HIGHER ORDER PREDICTION FOR SUB-PIXEL MOTION ESTIMATION

Damith J. Mudugamuwa[1,2], *Xiangjian He*[1,3], *Chung-Hyun Ahn*[1,4], *Jie Yang*[5]

[1]Centre for Innovation in IT Services and Applications, University of Technology Sydney, Australia
[2]APIIT Lanka, Colombo, Sri Lanka
[3]Lab of Biomedical Information Technology, University of Aizu, Japan
[4]Electronics and Telecommunication Research Institute, Korea
[5]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China

ABSTRACT

Estimating motion between two frames of a video sequence, up to sub-pixel accuracy, is a critical task for many image processing applications. Efficient block matching algorithms were proposed in [1, 4, 5, 6] for motion estimation up to pixel accuracy. Applying these fast block search algorithms to up-sampled and interpolated frames can produce good results but with significant increase in computations. To reduce the number of search points, and therefore the computational cost, quadratic prediction was proposed earlier [1, 2] to predict the location of minimum block matching error, and then to limit the search window to the vicinity of the predicted location. In this paper we investigate the typical behavior of block matching error surface and propose an improved higher order prediction that models the error surface more accurately, utilizing additional local image behavior. Initial experiments have proved promising results of about 50% more improvement in PSNR compared to quadratic prediction with only a marginal increase in the computational cost.

Index Terms - sub-pixel motion, image registration

1. INTRODUCTION

Motion estimation up to sub-pixel accuracy is a key task in many image processing applications. In a typical H.264/AVC encoder, 60-90% of total computing power was claimed to be consumed by motion estimation [3]. One of the key factors limiting the performance of super-resolution algorithms is accurate local motion estimation up to sub-pixel accuracy within an implementable computational complexity. Though regularization techniques used in super resolution can prevent undesirable artifacts due to poor motion estimation, these techniques can not compensate for the lost information resulted by poor motion estimation. Efficient sub-pixel motion estimation, both in terms of high accuracy and low computational complexity, is therefore of prime importance to these applications.

Motion estimation or image registration in its generic sense refers to mapping two or more images of the same scene to the same pixel grid, geometrically aligning the images. A wide range of models with varying complexity have been proposed to realize this mapping and [7] provides a good survey of these techniques. However these techniques are either deficient or computationally exhaustive to be applied for general video sequences with arbitrary motions, for instance in video encoding or super-resolution applications.

For super-resolution applications, computationally efficient frequency domain approaches [8, 9] and spatial domain gradient based approaches [10, 11] have been proposed for accurate image registration. However these techniques are limited to global translation and rotation and cannot be used for generic video with arbitrary local motion patterns. Though block matching techniques are often used for estimation of planar translational motion, Callico et.al. in [6] reported application of fast block matching algorithms in generic super-resolution reconstruction. In video coding, block search algorithms are popular for their easy implementation and simplicity. However computation complexity of block matching algorithms for sub-pel motion estimation often limits the scope of their applicability.

In block matching algorithms, motion estimation is carried out in two steps. In step 1, motion is estimated up to pixel accuracy by direct application of a fast block search algorithm such as new three step search algorithm [5]. In step 2, fractional pixel motions are estimated around the best integer pixel match of step 1. In estimating fractional pixel motion, algorithms first predict a sub-pel motion vector (SPMV) which is later refined by a limited local search around the predicted SPMV. The accuracy of this prediction affects accuracy of the sub-pel motion estimation or size of the local search area. One approach to predict the SPMV is to model the block matching error, sum of the absolute difference (SAD) surface as a polynomial function of pixel coordinates and to predict the SPMV as the distance to the minimum SAD point.
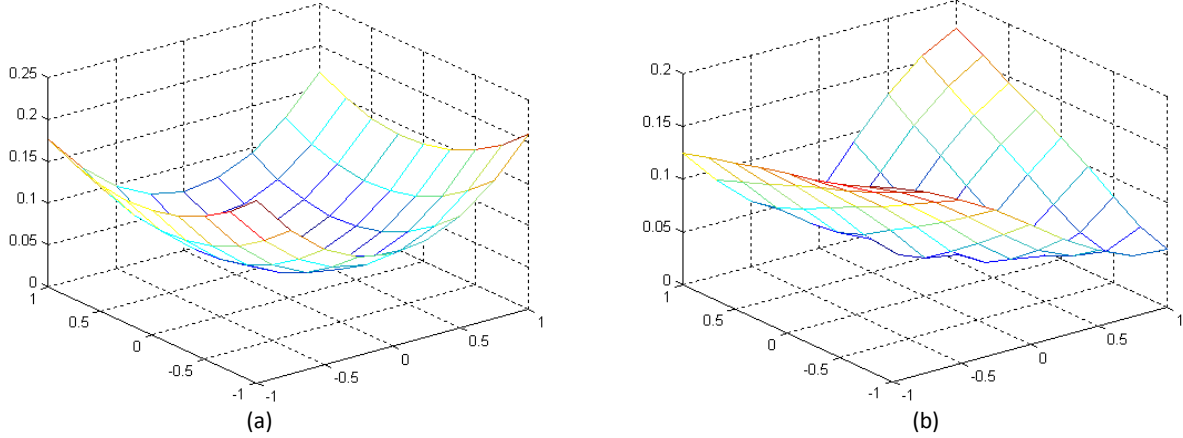
Figure 1 – SAD surface resulted by (a) QP and (b) expanded images, around the best integer pixel match.
(This is for a 3x3 block picked from the middle of the image frame of the "Suzie" sequence)

