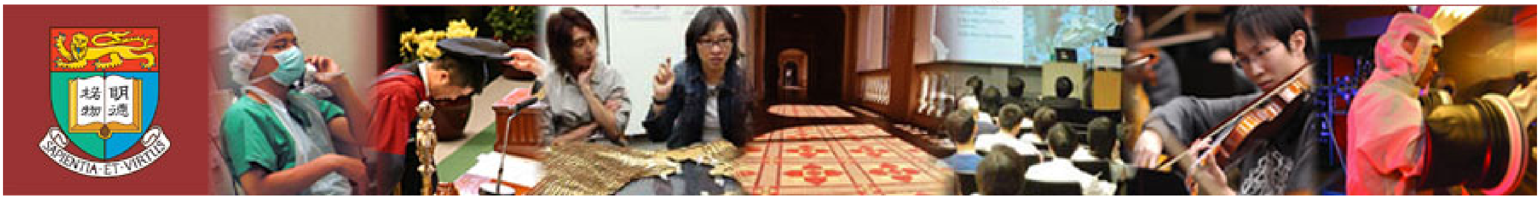

HIERARCHICAL MOTION ESTIMATION BASED ON VISUAL PATTERNS FOR VIDEO CODING

Sheng Zhong, Francis Chin, Y. S. Cheung*† *and Doug Kwan*

Department of Computer Science
†Department of Electrical and Electronic Engineering
University of Hong Kong, Pokfulam Road, Hong Kong
Email: {zhsh, chin}@cs.hku.hk

ABSTRACT

Block matching algorithms(BMAs) are often employed for motion estimation(ME) in video coding. Most conventional fast BMAs treat the ME problem as an optimization problem and suffer heavily from the problem of being trapped at local minima. The full search algorithm(FS), on the other hand, is very time-consuming. Few of them makes use of the information inherent in the images explicitly. We propose a new ME algorithm which can reduce the search range while guaranteeing global optimality in most cases, making use of the edge features. Microblock visual patterns are designed to extract edge information to guide block matching: searching is only carried out at places where the real match most likely happens. The motion field subsampling technique is further employed to get a hierarchical algorithm, which can further double the speed. The proposed algorithms obtain speeds about ten times faster than that of FS with comparable prediction quality.

1. INTRODUCTION

Video compression techniques have advanced a lot during the past several years, symbolized by the establishment of several international standards for coding of video and associated audio information. Those standards, such as MPEG-1, MPEG-2 and H.26x, all propose a block motion vector(MV) architecture. Thus, block matching algorithms(BMAs) are often employed for motion estimation(ME) in video coding. Most of the conventional BMAs treat the ME problem as an optimization problem and employ certain search schemes to find a solution. Except the full/exhaustive search algorithm(FS), fast algorithms such as the three step search(TSS), even though computationally efficient, cannot guarantee optimal solutions, i.e., the search is often trapped at local minima. The ME results are thus usually unsatisfactory. The FS, though being optimal in the sense of the error function such as MAE or MSE, on the other hand, is very time consuming. Few of these methods explicitly make use of the information inherent in the images.

As a matter of fact, when two blocks are matched, some important image features such as the intensity edges/object-contours tend to coincide in the current picture and the reference picture. And the matching of these key visual features should be accurate also because they are the most crucial parts that affect human eye perception. Thus intensity edge information can be extracted to give heuristics to block matching(BM). The search only happens at places where both the current block and reference block present similar edge features. Unnecessary computation can be avoided and the key features are retained by the prediction and very close to those in the original picture. This idea can be illustrated by Fig. 1 of the energy(MAE) distribution(in 1-D case). To get the global optimum, one only needs search the reduced range $D^r = \cup_{i=1}^{4} D_i$ instead of the full range. The reduced search range is represented by the similar edge feature area.

The idea of motion tracing by matching edge features is not unreasonable. Edge features play an important role in human eye scene understanding and object tracking. Many computer methods make use of edge information for these tasks too.

