OPEN ACCESS

University of BRISTOL

ISCAS 2000 - IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland

SIMPLEX MINIMISATION FOR MULTIPLE-REFERENCE MOTION ESTIMATION

Mohammed E. Al-Mualla, Nishan Canagarajah and David R. Bull

Image Communications Group, Centre for Communications Research, University of Bristol,
Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, United Kingdom,
Tel: +44-117-954 5126, Fax: +44-117-954 5206,
e-mail: {m.almualla, nishan.canagarajah, dave.bull}@bris.ac.uk

ABSTRACT

This paper investigates the properties of the multiple-reference block motion field. Guided by the results of this investigation, the paper proposes three fast multiple-reference block matching motion estimation algorithms. The proposed algorithms are extensions of the single-reference simplex minimisation search (SMS) algorithm. The algorithms provide different degrees of compromise between prediction quality and computational complexity. Simulation results using a multi-frame memory of 50 frames indicate that the proposed multiple-reference SMS algorithms have a computational complexity comparable to that of single-reference full-search while still maintaining the prediction gain of multiple-reference motion estimation.

1. INTRODUCTION

In video coding, the high correlation between successive frames can be exploited to improve coding efficiency. This is usually achieved using *motion compensated prediction* (MCP).

Among existing methods for motion estimation, the *block matching motion estimation* (BMME) algorithm has received considerable attention and has been, implicitly, incorporated into various video coding standards (e.g. MPEG 1–2, H.261 and H.263). This is mainly due to its simplicity and good compromise between prediction quality and motion overhead. In BMME, the current frame is divided into non-overlapping blocks and the motion of each block is represented by a single *two-dimensional* vector $\mathbf{d} = (d_x, d_y)$. This vector is the displacement between the current block and its *best match* block within a search window in a reference frame. This match is usually decided using a *block distortion measure* (BDM) such as the *sum of absolute differences* (SAD).

BMME has always been a bottleneck problem in many video applications, e.g. wireless video terminals and software-based video codecs, especially if real-time video coding is required. For a maximum displacement of $\pm d_m$ pixels, a *full-search* (FS) BMME algorithm performs $(2d_m + 1)^2$ block matches per block. This computational complexity is greater than that of all the remaining encoding steps combined. This has motivated the development of a number of fast BMME algorithms, e.g. [1]–[5].

Recently, a *multiple-reference* BMME (MR-BMME) algorithm has been reported in the literature [6]. In this algorithm, previous frames are assembled in a *multi-frame memory*. For each block in the current frame, BMME is *extended* to search over all reference frames. In addition to the *spatial displacements* (d_x, d_y), the estimated motion vector, $\mathbf{d}$, of the block will now include a *temporal displacement*, d_t, which is the value of the *index* into the multi-frame memory. The technique is reported to provide significantly improved prediction gain. This is, however, at the expense of a significant increase in computational complexity. For each block, *multiple-reference full-search* (MR-FS) performs $R(2d_m + 1)^2$ block matches, where R is the size (in frames) of the multi-frame memory. This increase in complexity calls for further research into the area of reduced complexity motion estimation.

In [4] and [5] we proposed a very efficient BMME algorithm called the *simplex minimisation search* (SMS). In this paper we extend our SMS algorithm to the multiple-reference case. The paper is organised as follows. Section 2 briefly describes the single-reference SMS algorithm. Section 3 investigates the properties of the multiple-reference block motion field. Guided by the results of this investigation, section 4 proposes three multiple-reference SMS algorithms. Section 5 presents the results of testing the proposed algorithms. Finally, section 6 gives some concluding remarks.

2. THE SIMPLEX MINIMISATION SEARCH (SMS)

BMME can be formulated as a *two-dimensional constrained optimisation problem* [4][5]. The two dimensions are the horizontal, d_x, and vertical, d_y, displacements, the function to be optimised (or minimised) is the BDM, and the independent variables are constrained within a limited range, $-d_m \leq d_x, d_y \leq +d_m$, and are usually evaluated to a certain accuracy, e.g. half- or full-pixel accuracy. In this case, the search window represents a *search space* and each possible block within this window is represented by a *search location*. The corresponding BDM values form an *error surface* and the best match block represents the *global minimum* within this surface. The motion vectors resulting from this process form a *block motion field*.

