1997 IEEE International Symposium on Circuits and Systems, June 9-12, 1997, Hong Kong

EFFICIENT HIERARCHICAL MOTION ESTIMATION ALGORITHM BASED ON VISUAL PATTERN BLOCK SEGMENTATION

Mei-Juan Chen, Liang-Gee Chen, Ro-Min Weng and Yung-Pin Lee
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan, R.O.C China

ABSTRACT

A new hierarchical block-matching motion estimation algorithm with the segmentation of variable block size and visual pattern is presented. The block partition with visual patterns which exploiting the properties of the human visual systems(HVS) is employed to get a more precise motion estimation for the detailed regions with complex motion in an image. The hierarchical search expedites the motion estimation and simplifies the processing of the stationary area. Simple control overhead and low side information are required due to the regular decomposition of visual pattern structures. The performance of proposed method is superior over the conventional full search block-matching motion estimation while reducing computation complexity drastically. Extensive experimental results are included in this paper.

1. INTRODUCTION

Motion estimation is a powerful technique to eliminate temporal redundancy for video compression. The block-matching approach[1] is popular and has been widely adopted by several video coding standards, such as H.261[2], H.263[3] and MPEG[4]. Because the removal of temporal redundancy between successive image frames relies heavily on the use of a block-matching motion estimation technique, the performance and efficiency of the video coding systems depend on the *accuracy* and the *speed* of the block-matching motion estimation algorithm(BMA).

Block-matching algorithms find the motion vector based on a block-by-block matching. Jain and Jain[1] originally proposed the motion estimation which involves the division of the image into fixed size blocks. The motion vector is the location that has the maximum correlation value between blocks in temporally adjacent frames. Each equal-size block is compared against candidate blocks in a search area on the previous frame to get the best-matched one. The success of BMA lies on the ability of prediction. Among various BMAs, the full search(or exhaustive search) is the most popular one. However, a massive computation effort is usually needed. In a typical hybrid coding system, there is 51% time spent in operation for the motion estimation. For the applications where power consumption and processing speed are critical, it is clear that the fast algorithms are necessary. In the past, various fast search algorithms[6][7] have been proposed to alleviate the computational burden imposed by full search. Basically, the majority of these fast search algorithms reduces the number of searched positions largely based on the assumption that the matching criterion increases monotonically as the search position moves away from the best match position and reduce the computation time at the expense of the accuracy.

For BMA, there is an implicit assumption that the motion within each block is uniform. It is not always valid if the fixed block size is too large compared to real object in an image. Then the block effect due to moving edge will be noticeable and the quality of the predict suffers. However, a decrease in the size of the blocks means the number of motion vectors to be transmitted increase. The large overhead will frustrate coding efficiency. In the past, the quad-tree segmentation has been used in image processing[12]. Historically, [9][10][11] segmented the image into variable block sizes based on bin-tree or quad-tree decompositions before BMA to make a compromise between performance and bit-rates. Quad-tree segmentation is one of popular techniques for variable block size coding. According to [10], the quad-tree is a tree structure representation in which each node, unless it is a leaf, generates four children. Each child occupies a quarter of the area of its parent. The subdivision of a parent node into its four children is guided by a homogeneity test. In this test, a decision is made whether four subblocks of temporal difference are homogeneous. If the test fails, the four subblocks are generated and they are regarded as four new independent parent blocks to be tested further. After the decomposition with quad-tree structure, the block-matching motion estimation is conducted on the variable size block.

Natural video images typically consist of regions with widely varying content and activity. Most motion estimation algorithms deal with each block equally in spite of the different complexity of motion for the various blocks. To use the computing resources efficiently, different computation efforts should be made for blocks containing different amount of motion. Thus the video image can be segmented into regions having widely different degree of motions. Especially for typical videoconferencing sequence, certain regions, like the speaker's eyes and mouth in the head-and-shoulders sequence, are critical to our subjective evaluation of quality, and relatively small errors can perceptually have a major degrading effect on the overall quality. Such regions tend to dominate the viewer's attention and are intrinsically more difficult compensated than the 'background' of the sequence.

