RESIDUAL ERROR CURVATURE ESTIMATION AND ADAPTIVE CLASSIFICATION FOR SELECTIVE SUB-PEL PRECISION MOTION ESTIMATION

Saverio G. Blasi and Ebroul Izquierdo

School of Electronic Engineering and Computer Science, Queen Mary University of London

ABSTRACT

We present a novel approach for adaptive precision motion estimation based on a classification of the residual error curvature. A fast algorithm is proposed to estimate the curvature of the interpolated residual surface using the error samples after integer precision motion estimation. We also propose an original technique to compute and successively update a set of thresholds using the information from previously coded frames. The optimal motion vector precision is then selected for each block according to the current thresholds. The approach is compared in terms of PSNR of the motion compensated reconstruction against conventional state of the art sub-pel motion estimation algorithms, and it is shown to efficiently reduce complexity and coding times of a typical video encoder with negligible effects on the prediction accuracy.

Index Terms— Motion estimation, video coding

1. INTRODUCTION

The aim of motion estimation ME is to reconstruct a video frame based on a previously encoded frame using some descriptors of the estimated motion (motion vectors MV) and the residual error (usually sum of absolute differences SAD) between prediction and original target. Very high compression ratios can be obtained transmitting MV and prediction instead of the original frame, and ME is in fact adopted in most of the common video standards (from MPEG-2 [1] to the current state-of-the-art H.264/AVC [2]). ME is usually obtained using block-matching algorithms [3] where MV are found independently for each processed block. Many techniques were introduced to improve accuracy or complexity of ME (as in [4], [5], [6] and [7]). Sub-pel precision ME techniques were introduced [8] to further refine the ME accuracy, allowing the MV to assume fractional values. In order to do so, intensities need be estimated at sub-pel locations interpolating the integer values, resulting in a higher resolution frame and correspondingly very high complexity.

Efficient methods have been proposed to fasten sub-pel ME. A widely adopted method (called *hierarchical search*) consists in assuming the SAD is approximately concave around the integer solution ([9]), hence looking for half-pel solutions only among few locations close to it and for quarter-pel solutions among locations close to the half-pel solution. In order to further reduce the complexity, two classes of algorithms were proposed, one based on fastening the search procedure and the other based on reducing the interpolation complexity. K.H. Lee et al. [11] proposed a two-step technique at half-precision accuracy for reducing the number of candidates using minimum error pairs. A similar idea is used in the *Predictive Zonal Search* [12], where vectors are forced in the direction of the minimum and second-minimum RE. An enhanced version *EPZS* is successfully implemented in the JM software of the H.264/AVC standard [13].

Techniques for reducing complexity are usually based on interpolating the SAD and finding its minima avoiding search among candidates. Suh et. al. [14] obtained good half-pel solutions from nine integer-located SAD values around the integer solution, using approximated numerical polynomial interpolation. Similar techniques were proposed to refine the interpolation or improve the coefficients calculation. The approach was improved in [15] and extended to quarter-pel in [16].

Due to its high computational complexity the role of sub-pel ME was extensively investigated. Girod in [17] proved that sub-pel refinements can improve the prediction only up to a certain limit. Higher precision levels only increase complexity with no effects on the accuracy. Hence, adaptive precision ME techniques have been validated (see [18]) and proposed to establish such precision limit in a dynamic way, mainly based on statistical characterizations of the target frame. Q. Zhang et al. ([19]) proposed an approach in which a coefficient, referred to as *Deviation from Flatness, DF*, is evaluated for each block using a 3×3 window of SAD values as an estimate of geometrical properties of the interpolated residual error (RE) shape. Based on some pre-calculated thresholds, sub-pel ME for the current block can either be skipped or performed at a given precision depending on the corresponding DF. The approach is innovative but presents some problems in classification accuracy and sequence biasing, as further illustrated in the next sections.

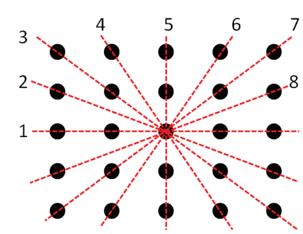

Fig. 1. The grid of RE values and the set of eight normal planes used for the sampled curvedness estimation.

