Workshop on Picture Coding and Image Processing, PCSJ2010/IMPS2010
December 7, 2010, Nagoya, Japan

A LINEAR MODEL-BASED SUB-PIXEL MOTION ESTIMATION METHOD FOR H.264/AVC STANDARD

Chang-Uk Jeong and Hiroshi Watanabe

Graduate School of Global Information and Telecommunication Studies, Waseda University

1. INTRODUCTION

H.264/AVC [1], which is a recently published video coding standard, was developed and has been updated by Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. The H.264 encoder, similar to that of the existing video standards such as ISO/IEC MPEG-1, MPEG-2, MPEG-4, ITU-T H.261, H.263 [2], is composed of a temporal model, a spatial model, and an entropy encoder. The runtime of the temporal model, which is named motion estimation (ME) part to eliminate the temporal redundancy between adjacent frames, occupies more than 80% of the whole runtime of the H.264 encoder [3]. Furthermore, due to various advanced encoding technologies such as variable block-size ME, multiple reference frames, and quarter-pixel ME and compensation, the computational complexity was increased dramatically. Although the advanced encoding technologies have significantly improved the quality performance of the H.264 standard, the high complexity imposes restrictions on real time video applications, especially for mobile devices.

The motion estimator in the H.264 encoder conducts the integer-pixel ME (IME), followed by the fractional-pixel ME (FME). Particularly, the runtime of IME is about half of the total encoding time [3]. In which, many fast IME algorithms including the unsymmetrical-cross multi-hexagon-grid search (UMHexagonS) [4] have been developed up to now to reduce integer search points. The proposed IME algorithms have different search strategies to satisfy the accuracy of motion estimation and the search speed, respectively.

On the other hand, the FME part also takes up over 40% of the total ME runtime [3]. After performing the FME process, about 1-3 PSNR improvement can be expected. The FME part, therefore, has been recently regarded as an important research topic in H.264/AVC. The conventional full fractional-pixel search always requires a fixed number of search points and consumes a large amount of memory due to the interpolation process for fractional-pixel search. Thus, in this paper, a linear model-based sub-pixel motion estimation method for H.264/AVC standard is proposed to efficiently reduce the computational complexity and maintain the quality performance without the interpolation process.

2. PREVIOUS FME ALGORITHM

The quadratic prediction-based FME (QPFPS) [5] was proposed in 2005. A degenerate quadratic prediction function used in this algorithm affects the determination of the best quantized predicted motion vector (MV) at quarter-pixel accuracy, which is shown below.

$$F(x,y) = Ax^2 + Bx + Cy^2 + Dy + E \tag{1}$$

As described in Equation 2, the differential operation is executed on the quadratic prediction function to assume the minimum cost.

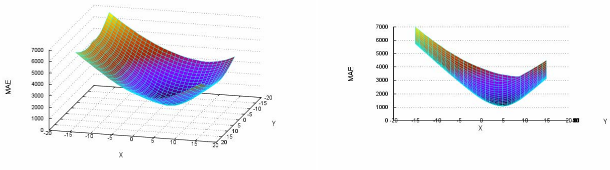
$$\begin{cases} 2Ax_p + B = 0 \\ 2Cy_p + D = 0 \end{cases} \tag{2}$$

In the final step, the predicted position is located at the center of the small diamond search pattern, and then the small diamond search algorithm is carried out. In this step case, however, the interpolation process may be needed to create a fractional-pixel search area.

3. ERROR SURFACE OF FME

We have simulated the error surface of FME directly. Unlike that of IME, the FME error surface is clearly unimodal because the sub-pixels are generated by the interpolation process using the existing integer-pixels. Therefore, the approach to the FME part should be different from IME.

Figure 1 (a) shows the FME error surface for the "Garden" sequence. Particularly, we take notice of Figure 1 (b). The shape looks like parabolic definitely. It means we can apply parabolic models including some quadratic functions for FME. In addition, the graph shows it has vertical symmetry with respect to *x*=5.



(a) (b)

Figure 1. (a) Error surface of FME for "Garden" at 1/16-pixel accuracy, (b) Another angle of (a).

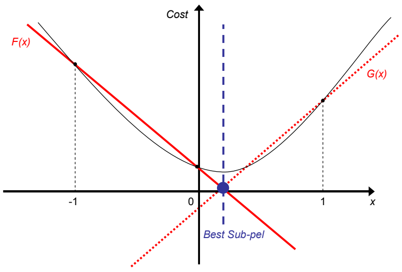


Figure 2. The basic concept of the proposed method.

From understanding Figure 1, a linear model graph can be drawn, as shown in Figure 2. Line $F(x)$ adjoins point (-1, $F(-1)$) and (0, $F(0)$). Here, we can introduce symmetry assumption for $F(x)$. Line $G(x)$, which has the negative slope value of $F(x)$, is passing point (1, $G(1)$). That is, the basic principle of the proposed FME method is to find the location where the two lines intersect.