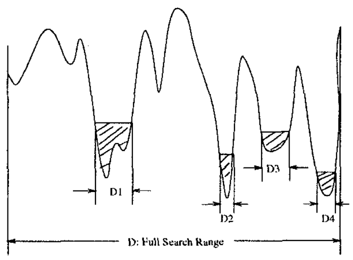

Figure 1: Full search range and reduced search range

2. EDGE-PATTERN-BASED ME

To locate the edge features, we employ the binary matching technique[4]. The edge information is extracted locally in 4x4 blocks(called microblock). Each microblock is binarized and then compared with some patterns with certain edge orientation and location. The microblock edge patterns(MIEPs) are then employed to guide BMA: a mac-

*The author is also with the Department of Information Science and the National Laboratory on Machine Perception, Peking University, Beijing, China.

roblock(MB) (16×16 or 16×8, as defined in MPEG-1 and MPEG-2) is matched against a reference MB only if their corresponding MIEPs do not differ by a lot in their orientations; and the locations of the corresponding MIEPs are employed to decide a narrow band area around the edges in the reference picture for searching.

The next reference MB is not the one with one pixel displacement as in FS, but rather the one with one microblock displacement, i.e., the next reference MB for matching is the one whose upleft corner is at that of next microblock. This microblock-by-microblock-moving searching is the first stage of the proposed edge pattern-based search(EPS) scheme: locate the reference MBs which have similar corresponding edge pattern orientations as the current MB. The second stage of EPS is a fine-tuning process: the current MB is moved according to the edge pattern locations to let the edge slide along the reference edges as illustrated in Fig. 2, and select the best position in the MAE sense as the matched position. In order to reduce the effect of noise, motions other than translation and inaccuracy of edge localization, we also allow searching at nearby pixels along the edges.

For those MBs whose 16 MIEPs are all uniform and where there is no inter-microblock edge in the current picture, a uniform reference MB can be used as the prediction. In the experiments in this paper, we select among several uniformly distributed pixels in the search range the one with smallest MAE as the predicted MB for the uniform MB. For most uniform MBs, this scheme works very well in the sense that MVs very close to the global optima can be found. In the occasional cases when it doesn't work well, we will not try to find a better match for a uniform MB because we think the residue after it is predicted is also uniform and smooth, and can be easily DCT+Quantization+VLC compensated.

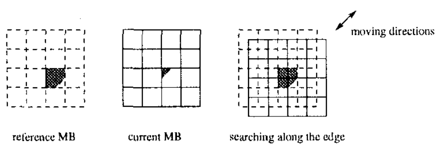

Figure 2: Macroblock matching

Furthermore, the current picture MIEPs are reserved for future usage as reference picture MIEPs. Thus, the MIEPs are computed only once. The EPS scheme is given in Fig. 3.

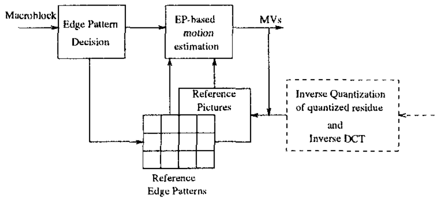
Macroblock
Edge Pattern Decision
EP-based motion estimation
MVs
Reference Pictures
Inverse Quantization of quantized residue and Inverse DCT
Reference Edge Patterns

Figure 3: Diagram for EP-based motion estimation

Fig.4 is an illustration of the EPS searching. To find the MV for the MB in the current picture, we use one of the 16 MIEPs in the MB as the pre-matching index(microblock A in Fig.8). Note that an MB consists of $4 \times 4 = 16$ microblocks and thus 16 MIEPs. We only need search the areas where there are similar corresponding MIEPs in the reference MBs within the search range(the bold-bounded window in Fig.8). Only the microblocks $A1, A2, A3, A4$ and $A5$ have similar MIEPs as A in Fig.8. We only need move the MB to let the edge in microblock A slide along the edges in the microblocks $A1, A2, A3$ and $A4$ to select the best match. And we need not search around the edge in microblock $A5$ even though $A5$ also has similar orientation as A, because when A is moved to match $A5$, the MVs have been out of the search range and would be invalid.

