ADAPTIVE MULTI-REFERENCE DOWNHILL SIMPLEX SEARCH BASED ON SPATIAL-TEMPORAL MOTION SMOOTHNESS CRITERION

Wei-Hau Pan, Chen-Kuo Chiang, Shang-Hong Lai

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
{whp, ckchiang, lai}@cs.nthu.edu.tw

ABSTRACT

Multi-reference frame motion estimation improves the accuracy of motion compensation in video coding. However, it also increases computational complexity dramatically. In this paper, we propose a different approach for multi-reference motion estimation via downhill simplex search. Additionally, an adaptive reference frame selection algorithm is developed based on spatial and temporal smoothness of motion vectors. We first apply single-reference downhill simplex search to the previous frame. Then, temporal smoothness of motion vectors in collocated blocks is calculated to decide the number of reference frames to be included for motion estimation. Spatial smoothness of motion vectors in the neighboring blocks is used as a criterion for termination. Experimental results show that the proposed algorithm provides better PSNR than that of original multi-reference downhill simplex search in all testing sequences with similar computational speed. In addition, it outperforms several representative single-reference frame block matching methods in terms of estimation speed and coding quality.

***Index Terms*—**Motion analysis, Motion measurement, Motion compensation, Video coding, Video codecs.

1. INTRODUCTION

In multi-reference frame motion estimation (ME), the motion vector of one block can be predicted from many reference frames to improve the coding quality. Recently, a number of algorithms have been proposed to reduce the computational complexity. Center-biased frame selection and recent-biased search perform ME with predefined search patterns in 3D space. Su and Sun [1] used composed motion vectors (MVs) to predict approximate results in multi-reference frames. In addition, a simplex minimization method [2] is applied in each previous frame to form an initial simplex for searching the minimal solution of the block distortion function.

Recently, an improved multi-reference downhill simplex search (MRDSS) algorithm [3] was proposed by regarding it as a function minimization problem in a finite 3D space. Experimental results show that fast block-based motion estimation can be achieved by using an efficient function minimization algorithm other than using a predefined search pattern.

In this paper, we provide a different approach to use the downhill simplex search in multiple reference frames and propose a reference frame selection algorithm to improve the accuracy and efficiency of motion estimation. The rest of this paper is organized as follow: Section 2 introduces the downhill simplex search method and the algorithm for multi-reference frame motion estimation. In section 3, we present the frame elimination method, including estimating spatial and temporal correlation and the algorithm to eliminate unnecessary motion search in reference frames. Section 4 gives experimental results and comparison of the proposed algorithm with other block matching algorithms. Finally, we conclude this paper in section 5.

2. DOWNHILL SIMPLEX SEARCH MOTION ESTIMATION ALGORITHM

Downhill simplex search [4] is a derivative-free multidimensional function minimization method. In the downhill simplex search, a collection of n + 1 points in n-dimensional space is called a simplex. In the iterative simplex update process, the point with the highest function value is iteratively replaced by a new point with a smaller function value until the stopping criterion is satisfied.

In [5], a single-reference downhill simplex search (DSS) finds MVs in 2-D space. Three points of the simplex form an initial simplex for the minimization process. In MRDSS, four points are selected to form the initial simplex. It finds motion vectors in 3-D space. Motion search algorithm contains two parts. In the first part, motion trajectories are traced to get approximated MVs in each reference frame. Four candidates with minimal block distortion values are chosen to form an initial simplex. In the second part, four actions; namely reflection, expansion, contraction, and shrinkage, are performed iteratively to search the best solution.

3. PROPOSED ALGORITHM

In this section, we give the details of the proposed motion estimation algorithm. The proposed method modifies the search behaviors in MRDSS and provides a frame elimination method to reduce the computational complexity. Our algorithm first searches MVs in each reference by single-reference downhill simplex search. Then we calculate the temporal and spatial correlation between MVs. Temporal smoothness of motion vectors in the collocated blocks is calculated to decide the number of reference frames to be searched. Spatial smoothness of motion vectors in the neighboring blocks is used as a criterion for termination.

3.1. Modified Multi-Reference Downhill Simplex Search

In [5], single-reference downhill simplex search gives fast search speed and better video quality than many well-known motion estimation algorithms. In MRDSS, a multi-reference downhill simplex search uses four points to search MVs in 3-D space. In order to get more precise results, the single-reference DSS algorithm can be used to individually search each frame in the multiple reference frames and produce a best match block from that frame. The overall best match is then chosen from this set of candidates. However, it increases search numbers in some degree. Here, we analyze the spatial and temporal property of MVs in the previous frames and propose an adaptive reference frame selection algorithm to reduce the search efforts on unnecessary reference frames.