Since BMME is an optimisation problem, then it can be solved with reduced complexity using a wealth of mature optimisation techniques. In [4] and [5] we proposed to solve this problem using the *simplex minimisation* (SM) optimisation method [7]. We called the resulting solution (or algorithm) the *simplex minimisation search* (SMS).

SM, as introduced by Nedler and Mead in [7], is a multidimensional unconstrained optimisation method. A *simplex* is a geometrical figure which consists, in N dimensions, of $N + 1$ vertices and all their interconnecting line segments, polygonal faces,

The authors wish to acknowledge the support of Emirates Telecommunications Corporation (ETISALAT), United Arab Emirates.

This work has been submitted as part of a regular paper to the IEEE Transactions on Circuits and Systems for Video Technology.

etc. A *nondegenerate* simplex is one that encloses a finite inner N-dimensional volume. To minimise a function of N variables, the SM method must be initialised with $N + 1$ search locations defining an *initial* nondegenerate simplex in the search space. The method then takes a series of steps, *reflecting*, *expanding* or *contracting* the simplex from the location where the function value is largest, in an attempt to find a better location. This is repeated until a termination criterion is satisfied. For more details the reader is referred to [7].

The SMS algorithm uses a *two-dimensional constrained* version of the SM method. The initial *three* vertices of the simplex are set to the motion vectors of the blocks to the left of and above the current block, and to the $(0, 0)$ vector. If this does not produce a nondegenerate simplex then a local search is applied around the best of those three. The initial nondegenerate simplex is then chosen from this new set of candidates. The search then proceeds as described in [7] subject to the constraints that any location produced by reflection, expansion or contraction must be *set to the required accuracy* and must also be *set to the nearest location within the search space*, before any BDM evaluation can take place. The search is terminated when *the three vertices of the simplex become neighbours*. The SMS algorithm is described in greater details in [4] and [5].

Simulation results showed that the SMS algorithm outperforms other fast BMME algorithms, providing a better prediction quality, a smoother motion field and a higher speed-up ratio.

3. PROPERTIES OF THE MULTIPLE-REFERENCE BLOCK MOTION FIELD

The design of the SMS algorithm was based on *two* important properties of the block motion fields of typical video sequences:

Property 1 *The distribution of the block motion field is centre-biased. This means that smaller displacements are more probable and the motion vector* $(0, 0)$ *has the highest probability of occurrence.*

Property 2 *The block motion field is smooth and varies slowly. This means that there is a high correlation between the motion vectors of adjacent blocks.*

Before extending the SMS algorithm to the multiple-reference case, it is, therefore, important to ensure that the above two properties still hold for the multiple-reference block motion field. Figures 1–3 summarise the statistical properties of the multiple-reference block motion field (with $R = 50$) and compare them to those of the single-reference field[1].

Figure 1 shows the distribution of the relative frequency of occurrence of the spatial displacement d, where d here refers to both d_x and d_y. The figure indicates that *Property 1* still holds for the multiple-reference case, although longer displacements are slightly more frequent than in the single-reference case.

Figure 2 shows the correlation coefficients between the motion vectors of a block and its top and left neighbours. Again, this figure indicates that *Property 2* still holds for the multiple-reference case.

[1]The figures were generated using the luminance components of 6 QSIF sequences with different motion content and different frame skips. Motion was estimated using full-search full-pixel block matching, SAD as the BDM, 16×16 blocks, ± 15 maximum displacement, restricted motion vectors and original reference frames.

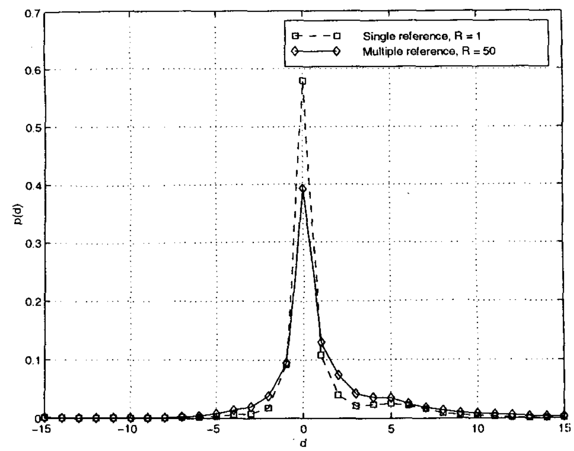


Figure 1: Frequency of occurrence of the spatial displacement, d, for the single-reference, $R = 1$, and multiple-reference, $R = 50$, cases.

