A NOVEL AND FAST FEATURE BASED MOTION ESTIMATION ALGORITHM THROUGH EXTRACTION OF THE BACKGROUND AND MOVING OBJECTS

Nelson H.C. Yung and W. H. Mok

Department of Electrical & Electronic Engineering, The University of Hong Kong,
Chow Yei Ching Building, Pokfulam Road, Hong Kong

ABSTRACT

This paper presents a novel and fast Feature Based Motion Estimation algorithm which is developed for typical video-phone scenario. In essence it combines the technique of object extraction with traditional block based motion estimation methods by estimating the background and extracting the moving object continuously in the first stage, then performs a block based motion estimation on the extracted. Simulation of the algorithm with full search as core shows that the estimation time can be reduced by as much as 50%, while the MSE and PSNR remain almost the same as the full search results.

1. INTRODUCTION

In existing video compression standards, perhaps the most complex stage of processing is motion estimation, where temporal correlation inherent in an image sequence is being dealt with. The ultimate performance of motion estimation is to achieve the best temporal prediction at the lowest possible overhead in terms of the compressed bit stream and the computing cost. Over the years, there are many motion estimation algorithms developed, addressing one or more of three aspects, and can be roughly classified in to a number of groups including the optical flow based methods; block based methods, pel-recursive methods and segmentation methods. The most popular is of course the block-based methods as they offer high speed and compatibility with the existing standards.

However, recent applications such as content-based image data access and object based manipulation interactivity have identified new demand for methods that are able to extract or segment the content or objects within an image or sequence, rather than just performing estimation across the sequence without bias as in the block-based approaches. This places a new emphasis on the group of segmentation methods and consideration is being made as to how these segmentation techniques can work in conjunction with the well established block-based methods. Under this context, this paper proposes a feature-based estimation algorithm that performs a background estimation and moving object extraction in the first stage, and then a full search motion estimation in its second stage. As the motion estimation is performed on the moving object only, the proposed method offers a reduction in computing cost without sacrificing accuracy. Our simulation shows that indeed the estimation time can be reduced by as much as 50% when full search is concerned, while the MSE and PSNR remain almost the same as the full search results.

2. FEATURE BASED MOTION ESTIMATION

For a typical video-conferencing or video-phone scenario using a fixed camera on a stationary background, only the moving person contributes to the difference between successive frames. Therefore, tracking the whole or part of the person and carrying motion estimation on it become a natural approach, as seen by many researchers[1-3]. Direct segmentation and region based motion estimation[4-6] seem to be the most popular approach of object tracking but they suffer from many problems such as high computational cost, failure to handle occlusion of objects and complex motion such as the blinking of eyes and movement of the mouth.

For this reason, the proposed Feature-based Motion Estimation (FME) algorithm extracts the background first and then obtains the moving object(s) from the currently extracted background on a block-basis and uses the block-based motion estimation to determine the motion vectors of the moving object only. In essence, the FME algorithm can be divided into: Background estimation (MBE) (Section 2.1), moving object extraction (Section 2.2) and selective motion estimation (Section 2.3).

2.1 MOTION-BASED BACKGROUND ESTIMATION

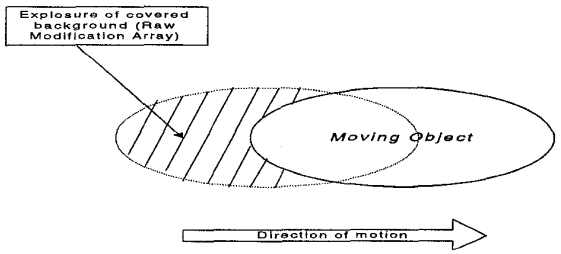


Figure 1 Conceptual diagram of the MBE algorithm

In the Motion-Based Background Estimation (MBE) algorithm, the background in an image sequence is defined as the image regions which are stationary over a period of time. The basic principle of the MBE algorithm is to estimate the background in a number of consecutive frames. This is based on the observation that when an object moves in a certain direction, it shades the background in front of it and uncovers the background behind it, as illustrated in Figure 1. Considering two consecutive images, if

0-7803-4455-3/98/$10.00 © 1998 IEEE

the uncovered background can be detected, it can be cut and paste over the moving object of the first image, reducing the amount of background covered by the object. Extending this concept, the background can be uncovered bit by bit as the object moves.