In this paper we investigate the local behavior of a typical SAD surface and propose an improved SPMV prediction using higher order functions. In section 2, performance of quadratic prediction is investigated. In section 3, higher order SPMV prediction is introduced. Section 4 provides the experimental results followed by conclusion.

2. QUADRATIC PREDICTION

Quadratic functions have been used to model the SAD surface for the purpose of SPMV prediction in video coding [1, 2]. Commonly used functions are given in (1) and (2) below.

$$SAD(x,y) = c_0 + c_1 x + c_2 y + c_3 x^2 + c_4 y^2 \qquad (1)$$
$$SAD(x,y) = c_0 + c_1 x + c_2 y + c_3 x^2 + c_4 y^2 + c_5 xy \qquad (2)$$

The coefficients c_i can be evaluated by applying the function to known SAD values (of integer pixel matches) around the best integer pixel match and SPMV is then predicted as the distance to minimum SAD value given by the function. Setting the derivatives of (1) or (2) in x and y directions to zero, a direct analytical solution is possible for the minimum SAD coordinates. A typical SAD surface modeled by (1) and the actual SAD surface calculated by expanded images are illustrated in figure 1.

The mismatch between the above two graphs in figure 1 indicates the potential for improved prediction resulting more accurate SPMVs. To further investigate the performance of quadratic prediction (QP), the test framework in section 2.1 was adopted.

2.1. Test framework

To evaluate the performance of SPMV prediction we used the SPMV calculated by expanded or interpolated images. Interpolated version of SPMV was obtained by applying a modified 3 step search (3SS) [4] on expanded images

around the corresponding location of the best integer match. The images were expanded to a higher resolution depending on the required sub-pel accuracy. Number of steps and the search points for each step were altered to cover the required search window. Though full block search was not employed to obtain the interpolated solution, a more generous search pattern was used at each step of the modified 3SS. However, the SPMV produced by QP or any higher order prediction is in continuous domain and the reference SPMV obtained by interpolation is discrete and is only a quantized version of the actual SPMV. Therefore these SPMV's cannot be directly compared.

To evaluate the performance of SPMV prediction, therefore, we obtained a motion compensated image sequence using predicted SPMVs and compared the original frames with the motion compensated frames. PSNR was used as the measure of comparison between two frames. Test scope was limited to adjacent frames of 3 test sequences: Mobile (and Calendar), Suzie and Car phone video sequences. The block size used for all tests was 3x3. Furthermore as real image sequences are used, due to occlusions, noise and other distortions, one-to-one mapping of pixels between two frames is not possible. Therefore we used the following criteria to attain the best registration possible. Given a pixel in a frame, we search for the best match in the reference frame and register that pixel at the best match. This results in both 'holes' (empty pixels) and 'collisions' (multiple matches to the same location) in the registered image. Collisions were resolved in favor of the least SAD value. Holes were left blank and disregarded from the PSNR calculation. Rationale for the latter is that the holes could be resulted by corresponding pixels not being present in the frame to be registered due to occlusions, noise and other distortions and a registration algorithm, at its best, should disregard such pixels, avoiding possible ill-registrations.

PSNR between motion compensated and reference images can also be calculated for integer motion vectors (IMV) and

Table 1 – Average PSNR (in dB) resulted from IMV and SPMV correction for interpolated and QP versions.