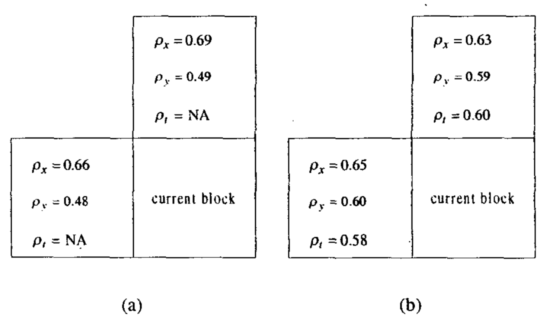


Figure 2: Correlation coefficients between the motion vector of a block and its top and left neighbours in (a) the single-reference, $R = 1$, case and (b) the multiple-reference, $R = 50$, case.

Figure 3 shows the distribution of the relative frequency of occurrence of the temporal displacement d_t. The temporal displacement $d_t = 1$ (which refers to the most recent reference frame in memory) has the highest frequency of occurrence. This property will be exploited in the next section.

4. MULTIPLE-REFERENCE SMS

The above investigation indicates that *moving from a single-reference to a multiple-reference system does not significantly change the properties of the block motion field.* Thus, the SMS algorithm can be extended to the multiple-reference case. We propose three different extensions (or algorithms) as follows.

MR-SMS This is a direct extension of SMS. For each block, the SMS algorithm is used to search each frame in the multi-frame memory and produce a best match from that frame.

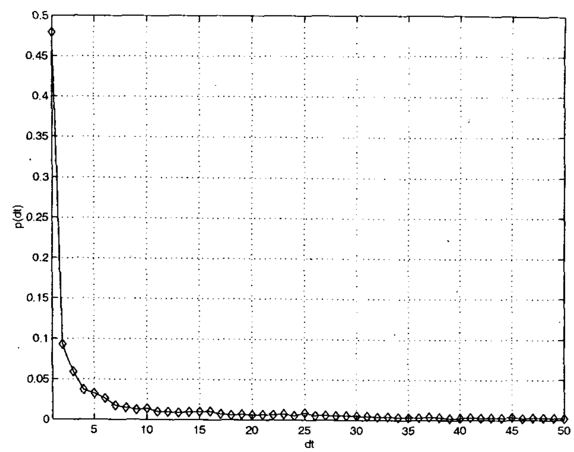


Figure 3: Frequency of occurrence of the temporal displacement, d_t, for the multiple-reference, $R = 50$, case.

The overall best match is then chosen from this set of R candidates.

MR-FS/SMS This is the same as MR-SMS but the most recent frame in memory (i.e. the frame for which $d_t = 1$) is searched using full-search instead of SMS. This is motivated by the observation, made earlier, that the most recent frame has the highest probability of selection. Thus, it must be given more importance.

MR-3DSM The single-reference SMS search is based on a two-dimensional version of the SM optimisation method. Algorithm MR-3DSM, however, is based on a *three-dimensional* version. For each block, the initialisation procedure described in section 2 is applied to each frame in the multi-frame memory. This will generate an initial simplex of three vertices for each frame. The best *four* vertices, in terms of BDM value, are selected from this set of $3R$ candidates. Those four vertices are then used as an initial simplex to a three-dimensional version of the SM optimisation method (the third dimension here is the temporal displacement). The same termination criterion described in section 2 is used, with the added condition that *the four vertices of the final simplex must have the same temporal displacement.*

