A ROBUST INTERPOLATION-FREE APPROACH FOR SUB-PIXEL ACCURACY MOTION ESTIMATION

Wei Dai, Oscar C. Au, Wenjing Zhu, Wei Hu, Pengfei Wan, Jiali Li

Department of Electronic and Computer Engineering
HKUST, Clear Water Bay, Kowloon, Hong Kong
{weidai, eeau, wzhuaa, huwei, leoman, jiali}@ust.hk

ABSTRACT

Motion estimation (ME) is one of the key elements in video coding standard which eliminates the temporal redundancy by using a motion vector (MV) to indicate the best match between the current frame and reference frame. A coarse to fine process is taken to find the best MV. First of all, integer-pixel ME finds a coarse MV and followed by the sub-pixel ME around the best integer-pixel point. The sub-pixel ME plays an important role in improving the coding efficiency. However, the computational complexity of searching one sub-pixel point is much higher than the integer-pixel point searching because of the interpolation and Hadamard transform operation. In this paper, an accurate optimal sub-pixel position prediction algorithm is presented. With the information of the 8 neighboring integer-pixel points, the optimal sub-pixel position is predicted directly without explicitly solving model parameters. Moreover, an outlier rejection scheme is applied to improve the robustness of the proposed algorithm. Experimental results show that the proposed algorithm outperforms the state of the art interpolation-free sub-pixel ME algorithms.

Index Terms— Sub-pixel motion estimation, video coding, interpolation-free

1. INTRODUCTION

Motion estimation is one of the most important parts in video coding standard which is dedicated to achieve high coding performance by removing the temporal redundancy inherent in the video. However, it is very time consuming that it nearly takes 60% to 80% of the total encoding time in the H.264 codec [1]. The traditional ME process are divided into two stages: integer-pixel ME within a predefined search range and sub-pixel ME around the best integer-pixel point. The importance of sub-pixel ME has been widely studied and recognized [2]. Moreover, it is becoming more and more crucial to develop fast and effective sub-pixel ME algorithms due to the following two reasons.

First of all, the computational overhead of sub-pixel ME has become relatively significant because the complexity of integer-pixel ME has been greatly reduced by a lot of fast algorithms such as three step search (TSS) [3], new three step search (NTSS) [4], PMVFAST [5], E-PMVFAST [6] and so on. These algorithms are very effective and some of them only need to search less than 10 integer-pixel points. However, the traditional sub-pixel ME needs to search 8 half-pixel points and 8 quarter-pixel points, 16 sub-pixel points in all, which is a comparatively large computational burden. Secondly, the sub-pixel ME also requires interpolation operation to get the in-between pixel values to compute the sum of absolute transform difference (SATD), which does another Hadamard transform on the residuals to improve the coding gain. If the number of sub-pixel searching point is reduced, the interpolation operation and Hadamard transform can also be saved.

In this paper, a robust interpolation-free sub-pixel ME approach is presented. 8 neighboring integer-pixel points are used to model the error surface around the best integer-pixel point. By exploring the properties of the modeled error surface, the best sub-pixel position can be derived without explicitly solving the model parameters. Moreover, an outlier rejection scheme is applied to improve the robustness of the proposed method.

The remainder of the paper is organized as follow. Section 2 reviews some of the existing interpolation-free sub-pixel ME algorithms. Section 3 provides the analysis on the properties of the modeled error surface and gives a very simple but accurate minimum position prediction approach. The detailed algorithm is described in Section 4 and Section 5 shows the experimental results. Finally, Section 6 concludes the paper.