Motivated by the above considerations, it is desirable to find a segmentation that allocates less computation efforts and bit-rates to homogeneous regions and more for the area with complex motion. In this paper, we devise a hierarchical BMA in which visual pattern structures are utilized to describe the segmentation of the blocks. The hierarchical procedure simplifies the motion estimation of the blocks with slow motion. For the regions experiencing complicated motions, the algorithm provides precise motion estimation by splitting them with visual pattern blocks which are designed to take advantage of human visual system characteristics for image coding[13]. The visual pattern design is developed using relevant psychophysical and physiological data. The objective of our approach is to obtain a uniform quality over the entire image, especially to improve the large degradation at the area with quick motions.

0-7803-3583-X/97 $10.00 ©1997 IEEE

2. PROPOSED HIERARCHICAL ALGORITHM

The hierarchical processing methodology is being proposed for increasing applications of image processing and video coding[8][11]. The success may be mainly attributed to the similar level of efficiency achieved in each hierarchy. The definition of hierarchy in our proposed algorithm is the block size. To obtain the flexibility of visual pattern block partition while avoiding the excessive overhead or side information needed to characterize more sophisticated image segmentation, we use a top-down(splitting) bit-quads and visual pattern blocks for image decomposition. From the top to the bottom level, the block sizes 32x32, 16x16, 8x8 and visual pattern block(VPB) are chosen. And the image segmentation doesn't involve upsampling or downsampling because the image has single resolution in each hierarchy.

The encoding procedure takes place in three stages. In the first stage, an initial segmentation of 32x32 blocks of the current frame is performed. The procedure starts from the largest possible block(32x32). The first stage consists of determining which bit-quads pattern(as seen in Figure 1) a 32x32 block should be mapped. For the four subblocks(16x16) in each 32x32 block, the average absolute frame difference(FD) is used as a segmentation rule. If this value is greater than some threshold, bit 1 is assigned to indicate the subblock needs further processing to get satisfied quality. And the *half* of average frame difference of this subblock is obtained as the adaptive threshold value of test of next hierarchy. In general, video statistics are not known a *prior*. For this reason, we develop adaptive algorithm that is capable of adjusting the thresholds to adapt to the statistics of the video activity. To avoid unnecessary division of blocks due to small luminance changes or random noise, the adaptive threshold is compared against a minimum threshold(min_thresh). If the adaptive threshold is smaller than min_thresh, min_thresh replaces it. However, if the threshold test is satisfied for the 16x16 block, the operation will be stopped and bit 0 is assigned for the subblock. We can use the bit-quads pattern which is used in binary image processing[5] to describe the condition of the four 16x16 subblocks in a 32x32 block. If bit 0 is corresponding to the subblock, the 16x16 block is treated as 'background' and without the need for the second stage. Otherwise, the bit 1 indicates that the 16x16 block requires the processing of the next hierarchy(stage 2). Since natural images usually contain numerous large approximately static regions, larger blocks are adequate. Four bits for pattern index are needed for each 32x32 block at stage 1. The purpose of the first stage of the coder is to find those stationary area to save further processing efforts and no operation for motion estimation is involved.

For each 16x16 block with bit 1 in the bit-quads pattern decided in stage 1, a full search block matching is performed at stage 2. And a search interval of only -2∼+2 pixels is sufficient. For each of the four 8x8 blocks, the displaced frame difference(DFD) is employed to determine whether the block should be processed further. If the adaptive threshold obtained at stage 1 is satisfied, the operation will be stopped for the 8x8 block. Otherwise, the half of DFD is obtained to be the adaptive threshold of next hierarchy. The min_threshold check is the same as in the first stage. The bit-quads pattern can indicate the decomposition structure of stage 2 as in stage 1.

