PREDICTION-BASED DIRECTIONAL FRACTIONAL PIXEL MOTION ESTIMATION FOR H.264 VIDEO CODING

Libo Yang[1*], *Keman Yu*[2], *Jiang Li*[2], *and Shipeng Li*[2]

[1]Department of Information and Electronic Engineering, Zhejiang University, Hangzhou, China
[2]Microsoft Research Asia, 3F Sigma Building, 49 Zhichun Road, Beijing 100080, China

ABSTRACT

In an H.264 video encoder, motion estimation (ME) is the most time-consuming component. The ME process consists of two stages: integer pixel search and fractional pixel search. Since the complexity of integer pixel search has been greatly reduced by numerous fast ME algorithms, the computation overhead required by fractional pixel ME has become relatively significant. To reduce the complexity of fractional pixel ME, we propose a prediction-based directional fractional pixel ME algorithm. We utilize more accurate motion vector predictions and directional search to achieve better computation reduction. We further propose an early termination method to decrease the amount of search. Experimental results show that compared to the full search sub-pel ME and the fast sub-pel ME proposed in H.264, the proposed method can reduce up to 84% and 74% of fractional pixel search points respectively, with a negligible degradation in quality.

1. INTRODUCTION

In general video coding systems, motion estimation (ME) can efficiently eliminate temporal redundancy between adjacent frames. At the same time, ME is also regarded as a vital component in a video coder as it consumes a large amount of computation resources. Especially, in the video coding standard H.264, ME accounts for most of the complexity of the encoder for its seven different block sizes and 1/4-pel accuracy fractional pixel search. Therefore, simplifying the ME process is essential for real-time applications.

In H.264, the ME process is divided into two steps: integer pixel ME and fractional pixel ME. Generally, integer pixel ME takes most of the computational cost of the whole ME. However, with the development of fast integer ME algorithms [1] [2] [3] [4], the computational cost of integer pixel ME has been greatly reduced. For instance, the fast algorithm based on merge and split procedures [4] decreases the average number of integer pixel search points to 6. However the conventional Full Fractional Pixel Search (FFPS) at 1/4-pel accuracy needs to check 16 sub-pel search points. Therefore the computational cost of fractional pixel ME becomes comparable to that of integer pixel ME, and even much higher in some cases. How to speed up the fractional pixel ME tends to be the key issue in simplifying the whole ME process.

To reduce the complexity of fractional pixel ME, several fast algorithms have been proposed. Their common idea is to simplify the search pattern. Based on the assumption that the error surface is monotonic, two search patterns are proposed in [5] to refine the motion vectors (MV) surrounding the point with only a minimal error. Similarly, based on the observation that the cost function is a smooth convex function in the prediction area, a Paraboloid Prediction based Fractional Pixel Search (PPFPS) algorithm [6] examines three 1/2-pel positions surrounding the best integer point and three 1/4-pel positions between the best and sub-optimal 1/2-pel positions only. However, both methods involve the costs of examining the neighboring integer positions surrounding the best integer-pel position. As a result, they can only work with the integer-pel ME algorithms that satisfy the monotonicity requirement. The Center Biased Fractional Pel Search (CBFPS) [3] solves this problem well by checking the predicted position directly and refining the MV surrounding it. However, it suffers from falling into a local minimum in early stages because the predicted MVs are not accurate enough. In addition, some early termination methods in [5] [7] are used to stop the sub-pel motion search process early when the cost is less than a given threshold.

In this paper, we propose a Prediction-based Directional Fractional Pixel Search (PDFPS) algorithm. It differs from the previous methods in three aspects. i) Two kinds of predicted MVs, median prediction and up-layer prediction, are used to predict the sub-pel MV, which make the MV prediction more accurate. ii) Only the predicted points or the points in the predicted direction, rather than all eight neighboring points, are examined, which greatly reduces the number of checking points. iii) Some unnecessary procedures are further eliminated by an early termination method. Moreover, our method does not require the costs of examining the neighboring integer pixel points, which allows it to work with any integer pixel ME algorithm.

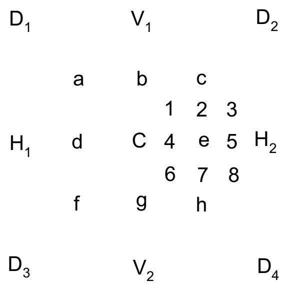
The rest of this paper is organized as follows. Section 2 describes the fractional pixel ME methods proposed in H.264. Section 3 presents our fast fractional pixel ME algorithm. Simulation results are shown in Section 4. Finally, we conclude the paper and give future directions in Section 5.