In this paper we introduce a novel technique for sub-pel ME with adaptive precision MV. The approach is based on the assumption that a characterization of the SAD (based on the curvature of the interpolated surface) can be used to predict the outcomes of ME, selecting the blocks where sub-pixel refinements can decrease the RE effectively. We introduce a fast technique for estimating such coefficient (called *sampled curvedness SC*) using a large set of 5×5 SAD values. SC is then classified in a per-block basis using a set of thresholds that are calculated and successively updated as more frames are coded in the sequence using an original adaptive technique. The approach is found to successfully reduce complexity compared with conventional sub-pel ME, avoiding unnecessary computations with

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

negligible impact on the accuracy.

The rest of this paper is organized as follows. A fast estimation of the RE curvature is proposed in Section 2 and compared with conventional adaptive sub-pel ME. An algorithm for calculating and adaptively updating the thresholds is illustrated in Section 3 along with the proposed encoder framework. Comprehensive tests and results are shown in Section 4 where the algorithm is compared with state-of-the-art approaches. Conclusions are drawn in Section 5.

2. SAMPLED CURVEDNESS

The curvature of a surface at a point is completely characterised by the maximum and minimum curvatures of the trajectories obtained as interceptions between the surface and the normal planes at the point at all directions (the principal curvatures of the shape [20]). We can calculate the curvature of these scalar functions as second order derivatives evaluated at the point of interest. This means that analytically such principal curvatures correspond to maximum and minimum eigenvalues of the Hessian matrix of the original surface calculated at the point. Instead of using two different coefficients we can combine the principal curvatures into a single parameter. In [21] Koenderink defines an original measure of the curvature of a shape at a point, referred to as *curvedness* and obtained as the square root of the sum of the squared principal curvatures. Curvedness is suitable for the ME case as it is able to detect local behaviours (as opposite to more common coefficients such as Gaussian curvature or mean curvature). We propose in this paper an original fast technique to estimate the curvedness of the SAD interpolated shape for characterizing the blocks in terms of geometric properties of the residual.

Assume that integer-precision ME is performed in the current block resulting in the MV (x_0, y_0). Consider a window of 5×5 SAD values around the integer SAD solution including ±2 horizontal and vertical displacements. Instead of considering continuous polynomial interpolation to find all the directions at (x_0, y_0), we consider a finite set of eight intercepting normal planes (as in Figure 1), namely the planes passing through existing integer locations. The SAD values within each plane are fitted using polynomial interpolation. Depending on the particular direction, a plane intercepts either three or five values. In the first case denote these values with A_3, B_3 and $SAD(x_0, y_0)$ and with t the continuous independent variable (the spatial direction). Denote with $SAD_3(t)$ the corresponding polynomial interpolation. We have:

$$
\begin{aligned}
&SAD_3(t) = \\
&= \frac{t(t-1)}{2}A_3 - (t-1)(t+1)SAD(x_0,y_0) + \frac{t(t+1)}{2}B_3 = \\
&= \left[\frac{A}{2} + \frac{B}{2} - SAD(x_0,y_0)\right]t^2 + \left[\frac{B}{2} - \frac{A}{2}\right]t + SAD(x_0,y_0)
\end{aligned}
$$

Similarly in the case of five values (denote these as A_5, B_5, C_5 D_5 and $SAD(x_0, y_0)$) we can find a continuous function $SAD_5(t)$.

Double-differentiating both functions at $t = 0$ we obtain a set of eight values:

$$\alpha_i = A_3 + B_3 - 2SAD(x_0, y_0)$$

where $i = 1, 3, 5, 7$ and:

$$\alpha_i = -\frac{1}{12}A_5 + \frac{4}{3}B_5 + \frac{5}{2}SAD(x_0,y_0) + \frac{4}{3}D_5 - \frac{1}{12}E_5$$

where $i = 2, 4, 6, 8$.

Denote now as α_{max} and α_{min} the maximum and minimum values. We approximate the curvedness of the original shape according to the definition in [21] with:

$$SC = \sqrt{\alpha_{max}^2 + \alpha_{min}^2}$$

which we refer to as Sampled Curvedness (SC).