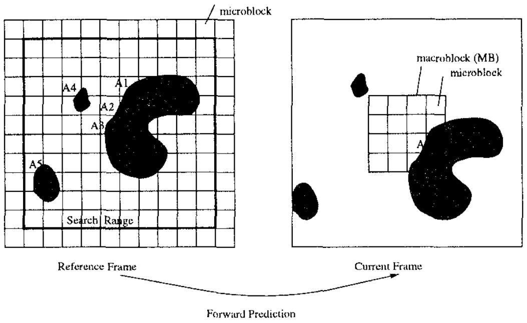

Figure 4: Edge pattern based search

As discussed in section 2 and illustrated in Fig.1, by confining the search to the small band areas around the similar features, we search the neighborhoods of the local minima. This can reduce unnecessary computation greatly and keep the global MAE minimum in most cases. The occasional loss of the global MAE minimum would not hurt the visual quality of predicted pictures, especially at the intensity edges.

The computational complexity of EPS is determined by the number of pixels searched, which can be estimated by the number of pixels along the similar edges in the search range. We find this number is usually small and comparable to (slightly larger that) that of TSS in usual implementing configurations, e.g., search range windows of size about 15×15, 31×31, or even larger.

2.1. MPEG-2 compatible EP-based BMA

The idea of EPS can be applied to produce MPEG-2 compatible motion vectors too. In MPEG-2 ME for interlaced video sequences, where moving objects may occur at very different positions in the two fields which are actually sampled at different time, the adaptive frame/field switch ME scheme is usually adopted to better gear the interlaced picture structure according to the correlations between the two fields. This strategy has been proven efficient to improve the prediction quality of the interlaced sequences. In field ME, 16×8_MVs for the upper and lower parts of the macroblock are also supported.

In our MPEG-2 compatible EP-based BMA, an adaptive frame/field edge pattern based search scheme is employed accordingly. Specifically, MIEPs are estimated for both a frame picture and its two field pictures. When estimating the frame MVs, we utilize the frame edge patterns and frame edge pattern references. When estimating the field MVs, we utilize the field edge patterns and field edge pattern references. 16x8_MVs can be similarly estimated.

3. HIERARCHICAL EPS ALGORITHM

It is known that neighboring blocks in a picture usually share similar motion, especially for those blocks which contain the same object. This kind of redundance can be used to further reduce the computational load for ME in a hierarchical way. There are many ways to realize a hierarchical ME. Multiresolutional method is often used. Here we adopted the motion field subsampling(MFS) technique. Liu[3], *et al* proposed a 2:1 MFS scheme which can result in a 2:1 speed-up for ME with graceful degradation of prediction quality.

We combine the EPS algorithm with the MFS scheme and get a hierarchical EPS(HEPS). By HEPS, first, only the MVs of the half of the MBs(the shaded MBs in Fig. 5) are to be estimated by the EPS method. The MVs of the rest half MBs are then estimated by selecting the best MV among the four MVs of the four neighboring shaded MBs. For example, the MV of the MB labeled by E in Fig. 5 is selected to be the best MV among the MVs of its four neighboring MBs. The computation only needs to search four pixels. Thus a nearly doubled speed is obtained.