At stage 3, for each 8x8 block with bit 1 in bit-quads pattern, a full search block matching around the motion vector obtained at stage 2 with -2∼+2 search range is performed. The displaced frame difference(DFD) of each 8x8 block is checked to see whether the threshold test is satisfied or not . If it is not satisfied, a full search block matching with -3∼+3 search range around the motion vector so far obtained is implemented. In this stage, we use the visual pattern blocks as segmentation model(as indicated in Figure 2). The size of the large block is 8x8 and that of the small block is 2x2. The design of visual pattern blocks in a 8x8 block is defined encompassing the orientation. A 8x8 block can be uniform(index 0), partitioned into two regions(index 1∼14) with respective orientation 90°, 45°, 0° and −45° or decomposed into four subblocks(index 15). The regions of different colors in Figure 2 represent different *objects* with homogeneous motion in the block respectively. The algorithm computes DFD of every subblock(2x2) for every displacement in search range. The minimum DFD is defined as the sum of individual minimum DFD of different parts in those visual pattern blocks. Although there are 16 possibilities for segmenting the block into separate objects, the computation complexity is the same as a 8x8 block-matching due to the requirement of memory elements. The implementation of pattern 15 needs more bit-rates for the coding of motion vectors. Thus we have a compromise between performance and bit-rate for pattern 15. The index between 0∼15 as seen in Figure 2 is transmitted by selecting the pattern whose minimum DFD is optimized. The patterns require only a small overhead rate by restricting the shape and possible numbers of the final regions from a predetermined set of options. The residual image is considerably concentrated in the boundary of moving object. The visual patterns are selected using a simple viewing geometry model in conjunction with measured properties of biological vision and suitable to describe the boundary of objects in a small block. The visual pattern decomposition provides an effective and economical solution to the problem of object segmentation in the application of block-matching motion estimation.

The bit-quads and visual patterns are not transmitted but instead are built independently by the receiver and transmitter only a few indexes must be communicated from the receiver to the transmitter to define the shape. The small overhead with the proposed method is due to its structural decomposition property, i.e. partitioning the picture frame into subblocks, whose sizes, shapes and locations are predetermined and thus are not transmitted. For hierarchical search, matching is first performed in the 16x16 block to obtain an initial estimation of the vector field; the computed vector field is then propagated to the next stage, where it is corrected and again propagated to the next stage until the necessary stage is reached. For each stage, the vector field to be transmitted will be composed of small vectors which might be efficiently coded using an MPEG-like Huffman coder. The hierarchical search incorporated with bit-quads and visual patterns reduces the overhead for coding motion vectors.

3. SIMULATION RESULTS

The performance of the proposed algorithm is demonstrated by five various video sequences, which are *Claire*, *Miss America*, *Susie*, *Salesman* and *Table Tennis*. These sequences were chosen for their different motions and characteristics. *Claire* and *Miss America* are typical head-and-shoulders sequences. *Susie* and *Salesman* involve local motions. The aforesaid four sequences have a speaker imposed on a static background. However, *Table Tennis* includes zooming and large displacement. Each sequence has 30 frames/sec, 60 frames and luminance only. The performance of proposed method is evaluated and compared with the conventional full search(block size 16x16), three-step fast search(block size 16x16) and variable block size quad-tree segmentation block matching[10]. For the quad-tree segmentation block-matching scheme, the largest possible block has 32x32 pixels, whereas the smallest one consists of 4x4 pixels. Then the full search is applied. All the schemes utilize mean absolute error(MAE) as matching criterion for simplicity. And the final search ranges of motion estimation are all set to -7~+7 with integer precision.

The methods aforementioned are also compared according to the PSNR, computation complexity, block number, total bit-rate and subjective quality. The PSNR gives an objective measurement for the quality of reconstructed image. The computation complexity, which is important for real-time video coding, is represented by the computation of one 16x16 block-matching. And the total bit-rates consists of two components. One is the entropy of the prediction error image. The other is the bit-rate for coding the motion vectors, which is obtained from the multiplication of the motion vector entropy and total block number. The measurement of the bit-rate of motion vector also includes the overhead for the transmission of pattern indexes used in our strategy or the code addressing quad-tree structure for quad-tree decomposition method.

