EFFICIENT MOTION ESTIMATION FOR BLOCK BASED VIDEO COMPRESSION

K. Panusopone† and K. R. Rao‡

Electrical Engineering Department
University of Texas at Arlington
Box 19016, Arlington, TX, 76019
Phone: (817) 272-3478 Fax: (817) 272-2253
e-mail: †krit@dip.uta.edu ‡rao@ee.uta.edu

ABSTRACT

This paper describes a new motion estimation algorithm for block based video compression. Unlike other fast algorithms, the proposed method works efficiently with the adaptive size of data depending on the local information. Motion vector estimated by this mechanism provides a high accuracy close to that of a full search. However, fewer search points are employed in the process leading to a lower complexity. Although the number of searches is greater than that of hierarchical or other fast algorithms, it is less vulnerable to the local minima and also yields an efficient tool for the existing compression standards since it only computes on a fixed block size. The simulation results show the closeness in performance between the new method and the full search BMA.

1. INTRODUCTION

Motion estimation (ME) is the most popular method to reduce/remove temporal redundancy among successive image frames. Block based motion estimation (BME) which performs matching between the block in the present frame and the blocks of adjacent frames in the memory (multiframe and bidirectional matching are possible at the cost of larger frame memory) works well under the assumption that all pels in the block (present) being processed are uniformly displaced with respect to the reference frame. In this situation, block in the reference frame that yields the least cost or highest correlation is normally selected as a prediction of the processed block and one displacement vector (so called motion vector) describing the location of the best matched block is the only information needed by the decoder to predict the processed block. With these properties, BME or specifically block matching algorithms (BMA) are widely employed as interframe image coding methods in a number of video compression standards [1].

It is also known that basic BMA is a computationally intensive task since all possible blocks in the search area have to be matched with the processed block. This basic method finds the best matched block by performing a full search i.e., the processed block is compared with reference blocks within the search window of adjacent frames for every possible displacement. To alleviate the computational load, several fast search techniques have been proposed [1]. The main concepts of these algorithms are to search for a main detail and refine the search in the limited range without getting trapped into a local minima. This paper describes an efficient BMA that performs the matching based on the crucial information in the block so fewer pels are involved in the computation while the quality is comparable to that of a full search. Moreover, this technique is fully compatible with the existing standards and the results of this technique as a part of H.263 [2] will be provided.

2. EFFICIENT BMA

Various simplified search alternatives for BME such as 3 step search [3], logarithmic search [4] spend less time to find the best match. Unlike the full search, these schemes rely on the results of the previous search and concentrate on the related area to reduce the number of search points. Unfortunately, motion field is not truly uniform and hence it sometimes fails to track the best matched block because of the local minima. This problem can be partially solved by using multiresolution technique. Although both of them perform a multiple search to find the best match, multiresolution or hierarchical search [5] utilizes a full search over a varying scale. With the lowest resolution scale at the beginning
stage, multiresolution technique needs a lower searching point to go through the entire search area. The precise estimate can be obtained as a refinement of the previous search by using a higher resolution.

Multiresolution approach has its own problem, i.e., aliasing error. When a region is smooth, the lower resolution representation is a pretty good estimate of that region. Nonetheless, the lower resolution of a detail region which contains a significant amount of high frequency spectrum cannot retain the fine information adequately. Under these circumstances, the lower resolution search means nothing to the actual result and multiresolution approach is not able to find the best match for the original resolution input block. The correct solution to this problem is to find a representation that can reproduce an acceptable quality of the entire region. The criterion for search is to find a block that is the best match with the given representation. Since data in an image is mostly redundant, the representation can be formed efficiently with a few pixels. Therefore, this mechanism can be a reliable search with fewer computations.

3. IMPLEMENTATION

The proposed algorithm attempts to search at a proper resolution. With this concept, we can find reliable motion under few searches because the best match with the distinct detail is guaranteed. For a complex detail region, a basic full search is necessary to avoid a local minima arising from any possible lower resolution blocks. Thus, the proposed scheme is a compromise (adaptive) between the basic full search when high detail spreads over the entire picture and the multiresolution search when picture contains only smooth detail. Since high or smooth detail has no obvious transition point, designer has ability to adjust the reliability level to suit with the complexity cost. The proposed algorithm analyzes a fixed size input to find the appropriate resolution which contains enough description for restoring the input. The sparser sampling grid is successively used to approximate input data and the lowest resolution that can sample the input with negligible loss is exploited.

Block diagram of the efficient BMA is shown in Fig. 1. It can produce one motion vector per macroblock similar to H.263 without options [2]. This structure can be easily adjusted to comply with other modes of H.263 [2] or other compression standards [1]. The system works on macroblock by macroblock basis. Sampling grids of size 8x8, 4x4, 2x2 and 1x1 are applied respectively to all 16x16 luminance data. For simplicity, range of intensity is used to check whether the lower resolution representation can retain the integrity of the input or not. Only the lower right pixel is used for matching purpose if the test (integrity) is satisfied. The test procedure can be formulated as

\[ x = X \quad \text{if} \quad \max \{|X - x| < \text{th} \quad \text{for} \quad x \in \hat{X} \} \]

where \( \hat{X} \) is the interpolation of \( X \) onto the same resolution as in \( x \), \( \text{th} \) is the range threshold and \( X \) is the sample at the lower resolution.