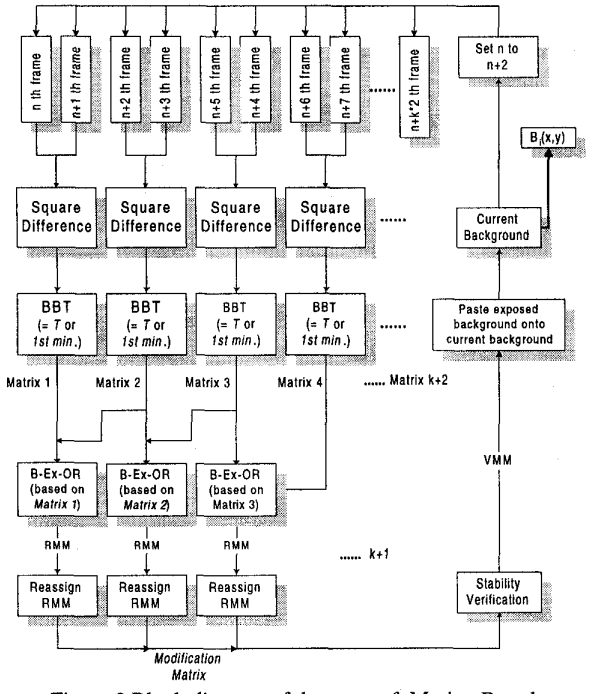


Figure 2 Block diagram of the stage of Motion Based Background Extraction, MBE

To do this, the block diagram of the MBE algorithm is depicted in Figure 2. Consider an image sequence of N frames, the first two frames $I_1(x,y)$ and $I_2(x,y)$ are used for making the initial estimation. They are first divided equally into ${}^{W}/_{S} \times {}^{H}/_{S}$ non-overlapping blocks, where WxH is the image size and S equals to 8 or 16. The square difference, P_k of block K at location (x,y) with size SxS between the first two frames is determined by:

$$P_K = \sum_{i}^{S} \sum_{j}^{S} \left| I_1(x+i, y+j) - I_2(x+i, y+j) \right|^2 \quad \ldots(1)$$

where P_k is then compared with a threshold, T. If $P_k < T$, block K is classified as stationary over the two frames, and assigned a value '0'. Otherwise, it is classified as containing moving objects and assigned a value of '1'. After all the blocks are processed, a matrix of ${}^{W}/_{S} \times {}^{H}/_{S}$ entries is formed with either '0' (stationary) or '1' (moving). This matrix is termed *Matrix 1*. With the same applied to the 3rd and 4th frames, a second matrix, *Matrix 2* is obtained. By comparing *Matrix 1* and *Matrix 2* in a biased-exclusive-OR manner which biased towards *Matrix 1*, a *Raw Modification Matrix (RMM)* is obtained using the set notation:

$$RMM = \{ \delta : \delta = \text{Matrix 1} \oplus \text{Matrix 2}, \ \forall \delta \in \text{Matrix 1} \} \quad \ldots(2)$$

The RMM contains '0' and '1' where a '1' represents an *exposed background.*

When assigning '1' and '0' to the matrix, an important factor here is the threshold T. To select a right value for T, the block square difference (Eqt. 2) of two successive frames is used as a reference. As depicted in Figure 3, the smooth, continuous portion can be interpreted as the variation of pixel intensity of stationary background owing to the imperfection or natural variance of the image capturing device (camera). The impulsive ripples are accounted by the movement of object. Therefore, the 1st local minimum of the profile is chosen as the threshold T for *Matrix 1* and *2*.