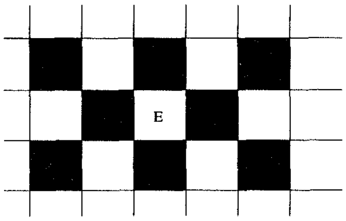

Figure 5: Motion field subsampled ME

If the EPS algorithm can find the true motion or the global optimum MV in most cases, then it can benefit from combining the MFS scheme and produce graceful degraded prediction quality with nearly doubled speed because there are usually a lot of neighboring blocks which share similar motion in most natural video images. In the cases where there are correct MV estimations among the four shaded neighbors and the central MB contains the same object as one of the neighboring MBs with correct MVs, HEPS reduces the possibility of mismatching by using EPS solely. So in most cases, the HEPS method will produce prediction quality comparable to EPS; and in the cases where there is quality degradation, the degradation is graceful. The MFS scheme fails only when block E, e.g., contains a small object(small enough to be contained in E completely) which has different MV from its neighbors.

This HEPS scheme can be used to produce MPEG-2 compatible MVs too.

4. EXPERIMENTAL RESULTS

We have tested the HEPS scheme by both MPEG-1 and MPEG-2 compatible video encoding. The experiments show that the HEPS algorithm obtains an ME speed which is more than 10 times as fast as that of FS with comparable prediction quality which is much better than that of TSS.

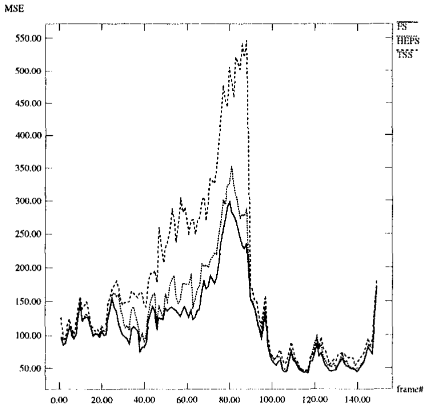

Figure 6: Prediction error comparison

Fig.6 shows the prediction errors(MSE) of HEPS, FS and TSS of the first 150 frames of the CIF tennis sequence. It is coded into bitstreams of the bitrate 1.15Mbits/s. The maximal displacement from the search center is 15. Only forward prediction is adopted to evaluate the prediction performance and thus the encoding only contains I- or P-pictures. Each group of pictures contains 12 frames. The macroblock size is 16 × 16. The motion vectors(MVs) are of half-pixel precision. First, full-pixel precision search is carried out using original pictures as references. Then the half-pixel precision search is done using the reconstructed pictures as references.

The HEPS gets a much better result than TSS in all cases. And in almost all cases, the increment of MSE by HEPS is less than 10% compared to FS. Except a few MBs in each picture, the MVs are quite similar to those found by FS; many are exactly the same.

Fig. 7 shows predicted pictures of 63 of the Tennis sequence. Comparing the tennis net, table border, characters on the wall and the background, the superiority of the prediction quality by HEPS to that by TSS is very obvious. It is very close to that by FS too. On the whole, the HEPS searching scheme seems to be able to tracking the true motion by matching only at similar edge features.

The speed comparisons with FS are presented in Table 1, which lists the numbers of the searched pixels by the two schemes. We find HEPS searched less than one-tenth of the pixels as FS did. The time consumed by the edge pattern extraction, which is less than 1% of the total encoding time, is quite small and negligible. HEPS, being slightly slower than TSS, is much faster than FS with comparable prediction quality.

	FS	EPS
tennis (CIF, MPEG-1)	969	80
football (CIF, MPEG-1)	969	81

Table 1: Number of searched points for each macroblock by different schemes. The number for HEPS is the average value over the whole sequence.

When applied to the MPEG-2 compatible ME, it gives similar results.

5. DISCUSSION AND CONCLUSION

We propose fast ME algorithms using visual features for matching and get ME speeds which can be more than 10 times faster than that of FS with prediction quality comparable to that of FS. This edge pattern-based ME algorithm can be impelmented in parallel and can produce MPEG-1 compatible or MPEG-2 compatible motion vectors as well.