Table 1. Relationship between samples in the original data and its lower scale (position 0 is at the top or leftmost position).

<table>
<thead>
<tr>
<th>scale</th>
<th>pixel position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/16</td>
<td>15</td>
</tr>
<tr>
<td>1/8</td>
<td>7, 15</td>
</tr>
<tr>
<td>1/4</td>
<td>3, 7, 11, 15</td>
</tr>
<tr>
<td>1/2</td>
<td>1, 3, 5, 7, 9, 11, 13, 15</td>
</tr>
<tr>
<td>1/1</td>
<td>0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15</td>
</tr>
</tbody>
</table>

Similar concept is also applied in the search area. It is assumed that search dimension is within \( \pm 16 \) pixels which covers exactly 3x3 blocks of 16x16 pixels (Fig. 2 with \( W=48 \) and \( N=16 \)). To increase representation efficiency in the search area, each block is subsampled independently and the whole search area is irregularly sampled. The processed input and search area are exploited in the matching process. Interpolation is required in some areas to avoid aliasing when
the resolution of input and output are not matched. This estimation is summarized as follows:

1. The input block is subsampled based on its local redundancy. Only pixels at \((p, q)\) which are specified in Table 1 are used in the lower resolution.

2. The search area is split into \(3 \times 3\) regions and each of them is applied to the same subsampling process as in step 1 (assume that each region is in the same size as in the input block).

3. Full search BMA is performed on the subsampled data and motion vector is obtained from the displacement that gives the lowest weighted MAD (mean absolute distortion).

4. COMPLEXITY

The computational complexity is obtained in terms of the number of testing operations before subsampling, the number of search points, the number of operations used to find MAD of each search point and the conversion to weighted MAD in the reference resolution. Testing operations before subsampling are negligible since each \(16 \times 16\) pixels block is subsampled only once. In the worst case scenario, no subsampling is possible in both search area and input data. The complexity in this case is about the same as complexity of full search scheme or

\[
C_{fs} = \frac{PO}{N^2}(2W + 1)^2(2N^2 - 1)
\]

where \(Q\), \(P\) is the size of image frame, \(N\) is the size of square input block and \(W\) is the size of square search area as shown in Fig.2.

The tighter bound is determined by assuming that all data are interpolated onto the highest resolution appeared in the search area. The higher subsampling ratio results in the lower \(N\) and \(W\) thereby reducing complexity of the algorithm. The most efficient way and alias-free is complicate. Either input or searching data are interpolated to match the resolution of its counterpart. For example, \(1/16\) scale input are upsampled by factor of 4 in both directions when searching in \(1/4\) scale search area. To make it even more difficult, search points are placed irregularly in the search area and some portions of the search area may have to be interpolated. Fig.3 is an example of irregular samples in the search area. If the input data is assumed to be in the lower resolution than that of the search area, configuration of all necessary search points is shown in Fig.4.

5. SIMULATION RESULTS

The proposed method is simulated based on H.263 standard [6]. Input block is fixed at \(16 \times 16\) pixels and the search area is within \(\pm 16\) pixels from block boundary. MAD criterion is used to find the best motion vector at half pixel accuracy resolution. The first 30 frames in QCIF format from “Miss America” sequence are used to examine the proposed system performance. In this simulation, the performance is measured in terms of the average PSNR which is the average value of PSNR of 3 components (Y, Cb, Cr). Fig.5 shows the difference in average PSNR of the motion compensated frame between full search and the proposed method. Performance of H.263 incorporating with full search BMA and the proposed estimation at 64 Kbps without the options is shown in Fig.6. Due to the space limitation, the upper bound on the operations required for the proposed algorithm is not shown in this paper.

Interestingly, Fig.5 shows that motion compensated
frame from both BMA and the proposed method have a similar quality. The discrepancy in chrominance images show more fluctuation because motion estimation relies data entirely on luminance image and the scaled motion vectors are used directly in chrominance images. Threshold shows a big impact in the complete system. Performance of H.263 at th=10 is within 1 dB from the full search. The gap with the full search, however, extends with increasing threshold. More than 3 dB can occur at very high thresholds. This result is predictable since it can hardly maintain much information at this threshold. This situation is similar to coding at a low resolution only. Improvements to the proposed method are possible. In this paper, only simple testing mechanism for subsampling is used. Higher performance of the system is expected with better representation. Moreover, all interpolation for matching is done with the zero order hold and a more sophisticated technique is also applicable for the proposed scheme.

6. CONCLUSIONS

A new motion estimation algorithm is presented in this paper. This technique provides an efficient way to perform a block based motion estimation suitable for block based video compression. It needs fewer operations compared to the full search while keeping the error minimal. Higher performance of this scheme is possible at the cost of the increasing complexity. The most important factor for this technique is to control the threshold effectively. We want to keep threshold high to gain lower complexity. On the other hand, we do not want to loose important information and get a wrong motion vector. This paper shows the performance of the proposed system and its application to H.263 standard. It is seen that more efficient motion estimation than a conventional full search BMA is realizable with block based video compression.

7. ACKNOWLEDGEMENT

Authors are thankful to Telenor R&D who made the software for H.263 standard available in the public domain.

8. REFERENCES


