_r = 30Hz$, $W = 720$, $H = 480$.

Algorithms	BS($M = 128$)	BS($M = 64$)
$FBPME$	8.73	8.21
$FullSearch$	238.93	109.90

sequences "Jeep", "Mask"of size 1920×1080 consisting of 105 frames each were used in this simulation. In our experiment, the size of group of the pictures (GOP) was set to 15. The prediction distance between I frame and P frame was 3. The effective search range was set to ±128. The coding rate was 4Mbps for SDTV and 19Mbps for HDTV. The buffer size for SDTV and HDTV was set to 1.79Mbits and 7.81 Mbits, respectively.

Our motion estimator computes all encoding modes including forward, backward, bi-directional and dual prime but only one mode will be selected. The mode decision for each macro-block is critical for the encoder performance. In our experiment, we use an optimized mode selection that considers the rate-distortion behavior of each coding mode. For example, the bi-directional interpolative mode uses more motion vectors although it typically yields less residual. Thus, our experimental results will give a better representation of the encoding results of an optimized encoder. The comparison between full search and FBPME uses the same optimized mode selection.

The PSNR results using the proposed FBPME and full search for the "Flowergarden","Mobi", "Jeep", and "Mask" are shown in Figures 15-18, respectively. Table II shows the average PSNR of the MPEG-2 with FBPME and full search.

TABLE II

PSNR (dB) of MPEG-2 combined with FBPME method and full search.

Methods	PSNR (dB)			
	Flowergarden	Mobi	Jeep	Mask
FBPME	30.63	27.22	39.74	42.75
Full Search	30.84	27.41	39.83	42.78

One can see from Table II, Figures 15-18 that the proposed FBPME achieves comparable performance with full search but with much less complexity.

VIII. Conclusions

In this paper, we presented a fast binary pyramid motion estimation algorithm that not only significantly reduces the computational complexity, but also greatly reduces the bus bandwidth requirement. The FBPME takes advantages of XOR Boolean matching and *N-scale* tiling multi-path search scheme. It achieves the same level of PSNR and visual quality at less than 1/4000 of the computational load of the conventional full-search algorithm.

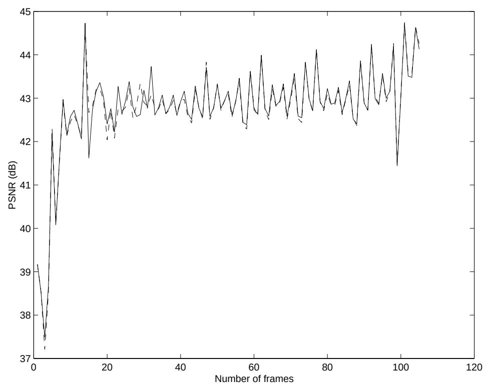

Fig. 18. PSNR versus frame number for an MPEG-2 encoder. Test sequence: Mask. Solid line represents Full Search. Dashed line represents FBPME.

Its binary nature allows it to be implemented very easily in hardware. The proposed scheme was implemented in an MPEG-2 MP@ML and ATSC HDTV encoder framework. Extensive test results have indicated the superior performance of the proposed FBME in speed, memory requirement, as well as data throughput.

Acknowledgments

The authors would like to thank the suggestions of the anonymous reviewers.

References

[1] ITU-T H263, *Recommendation H.263 video coding for low bit rate communication* , January 1998.

[2] ISO/IEC 11172-2, *Information technology - coding of moving pictures and associated audio - for digital storage media at up to about 1.5 Mbit /s - Part 2: Video,* 1990.

[3] ISO/IEC International Standard 13818-2, *Information technology - generic coding of moving pictures and associated audio information - Part 2: Video,* 1994.

[4] MPEG-4 FDIS ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio ISO/IEC 14496-2, March 1999.

[5] ATSC A/53, *ATSC Digital Television Standard, ATSC Specifications, USA,* Sept. 1995.

[6] T. Koga et al, "Motion-compensated interframe coding for video conferencing," in *Proc. N Nat. Telecom. Conf.* , pp. G 5.3.1-G 5.3.5, Nov./Dec. 1981.

[7] J.R. Jain and A.K. Jain, "Displacement measurement and its application in interframe image coding, " *IEEE Trans. Commun.* , COM-29, pp. 1799-1808, Dec. 1981.

[8] R. Srinivasan and K.R. Rao, "Predictive coding based on efficient motion estimation, " *IEEE Trans. Commun.* , COM-33, pp. 1011-1014, Sept. 1985.