4. PROPOSED FME METHOD

The proposed linear model-based sub-pixel motion estimation method reuses the matching error cost of nine integer-pixel points. Basically, a linear prediction function applied in the proposed method is described as follows.

$$\begin{cases} F(x) = (-H_1 + C)x + C \\ G(x) = (-H_1 + C)(1 - x) + H_2 \end{cases} \tag{3}$$

In function $F(x)$, for example, $-H_1+C$ is the slope value of the linear equation for a group (H_1, C, H_2). Function $G(x)$ represents a symmetrical linear function with the negative slope value corresponding to $F(x)$. The intersection point x between $F(x)$ and $G(x)$ can be shown below.

$$x = \frac{(H_2 - H_1)}{2(C - H_1)} \tag{4}$$

In addition to the linear-based prediction, we propose a grouping strategy to enhance the accuracy. The nine integer pixel search points as shown in Figure 3 are grouped according to the close proximity. They are divided into three groups, the horizontal, the vertical, and the both horizontal and vertical, of three points. For instance, (S_1, V_1, S_2), (V_1, C, V_2), and (S_1, C, S_4) are members of the above mentioned groups, respectively. Each group is calculated to predict the matching error cost of FME, and then the x- and y-coordinate of the location which produces the minimum prediction cost are regarded and selected as the best fractional-pixel position.

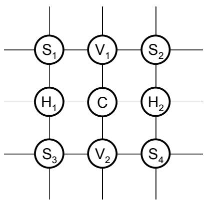


Figure 3. The nine integer-pixel search points reused in the proposed method.

5. EXPERIMENTAL RESULTS

The proposed method has been evaluated based on the H.264/AVC reference software JM12.4 [6]. The simulation is conducted with the default settings of search range=16, quantization parameter=28, rate distortion optimization=off, and baseline profile. After carrying out UMHexagonS for IME, the FME module is applied at quarter-pixel accuracy. The four sequences – QCIF sequence “Claire” and CIF sequences “Football”, “Mobile”, and “News” are used. They include a variety of motion contents and activities, respectively. 100 frames for each sequence are encoded. We compare our method with QPFPS algorithm. To assess only the prediction of them without the interpolation process, the small diamond search process used in QPFPS is skipped.

From Table 1, we can see that the PSNR performance of the proposed method shows a little better improvement compared with the quadratic prediction method. As shown in Table 2, the bit rate is also lower than the quadratics’. In this test, the ME time is not considered because all of them do not use any search points at sub-pixel accuracy.

Table 1. PSNR [dB] performance comparison.

Method	QCIF	CIF		
	Clair	Football	Mobile	News
QPFPS	39.631	36.171	33.792	37.981
Proposed	39.649	36.175	33.794	37.982

Table 2. Average bit rate [bps] comparison.

Method	QCIF	CIF		
	Clair	Football	Mobile	News
QPFPS	34802	1599065	2062130	232426
Proposed	34627	1579102	2059097	231466

6. CONCLUSION

In this paper, we proposed a linear model-based sub-pixel motion estimation method which is based on a simple linear equation reusing the matching error cost of integer-pixel points and expanded by a grouping strategy. Particularly, the proposed FME method was designed to implement in the H.264 encoder without respect to the interpolation process. The results of the experiment show that our method can bring about a slightly better performance in terms of PSNR compared with a quadratic prediction-based algorithm, whereas the bit rate is decreased.

7. REFERENCES

[1] “Draft ITU-T Rec. and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264-ISO/IEC 14 496-10 AVC),” Joint Video Team (JVT) of ITU-T and ISO/IEC JTC1, Geneva, JVT of ISO/IEC MPEG and ITU-T VCEG, JVT-G050r1, Mar. 2003.

[2] K. R. Rao and J. J Hwang, *Techniques and Standards for Image, Video and Audio Coding*. Englewood Cliffs, NJ: Prentice Hall, 1996.

[3] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, “Motion compensated interframe coding for video conferencing,” in *Proc. Nat. Telecommun. Conf.*, New Orleans, LA, USA, Nov. 29–Dec. 3 1981, pp. G5.3.1–G5.3.5.

[4] Z. Chen, Y. He, and J. Xu, “Hybrid unsymmetrical-cross multi-hexagon-grid search strategy for integer pel motion estimation in H.264,” in *Proc. Picture Coding Symp.*, Saint Malo, Apr. 2003, pp. 17–22.

[5] J. F. Chang and J. J. Leou, “A quadratic prediction based fractional-pixel motion estimation algorithm for H.264,” in *Proc. Seventh IEEE Int. Symp. on Multimedia.*, pp. 491–498, Dec. 2005.

[6] JVT H.264/AVC Reference Software Joint Model (JM), Download: http://iphome.hhi.de/suehring/tml/