Table 1 provides the average performance comparison of the proposed hierarchical algorithm with other approaches. The superiority over the previous methods is obvious. Regardless of picture type, for the objective quality(in terms of PSNR), our proposed method outperforms than conventional schemes, even though the full search. A massive computation effort is needed by full search algorithm. Even for the fast-moving sequence(*Table Tennis*), the computation of our method is lower than the three-step search. The performance improvement and computation reduction of our proposed algorithm result from that for those large blocks which exhibit low picutre activity, like backgrounds, are usually quite successfully compensated with only small vectors; however, for the regions with complicated motion, the algorithm divides the blocks to lower hierarchy or to visual pattern blocks to make accurate motion estimation and thus attains improvement on accuracy of reconstructed image and lower bit-rate for error images. Due to the advantage of hierarchical procedure, there is no large amount computation needed for visual pattern blocks and the vector distribution is more smooth and thus the lower bit-rate needed for coding motion vectors. In respect of bit-rate, the proposed algorithm requires very little overhead for coding the bit-quads and visual patterns indexes. With fewer blocks and little overhead, the bit-rate of proposed method is competable to other methods. In [14], the edge information in a block is extracted to guide block-matching. Although the scheme makes use of the edge features, it gets inferior performance than full search.

Figures 3 illustrates the subjective quality comparison of the reconstructed pictures with motion-compensated vectors of proposed algorithm and full search for *Table Tennis* sequence. The figures show that the proposed algorithm applied to visual pattern blocks so accurately compensates the motion that there is little need for coding the error signal. Both uniform areas and many edges and detailed in the image are well compensated. The image quality is clearly better when the proposed algorithm is used. Moreover, the blocking artifact, present with fixed-size block matching, is almost nonexist. In particular, observe the small detailed such as the player's hands, the boundary of the banner and the corner of the table. The quad-tree segmentation using the homogeneity test rule of temporal difference tends to get smaller blocks in a cluster. Although the use of smaller blocks results in higher adaptivity, but the correlation among blocks can't be exploited and therefore limits the accuracy and the compression ratio achieved. However, the use of bit-quads and visual patterns have suitable segmentation to fit the real objects. In particular, observe the partition of visual patterns used at the player's left ear, hands and the ball. From the simulation, our method gives more reliable motion estimation with much lower computation.

4. CONCLUSION

A new hierarchical algorithm for block-matching motion estimation has presented which combines the concepts of variable block size and visual pattern segmentation. To make use of the segmentation of variable block size, different computational efforts may be made for regions having different complexity of motion. Particularly, we utilize the visual patterns of single-image coding to the block decomposition for smaller blocks to get very accurate motion compensation for those regions with complex motions. We develop the adaptive algorithm that is capable of adjusting the thresholds to adapt to the statistics of the video activity. Extensive experiments have shown that the proposed methods reduces computation complexity while providing better performance than conventional block-matching methods.

REFERENCES

[1] J.R.Jain and A.K.Jain, "Displacement measurement and its application in interframe image coding", *IEEE Trans.Commun.*, Vol.COM-29, No.12, pp.1799-1808, December 1981.

[2] "CCITT Standard H.261, Video codec for audio visual services at p×64 kbps", July 1990.

[3] ITU-T Study Group 15, Working Party 15/1, "Draft ITU-T Recommendation H.263 - Video Coding for Low Bit-Rate Communication", July 1995.

[4] "ISO/IEC 13818-2 Generic Coding of Moving Pictures and Associated Audio: Recommendation H.262", November 1993.