It can be proved that the SC obtained as above is very close to the curvedness obtained using complete smooth interpolation, in the sense that they both share the same structure as linear combinations of integer SAD values. We prove here that the SC is also highly correlated with outcomes of conventional ME. To show this, we performed full integer precision ME and hierarchical sub-pel ME (using 6-tap half-pixel interpolation and bilinear quarter-precision interpolation, as in typical H.264/AVC encoders [2]) on some test sequences using 16×16 macroblock mode. We denote a block as *integer-displaced* if the fractional components of the MV resulting from sub-pel ME are equal to zero, and as *quarter-displaced* otherwise (that is if MV have non-zero fractional components). We computed SC and DF (as proposed in [19]) for all blocks and calculated the average value of such coefficients for all the integer-displaced blocks and quarter-displaced blocks per sequence, as shown in Table 1.

Table 1. Average SC and DF of integer-displaced (Int-D.) and quarter-displaced (Quar-D.) Blocks

Sequence	**Average SC**		**Average DF**	
	Int-D.	Quar-D.	Int-D.	Quar-D.
Akiyo	3.4	1.1	3.6	2.0
Bowing	1.9	0.6	1.4	1.1
Coastguard	4.3	3.5	4.4	6.9
Moth.& Daught.	1.7	0.3	1.0	0.5

Quarter-displaced blocks show much higher SC values compared with the values of integer-displaced blocks, proving that using SC we can classify blocks according to the outcomes of sub-pel ME without actually performing it. We do not obtain the same results using DF, in fact average DF values of quarter-displaced blocks are generically closer to values of integer-displaced blocks making the classification a more difficult task. In the *Coastguard* sequence average DF of integer-displaced blocks is even higher than average DF of quarter-displaced blocks which means the separation fails in most of the blocks. SC is able to classify most of the blocks of the *Coastguard* sequence correctly. Such results show that if we can define a set of efficient thresholds, we can use the SC values to select the blocks according to ME outcomes avoiding unnecessary SAD calculations for finding fractional components in the integer-displaced blocks. We could save complexity and computational time selecting the blocks efficiently.

To do so we first need to find suitable threshold values. Adaptive precision approaches already exist that classify blocks based on a set of threshold on previously estimated parameters. Such approaches usually use statistically-based techniques to compute the thresholds (such as the mean of values of quarter-displaced blocks and values of integer-displaced blocks). Similar techniques lead generically to very unsatisfactory results in that the thresholds do not take into account the particular features of the actual part of the sequence being encoded (for instance the presence of sudden fast-motion or slow-motion portions of the sequence). A different, novel approach is proposed here where thresholds are adaptively updated as more frames are coded based on the outcomes of previous predictions. Much better results can be expected using such approach.

3. SELECTIVE PRECISION AND ADAPTIVE THRESHOLDS CALCULATION

Let us assume we encode the first inter-frame in the sequence using conventional sub-pel techniques and compute the SC for each block. Consider the blocks with fractional MV components and denote with $SC_q(0)$ the average SC of such blocks. Denote similarly with $SC_i(0)$ the average SC of blocks with integer MV components and consider $mean = \frac{1}{2}(SC_i(0) + SC_q(0))$. If $\frac{1}{4}T \leq mean \leq 4T$, $T(0) = mean$. Instead if $mean < \frac{1}{4}T$, then $T(0) = \frac{1}{4}T$ or if $mean > 4T$, then $T(0) = 4T$, where T is a fixed parameter calculated using statistical analysis and used to correct the thresholds if too little information is extrapolated from previous blocks. Consider eventually $T_q = T(0)$ and $T_h = \frac{1}{2}T(0)$. For each block in the next inter-frame integer-precision ME is performed and SAD are used to compute the SC. If $SC < T_h$ no further action is required, otherwise if $SC \geq T_h$ the half-precision MV is calculated. Eventually if $SC \geq T_q$ the quarter-precision MV is also calculated (block is classified as quarter-displaced). The MV of quarter-displaced blocks are used to update the thresholds as follows. Four indexes N_i, C_i, N_q and C_q are initialised to zero before processing the next frame. Consider the SC of current block, SC_b. Then:

1. If MV of a quarter-displaced block is integer valued, $N_i = N_i + 1$ and $C_i = C_i + SC_b$.
2. If MV of a quarter-displaced block is fractional, $N_q = N_q + 1$ and $C_q = C_q + SC_b$.