5. SIMULATION RESULTS AND DISCUSSIONS

The proposed algorithms where tested using the luminance components of three QSIF sequences: AKIYO, FOREMAN and TABLE TENNIS[2]. Note that each sequence represents one of the classes defined by MPEG. AKIYO represents class A: low amount of movement and low spatial detail, FOREMAN represents class B: medium amount of movement and low spatial detail or vice versa, and TABLE TENNIS represents class C: high amount of movement and medium spatial detail or vice versa. The BDM was defined to be the SAD, the block size was set to 16×16 pixels, and the maximum allowed motion displacement was assumed to be ± 15 pixels. The search was performed to full-pixel accuracy and motion vectors were restricted so that they don't point outside the frame. Motion was estimated and compensated using original previous frames. In addition to the proposed multiple-reference SMS algorithms, the single-reference full-search (FS) and the multiple-reference full-search (MR-FS) algorithms were also simulated. A multi-frame memory of size $R = 50$ frames was employed.

Table 1: Comparison between different multiple-reference block matching algorithms in terms of prediction quality (average PSNR in dB) with a multi-frame memory of 50 frames.

	AKIYO	FOREMAN	TABLE
FS	45.93	32.20	32.17
MR-FS	46.55	33.97	32.87
MR-SMS	46.55	33.87	32.67
MR-FS/SMS	46.55	33.92	32.80
MR-3DSM	46.55	33.51	32.46

Table 2: Comparison between different multiple-reference block matching algorithms in terms of computational complexity (average searched locations/frame) with a multi-frame memory of 50 frames.

	AKIYO	FOREMAN	TABLE
FS	65,621	77,439	65,621
MR-FS	3,012,200	3,554,700	3,012,200
MR-SMS	38,880	106,830	69,443
MR-FS/SMS	103,820	183,240	134,270
MR-3DSM	35,867	66,357	45,518

Tables 1 and 2 summarise the results of the simulation. It is immediately evident that the proposed multiple-reference SMS algorithms provide significant reductions in computational complexity compared to the MR-FS algorithm. The proposed algorithms represent different degrees of compromise between prediction quality and computational complexity. At one extreme is the MR-3DSM. Compared to MR-FS, the MR-3DSM provides significant reductions in computational complexity (a speed up ratio of about 54–66) at the expense of a moderate reduction in prediction quality (about 0.41–0.46 dB loss). At the other extreme is the MR-FS/SMS algorithm. It uses full-search on the most recent frame in memory to provide a prediction quality which is almost identical to that of MR-FS (about 0.05–0.07 dB loss) and still achieves moderate reductions in computational complexity (a speed up ratio of about 19–22). Between the two extremes is the MR-SMS. Compared to MR-FS it achieves a speed-up ratio of about 33–43 with only a slight loss in prediction quality (about 0.1–0.2 dB loss). The above observations are further emphasised using Figure 4 which shows the prediction quality for the FOREMAN sequence at different frame skips.

A very interesting point to note is that the computational complexity of the proposed algorithms is comparable (and in some cases less than) that of single-reference FS and yet they still maintain the improved prediction gain of multiple-reference motion es-

[2]Each sequence included 300 frames. AKIYO and TABLE TENNIS have luminance components of 176×120 @ 30 fps, whereas FOREMAN has a luminance component of 176×144 @ 25 fps.