Video Sequence	Average PSNR (in dB)				Average PSNR improvement above IMV correction (in dB)		
	IMV	Interpolated SPMV	QP(1) SPMV	QP(2) SPMV	Interpolated SPMV	QP(1) SPMV	QP(2) SPMV
Mobile	28.548	30.301	29.190	28.992	1.753	0.642	0.444
Suzie	37.870	40.247	38.663	38.517	2.377	0.793	0.646
Car Phone	33.226	34.891	33.682	33.449	1.665	0.456	0.222

for SPMVs resulted from interpolation. Any sub-pixel refinement in motion estimation should increase the PSNR above the IMV corrected PSNR. Full block search on expanded images provides the optimal PSNR for a given sub-pel accuracy. Therefore IMV corrected and interpolated SPMV corrected, PSNR figures provide lower and upper bounds, respectively, for any SPMV prediction algorithm.

The usual local search to improve SPMV prediction was not applied in these tests as the purpose was to evaluate the performance of prediction. Furthermore, in all experiments sub-pel accuracy was calculated to a quarter pixel.

2.2. Quadratic prediction evaluation results

By adopting the above test framework, PSNR was calculated for QP using both equation (1) and (2) and the results are summarized in table 1, under QP(1) and QP(2) columns. PSNR for IMV corrected frames and interpolated SPMV corrected frames are also calculated as these provide lower and upper bounds to bench mark QP. From each video sequence 20 frames extracted from the middle of the sequence were only considered. Table 1 also illustrates the PSNR improvement above the IMV correction in dB. For the tested 3 sequences, the PSNR improvement of QP varied from 0.222dB to 0.793dB above the lower bound of IMV correction. The maximum possible improvement achieved for interpolated SPMV varied from 1.665dB to 2.377dB above the lower bound of IMV correction.

The results produced by QP(2) were inferior to the results of QP(1), even though additional term was included in QP(2) to better model the SAD surface. This observation was consistent across the summarized results and also across other experiments carried out. This can be contributed to the asymmetric nature of the SAD points considered for the coefficient calculation. The five coefficients of equation (1) were calculated by considering the SAD values resulted from the best integer match and at its 4-neighbors. However for equation (2) to calculate the additional coefficient, c_5, an additional SAD point from the group of 4 diagonal neighbors had to be considered and this has biased the modeled SAD surface towards the selected additional neighbor pixel. This asymmetric nature of the modeled SAD surface has adversely affected the performance of QP(2).

3. HIGHER ORDER PREDICTION

To improve the SAD surface in figure 1a and average PSNR more closer to the results produced by interpolation, we propose to model the SAD surface using a functional with higher order terms. We propose the function in (3), arbitrarily, but ensuring all 8-neighbours are considered in coefficient calculation, to avoid any adverse effects observed in using function (2).

$$SAD(x, y) = c_0 + c_1 x + c_2 y + c_3 x^2 + c_4 y^2 + c_5 xy + c_6 x^2 y + c_7 xy^2 + c_8 x^2 y^2 \tag{3}$$

By applying the function (3) to the best integer match and the surrounding 8 neighbors, all 9 coefficients can be calculated analytically. To find the location with minimum SAD, we set the derivatives of (3) to zero, resulting,

$$c_1 + 2c_3 x + c_5 y + 2c_6 xy + c_7 y^2 + 2c_8 xy^2 = 0 , \tag{4}$$
$$c_2 + 2c_4 y + c_5 x + c_6 x^2 + 2c_7 xy + 2c_8 x^2 y = 0 . \tag{5}$$

The two equations (4) and (5) were solved numerically by the method of successive approximations using iteration (6).

$$x_{n+1} = -\frac{c_1 + c_5 y_n + c_7 y_n^2}{2c_3 + 2c_6 y_n + 2c_8 y_n^2}$$

$$y_{n+1} = -\frac{c_2 + c_5 x_n + c_6 x_n^2}{2c_4 + 2c_7 x_n + 2c_8 x_n^2} \tag{6}$$

Initial approximations for x_0 and y_0 were estimated by the same equations (4) and (5) but setting the higher order coefficients c_5 and above to zero. This estimation corresponds to the solution of QP provided by equation (1) as in both cases it is to solve an identical system of equations to determine the coefficients, c_1 to c_4. Initial approximations for x and y were improved by (6) for a fixed number of 5 iterations and in most cases it was observed that the solution converged within 3 to 4 iterations.