This is repeated for each block in each inter-frame until the last block in the $(K-1)$-th frame, where K is the period (in frames) at which thresholds are updated. Then we obtain:

$$SC_i(K-1) = \frac{C_i}{N_i}$$

$$SC_q(K-1) = \frac{C_q}{N_q}$$

and

$$mean = \frac{1}{2}(SC_i(K-1) + SC_q(K-1))$$

Thresholds are updated correspondingly (if $\frac{1}{4}T \leq mean \leq 4T$, $T(0) = mean$; if $mean < \frac{1}{4}T$, $T(0) = \frac{1}{4}T$; if $mean > 4T$, $T(0) = 4T$). N_i, C_i, N_q and C_q are set to zero and we proceed to next frame until time instant $2K-1$ (when thresholds are updated again), and so on to the end of the sequence.

The algorithm is very efficient and presents several advantages with respect to existing approaches, as shown in the next section.

4. RESULTS

The approach is tested on some common slow-motion, mid-motion and fast-motion video sequences at CIF (352×288 pixels) and QCIF (176×144 pixels) resolution in terms of PSNR of the reconstructed against original target frame. Each tested frame is reconstructed using motion compensation on the immediately previous reference frame. ME is performed on the reference using the proposed framework and conventional sub-pel techniques with fixed block partition size. Experiments are conducted using fixed 16×16, 8×8 or 4×4 partition sizes.

The first 50 frames of the slow-motion *Container* sequence are tested as above with fixed 16×16 partition size, resulting in the plot in Figure 2. The PSNR values obtained using the proposed approach are in general very close to those obtained using conventional quarter-pel ME, indicating that the framework does not affect the overall estimation accuracy. The next plot in Figure 3 shows the percentage of blocks per frame where quarter-pel refinements or half-pel refinements are skipped from the ME algorithm (due to SC values below the current thresholds) for the same 100 test frames. We can see that quarter-pel ME is skipped for around 70% of the blocks in each frame. We can expect correspondingly high gains in terms of speed and computational complexity of the overall encoder. Very similar results are obtained using smaller partition sizes: 4×4

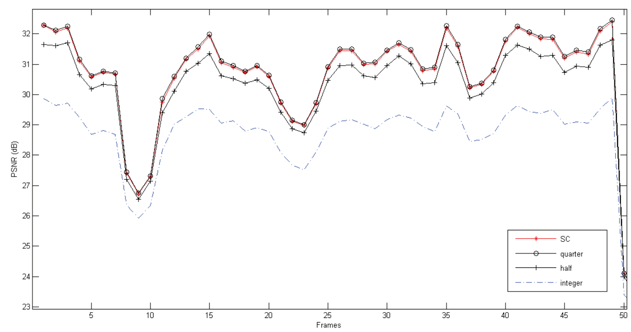

Fig. 2. PSNR of reconstructed target for first 50 test frames of the *Container* sequence.

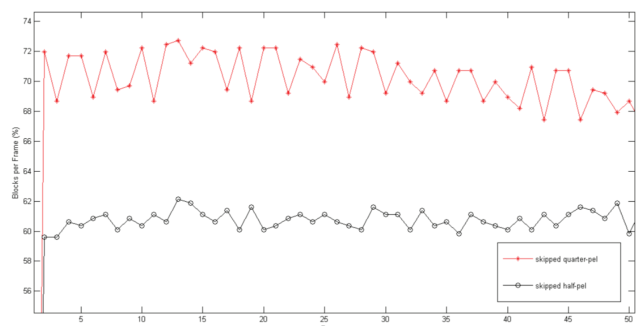

Fig. 3. Percentage of blocks per frame (with fixed 16x16 partition size) where sub-pel ME is skipped for first 50 test frames of the *Container* sequence.

size is used in Figure 4 (PSNR of reconstructed target) and Figure 5 (percentage of blocks per frame where sub-pel ME is skipped) for the first 50 frames of the fast-motion *Bus* sequence. Notice that in some frames SC falls below the quarter-pel threshold in up to 90% of the blocks per frame. Other tests on different sequences and partition size are performed but cannot be displayed due to space constraints.