2. FRACTIONAL PIXEL MOTION ESTIMATION ALGORITHMS PROPOSED IN H.264

Fig.1 shows a typical process of the conventional FFPS method in H.264. First, it examines eight 1/2-pel positions (represented by a ~ h) surrounding the best integer pixel position *C* and then obtains the best 1/2-pel MV, supposing *e*. Then, it checks eight

[*] The work presented in this paper was carried out at Microsoft Research Asia

1/4-pel positions (represented by 1 ~ 8) surrounding *e* to obtain the best 1/4-pel MV.



Capital letters (C, H1, H2...): Integer pixel positions
Low case letters (a, b, c…): 1/2-pel positions
Arabian numbers (1, 2, 3…): 1/4-pel positions
Figure 1: A typical process of full fractional pixel search

A fast sub-pel ME algorithm, Center Biased Fractional Pel Search (CBFPS) [3], has been adopted by H.264 and integrated into the latest reference software. First, it predicts the sub-pel MV of the current block using the median MV of neighboring blocks. Then, it examines the zero sub-pel MV and the predicted sub-pel MV. The one possessing the minimum matching error is chosen to be the start position. Finally, the algorithm employs a diamond search method with a distance of 1/4-pel to refine the fractional pixel MV. However, for large blocks, the predicted MVs are not accurate enough in some cases, which results in a relatively large degradation in PSNR. Therefore, in implementation, for blocks not larger than 8x8 H.264 reference software adopts this method, while for the larger blocks (8x16, 16x8, 16x16) it stills uses full search. Experimental results shows that compared with full search, this method can reduce the computational cost by 33% [3].

3. PREDICTION-BASED DIRECTIONAL FRACTIONAL PIXEL SEARCH (PDFPS) ALGORITHM

Many fast motion estimation algorithms use the predicted MV as their initial search point [4] [8] [9]. A predicted MV includes the information of predicted integer pixel MV and predicted fractional pixel MV. If the predicted integer pixel MV is exactly the best MV after integer pixel ME, the predicted fractional pixel MV is also likely to be the best fractional pixel MV. We can start search at 1/4-pel accuracy directly from the predicted position. Otherwise, we can still take advantage of the directional information of the predicted sub-pel MV. We check the points in the direction of the predicted MV only rather than all possible points to reduce the number of search points. Sections 3.1 and 3.2 describe these two search strategies respectively. The entire algorithm is given in section 3.3.

3.1 Prediction-based search

3.1.1 MV prediction

In H.264, the predicted MV is obtained from the median MV value of the adjacent blocks on the left, top and top-right (or top-left) of the current block, as illustrated in Figure 2. The predicted MV is generally used as the initial search position in ME.

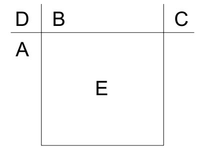


Figure 2: Median prediction

In addition, there are seven block sizes varying from 16x16 to 4x4 in H.264. Fig.3 illustrates two example structures of a macroblock. It has been found that large blocks and the small blocks in them are likely to have similar motion. To utilize the relationship of motion among various blocks, we use the MV of large blocks as the prediction of small blocks in it, as described in [4] [8]. We call this kind of prediction up-layer prediction. For instance, a 16x16 block is the up-layer of the 16x8 or 8x16 block in it, while a 8x16 block is the up-layer of the 8x8 block, and so on.

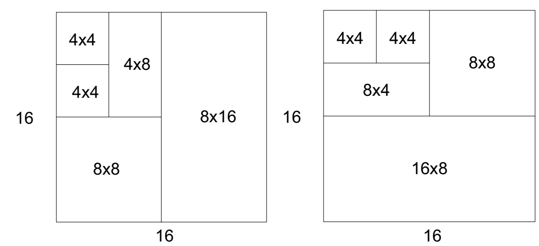


Figure 3: Examples of various blocks in a macroblock

We note that a block of size 16x16 does not have an up-layer. Since the motion trajectory of a moving object is usually continuous in natural video sequences, we use the MV of the block at the same position in the previous frame as the up-layer of the 16x16 block.