4. EVALUATION RESULTS

The same test framework described in section 2.1 was adopted to evaluate the performance of the higher order prediction (HP) described above in section 3. The results are

Table 2 – Average PSNR and PSNR improvement above IMV correction for HP (in dB).

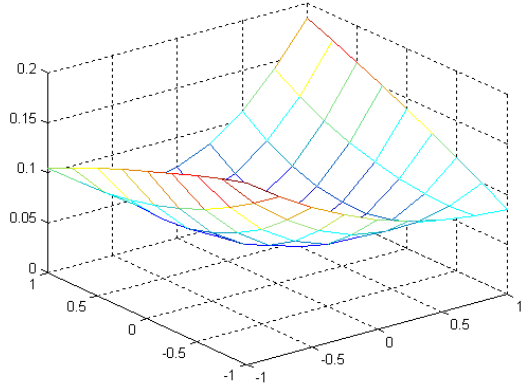| Video Sequence	Average PSNR	Average PSNR improvement above IMV correction		
	HP	QP(1)	HP	Upper bound
Mobile	29.365	0.642	0.816	1.753
Suzie	38.991	0.793	1.120	2.377
Car Phone	33.914	0.456	0.687	1.665

Figure 2 – SAD surface modeled by HP for the same scenario that corresponds to figure 1

summarized in table 2 along with the results of QP(1) and interpolated version for comparison. Table 2 presents average PSNR attained for HP and the PSNR improvement above IMV correction in dB. HP has attained 0.687dB to 1.120dB PSNR improvement above IMV correction, compared to 0.222dB to 0.793dB in QP(1). Table 2 also tabulates the maximum possible PSNR improvement (using interpolated version of SPMV) above IMV correction, for comparison.

5. DISCUSSION AND CONCLUSION

Accurate estimation of SPMV without a block search on expanded images greatly reduces the computational requirement for sub-pel motion estimation. Estimated number of operations per SPMV (calculated to a quarter pixel accuracy) is 183 operations for HP compared to 32 operations for QP and 7920 operations for full block search (around IMV) on expanded images.

The selection of function (3) in HP was quite arbitrary in our proposal. Figure 2 illustrates, however, the SAD surface modeled by function (3) for the same block considered in the illustration of figure 1. It is clear that figure 2 approximates to figure 1b much closely than figure 1a. We believe that by careful selection of the terms in function (3) further improvement in PSNR is possible. The iterative solution for minimizing the derivatives (4) and (5) was also a quite arbitrary choice. More than 50% of the operations of

HP are consumed by this iterative solution and there is scope to much reduce the cost of these computations.

As the results indicate there is much potential to enhance SAD prediction for accurate sub-pel motion estimation and successfully apply for super resolution and video coding applications avoiding expensive block search on expanded images.

6. REFERENCES

[1] W. L. Baylon, D. Panusopone and K. Ming-Ting Sun, "Fast sub-pixel motion estimation and mode decision for H.264," *IEEE International Symposium on Circuits and Systems*, Seattle, WA, pp. 3482-3485, 2008.

[2] J.F. Chang and J.J. Leou, "A Quadratic Prediction Based Fractional-Pixel Motion Estimation Algorithm for H.264," *IEEE International Symposium on Multimedia*, pp. 491-498, 2005.

[3] M.R.H. Fatemi, H.F Ates and R. Salleh, "A Cost-Efficient Bit-Serial Architecture for Sub-pixel Motion Estimation of H.264/AVC," *Int. Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Harbin, pp. 818 – 821, 2008.

[4] T. Koga, K. Iinuma, A. Hirano, Y. Iijima and T. Ishiguro, "Motion compensated inter-frame coding for video conferencing," *Proceedings of the National Telecommunications Conference (NTC)*, Dec. 1981.

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

[6] G. Callico, S. Lopez, O. Sosa, J.F. Lopez and R. Sarmiento, "Analysis of fast block matching motion estimation algorithms for video super-resolution systems," *IEEE Transactions on Consumer Electronics*, vol.54, no.3, pp. 1430 – 1438, Aug. 2008.

[7] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Nov. 2003.

[8] M. Balci and H. Foroosh, "Sub pixel Registration Directly from the Phase Difference," *EURASIP Journal on Applied Signal Processing*, pp. 1–11, 2006.

[9] P. Vandewalle, S. Susstrunk and M. Vetterli, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution," *EURASIP Journal on Applied Signal Processing,* 2006.

[10] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

[11] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Proceeding of IEEE Computer Society Conference on Computer Vision and Patter Recognition (CVPR '88)*, pp. 742–746, June 1988.