Average SC values per frame is shown in Figure 6 for the first 50 frames of 4 test sequences (high-motion, mid-motion and slow-motion) at QCIF resolution using fixed 16×16 partition size.

5. CONCLUSIONS

We propose an original approach for adaptive precision sub-pel ME where the MV precision is selected according to a geometrical characterization of the residual error. A fast algorithm is proposed for

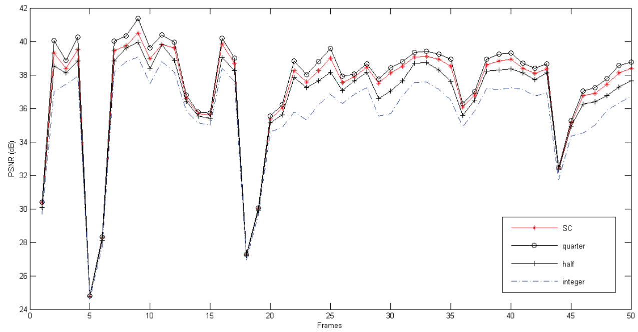

Fig. 4. PSNR of reconstructed target for first 50 test frames of the *Bus* sequence.

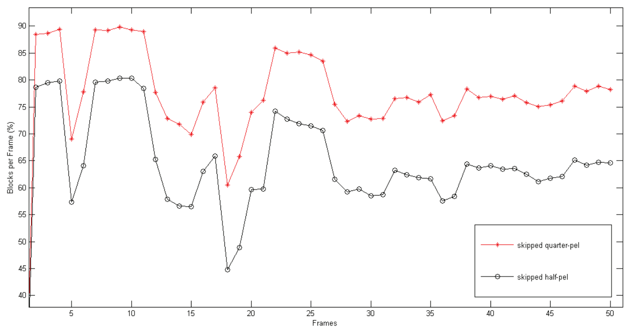

Fig. 5. Percentage of blocks per frame (with fixed 4x4 partition size) where sub-pel ME is skipped for first 50 test frames of the *Bus* sequence.

estimating the curvature of the residual error interpolated surface in terms of an original parameter, called sampled curvedness, which depends on the SAD at integer locations. The parameter is validated against similar existing approaches and shown to predict efficiently the outcomes of conventional sub-pel ME techniques. A novel approach is proposed to estimate and successively update a set of thresholds for classifying the blocks according to the sampled curvedness, selecting the optimal MV precision for each block. Thresholds are updated at a fixed period in time as new frames are coded in the sequence. The approach is tested against conventional sub-pel ME techniques in terms of PSNR of the motion compensated target using different partition sizes and resolutions. The framework is shown to achieve accuracy comparable with quarter-pel techniques even though sub-pel ME is effectively skipped in the large majority of the blocks (up to 90% in some slow-motion frames). The approach is promising in terms of reducing complexity and encoding times of a typical video encoder.

6. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 MPEG-2 Standard TM 5 (1995)

[2] T.Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Transactions on Circuits and Systems for Video Technology*, v. 13 n.7, pp. 560-576, (2003)

[3] G.J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression", *IEEE Signal Processing Magazine*, v. 15 n. 6, (1998)

[4] W. Li and E. Salari, "Successive Elimination Algorithm for Motion Estimation", *IEEE Transactions On Image Processing*, v. 4 n. 1, pp. 105-107, (1995)

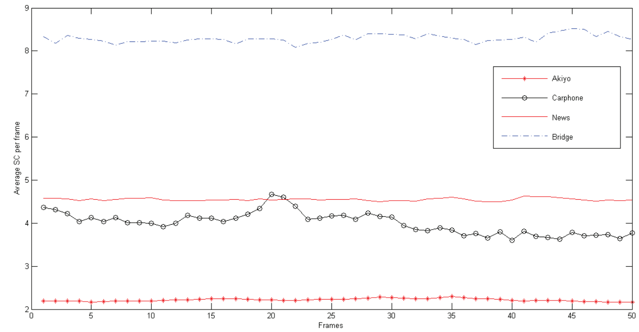

Fig. 6. Average SC per frame of first 50 test frames of 4 QCIF sequences (high-motion *Bridge*, mid-motion *News* and *Carphone* and slow-motion *Akiyo*.