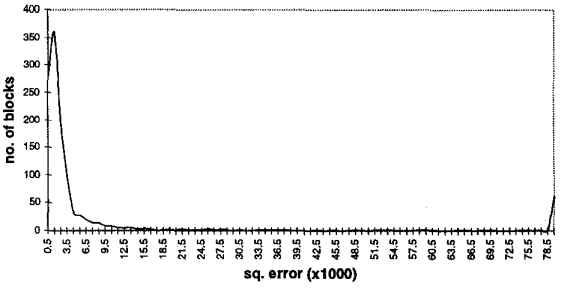


Figure 3 Typical distribution trend of block square difference of 2 successive frames N and N+1

If the color of the moving object is homogeneous, the overlapping portion of the object between two frames may be treated as stationary and incorrectly classified as an exposed background. To relief the problem to some extent, the RMM values are reassigned to generate a *Modification Matrix* (MM). The reassignment is such that if the number of neighbors of an exposed background is small, the background is assigned a '0', and if the no. of neighbors of a '0' is small, it is assigned a '1' instead.

After the MM is determined, a Verified Modification Matrix (VMM) is determined by checking whether the same block has been classified as exposed background before. If it has, then it would be changed to '0'. The VMM contains the exposed background blocks to be pasted onto the current estimated background. The current estimated background $B_O(x,y)$, is given by:

$$B_O = \{ B_O(x,y): B_O(x,y) = I_K(x,y) \mid VMM_{K/2}(x,y) = 1 \ \& \ VMM_{W/2}(x,y) = 0, \quad \forall O \geq K/2, O \geq W/2 > K/2 > 0 \} \quad \ldots(3)$$

2.2 OBJECT EXTRACTION WITH CURRENT BACKGROUND

With a current estimated background $B_o(x,y)$, it is possible to extract the moving object by comparing the current frame with the current background. As depicted in Figure 4, a linear block difference is calculated according to Equation (4).

$$L_K = \sum_{i}^{S} \sum_{j}^{S} \left| I_1(x+i, y+j) - I_2(x+i, y+j) \right| \quad \ldots(4)$$

These values are threshold using the 1st minimum, to generate an MM. From the MM, the output matrix describes the location of the object.

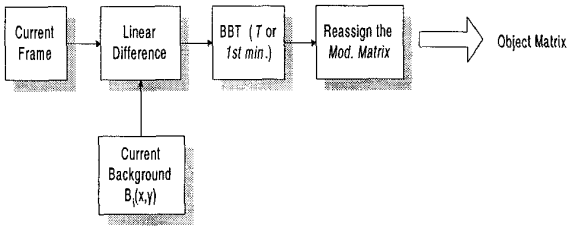


Figure 4 Block diagram of the stage of object extraction

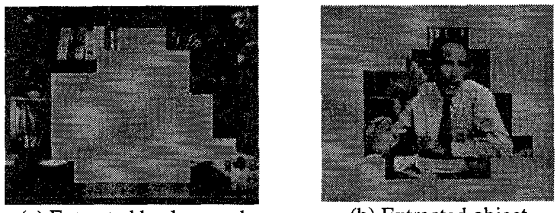
In the FME algorithm, there are two parameters to be considered, the first one is the number of frame pairs, k for looking ahead in the process of *'Stability Verification'* (Fig. 2). The second one is the sampling rate, S_p for selecting the frame pairs for background extraction. If Sp=3, frames 1,4,7,10... are used

2.3 MOTION ESTIMATION ON EXTRACTED OBJECTS

When all the moving objects are extracted from an image, no motion estimation is performed on the stationary parts, and their corresponding motion vectors are all zero. For the moving objects, any block-based estimation methods such as full search, three-step, cross search can be used. In our simulation, full search was chosen for convenience. As block-based method is used, the blocks chosen for estimation usually cover a slightly bigger area than the moving object such that no major parts of the object are missed out.