[9] K. Chow and M.L. Liou, "Genetic motion search algorithm for video compression," *IEEE Trans. Circuits and Systems for Video Technology*, No. 6, pp. 440-445, Dec. 1993.

[10] C.H. Lin and J.L. Wu, "A lightweight genetic block-matching algorithm for video coding" *IEEE Trans. Circuits and Systems for Video Technology*, No. 4, pp. 386-392, Aug. 1998.

[11] J.Y. Tham, S. Ranganath, M. Ranganath, and A. Kassim, "A novel unrestricted center-biased Diamond search algorithm for block motion estimation ," *IEEE Trans. Circuits and Systems for Video Technology*, No. 4, pp. 369-377, Aug. 1998.

[12] J.S. Kim and R.H. Park, "A fast feature-based block matching algorithm using integral projections," *IEEE Journal on Selected Areas in Comm.*, Vol. 10, pp. 968-971, June 1992.

[13] B. Liu and A. Zaccarin, "New fast algorithm for estimation of block motion vectors," *IEEE Trans. Circuits and Systems for Video Technology*, No. 2, pp. 148-157, April 1993.

[14] M. Bierling, "Displacement estimation by hierarchical block-

matching," *SPIE Visual Comm. Image Process.*, Vol. 100, pp. 942-951, 1988.

[15] Y.-Q. Zhang and S. Zafar, "Motion-Compensated wavelet transform coding for color video compression", *IEEE Trans. Circuits and Systems for Video Technology*, No. 3, 2, pp. 285-296, Sept. 1992.

[16] B.B. Paul and E. Viscito, "Hierarchical motion estimation with 2-scale tilings," *Proc. IEEE International Conference on Image Processing*, pp . 260-264, Nov. 1994.

[17] X. Song, T. Chiang, and Ya-Qin Zhang, "A hierarchical motion estimation algorithm using nonlinear pyramid for MPEG-2," in *Proc. IEEE International Symposium on Circuits and Systems*, pp. 1165-1168, June 1997.

[18] J. Chalidabhongse and C.-C. J. Kuo, "Fast motion vector estimation using multiresolution-spatio-temporal correlations," *IEEE Trans. Circuits and Systems for Video Technology*, No. 3, pp. 477-488, June 1997.

[19] P. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communication*, Vol. COM-31, no 4, April 1983.

[20] H. Gharavi and M. Mills, "Blockmatching motion estimation algorithms - New results," *IEEE Trans. on Circuits and Systems* Vol. 37, No. 5 pp. 649-651, May 1990.

[21] X. Lee, "A fast feature matching algorithm of motion compensation for hierarchical video codec,,"in *SPIE Con. Visual Communication and Image Processing*, Boston, MA, vol. 1818, pp. 1462-1474, Nov. 1992.

[22] X. Lee and Ya-Qin Zhang, "A fast hierarchical motion-compensation scheme for video coding using block feature matching," *IEEE Trans. on Circuits and Systems for Video Technology* Vol. 6, pp. 627-635, Dec. 1996.

[23] B. Natarajan, V. Bhaskaran, and K. Konstantinides, "Low-Complexity block-based motion estimation via one-bit transforms," *IEEE Trans. on Circuits and Systems for Video Technology* Vol. 7, pp. 702-706, Aug. 1997.

[24] X. Song, Ya-Qin Zhang, and T. Chiang "Hierarchical motion estimation algorithm using binary pyramid with 3-scale tilings," in *Proc. of SPIE Visual Communications and Image Processing*, pp. 80-87, January 1998.

[25] X. Song, T. Chiang, and Ya-Qin Zhang "A scalable hierarchical motion estimation algorithm for MPEG-2 ," in *Proc. of IEEE International S ymposium on Circuits and Systems*, June 1998.

[26] A. Zakhor and F. Lari, "Edge-Based 3-D camera motion estimation with application to video coding," *IEEE Trans. on Image Processing* Vol. 2, No. 4, pp. 481-498, Oct. 1993.

[27] J. Feng, K.-T. Lo, H. Mehrpour, and A. E. Karbowiak, "Adaptive block matching motion estimation algorithm using bit-plane matching," in *IEEE Int. conf. Image Processing,* Washington, DC, 1995, pp. 496-499.

Xudong Song (M'96) received the M.S. degree in the Institute of Information Science from Northern Jiaotong University, Beijing, China, in 1988, and the Ph.D degree in electrical engineering from Tampere University of Technology, Tampere, Finland, in 1993.