[5] W.K. Pratt, *Digital Image Processing*, A Wiley-Interscience Publication, 1991

[6] T.Koya, K.Iinuma, A.Hirano, Y.Iiyima, and T.Ishiguro, "Motion compensated interframe coding for video conferencing", *Proc.NTC81*, pp.G5.3.1-G5.3.5, New Orleans, LA, December 1981.

[7] M.J. Chen, L.G. Chen and T.D. Chiueh, "One-dimensional Motion Estimation Algorithm for Video Coding", *IEEE Trans. CAS for Video Technology*, Vol.4, No.5, pp.504-509, October 1994.

[8] M.Bierling, "Displacement estimation by hierarchical block-matching", *SPIE, Vol.1001 Visual Communications and Image Processing*, pp.942-951, 1988.

[9] M.H.Chan, Y.B.Yu, A.G.Constantinides,"Variable size block matching motion compensation with applications to video coding", *IEE Proceedings*, Vol.137, No.4, August 1990.

[10] V.Seferidis, M.Ghanbari, "Generalised block-matching motion estimation using quad-tree structured spatial decomposition", *IEE Proceedings-Vision Image Signal Processing*, Vol.141, No.6, pp.446–452,December 1994.

[11] X.Xia and Y.Q.Shi, "A thresholding hierarchical block matching algorithm for motion estimation", *ISCAS*, pp.624–627, 1996.

[12] E.Shusterman and M.Feder, "Image compression via improved quadtree decomposition algorithms", *IEEE Trans. Image Processing*, Vol.3, No.2, pp.207–215, March 1994.

[13] D.Chen and A.C.Bovik, "Visual pattern Image Coding", *IEEE Trans. Communications*, Vol.38, No.12, pp.2137–2145, December 1990.

[14] S.Zhong, F.Chin,Y.S.Cheung and D.Kwan, "Hierarchical motion estimation based on visual patterns for video coding", *ICASSP*, pp.2323–2326, 1996.

Algorithms	PSNR (dB)	Match	Block Number	Entropy(bits/pixel) DFD	MV+Over.	Total
Claire						
Proposed	42.95	1216	124	2.15	0.010	2.160
Full Search	42.28	64148	308	2.17	0.006	2.176
Three-Step	42.21	7954	308	2.18	0.006	2.186
Quadtree	42.43	15392	249	2.17	0.012	2.182
Miss America						
Proposed	38.78	4118	218	3.52	0.023	3.543
Full Search	38.52	64148	308	3.50	0.022	3.522
Three-Step	38.15	7931	308	3.56	0.021	3.581
Quadtree	38.28	13739	154	3.56	0.010	3.570
Susie						
Proposed	36.18	6123	325	3.82	0.039	3.859
Full Search	35.65	64148	308	3.84	0.012	3.852
Three-Step	35.14	7902	308	3.90	0.012	3.912
Quadtree	35.45	19093	437	3.87	0.029	3.899
Salesman						
Proposed	35.52	3660	206	3.98	0.024	4.004
Full Search	34.64	64148	308	4.02	0.007	4.027
Three-Step	34.41	7878	308	4.04	0.008	4.048
Quadtree	34.35	14873	222	4.07	0.017	4.087
Table Tennis						
Proposed	25.96	7458	486	5.35	0.069	5.419
Full Search	25.45	64148	308	5.28	0.023	5.303
Three-Step	24.02	7961	308	5.49	0.025	5.515
Quadtree	25.48	22969	640	5.37	0.062	5.432

Table 1. The performance comparison for various algorithms and video sequences

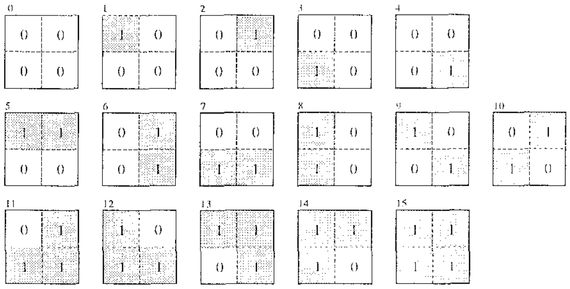


Figure 1. The bit-quads pattern used in stage 1(segmenting a 32x32 block into four 16x16 blocks) and stage 2(segmenting a 16x16 block into four 8x8 blocks)