3.1.2 Predicted fractional pixel MV accuracy analysis

Given the predicted MV *pred_mv* and the best integer pixel MV *mv* at a 1/4-pel unit, we can extract the predicted fractional pixel MV *frac_pred_mv*:

$$frac\_pred\_mv = (pred\_mv - mv) \bmod 4 \tag{1}$$

In order to measure the accuracy of the predicted sub-pel MVs, we use the method described in [4] to check the Cumulative Distribution Function (CDF). It is defined as "the delta MVs between the predicted MVs and the MVs obtained by full search and is less than or equal to *d* pixels":

$$P(delta\_MV \le d) = P\left(\bigcup_{i=1}^{7} \left(\left|PMV_{ix} - MV_{ix}\right| + \left|PMV_{iy} - MV_{iy}\right|\right) \le d\right) \tag{2}$$

where PMV_{ix} and PMV_{iy} are the horizontal and vertical components of the predicted fractional pixel MV for block type *i* respectively. (MV_{ix}, MV_{iy}) represent the fractional pixel MV obtained by full search.

We calculate CDF under the condition that the predicted integer pixel MV of median prediction or up-layer prediction is exactly the best MV after integer pixel ME. Some results are shown in Table 1. Here we use the Akiyo and Foreman sequences to represent small and relatively large motion scenes respectively. *delta_MV* is defined as a 1/4-pel unit.

Table 1: CDF of MV difference between predicted MV and MV obtained by full search

Sequence	Predicted MV	$d = 0$	$d = 1$	$d = 2$	$d = 3$
Akiyo	Median	80.65%	95.03%	98.28%	99.41%
	Up-layer	79.44%	94.94%	98.50%	99.51%
	Median + Up-layer	86.33%	97.40%	99.29%	99.77%
Foreman	Median	59.13%	77.99%	86.74%	93.14%
	Up-layer	50.46%	70.77%	83.69%	90.87%
	Median + Up-layer	72.07%	89.08%	95.38%	98.20%

From Table 1, we can see that for the Foreman sequence, the probability of median prediction and up-layer prediction being the best MV is 59.13% and 50.46% respectively. If we combine the two predictors, the probability is improved to 72.07%. Although the probability is still not high enough so that we can use the predicted MV as the final MV, if we use median prediction and up-layer prediction as the initial search points, followed by one-step 1/4-pel Small Diamond Search Pattern (SDSP) search [10] (d=1), we can get 89.08% of the best MVs. With a three-step SDSP (d=3), we can improve the accuracy to 98.20%. For the Akiyo sequence which contains little movement, the accuracy is even higher. Thus, we can choose the predicted point with the minimum distortion as the initial search point and apply three steps of SDSP to get the final MV.

Fig. 4 illustrates a typical process of SDSP search. At each step, the algorithm examines four points surrounding the current position. The point with the minimum distortion is regarded as the center of the next iteration. The search process iterates until the point with the minimum distortion is the SDSP center or the number of iterations reaches a given number.

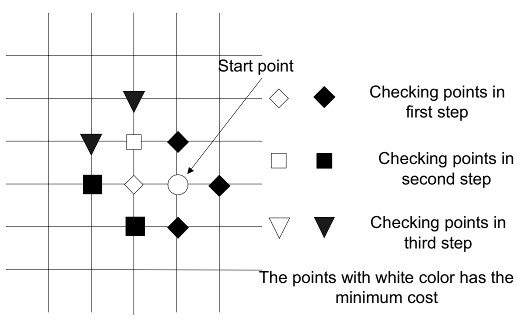


Figure 4: A typical process of SDSP search

3.1.3 An early termination method

From Table 1, we can also observe that in the Akiyo and Foreman sequences, for 86.33% and 72.07% of all blocks respectively, the best sub-pel MV is one of the predicted MVs. For convenience of description, we define the blocks whose best MV is exactly the predicted vector as predicted vector blocks (PVB). If we can predict PVBs, we can skip SDSP and eliminate a significant portion of the computation.