From 1993 to 1995, he worked as a Research Associate in the Department of Neurosurgery, the University of Illinois at Chicago, in the area of medical imaging. From 1996 to 2000, he was with the Sarnoff Corporation (formerly David Sarnoff Research Center), Princeton, NJ, as a Member of the Technical Staff where he was actively engaged in research and development of Electronic Cinema, H.263, MPEG, real time HDTV encoding, and digital television. He joined the IVAST in 2000 as a Senior Technical Staff where he works on multimedia. In 1999, he received sarnoff technical achievement award. He has six US patents granted or pending. His current research interests include multimedia, video compression over Internet Protocol based networks.

Tihao Chiang (S'90-M'95-SM'99) received the B.S. degree in electrical engineering from the National Taiwan University in 1987, and the M.S. degree in electrical engineering from Columbia University in 1991. He receved his Ph.D. degree in electrical engineering from Columbia University in 1995. In 1995, he joined David Sarnoff Research Center as a Member of Technical Staff. Later, he was promoted as a technology leader and a program manager. While at Sarnoff, he led a team and developed an optimized MPEG-2 software encoder. For his work in the encoder and MPEG-4 areas, he received two Sarnoff achievement awards and three Sarnoff team awards.

Since 1992 he actively participated in ISO's Moving Picture Experts Group (MPEG) digital video coding standardization process. He is currently the co-chairman for encoder optimization on the MPEG-4 committee. He has made more than 40 contributions to the MPEG committee. He has co-authored the rate control technology that was adopted as part of the MPEG-4 International Standards in 1998. His main research interests are compatible/scalable video compression, stereoscopic video coding, and motion estimation. In September 1999, he joined National Chiao-Tung University as an assistant professor in Taiwan, R.O.C. Dr. Chiang is currently a senior member of IEEE and holder of three US patents and more than ten pending patents. He published over 20 technical journal and conference papers in the field of video and signal processing.

Xiaobing Lee received the B.S.E.E from Beijing University of Post and Telecommunications, Beijing, China, in 1981, the M.S.E.E. from Columbia University, NY, in 1984, and Ph.D. degree in Electrical Engineering from Clarkson University, Potsdam, NY, in 1989.

Lee is director of research at SeaChange Systems, Inc. His research interests include digital video streaming and video servers, video processing/compression, cable and satellite Multimedia IP systems. He was with Sarnoff Corporation (formerly David Sarnoff Research Center) Princceton, NJ, in 1998 to 1999. He was principle engineer at Wegener Communications Corp, Atlanta, GA, in 1994 to 1998, and senior research engineer at Tee-Comm Electronics, Canada, in 1993 to 1994. He was research scientist at University of Toronto, Canada, in 1991 to 1993, and research technologist at SouthernWestern Bell Corp., St. Louis, MO, in 1989 to 1990.

His current responsibility is research and develop the integrated MPEG multiplex and QAM modulation / OC3-C line-card devices, Fiber-Channel backplanes and Gigabit networks for video servers and multimedia IP systems.

Ya-Qin Zhang joined Microsoft Research in China in January 1999, leaving his post as the Director of Multimedia Technology Laboratory at Sarnoff Corporation in Princeton, NJ (formerly David Sarnoff Research Center, and RCA Laboratories). His Laboratory is a world leader in MPEG-2/DTV, MPEG4/VLBR, and multimedia information technologies. He was with GTE Corp. in Waltham, MA and Contel Technology Center in Virginia from 1989 to 1994. He has authored and co-authored over 200 refereed papers and 30 US patents granted or pending in digital video, Internet, multimedia, wireless and satellite communications. Many of the technologies he and his team developed have become the basis for start-up ventures, commercial products, and international standards. He serves on the board of directors of five companies.

Dr. Zhang served as the Editor-In-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from July 1997 to July 1999. He is the Chairman of Visual Signal Processing and Communications Technical Committee of IEEE Circuits and Systems Society. He serves on the Editorial boards of seven other professional journals and over a dozen conference committees. He has been an active contributor to the ISO/MPEG and ITU standardization efforts in digital video and multimedia.

Dr. Zhang is a Fellow of IEEE. He received numerous awards, including several industry technical achievement awards and IEEE awards. He was awarded as the Research Engineer of the Year in 1998 in New Jersey. He recently received the national Eta Kappa Nu award as The Outstanding Young Electrical Engineering of 1999 in US.

He received his B.S. and M.S. in Electrical Engineering from the University of Science and Technology of China (USTC) in 1983 and 1985. He received his Ph.D in Electrical Engineering from George Washington University in 1989.