It is worth studying whether other edge extraction methods work better than binary matching. A natural idea is to adopt certain edge detector such as Sobel, Zero-crossing, or etc.. Some key points for this task are (1)it can reduce the effect of noise and produce stable edge information. For example, edge locations should not float by a lot under reasonable intensity variations; (2)edges should be easy for description to guide BMA, i.e., easy to gear the block structure of the BMA,

Finally, the edge-pattern-based search scheme is very close to model-based ME methods in nature. The edge patterns are employed locally as models for matching. This method can be extended to models in a global sense and applied to video coding which also supports content-based data manipulation such as VOD(video on demand). This is a very prospective direction in current video coding research.

References

[1] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: a review and a new contribution," *Proceedings of IEEE*, vol.83, no. 6, pp. 858-875, June 1995.

[2] T. Koga, K. Iinuma, A. Hirano, Y. Iijima and T. Ishguro, "Motion-compensated interframe coding for video conferencing," in *Proc. NTC 81*, pp. C9.6.1-9.6.5, New Orleans, LA, Nov./Dec. 1981.

[3] Bede Liu and Andre Zaccarin, "New fast algorithms for estimation of block motion vectors", *IEEE Trans. CAS for Video Tech.*, Vol. 3, no.2, pp. 148-157, April 1993.

[4] Sheng Zhong and Francis Chin, "Visual pattern-based motion estimation for video coding," to appear (*Proc. IS&T/SPIE Symposium on Electronic Imaging: Science and Technology, Digital Video Compression: Algorithms and Technologies 1996*, San Jose, California, Feb. 1996).

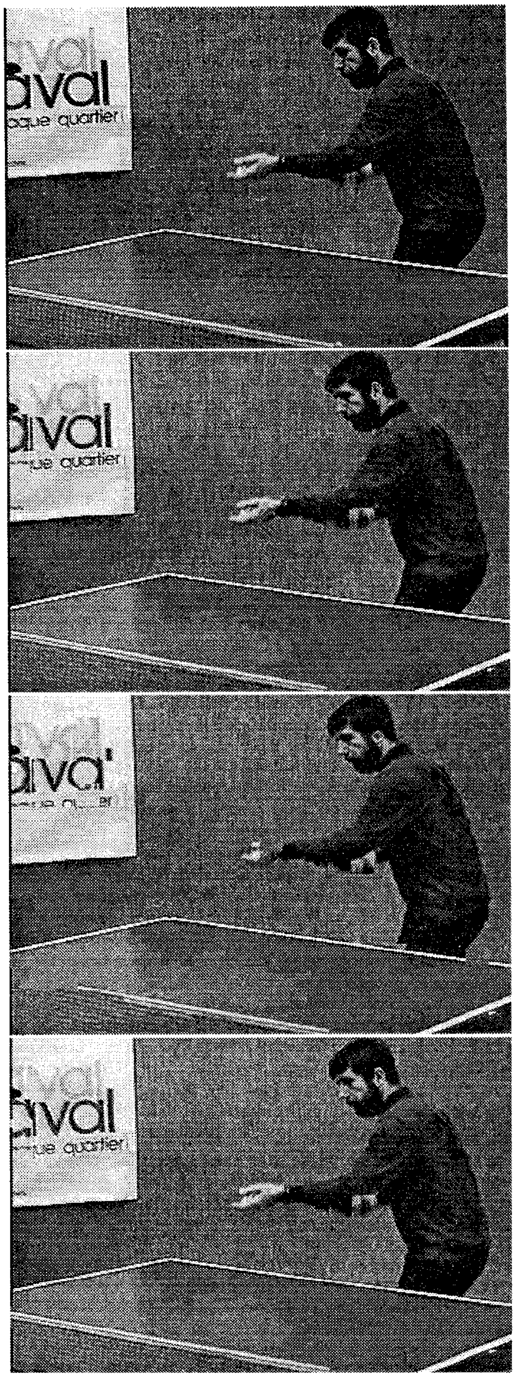

Figure 7: Original(top) and predicted pictures by FS(Second), TSS(Third) and HEPS(fourth)). Tennis (CIF format). MPEG-1 compatible motion vectors were produced.