2. RELATED WORK

The properties of the sub-pixel ME error surface are studied in [1]. Results show that the unimodal error surface of the sub-pixel ME assumption holds in most cases because of the small search range and the strong correlation between sub-pixel points due to the sub-pixel interpolation operation. Several surface models have been proposed to approximate the error surface around the integer-pixel point, including the 5-

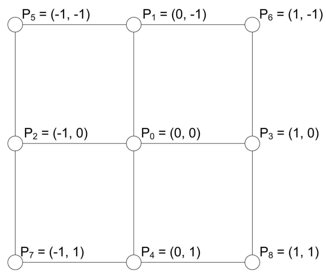

Fig. 1. Illustration of the location of best integer-pixel point P_0 and its 8 neighboring points.

term model (1) in [7], 6-term model (2) in [8], and 9-term model (3) in [9] models.

$$f_5(x,y) = Ax^2 + Bx + Cy^2 + Dy + E \quad (1)$$

$$f_6(x,y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F \quad (2)$$

$$f_9(x,y) = Ax^2y^2 + Bx^2y + Cxy^2 + Dx^2 + Exy + Fy^2 + Gx + Hy + I, \quad (3)$$

where parameters $A, B, \ldots, I$ are estimated by fitting the block distortion measurement (BDM) of integer-pixel points on the given models. A typical BDM is the sum of absolute difference (SAD) which is defined as:

$$SAD(x,y) = \sum_{i=1}^{m}\sum_{j=1}^{n} |s(i,j) - r(i+x, j+y)|,$$

where $s(i,j)$ denotes pixel value of place (i,j) in the current block, (x,y) represents the motion vector of the current block and $r(i+x, j+y)$ denotes the pixel value of the block in the reference frame. The length and width of the block is m and n respectively. Meanwhile, rate-distortion (RD) cost will also be used as BDM in the parameter estimation which is defined as:

$$RDCost(x,y) = SAD(x,y) + \lambda R(x,y),$$

where $R(x,y)$ denotes the number of bits needed to encoded the motion vector (x,y) and λ is the *Lagrange Multiplier*. The location of the best integer-pixel point and its 8 neighboring points are given in Fig. 1.

The minimum position of model (1) could be derived very easily, and it was widely used in a lot of papers to predict the best sub-pixel position, such as [7, 10, 11, 12]. Firstly, these algorithms predicted the best sub-pixel point by model (1) and then performed a small refined search around the predicted position to achieve the trade-off between the coding performance and computational complexity.

Hill *et al.* in [8] used the model (2) to predict the best sub-pixel position. The error surface of this model could well describe the true error surface. Since only 6 parameters were needed to be estimated, a minimization process was taken to choose one out of four corner integer-pixel points (P_5, P_6, P_7 and P_8 in Fig. 1). In [9], model (3) was proposed to model the error surface around the best integer-pixel point. All of the nine integer-pixel points were used to estimate the parameters. However, this model could not guarantee to be an unimodal. One problem of model (2) and (3) was that after solving the model parameters, unlike model (1), the minimum position could not be derived directly from the model parameters, several methods were proposed to find the minimum position in [8], including the simple 4-connected gradient descent search, hierarchical two-stage search and so on. These search methods required a lot of multiplications to calculate the estimated BDM.

However, sometimes there will be some outliers exist in the 8 neighboring points. When this happens, the model parameter will be greatly affected by the outlier and will result in a wrong prediction. Since the algorithms in [8] and [9] would both estimate the parameters using some fixed neighboring points, the robustness of the algorithms were not very strong. For the method in [8], there are 6 unknown parameters, so six integer-pixel points are needed. The author firstly selected the the best integer-pixel point P_0 and its left, right, up and bottom integer-pixel points (P_1, P_2, P_3 and P_4). The sixth point was chosen from the rest of the four points which gave the smallest matching error and ignored the other 3 points to avoid the effect of outliers. However, it still could not guarantee that P_1, P_2, P_3 or P_4 was not the outlier. If the points of P_1, P_2, P_3 and P_4 have one or more outliers, the prediction will fail. For the method in [9], when either one of the 8 neighboring points is outlier, the prediction accuracy will be affected.

In this paper, model (2) is used to model the error surface around the best integer-pixel point for its simplicity and generality. Moreover, model (2) has some nice properties that can let us derive the minimum point directly without solving the model parameters. This can help improving the prediction accuracy because it is very easy to introduce the error when solving the model parameters. The properties of model (2) are analyzed in the next section.

