Hybrid motion estimation algorithm for secondary SP-frame coding based on inter-frame correlation


A hybrid motion estimation algorithm for secondary SP-frame coding is proposed. This hybrid algorithm effectively combines two existing techniques: one provides high coding efficiency and the other is computationally less intensive. The combination is based on inter-frame correlation, which is measured using the bit-count of its corresponding primary SP-frame. Experimental results show that this is an effective measure of selecting an appropriate domain for motion estimation in secondary SP-frame coding.

Introduction: SP-frames have been defined in H.264 to facilitate efficient switching between two video streams with different bitrates, say B1 and B2 in Fig. 1. In this figure, two primary SP-frames (PSP-frames), PSP1,t and PSP2,t, are placed at time t (switching point). Also, a secondary SP-frame (SSP-frame, SSP12,t) is used to allow drift-free switching. They can exploit temporal redundancy with predictive coding, but use different reference frames. Even if different reference frames are used, an identical reconstruction can still be guaranteed. For instance, when switching from B1 to B2 needs to be performed at time t, SSP12,t is transmitted to identically reconstruct PSP3,t. This property aims at replacing I-frames for seamless switching between compressed bitstreams of different bitrates. The coding efficiency of an SP-frame is much better than that of an I-frame and is slightly worse than that of a P-frame [1-3]. However, the storage cost of SSP-frames is tremendous [4]. In [4], a new motion estimation algorithm, which is operated in the quantised-transform (QDCT) domain instead of the pixel domain, which is consistent with the transform coefficients of an SSP-frame [1, 5]. Both the reference frames at time t from B1 and B2, respectively. They share the same video content and the only discrepancy is the quantisation parameters. It implies that corrSSP,t is a good measure to determine a proper domain for motion estimation in SSP-frame coding. In other words, a smaller value of corrSSP,t tends to use QDCT-ME; otherwise, Pixel-ME is employed to reduce the required computational complexity. The number of bits required for encoding SSP12,t can be used directly as a measure of corrSSP,t. However, this bit-count of SSP12,t cannot be obtained prior to SSP-frame encoding. Fig. 2 then shows the bit counts for SSP12,t and PSP1,t of ‘Mobisode2’. It is noted that the SSP-frame is available when its corresponding SSP-frame is encoded. From Fig. 2, it can be easily seen that the general trends of the two curves are very similar. This is due to the fact that the current frames used for encoding SSP12,t and PSP1,t are the frames at time t+1 and current frame at time t, respectively. The new SAQTD is capable of finding the motion vectors with more zero values of quantised-transform coefficients, which benefits the entropy coding of an SSP-frame and provides a remarkable size reduction. For QDCT-ME, each current block Bc is transformed and quantised to Qc[T(Bc)]. A search window in the reference frame centred on the current block position is set. QDCT-ME starts with transforming and quantising a block, Qc[T(Bc)], in the top right-hand corner of the search window. After obtaining Qc[T(Bc)] and Qc[T(Bc)], SAQTD(Bc, Br) can then be calculated for this candidate. Afterwards, the next candidate is another block shifted by 1 pixel in the horizontal direction, and this block is also transformed and quantised for computing its SAQTD(Bc, Br). These procedures continue for all possible candidates within the search window. It means that all possible candidates within the search window also need to be transformed and quantised in the QDCT domain. Consequently, SAQTD is computationally very intensive though it can achieve higher coding efficiency compared with SAD.

Proposed algorithm: In this Letter, a hybrid motion estimation algorithm is proposed to make good use of the high coding efficiency of QDCT-ME and the low computational complexity of the pixel-domain ME (Pixel-ME). The hybrid algorithm is based on inter-frame correlation between the current frame and the reference frames. In [4], we revealed that improvement in coding efficiency of SSP-frames highly relies on the degree of inter-frame correlation between PSP1,t and P1,t-1 as shown in Fig. 1, and it is denoted by corrSSP,t. The weaker the inter-frame correlation, the better the coding efficiency of SAQTD. Therefore, corrSSP,t is a good measure to determine a proper domain for motion estimation in SSP-frame coding. In other words, a smaller value of corrSSP,t tends to use QDCT-ME; otherwise, Pixel-ME is employed to reduce the required computational complexity. The number of bits required for encoding SSP12,t can be used directly as a measure of corrSSP,t. However, this bit-count of SSP12,t cannot be obtained prior to SSP-frame encoding. Fig. 2 then shows the bit counts for SSP12,t and PSP1,t of ‘Mobisode2’. It is noted that the SSP-frame is available when its corresponding SSP-frame is encoded. From Fig. 2, it can be easily seen that the general trends of the two curves are very similar. This is due to the fact that the current frames used for encoding SSP12,t and PSP1,t are the frames at time t from B2 and B1, respectively. They share the same video content and the only discrepancy is the quantisation parameters. It implies that corrSSP,t is reasonably approximated by the bit-count of SSP1,t, denoted by BCSSP,t. BCSSP,t is then used to select a proper domain for motion estimation in secondary SP-frame coding with the proposed hybrid algorithm. When BCSSP,t is larger than a predefined threshold Th, corrSSP,t becomes lower and QDCT-ME is carried out for SSP-frame coding in order to offer higher coding efficiency. In contrast, Pixel-ME is good enough when BCSSP,t ≤ Th, which can relieve the computational burden of SSP-frame coding. In general, more frames in a sequence with complex or fast motion activities will be encoded using QDCT-ME. In our implementation, Th is set to αc, where C is a universal constant with a unit of bits and α is a scaling factor. The constant C can be applied to all video sequences and it is equal to 50000. Also, α is sequence-dependent, and is considered as a parameter to control the trade-off between the computational complexity and the accuracy of domain selection of the proposed hybrid algorithm.

