A HIGH-THROUGHPUT MODULAR ARCHITECTURE FOR THREE-STEP SEARCH BLOCK MATCHING MOTION ESTIMATION

Hangu Yeo
Yu Hen Hu
Department of Electrical and Computer Engineering
University of Wisconsin, Madison WI 53706

ABSTRACT
The three-step hierarchical search block matching motion estimation algorithm has played an important role in low bit rate video coding because of low computation complexity compared to the full search block matching motion estimation algorithm (FBMA). In this paper, a modular architecture for the three-step hierarchical search BMA is presented, which features a throughput rate as high as \( \frac{1}{2} \) block per clock cycle and low memory bandwidth with random access on-chip local memory. Furthermore, 100% processor utilization has been achieved by using a method called pipeline interleaving. As such, this architecture offers a feasible solution for the Grand Alliance HDTV picture format with large search range.

1. INTRODUCTION
Low bit rate coding has played a key role in compressing the large amount of image data for video technologies such as multi-media computing, digital publishing, video-on-demand (VOD), and high-definition television (HDTV), etc. Block matching algorithm is the most popular motion estimation algorithm which divides an image frame into macro blocks and matches each block in an intermediate image frame to the most similar block within a search range in a reference image frame [1].

A number of researchers have reported various architectures for FBMA [2]-[4] and three-step hierarchical search BMA [5] with the throughput rate as high as \( \frac{1}{2} \) block per clock cycle for an \( N \times N \) block with the search range of \( \frac{N}{8} \) horizontally and vertically. Nonetheless, these results are still not fast enough to handle real-time high volume video processing such as HDTV.

According to the Grand Alliance HDTV specification [6], the system should support the search range of ±128 pixels horizontally and ±32 pixels vertically for forward predicted (P) frames. This large search range may increase the hardware complexity and decrease the throughput rate of the existing architectures.

In this paper, we propose an architecture for the three-step hierarchical search BMA which features a throughput rate as high as \( \frac{1}{2} \) block per clock cycle with the search range of \( d_m = \frac{N}{16} - 1 \) pixels, simple interconnection, and very low memory bandwidth. First, the five-level nested do loop algorithm has been transformed to three-level nested do loop algorithm by algorithm transformation. With the transformed three-step hierarchical search BMA, the data dependency between the two neighboring candidate block data within the search area has been considered. With the search area data dependency, we could simplify the interconnection between processing elements (PE) and memory module. Furthermore, by pipelining the three stages, high throughput rate has been achieved. Preliminary estimate indicates that this high throughput architecture is capable of handling the previously mentioned progressive-scan HDTV format with large search range with a clock rate as low as 50M-Hz.

2. THREE-STEP HIERARCHICAL SEARCH BMA ALGORITHM TRANSFORMATION
2.1. Three-step hierarchical search BMA

\[
MV1 = MV2 = 0
\]

\[
do \ s = 0 \ to \ \lceil \log_2 d_m \rceil - 1
\]

\[
D_{\min} = \infty
\]

\[
d(s) = \left[ \frac{d_{s+1}}{d_{s+1}} \right]
\]

\[
do \ m = -d(s) \ to \ d(s); \ d(s)
\]

\[
do \ n = -d(s) \ to \ d(s); \ d(s)
\]

\[
MAD(m, n) = 0
\]

\[
do \ i = 0 \ to \ N - 1
\]

\[
do \ j = 0 \ to \ N - 1
\]

\[
MAD(m, n) = MAD(m, n) + |z(i, j) - y(i + m + MV1, j + n + MV2)|
\]

\[
enddo \ j
\]

\[
enddo \ i
\]

If \( D_{\min} > MAD(m, n) \)

\[
D_{\min} = MAD(m, n)
\]

\[
MV1(s + 1) = m
\]

\[
MV2(s + 1) = n
\]

\[
endif
\]

\[
enddo \ n
\]

\[
MV1 = MV1 + MV1(s + 1)
\]

\[
MV2 = MV2 + MV2(s + 1)
\]

\[
enddo \ s
\]

Figure 1. Three-step hierarchical search BMA.
Assuming that an intermediate image frame is divided into \( N \times N \) macro blocks with maximum search range \( d_m \), the number of search steps can be expressed as \( \lceil \log_2 d_m \rceil \). \( \lceil x \rceil \) is the ceiling function.

At step \( s (0 \leq s < \lceil \log_2 d_m \rceil) \), the coarsely spaced search
points around the approximation of the previous step \( s - 1 \) with distance \( d(s) = \left\lceil \frac{d_m}{2} \right\rceil \) are evaluated with the mean absolute difference (MAD) matching criterion.