3. ANALYSIS OF THE MODEL PROPERTIES

Instead of firstly solving the model parameters and then calculating the optimal sub-pixel position based on the parameters, this paper provides a method that can directly derive the best sub-pixel point position without estimating the model parameters. First of all, the properties of model (2) is studied.

Let $f_6(x,y)$ to be an arbitrate paraboloid error surface. Suppose three planes $x = -1, x = 0$ and $x = 1$ intersect with this surface. The intersection of these three planes are three parabolas which are:

$$
\begin{aligned}
f_6(-1,y) &= Cy^2 + (E-B)y + A - D + F \\
f_6(0,y) &= Cy^2 + Ey + F \\
f_6(1,y) &= Cy^2 + (E+B)y + A + D + F,
\end{aligned}
$$

so the minimum points of these three parabolas are

$$
\begin{aligned}
p_x^{-1} &= (-1, -\frac{E-B}{2C}) \\
p_x^{0} &= (0, -\frac{E}{2C}) \\
p_x^{1} &= (1, -\frac{E+B}{2C}).
\end{aligned}
$$

It can be easily proved that these three minimum points lie within a straight line which is:

$$y = -\frac{1}{2C}(Bx + E). \tag{4}$$

Moreover, it also can be proved that the minimum point of all the parabolas which are the intersection of the paraboloid with the plane $x = c$ also lies on this line, where c is an arbitrate constant. This means that the minimum point of model (2) also lies on this line. We define this kind of line as minimum line.

Now, let's consider the minimum point of the parabolas which are the intersection of the paraboloid with the plane $y = c$. It can also be proved that the all the minimum points lie within a line which is:

$$x = -\frac{1}{2A}(By + D). \tag{5}$$

Thus, the minimum point of the paraboloid can be determined by the intersection of two minimum lines (4) and (5).

4. ROBUST INTERPOLATION-FREE SUB-PIXEL ME ALGORITHM

By the analysis in the Section 3, the minimum point of the model (2) can be derived by the intersection of two minimum lines. These two lines are deterministic when the model parameters are decided. However, as the analysis in Section 2, the way to solve the 6 parameters is not robust enough because it can not deal with the situation when P_1, P_2, P_3 or P_4 is outlier. So in this paper, instead of predicting the optimal position by solving the model parameters, the optimal position is predicted by the intersection of two minimum lines and these two minimum lines are derived without solving the model parameters explicitly. Moreover, the outliers can be detected and rejected by the algorithm easily. Detailed algorithm is discussed below.

4.1. Minimum Line Derivation

Since the minimum lines (4) and (5) are the set of all the minimum points of the parabolas which are the intersection of the paraboloid with a group of parallel planes, these two lines can also be derived easily by finding two minimum points on the line. So in order to derive minimum line (4), the 9 integer-pixel points in Fig. 1 are divided into three groups, which are $(P_5, P_2, P_7), (P_1, P_0, P_4)$ and (P_6, P_3, P_8). In each group, suppose group (P_5, P_2, P_7), three integer-pixel points are used to model the parabola, which is defined as:

$$f(y) = ay^2 + by + c,$$

where a, b and c is the parameters of this parabola which can be easily solved by the method in [?]. After solving a, b and c, the minimum point can be easily derived as $p_x^{-1} = (-1, -\frac{b}{2a})$. Another two minimum points p_x^0 and p_x^1 can also be derived in this manner. Since all the minimum points tend to lie within a line, the line (4) can be derived by picking two minimum points on this line. Then, those 9 integer-pixel points are regrouped into three new groups which are $(P_5, P_1, P_6), (P_2, P_0, P_3)$ and (P_7, P_4, P_8), the same procedure applies to derive three minimum points p_y^{-1}, p_y^0 and p_y^1 and the line (5) is derived. Then, the global minimum point can be obtained by the intersection of the two minimum lines.