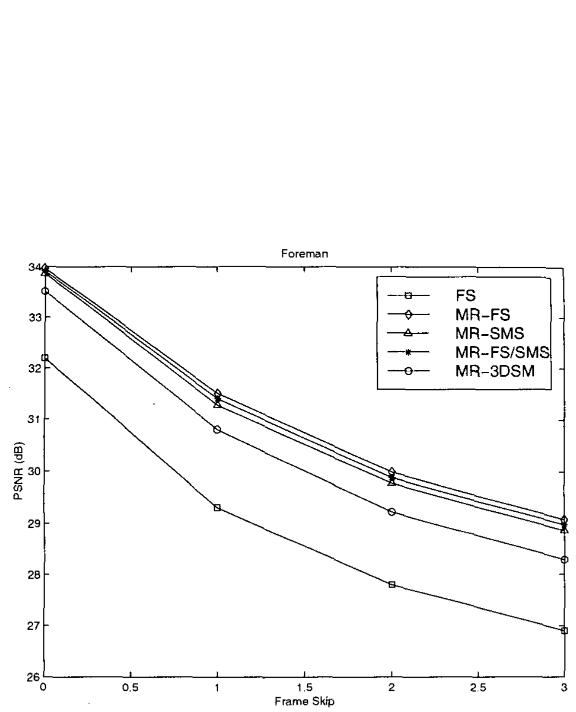
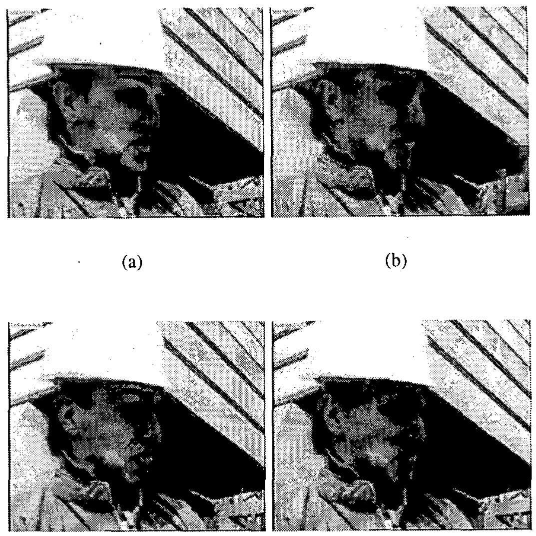


Figure 4: Prediction quality of the FOREMAN QSIF sequence at different frame skips. The multiple-reference (MR) algorithms use a multi-frame memory of 50 frames.



(a) (b)

(c) (d)

Figure 5: Subjective quality of the motion compensated 158^{th} frame of the QSIF FOREMAN sequence @ 25 fps: (a) original frame, (b) compensated using single-reference **FS**; 28.24 dB and 77,439 searched locations, (c) compensated using **MR-FS** with $R = 50$; 31.31 dB and 3,871,950 searched locations, and (d) compensated using **MR-3DSM** with $R = 50$; 31.04 dB and 72,532 searched locations.

timation. This is also illustrated in Figure 5 which shows the subjective quality of the motion compensated 158^{th} frame of FOREMAN. The uncovered background at the bottom-right corner of the frame is poorly compensated using single-reference FS (Figure 5(b)). This uncovered background is compensated with higher quality using the multiple-reference algorithms (Figures 5(c) and 5(d)). While the MR-FS achieves this improved prediction quality at the expense of about 50 orders of magnitude increase in computational complexity, our proposed MR-3DSM algorithm provides a similar improvement at no increase in computational complexity.

6. CONCLUSIONS

In this paper we investigated the properties of the multiple-reference block motion field. We found that moving from a single-reference system to a multiple-reference system does not significantly change the properties of the block motion field. Guided by the results of this investigation we extended the single-reference SMS algorithm to the multiple-reference case. Three multiple-reference SMS algorithms were proposed providing significant reductions in computational complexity compared to the multiple-reference full-search. The proposed algorithms represent different degrees of compromise between prediction quality and computational complexity. Simulation results using a multi-frame memory of 50 reference frames indicated that the proposed multiple-reference SMS algorithms have a computational complexity comparable to that of single-reference full-search while still maintaining the prediction gain of multiple-reference motion estimation.

7. REFERENCES

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

[2] T. Koga, K. Linuma, A. Hirano, Y. Lijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing," in *the National Telecommunications Conference (NTC)*, New Orleans, L. A., USA, November 29–December 3 1981, pp. G.5.3.1–G.5.3.5.

[3] M. Ghanbari, "The cross search algorithm," *IEEE Transactions on Communications*, vol. 38, no. 7, pp. 950–953, July 1990.

[4] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "A fast block matching motion estimation algorithm based on simplex minimisation," in *the IX European Signal Processing Conference EUSIPCO-98*, Island of Rhodes, Greece, 8–11 September 1998, vol. III, pp. 1565–1568.

[5] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "Simplex minimisation for fast block matching motion estimation," *IEE Electronics Letters*, vol. 34, no. 4, pp. 351–352, 19th of February 1998.

[6] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, Feb. 1999.

[7] J. A. Nedler and R. Mead, "A simplex method for function minimisation," *The Computer Journal*, vol. 7, pp. 308–313, 1965.