Fig. 1 Switching from B1 to B2 using SP-frames

SSP-frame encoding in QDCT domain: The way of encoding an SSP-frame is similar to that of encoding a P-frame, except additional quantisation/dequantisation steps with the quantisation level Qc are applied to the transform coefficients of an SSP-frame [1, 5]. Both the reference frame at time t-1 and the current frame at time t are first transformed and quantised using Qc. The quantisation-transform coefficients of the reference and current frames are then subtracted and there is no further quantisation for the residue. This can ensure perfect reconstruction since the residual computations are operated in the QDCT domain. However, motion estimation (ME) and compensation processes, which find the best-matched candidate block Bc in the reference frame for each block Br in the current frame, are performed in the pixel domain. This mismatch induces the bulky size of an SSP-frame. In [4], quantised-transform domain ME (QDCT-ME) can improve the coding efficiency of the SSP-frame. The technique minimises the residue in the QDCT domain rather than the pixel domain, which is consistent with the coding structure of an SSP-frame encoder. The traditional Lagrangian cost function, which computes the sum of absolute differences between Br and Bc (SAD) in the pixel domain, is then changed to calculate the sum of absolute differences in quantised-transform coefficients (SAQTD), and it can be defined as

\[ SAQTD(Bc, Br) = \sum [Qc[T(Bc)] - Qc[T(Br)]] \]

Fig. 2 Sizes of SSP-frames and corresponding PSP-frames of ‘Mobisode2’
**Simulation results:** The proposed hybrid algorithm has been implemented based on the JVT JM 11 encoder [6]. The luminance components of the first 100 frames of ‘Mobiisode1’ (832 x 480), ‘Mobiisode2’ (832 x 480) and ‘Crew’ (1280 x 720) were used. For ME, a search range of 32 was set for both P-frames and SP-frames. The sequences were encoded into two bitstreams with two different quantisation levels (Q_p = 24 and 20). For the proposed algorithm, taking motion activity into consideration, a was set to 1 and 10 for Mobiisode1/Mobiisode2 and Crew, respectively. Fig. 3 depicts the switching scenario of ‘Mobiisode2’ performed from Q_p = 24 to Q_p = 20. Only the first frames were encoded as I-frames and all other frames were encoded in turn as SP-frames while non-switching frames were encoded as P-frames. Fig. 3 shows the number of required bits of all possible SSP-frames by employing the Pixel-ME, the QDCT-ME and the proposed hybrid algorithm. The required bit-counts of the proposed algorithm are very similar to those of the QDCT-ME, but it requires fewer bit-counts compared with the Pixel-ME. Table 1 further illustrates the performance of the proposed hybrid algorithm when compared to the QDCT-ME in the three testing sequences. In the Table, Δbit-count and Δtime represent the percentage changes of the bit-count and encoding time of SSP-frames, respectively. The positive values mean increments, whereas negative values are decrements. Again, the Table shows that the bit-count increase of the proposed algorithm is negligible. However, it is observed that the proposed algorithm can substantially reduce computational complexity compared with the QDCT-ME.

![Fig. 3 Sizes of SSP-frames by applying Pixel-ME, QDCT-ME and proposed hybrid algorithm with Q_p = 24 to Q_p = 20 switching in 'Mobiisode2'.](image)

<table>
<thead>
<tr>
<th></th>
<th>Δbit-count</th>
<th>Δtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobiisode1</td>
<td>+0.25%</td>
<td>−9.69%</td>
</tr>
<tr>
<td>Mobiisode2</td>
<td>+0.85%</td>
<td>−32.77%</td>
</tr>
<tr>
<td>Crew</td>
<td>+0.63%</td>
<td>−41.92%</td>
</tr>
</tbody>
</table>

**Conclusion:** A hybrid motion estimation algorithm for secondary SP-frame coding is proposed. It adaptively selects QDCT or pixel domain for motion estimation according to the bit-count of its corresponding primary SP-frame. The hybrid algorithm has the advantages of utilising both domains in encoding secondary SP-frames. Experimental results show that the proposed new measure for selecting an appropriate domain is effective. Compared with the QDCT-domain algorithm, the proposed algorithm can achieve remarkable computational savings while maintaining a similar size of an SSP-frame.

**Acknowledgments:** The work is partially supported by CSP, EIE, PolyU and a grant from the Research Grants Council of the HKSAR, China (PolyU 5122/08E). K.K. Lai acknowledges the research studentships provided by the University.

© The Institution of Engineering and Technology 2011

4 November 2010
doi: 10.1049/el.2010.3102

K.K. Lai, Y.L. Chan, C.H. Fu and W.C. Siu (Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, People’s Republic of China)

E-mail: enylechan@polyu.edu.hk

**References**