4.2. Outlier Rejection Scheme

The most important procedure before estimating is outlier rejection. Only when the outliers are rejected, the accuracy of the model can be guaranteed. After six minimum points are derived, the locations of those points are checked. Under the assumption that the optimal sub-pixel position is around the integer-pixel point, several criteria are applied to reject the outliers. Takeing the points p_x^{-1} and p_x^1 as the example:

1. If the absolute value in y axis of p_x^{-1} or p_x^1 is larger than 1.25, there is a high probability that one of the three integer-pixel points in that group is an outlier, which means the derived minimum point is not trustworthy. So this minimum point will not be used to derive the minimum line. However, if all of these two minimum points are not used, the line which parallels to the x axis and pass through point p_x^0 will be used to approximate the minimum line.

2. If both of these two values are smaller than 0.75, there is a very high probability that these two points and the point p_x^0 lie within a straight line, so these two points will be used to derive the minimum line.

3. Otherwise, the point which is closer to 0 together with the point p_x^0 will be used to derive the minimum line.

With this outlier rejection scheme, the prediction accuracy can be improved. The best sub-pixel point can be predicted

Table 1. Performance Comparison of Several Interpolation-Free Algorithms.

Sequence Class	Free_5	Free_6 in [8]	Free_9 in [9]	Direct in [13]	Proposed
Class B	3.2%	3.9%	3.4%	2.9%	**2.6%**
Class C	4.9%	5.8%	5.1%	4.9%	**4.3%**
Class D	7.1%	8.7%	7.6%	6.9%	**6.2%**
Class F	4.0%	4.1%	3.7%	3.4%	**3.1%**
Average	4.7%	5.6%	4.9%	4.5%	**4.0%**

without the interpolation operation and Hadamard transform process, which can reduce a lot of computation power.

5. EXPERIMENTAL RESULTS

The proposed interpolation-free sub-pixel ME algorithm is implemented on the HEVC reference software HM6.0. Lowdelay_P_main condition is used for simulation which is specified in [14]. Sequences in class B, C, D and E are used to run the simulation. The name of sequences in each class is specified in [14]. BD-rate [15] is calculated as the comparative measurement of the coding performance. The interpolation-free scenario is used to test the prediction accuracy of the proposed method compared to other interpolation-free algorithms. The methods in [8] (called Free_6 in this paper) and [9] (called Free_9) are used as the comparative methods which are actually using model (2) and (3) to model the error surface. The optimal sub-pixel position is predicted by the hierarchical two-stage search which first calculates the estimated BDM at 8 half-pixel positions and then 8 quarter-pixel positions. Also, model (1) (called Free_5) is implemented as another interpolation-free scheme as the comparative method. Moreover, the state of the art interpolation-free sub-pixel prediction method [13] (called Direct) is implemented. The traditional ME algorithm which uses hierarchical sub-pixel ME algorithm is regarded as the anchor. Experimental results for the interpolation-free scheme is list in Table 1.

From the BD-rate comparison in Table 1, it can be concluded that the proposed method outperforms all of the other interpolation-free methods. The proposed method avoid solving the model parameters which would be affected by the outlier points easily and can derive the optimal sub-pixel position directly with a very safe outlier rejection scheme. Moreover, when comparing the complexity of deriving the minimum point, although the parameters in model (2) and (3) can be derived by only add and bit shift operation, the optimal point derivation still requires a lot of multiplications. For the proposed method, after 6 minimum points have be derived, 4 points are selected to calculate the minimum points, suppose the four points are $(x_1, y_1), (x_2, y_2)$ for (4) and $(x_3, y_3), (x_4, y_4)$ for (5), the optimal point (x_{min}, y_{min}) can be calculated as:

$$x_{min} = \frac{cd - af}{bd - ea}$$
$$y_{min} = \frac{bf - ce}{bd - ea},$$

where

$$a = x_1 - x_2$$
$$b = y_1 - y_2$$
$$c = x_1y_2 - x_2y_1$$
$$d = x_3 - x_4$$
$$e = y_3 - y_3$$
$$f = x_3y_4 - x_4y_3,$$

which only needs 10 multiplications, 9 additions and 2 divisions to get the minimum point. While for the Free_6, 8 multiplications and 5 additions is needed for one point BDM estimation, for the Free_9, 18 multiplications and 8 additions is needed for one point estimation and there are totally 16 points needed to be estimated.