The matching criterion of the three-step hierarchical search BMA can be expressed as:

\[
MAD(s, m(s), n(s)) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |x(i,j) - y(i + m(s)) + \sum_{k=0}^{s-1} MV1(k), j + n(s) + \sum_{k=0}^{s-1} MV2(k)|
\]

\[
m(s), n(s) = \left\lceil \frac{d_m}{2} \right\rceil, 0, \left\lceil \frac{d_m}{2} \right\rceil \text{ and } 0 \leq s \leq [\log_2 d_m] - 1
\]

\[
MV1(0) = MV2(0) = 0, \text{ and } (MV1(s), MV2(s)) \text{ is the displacement vector of step } s - 1. \text{ } x(i,j) \text{ is the luminance value of the intermediate frame and } y(i,j) \text{ is the luminance value of the reference frame. The final motion vector (MV) is}
\]

\[
MV = \sum_{s=0}^{[\log_2 d_m]-1} \arg\min\{MAD(s, m(s), n(s))\}
\]

\[
= \sum_{s=0}^{[\log_2 d_m]-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |x(i,j) - y(i + m(s)) + \sum_{k=0}^{s-1} MV1(k), j + n(s) + \sum_{k=0}^{s-1} MV2(k)|
\]

The three-step hierarchical search BMA is expressed in 5-nested do loops as depicted in figure 1, where \( s \) is search step index, \( m \) and \( n \) are search vector indices, and \( i \) and \( j \) are pixel coordinate indices.

### 2.2. Algorithm transformation

\[
MV1 = MV2 = 0
\]

\[
do s = 0 \text{ to } [\log_2 d_m] - 1
\]

\[
D_{min} = \infty
\]

\[
d(s) = \left\lceil \frac{d_m}{2} \right\rceil
\]

\[
do l = 0 \text{ to } 8
\]

\[
MAD(l) = 0
\]

\[
do k = 0 \text{ to } N^2 - 1
\]

\[
MAD(l) = MAD(l) + |x(l,k) - y(s,l,k)|
\]

\[
\text{enddo } k
\]

\[
\text{if } D_{min} > MAD(l)
\]

\[
D_{min} = MAD(l)
\]

\[
MV(s+1) = l
\]

\[
\text{endif}
\]

\[
\text{endo } l
\]

\[
MV1 = MV1 + (MV(s+1) - 1)d(s)
\]

\[
MV2 = MV2 + (MV(s+1) \mod 3 - 1)d(s)
\]

\[
\text{endo } s
\]

**Figure 2.** Three-level nested do loops after algorithm transformation.

Assuming that the reference block data are loaded in block scan order, and the search is performed sequentially in a column first order for the nine search points of each step, we can combine each pair of indices \( m(s), n(s) \) and \( i, j \) in the original formulation into a composite index; \( l \) and \( k \), respectively. \( x \) is the original function.

\[
\left\{ \begin{array}{l}
\frac{3m(s)}{d_m} + 1 + \frac{3k}{d_m} + 1 & 0 \leq l \leq 8 \\
0 & l = 9 \\
\end{array} \right.
\]

\[
n(s) = (l \mod 3 - 1)d(s)
\]

\[
\left\{ \begin{array}{l}
k = iN + j & 0 \leq k \leq N^2 - 1 \\
j = k \mod N
\end{array} \right.
\]

The three-level nested do loops after algorithm transformation is depicted in figure 2.

The dependency of search area data \( y_s(s,l,k) \) can be summarized in the following lemma.

**Lemma 1** Assuming that \( d(s) \leq N \), the dependency of \( y_s(s,l,k) \) can be expressed as:

\[
y_s(s,l,k) = y_s(s,l - 3c', k + c'd(s))N' \quad (I-a)
\]

if \( 0 \leq k < N(N - c'd(s)) \) and \( 3c' \leq l < 3(c' + 1) \).

\[
y_s(s,l,k) = y_s(s,l - c', k + d(s)c') \quad (I-b)
\]

if \( 0 \leq k \mod N < N - c'd(s) \) and \( l \mod 3 \neq 0 \), where \( c \) and \( c' \) are integers satisfying \( 1 \leq c' \leq c < \min(3, N) \).

### 3. PARALLEL ARCHITECTURE FOR THE THREE-STEP SEARCH HIERARCHICAL BMA

#### 3.1. Parallel architecture implementation

**Definition 1** (Blockwise Hankel Matrix) Let \( Y \) be a \( k^2 \times n^2 \) matrix as:

\[
Y = \begin{bmatrix}
Y_{0,0} & Y_{0,1} & \cdots & Y_{0,n-1} \\
Y_{1,0} & Y_{1,1} & \cdots & Y_{1,n-1} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{k-1,0} & Y_{k-1,1} & \cdots & Y_{k-1,n-1}
\end{bmatrix}
\]

Matrix \( Y \) is called a Blockwise Hankel Matrix of Type I if \( Y_{i,j} = Y_{i+1,j+1} \) for \( 1 \leq i < \min(n, k) \). Matrix \( Y \) is called a Blockwise Hankel Matrix of Type II if \( Y_{i,j} 's \) are Hankel Matrix of order \( k \times n \) and \( Y_{i,j} = Y_{i-n,j+n} \) for \( 1 \leq i < \min(n, k) \).
The matching criterion in section 2.1 is expressed as:
\[ MAD_s = Y_s \odot X. \]
\( MAD_s \) is a vector of nine components, \( Y_s \) is a \( 9 \times N^2 \) matrix, and \( X \) is a vector of \( N^2 \) components.