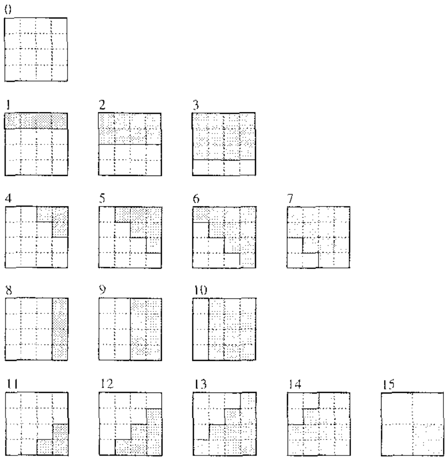
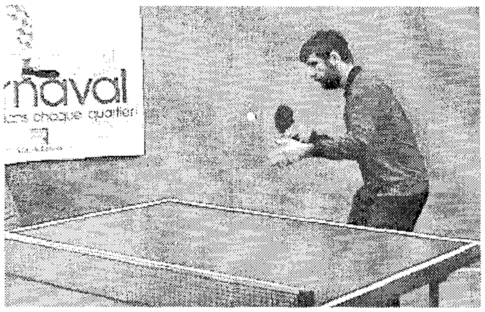
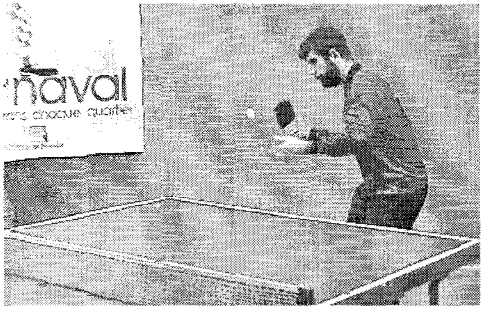


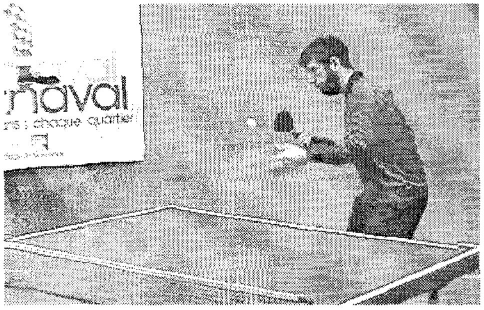
Figure 2. Set of visual pattern blocks used in stage 3(segmenting a 8x8 block into visual pattern blocks)



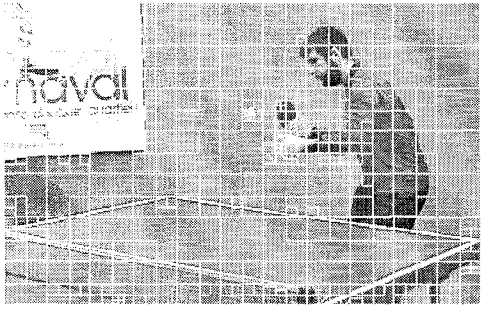
Original Picture



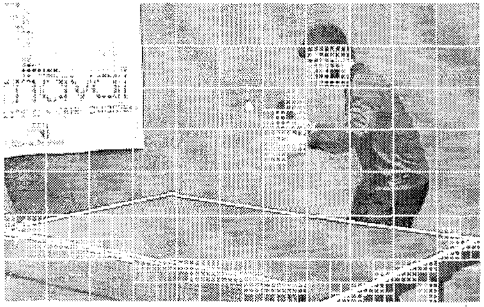
Proposed Algorithm



Full Search



Segmentation of Proposed Algorithm



Quad-tree Segmentation

Figure 3. Subjective comparison of proposed method and other approaches for Table Tennis sequence