If a block is a PVB, it is likely to have a small cost at one of the predicted MVs. Therefore, we define thresholds TH_i (i = 1,…,7) for seven block sizes respectively. During fractional pixel motion search, two predicted MVs are first examined. If either of their costs is less than TH_i, the block is regarded as a PVB, and then the remaining search can be skipped.

$$Cost_i < TH_i \qquad for\ i = 1, \cdots, 7 \tag{3}$$

3.2 Directional search

Even if the predicted integer-pel MV is not the best MV after integer pixel ME, we can still exploit the directional information of the predicted sub-pel MVs. As shown in Fig. 5, we examine the 1/2-pel points in the direction of the predicted sub-pel MVs.

1) If the predicted MV is zero, check the search center C only;
2) If the direction of the predicted MV is 1, check 1/2-pel positions c and e;
3) If the direction is 2 or 3, check c or e accordingly.

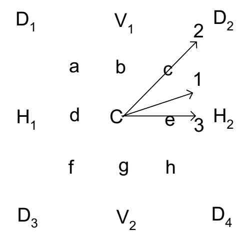


Figure 5: 1/2-pel points in the direction of predicted MV

Other directions are processed in a similar way. After checking the 1/2-pel MVs as described above, we select the one with the minimum cost as the best 1/2-pel MV, and then refine the MV using the previously described SDSP to obtain the final 1/4-pel MV.

3.3 Algorithm description

Based on the above observations and analysis, we propose a Prediction-based Directional Fractional Pixel Search (PDFPS) algorithm. For convenience of description, we define the blocks whose best integer pixel MV is equal to the integer part of the median predicted MV or the up-layer predicted MV as Median Predicted Blocks (MPB) or Up-layer Predicted Blocks (UPB) respectively. The median and up-layer predicted sub-pel MVs are calculated by Equation (1). TH_i is the early termination threshold defined in 3.1.3. The entire procedure is outlined as follows.

If the block is an MPB or UPB
    Check the search center (0, 0)
    If the block is an MPB
        Check the median predicted position
    If the block is a UPB
        Check the up-layer predicted position
    Select the point with the minimum cost $Cost_{min}$
    If $Cost_{min} < TH_i$
        Set the corresponding MV as the final MV and stop search.
    Else
        Iterate the SDSP search around the point with the $Cost_{min}$ until the minimum cost point locates at the SDSP center or the number of search steps reaches 3.
Else
    Check the search center (0, 0)
    Check the 1/2-pel positions in the directions of median and up-layer predicted MVs respectively
    Decide the best 1/2-pel position with the minimum cost
    Iterate the SDSP search around the best 1/2-pel point until the point with the minimum cost locates at the SDSP center.
Choose the minimum cost point as the final MV and stop search.

4. SIMULATION RESULTS

We apply the proposed method to JM8.0 of the H.264 reference software [11]. We set the motion search range to 16, and set the number of reference frames to 1. The RD optimization is disabled and the Hadamard transform and the CAVLC entropy coding are enabled in our experiments.

To examine the effectiveness of our method in different experimental conditions, we choose eight sequences with motion activities varying from small to large. They are Akiyo, Salesman, News, Silent, Coastguard, Foreman, Container and Football sequences. All sequences are in a QCIF format and encoded at 30 frames per second. All frames except for the first one are encoded as P-frames with a Quantization Parameter (QP) value of 32.

The thresholds TH_i (i = 1,…,7) for the seven block sizes are self-adaptively obtained during encoding. We use the average minimum fractional pixel cost of each block size in the previous frame as a reference for the correlation between adjacent frames. Many experiments have shown that when the threshold is set as the reference value multiply 1.1, a good tradeoff between complexity and performance can be achieved.

We compare the proposed method with Full Search and CBFPS under the same conditions. The number of average search points per block size (SP), PSNR and bit-rate for each sequence are shown in Table 2.

Table 2: Comparison of different methods

Sequence	Method	SP	PSNR (dB)	Bit-rate (kbps)
Akiyo	Full Search	17.00	35.29	15.29
	CBFPS	10.48	35.29	15.18
	Proposed Method	2.76	35.28	15.46
Salesman	Full Search	17.00	32.55	31.55
	CBFPS	10.69	32.53	31.50
	Proposed Method	3.36	32.55	31.75
News	Full Search	17.00	33.51	46.52
	CBFPS	10.58	33.54	46.69
	Proposed Method	3.20	33.55	47.23
Silent	Full Search	17.00	32.81	53.40
	CBFPS	10.80	32.79	53.56
	Proposed Method	3.86	32.81	53.91
Coastguard	Full Search	17.00	30.96	105.35
	CBFPS	10.86	30.97	105.15
	Proposed Method	4.54	30.98	105.81
Foreman	Full Search	17.00	33.28	91.31
	CBFPS	11.04	33.24	90.93
	Proposed Method	5.59	33.20	91.84
Container	Full Search	17.00	33.07	22.76
	CBFPS	10.36	33.08	22.60
	Proposed Method	2.80	33.10	22.84
Football	Full Search	17.00	32.15	297.12
	CBFPS	11.45	32.12	296.99
	Proposed Method	6.23	32.12	299.41