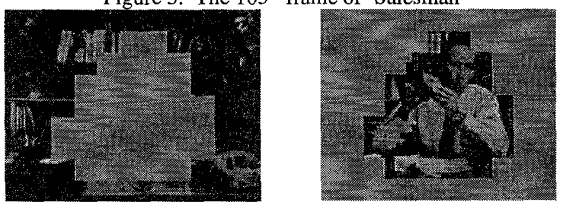
3. SIMULATION RESULT

The performance of the FME algorithm is demonstrated by the image sequence 'Salesman'. This sequence represents a typical video-phone scenario, coded with H.261 and run on SGI R4600 platform. The parameters chosen for the FME algorithm are S=16, Sp=15, K=6 and Threshold =1st min., where full search is used for the motion estimation. The FME algorithm and traditional full search algorithm are compared over 400 frames. Figures 5, 6, 7 show the extracted background and object at the 105th, 180th, 300th frames of 'Salesman' respectively. It can be seen that the object 'Salesman' can be segmented out successfully in most of the cases. Upon close inspection, the extracted object in the 105th frame is perhaps a little less than complete as part of the object held by the man is missing. On the other hand, the 180th and 300th frames are not perfect either, with part of the forehead missing in both frames. This problem is expected to persist because of the problem that the part of the forehead lies in a block that consists of a large portion of background.



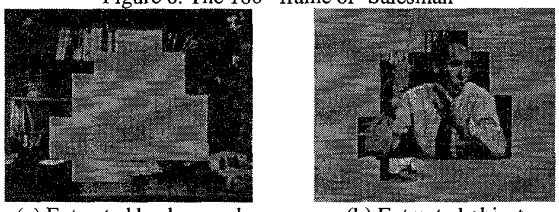
(a) Extracted background (b) Extracted object

Figure 5. The 105th frame of 'Salesman'



(a) Extracted background (b) Extracted object

Figure 6. The 180th frame of 'Salesman'



(a) Extracted background (b) Extracted object

Figure 7.The 300th frame of 'Salesman'

Figure 8 shows the motion estimation time of traditional full search and the FME algorithm. The FME starts to estimate the background from frame 0 and starts to carry out motion estimation on the extracted object after the 135th frame, where the curve *FME(obj)* starts to roll off. The *FME(MBE)* curve represents the background extraction process and its pulse-like behavior is expected since it estimates the background on every 15 frames. Figure 9 shows the reduction in the total motion estimation time of the FME algorithm when compared with full search. The average reduction is about 50% (after 135 frames). However, it should be noted that the FME algorithm also caused a 60% increase in delay at around 105th frame, which is the price to pay in this case.

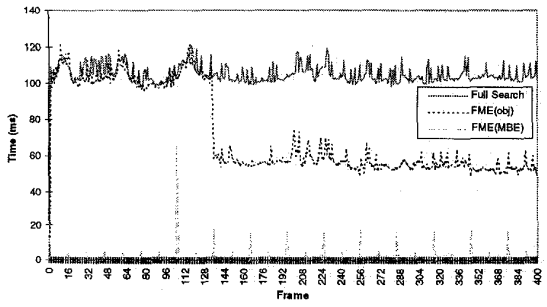


Figure 8 ME time of 'Salesman' running on R4600 (Sp=15, K=6, S=16)

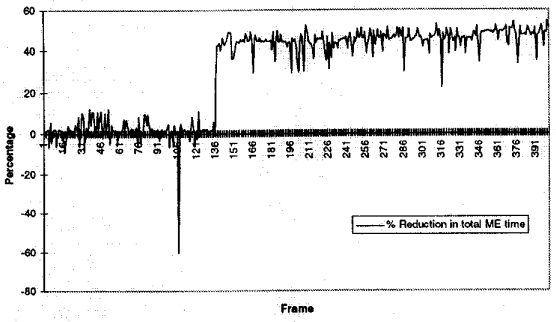


Figure 9 Reduction in total ME time of 'Salesman' when compared with full search(Sp=15, K=6, S=16)