3.2. Motion Smoothness Factor

In [6], motion smoothness factor is used as a measure to see if the MVs are similar. *Spatial Smoothness Factor* (SSF) measures the variation of the MV and its neighboring left, right, top and down MVs. *Temporal Smoothness Factor* (TSF) measures the variation of the MVs in the collocated blocks in all reference frames. The smoother the MV field, the higher motion correlation exists. In the spatial domain, the probability that the current block's MV has a similar value as its neighboring blocks is high when the MVs are very smooth.

3.3. The Reference Frame Determination Algorithm

The goal of frame elimination algorithm is to determine if it is useful to search MVs with more reference frames. Temporal and spatial smoothness factor are exploited to help make decisions. First, for every block on the current frame F_n, DSS is applied on F_{n-1}. The motion vector (mv_x, mv_y) denotes the search result for block matching by using DSS. Then, TSF α is calculated from Θ . Θ is the set of MVs of collocated blocks in each reference frame. mv(n-1, x, y) denotes motion vector of a block at frame n-1 positioned at (x, y). TSF α determines whether the motion in the temporal domain is smooth or not as shown in Figure 1.

$$\alpha = \max\{\, |mv_x - mv'_x| \,,\, |mv_y - mv'_y| \,\} \,,\, (mv'_x, mv'_y) \in \Theta$$

Θ : { mv(x, y, n-1), mv(x, y, n-2), mv(x, y, n-3), mv(x, y, n-4), mv(x, y, n-5)} (1)

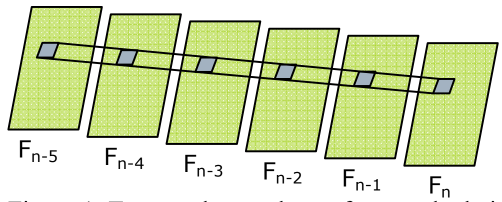

Figure 1. Temporal smoothness factor calculating

Then we decide how many frames will be referred according to TSF α. Ω_b, represents the degree of temporal correlation of MVs, is classified to low, medium or high, as given in equation (2). If Ω_b is high, the MVs in all reference frames are highly correlated which means the motion is very smooth. Therefore, it is enough to search MVs with just one reference frame. If Ω_b is low, it means the motion is chaotic. The block may be on the boundary of objects or occluded in previous frames. It is helpful to search MVs with more reference frames.

$$\Omega_b = \begin{cases} low & \alpha > L_2 & RefNum = 4 \\ medium & L_1 < \alpha \le L_2 & RefNum = 2 \\ high & \alpha \le L_1 & RefNum = 1 \end{cases} \quad (2)$$

After *RefNum* is decided, DSS is applied in reference frame F_{n-2}. SSF β is then calculated as a criterion for termination. Φ is the set of MVs of collocated block and its top, down, left and right blocks as shown in Figure 2.

$$\beta = \max\{\, |\overline{mv_x} - mv'_x| \,,\, |\overline{mv_y} - mv'_y| \,\} \,,\, (mv'_x, mv'_y) \in \Phi$$

$\overline{mv}$ is the average MV of Φ

Φ: collocated and neighboring MVs in the previous frame (3)

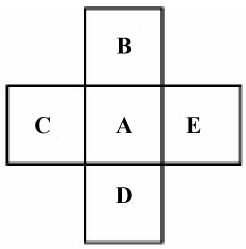

Figure 2. Illustration of Φ

According to β, three actions are taken :

1. *Continue_to_do* : High β value means the MVs of spatial neighboring blocks are chaotic. The spatial correlation of MVs is low. It is helpful to continue DSS in the next reference frame.

2. *Do_one_more* : The spatial smoothness factor of Φ are medium. One more chance to do DSS is given in the next reference frame.
3. *Early_terminate* : The spatial correlation of MVs are high. DSS in more reference frames can be terminated since the MVs of Φ are regular in the next reference frame.

$$\Omega_t = \begin{cases} low & \beta > L_4 & act = Continue\_to\_do \\ medium & L_3 < \beta \leqq L_4 & act = Do\_one\_more \\ high & \beta \leqq L_3 & act = Early\_terminate \end{cases} \quad (4)$$

After finishing motion search, the best MV candidate are chosen by minimal MSE. Figure 3. depicts the flowchart of Motion Smoothness Downhill Simplex Search (MSDSS).