[5] C. Zhu, W.S. Qi and W. Ser, "Predictive Fine Granularity Successive Elimination for Fast Optimal Block-Matching Motion Estimation", *IEEE Transactions On Image Processing*, v. 14 n. 2, pp. 213-221, (2005)

[6] R. Li, B. Zeng and M.L. Liou, "A new Three-Step Search Algorithm for Block Matching Motion Estimation", *IEEE Transactions On Circuits and Systems for Video Technology*, v. 4 n. 4, (1994)

[7] S. Zhu and K.K. Ma, "A New Diamond Search Algorithm for Fast Block Matching", *IEEE Transactions On Image Processing*, v. 9 n. 2, (2000)

[8] K. Minoo and N. Truongt, "Reciprocal Subpixel Motion Estimation: Video Coding With Limited Hardware Resources", *IEEE Transactions On Circuits and Systems for Video Technology*, v. 7 n. 6, (2007)

[9] J. Jeong and W. Ahn, "Sub-pixel Accuracy Motion Estimation using a Model for Motion Compensated Error", *Proceedings of PCS93*, (1993)

[10] E. Chan and S. Panchanathan, "Review of Block Matching based Motion Estimation Algorithms for Video Compression", *Canadian Conference on Electrical and Computer Engineering*, (1993)

[11] K.H. Lee, J.H. Choi, B.H. Lee and D.G. Kim "Fast Two-Step Half-Pixel Accuracy Motion Vector Prediction", *Electronics Letters* , v. 36 n. 7, pp. 625-626, (2000)

[12] A.M. Tourapis, O.C. Au and M.L. Liou, "Highly Efficient Predictive Zonal Algorithms for Fast Block-matching Motion Estimation", *IEEE Transactions On Circuits and Systems for Video Technology*, v. 12 n. 10, pp. 934, (2002)

[13] M. Von Dem Knesebeck and P. Nasiopoulos, "An Efficient Early-termination Mode Decision Algorithm for H.264", *IEEE Transactions On Consumer Electronics*, v. 55 n. 3, pp. 1501, (2009)

[14] J.W. Suh and J. Jeong, "Fast Sub-pixel Motion Estimation Techniques having Lower Computational Complexity", *IEEE Transactions On Consumer Electronics*, v. 50 n. 3, pp. 968, (2004)

[15] J. Cho, G. Jeon, J. Suhand and J. Jeong, "Sub-Pixel Motion Estimation Scheme Using Selective Interpolation ", *IEICE Transactions*, v. 91-B n. 12, pp. 4078-4080, (2008)

[16] J.W. Suh, J. Cho and J. Jeong, "Model-based Quarter-pixel Motion Estimation with Low Computational Complexity", *Electronics Letter*, v. 45 n. 12, pp. 618, (2009)

[17] B. Girod, "Motion-compensating Prediction with Fractional-pel Accuracy", *IEEE Transactions On Communications*, v. 41 n. 4, pp. 604, (1993)

[18] J. Ribas-Corbera and D. L. Neuhoff, "Optimizing Motion-Vector Accuracy in Block-Based Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, v. 11 n. 4, pp. 497-511, (2001)

[19] Q. Zhang, Y. Dai and C. C. J. Kuo, "Direct Techniques for Optimal Sub-Pixel Motion Accuracy Estimation and Position Prediction", *IEEE Transactions on Circuits and Systems for Video Technology*, v. 20 n. 12, pp. 1735-1744, (2010)

[20] J. J. Koenderink, "Solid Shape", *MIT Press*, Cambridge, MA (1990)

[21] and A. J. van Doorn., "Surface Shape and Curvature Scales", *Image and Vision Computing*, v.10 n.8, (1992)

[22] Joint Model (JM), "H.264/AVC Reference Software", *Artech House Inc.*, http://iphome.hhi.de/suehring/tml/download