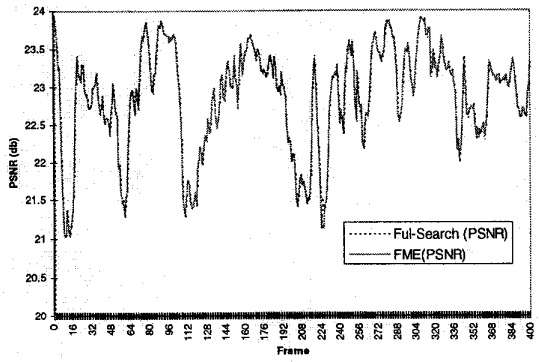


Figure 10 PSNR of the sequence 'Salesman' (Sp=15, K=6, S=16)

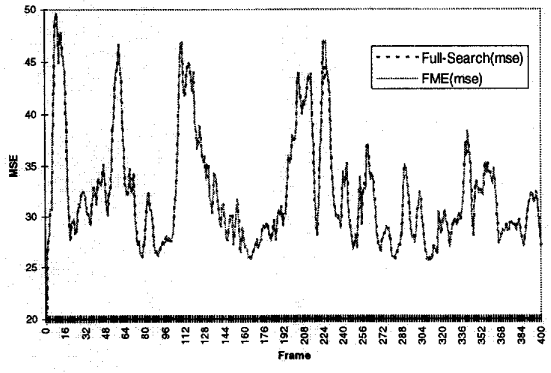


Figure 11 MSE of the sequence 'Salesman' (Sp=15, K=6, S=16)

The PSNR and MSE of 'Salesman' are shown in Figures 10 and 11. The PSNR and MSE obtained by the FME algorithm is very close to that obtained by the traditional full search. Theoretically, if all the moving objects in the picture are extracted accurately, the PSNR or MSE after motion estimation would be the same as the traditional approach. In other words, the saving offered by the FME algorithm lies in the computational reduction in the background blocks with the tradeoff of additional processes in finding the background and extracting the objects.

By running the FME algorithm with full search as core through the other sample sequences, the average reductions in total ME time are depicted in Table 1. Again, the PSNR/ MSE obtained by the FME algorithm is very close to that obtained by the traditional full search.

	Miss America	Claire
FME-FS	34%	40%

Table 1 - Percentage reduction in ME time for the FME algorithms with FS as core

4. CONCLUSION

In conclusion, a novel Feature Based Motion Estimation Algorithm has been proposed and discussed in this paper. The FME algorithm combines the merits of conventional block-based techniques and the knowledge of object/background. In terms of accuracy, which is determined by the motion estimation algorithm used, there is no obvious degradation due to the background estimation or object extraction. In terms of speed, it achieves a 50% reduction in delay. The drawback of this algorithm is the set up time required for the FME to be fully functional (135 frames in our simulation). For real-time video, this is equivalent to about 5 seconds. Therefore, our future research effort will be directed to reducing this to a minimum.

5. REFERENCES

1. Douglas Chai, K.N. Ngan, "Foreground/Background Video Coding Scheme." IEEE International Symposium on Circuits and Systems, vol 2, pp.1448-1451, June, 1997.
2. A. Neri, S. Colonness and G. Russo, "Video Sequence Segmentation for Object-based Coders using Higher Order Statistics." IEEE International Symposium on Circuits and Systems, vol. 2, pp.1245-1248, June, 1997.
3. D. Zhong and S.F. Chang, "Video Object Model and Segmentation for Content-Based Video Indexing." IEEE International Symposium on Circuits and Systems, vol. 2, pp.1492-1495, June, 1997.
4. J.G Choi, S.W Lee and S.D Kim, "Segmentation and motion estimation of moving objects for object-oriented analysis-synthesis coding," Proc. ICASSP95', pp2431-2434, 1995.
5. T. Ebrahimi, H. Chen and B.G. Haskell, "Joint motion estimation and segmentation for very low bitrate video coding," Proc. SPIE VCIP 95', pp787-797, 1995.
6. J Konrad, A.R Mansouri, E Dubois, V.N Dang and J.B Chartier, "On motion modeling and estimation for very low bit rate video coding," Proc. SPIE VCIP 95', pp262-273, 1995.