From the simulation results, we can see that our proposed method achieves good results. Compared with Full Search, our method reduces 63.38% to 83.75% of the sub-pel search points while the PSNR degradation is less than 0.01dB on average. In addition, compared with CBFPS, our method reduces 45.64% to 73.63% of search points while the average PSNR changes slightly. At the same time, the bit rates remain almost unchanged.

We also observe that for some sequences, such as News, Coastguard and Container, the PSNR values of our method and CBFPS are even a little higher than those of full search. This is possible because full search examines only eight 1/4-pel points surrounding the best 1/2-pel points instead of all the possible 1/4-pel points in the search area, which may not reach the global minimum in some cases. Our method and CBFPS, which use iterated diamond search, may work better in such cases.

5. CONCLUSIONS

In this paper, a fast fractional-pixel search algorithm is proposed for H.264 video coding. It predicts the sub-pel MV by using median prediction and up-layer prediction, together with 1/4-pel diamond search refinement. An early termination method is further applied to stop search early when the cost of the predicted MV is less than a given threshold. The proposed method greatly speeds up fractional pixel motion estimation. Experimental results show that compared with full search and the fast sub-pel ME algorithm proposed in H.264, our method can reduce up to 84% and 74% of sub-pel search points respectively, while the performance degradation is negligible.

Further directions may include improving the accuracy of prediction and improving the reliability of threshold decision.

6. REFERENCES

[1] C. Zhu, X. Lin, L.P. Chau, K.P. Lim, H.A. Ang, and C.Y. Ong, “A novel hexagon-based search algorithm for fast block motion estimation,” In Proc. of ICASSP 2001, vol.3, pp1593-96, May 2001.

[2] A. Tourapis, “Enhanced predictive zonal search for single and multiple frame motion estimation,” In *Proc. of VCIP 2002*, pp.1069-79, Jan. 2002.

[3] Z.B. Chen, P. Zhou, and Y. He, “Fast integer pel and fractional pel motion estimation for JVT,” *JVT-F017*, 6th meeting: Awaji, Japan, 5-13 Dec. 2002

[4] Z. Zhou, M.T. Sun and Y.F. Hsu, “Fast variable block-size motion estimation algorithms based on merge and split procedures for H.264/MPEG-4 AVC,” In *Proc. of ISCAS 2004*, pp.725-728, May 2004

[5] P. Yin, H.Y. C. Tourapis, A.M. Tourapis and J. Boyce, “Fast mode decision and motion estimation for JVT/H.264,” in *Proc. of ICIP 2003*, pp.853-856, Sep.2003

[6] Z.B. Chen, C. Du, J.H. Wang and Y. He, “PPFPS – A paraboloid prediction based fractional pixel search strategy for H.26L,” In *Proc. of ISCAS 2002*, pp. 9-12, May 2002

[7] Z.B. Chen and Y. He, “Prediction based Directional Refinement (PDR) algorithm for fractional pixel motion search strategy,” *JVT-D069*, 4th meeting: Klagenfurt, Austria, 22-26 July, 2002

[8] J.F. Xu, Z.B. Chen and Y. He, “Effective fast ME prediction and early-termination strategy based on H.264 statistical characters,” *PCM 2003*, pp.218-222, Dec. 2003.

[9] H.C. Tourapis, “Fast motion estimation with the JVT codec,” *JVT-E023*, 5th meeting: Geneva, Switzerland, 09-17 October, 2002.

[10] S. Zhu and K.K. Ma, “A new diamond search algorithm for fast block-matching motion estimation,” In *IEEE trans. Image Proceeding*, vol.11, pp.1442-49, Dec, 2002

[11] JVT reference software version 8.0, http://bs.hhi.de/~suehring/tml/download/old_jm/