**Corollary 1** Let \( \alpha \in \mathbb{Z}^+ \) with \( 1 \leq \alpha < \min(N, 3) \), then the search area matrix \( Y \) of order \( 9 \times N^2 \) has the following properties by Lemma 1:

\[ Y_{i,j} = Y_{i+\alpha,j+\alpha} \quad (\text{II-a}) \]
\[ y_{i,j} = y_{i+\alpha,j+\alpha} \quad (\text{II-b}) \]

for all \( \alpha \leq i, m \leq 2 \) and \( 0 \leq j, n < N - d(s) \alpha \).

\[
Y_s = \begin{bmatrix}
0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0
\end{bmatrix}
\]

\[
Y_{i,j} = \begin{bmatrix}
Y_{0,0} & \cdots & Y_{0,N^2-1} \\
Y_{1,0} & \cdots & Y_{1,N^2-1} \\
\vdots & \ddots & \vdots \\
Y_{N-1,0} & \cdots & Y_{N-1,N^2-1}
\end{bmatrix}
\]

**Figure 5.** Parallel architecture for Type II block-wise hankel matrix.

structure, and \( \frac{N^2}{4} \) number of search area data are loaded sequentially through each input port.

Figure 5 depicts a parallel architecture for Type II block-wise hankel matrix. Each module of nine processors handles each subregion of \( \frac{N^2}{16} \) pels depicted in figure 3 (b)-(c), i.e., the \( \frac{N^2}{16} \) number of reference block data are loaded sequentially through each input port (input port I-1 - IV-4). Furthermore, this architecture has \( 3 \times 12 = 36 \) input ports for the search area input data because of the Type II block-wise hankel matrix structure, and \( \frac{N^2}{4} \) number of search area data are loaded sequentially through each input port.

Type I and Type II architectures evaluate nine search points in parallel and generate motion vectors for each step every \( \frac{N^2}{4} \) and \( \frac{N^2}{16} \) clock cycles, respectively. High throughput rate can be achieved by pipelining the three stages. As such, the throughput rate as high as \( \frac{1}{4} \) block per clock cycle can be achieved as depicted in figure 6.

**3.2. Pipeline Interleaving for Larger Search Range**

According to the Grand Alliance HDTV specification, the architecture should support the search range of \( \pm N \) pixels horizontally and \( \pm 2N \) pixels vertically for the forward predicted frames with block size \( N = 16 \).

The problem of large search range can be solved by partitioning the search points into four subregions of \( (4N)^2 \) search points and interleaving the execution of four subregions to remove the delays caused by idle clock cycles. Each subregion of \( (4N)^2 \) search points can be handled by repeating the three-step search four times.

Figure 7 depicts the data flow diagram when the search range is \(-12N \) to \( 12N - 1 \) pixels horizontally and \(-2N \) to \( 2N - 1 \) pixels vertically, and shows that the throughput rate of the Type I and Type II architectures are \( \frac{1}{32} \) and \( \frac{1}{64} \) block/clock cycle, respectively when \( N = 16 \). Therefore, Type II architecture can support the Grand Alliance HDTV.
Table 1. Grand Alliance HDTV picture format: Block size: $16 \times 16$, Search Range: $d_m = 7$.

<table>
<thead>
<tr>
<th>Search Method</th>
<th>Architecture</th>
<th># PE</th>
<th>Throughput rate (block/clock cycle)</th>
<th># Data access/block</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBMA</td>
<td>1-D array [3]</td>
<td>16</td>
<td>12,288</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-D array [4]</td>
<td></td>
<td></td>
<td>768</td>
</tr>
<tr>
<td>3-SHS</td>
<td>Jong et al [5]</td>
<td>9</td>
<td>1,280</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(pipelined)</td>
<td></td>
<td></td>
<td>1,280</td>
</tr>
<tr>
<td></td>
<td>Type I (pipelined)</td>
<td>108</td>
<td>1,280</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Type II (pipelined)</td>
<td>402</td>
<td>1,280</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. Time diagram of the pipelined three-stage architecture.

4. COMPARISON

The comparison of our proposed architecture with the other existing architectures for the two different search methods is presented in Table I. Grand Alliance HDTV picture format with search range of $d_m = 7$ has been considered. By pipelining the three stages, our proposed Type I and Type II architectures could achieve the throughput rate of $\frac{5}{12}$ and $\frac{15}{12}$ block/clock cycle, respectively when the block size is $N \times N$ and search range $d_m = \frac{N}{2} - 1$.

5. CONCLUSION

In this paper, a high-throughput modular architecture for the three-step hierarchical search BMA has been described. By pipelining the three stages and utilizing the search area data dependency, highest throughput rate and simple interconnection between motion estimator and memory module have been achieved with 100% processor utilization.

It has been shown that only the proposed architecture is an optimal one for the Grand Alliance HDTV picture format with large search range by pipeline interleaving.

REFERENCES


[^1]: Search range of $-8$ to $7$ has been considered for 1-D and 2-D arrays.