6. CONCLUSION

In this paper, a robust optimal sub-pixel position prediction scheme is presented. The algorithm uses the information of 9 integer-pixel points to derive the best sub-pixel point directly without solving the modeling parameter explicitly. An outlier rejection scheme is applied to remove the outliers among the 9 integer-pixel points to improve the robustness of the algorithm. Experimental results show that the proposed algorithm outperforms the state of the art interpolation-free sub-pixel ME algorithms.

Acknowledgement

This work has been supported in part by the Research Grants Council (GRF Project no. 610112) and HKUST (HKUST Project no. FSGRF12EG01) of the Hong Kong Special Administrative Region, China.

7. REFERENCES

[1] Z. Chen, J. Xu, Y. He, and J. Zheng, "Fast integer-pel and fractional-pel motion estimation for H.264/AVC," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 264–290, 2006.

[2] J. Jain and A. Jain, "Displacement measurement and its application in interframe image coding," *Communications, IEEE Transactions on*, vol. 29, no. 12, pp. 1799–1808, 1981.

[3] T. Koga, "Motion-compensated interframe coding for video conferencing," in *Proc. NTC 81*, 1981, p. 9.

[4] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 4, no. 4, pp. 438–442, 1994.

[5] A. M. Tourapis, O. C. Au, and M. L. Liou, "Predictive motion vector field adaptive search technique (PMVFAST)-enhancing block based motion estimation," in *Proceedings of SPIE*, 2001, vol. 4310, pp. 883–892.

[6] H. M. Wong, O. C. Au, C. W. Ho, and S. K. Yip, "Enhanced predictive motion vector field adaptive search technique (E-PMVFAST)-based on future mv prediction," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 4–pp.

[7] J. F. Chang and J. J. Leou, "A quadratic prediction based fractional-pixel motion estimation algorithm for H.264," *Journal of Visual Communication and Image Representation*, vol. 17, no. 5, pp. 1074–1089, 2006.

[8] P. R. Hill, T. K. Chiew, D. R. Bull, and C. N. Canagarajah, "Interpolation free subpixel accuracy motion estimation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 12, pp. 1519–1526, 2006.

[9] S. Dikbas, T. Arici, and Y. Altunbasak, "Fast motion estimation with interpolation-free sub-sample accuracy," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 7, pp. 1047–1051, 2010.

[10] W. Dai, O. C. Au, S. Li, L. Sun, and R. Zou, "Fast sub-pixel motion estimation with simplified modeling in HEVC," in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 1560–1563.

[11] W. Dai, O. C. Au, C. Pang, L. Sun, R. Zou, and S. Li, "A novel fast two step sub-pixel motion estimation algorithm in HEVC," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1197–1200.

[12] J.W. Suh, J Cho, and J Jeong, "Model-based quarter-pixel motion estimation with low computational complexity," *Electronics letters*, vol. 45, no. 12, pp. 618–620, 2009.

[13] Q. Zhang, Y. Dai, and C. C. Kuo, "Direct techniques for optimal sub-pixel motion accuracy estimation and position prediction," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 12, pp. 1735–1744, 2010.

[14] F. Bossen, "HM 8 common test conditions and software reference configurations," in *Proc. 10th JCT-VC Meeting, Stockholm, Sweden, No. JCTVC-J1100*, 2012, pp. 1–3.

[15] G. Bjøntegaard, "Improvements of the BD-PSNR model," *document VCEG-AI11, ITU-T SG16*, 2008.