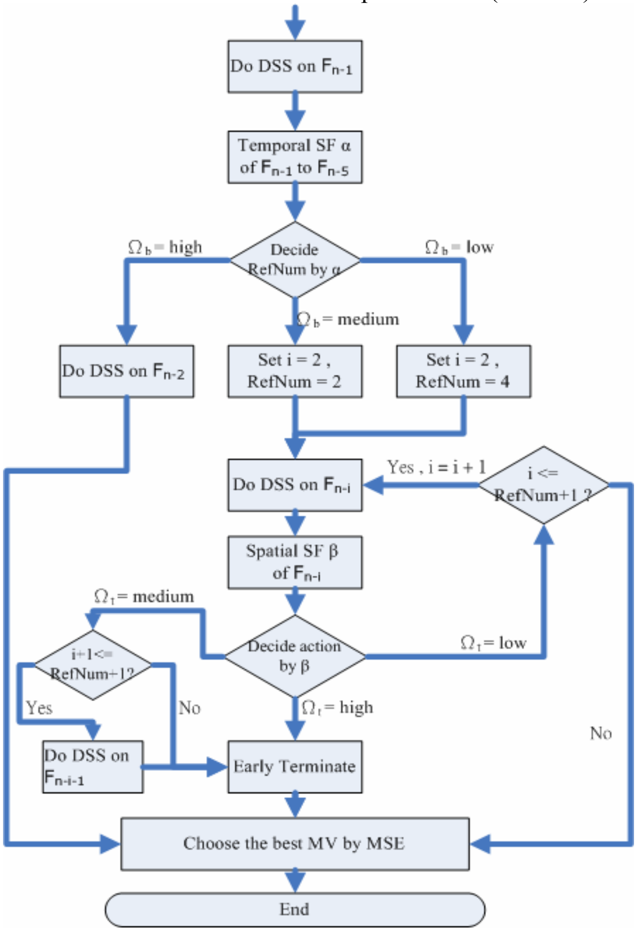

Figure 3. Flowchart of Motion Smoothness Downhill Simplex Search (MSDSS) algorithm

4. EXPERIMENTAL RESULTS

We compared six block matching algorithms, including full search (FS), four step search (FSS), diamond search (DS), simplex minimization search (SMS), downhill simplex search (DSS), multi-reference full search (MRFS), multi-reference frame downhill simplex search (MRDSS) with the proposed adaptive multi-reference frame downhill simplex search using motion smoothness (MSDSS) through experiments on four benchmarking video sequences (mobile, coastguard, news and stefan).

The mobile sequence contains complicated background textures and medium speed motions. The stefan sequence is a popular video because it contains different motion directions and large motions in the video. The coastguard sequence contains fast movement through the whole sequence. The news sequence almost remains static in most areas except the small area around the human face. The test video sequences are selected because they represent different types of video motions. In our experiments, all the video sequences are of QCIF format. For different sequence length, we compute the average number of search locations and the PSNR for each frame. Before evaluating the performance of the proposed algorithms, the reflection coefficient α, the contraction coefficient β and the expansion coefficient γ in DSS are set to 1 、0.5 、2. The threshold L1 、L2 、L3 and L4 are set to 1 、7 、4 and 8, respectively. In our experiments, the block distortion measurement was defined to be SSE, the block size was set to 16-by-16 pixels, the maximal allowed motion displacement was set to 16 pixels horizontally and vertically, and the search was performed to full-pixel accuracy. The experiments focus on the estimation speed and the prediction quality.

The term "Numbers of search locations" is calculated to indicate the motion estimation speed. In the search of motion vectors for one macroblock, the number of search locations is increased by 1 when a candidate block is chosen and the SSE is accumulated for this block. At last, the average number of search locations for all blocks per frame shows the estimation speed for a specific search algorithm. In addition, PSNR is calculated for the compressed video quality assessment. In other words, motion is estimated and compensated using the original, rather than the reconstructed, reference frame for each frame. This provides a particularly fair comparison between the algorithms on a frame-by-frame basis since poor prediction of one frame does not propagate to the next frame.

In our experiment, the total number of reference frames is set to 5, which complies with H.264 coding standard. As shown in Table 1, our proposed MSDSS provides significant reduction in computational cost. Compared with MRFS, a speed-up ratio by using our algorithm ranges from 217 to 360 for different types of video sequences in our experiments. The speed-up ratio R is defined as the ratio between the total number of search locations in MRFS and the total number of search locations in MSDSS. Table 2 shows the PSNR. The proposed MSDSS method outperforms the original MRDSS in all sequences. It is reasonable because applying DSS individually to each

Table 2. Comparison of PSNR values for different motion estimation methods

BMA	Mobile		Coastguard		News		Stefan	
	PSNR	Diff	PSNR	Diff	PSNR	Diff	PSNR	Diff
FS	26.39	0	33.25	0	37.64	0	27.53	0
FSS	26.31	-0.08	33.13	-0.12	37.61	-0.03	27.10	-0.43
DS	26.29	-0.1	33.17	-0.08	37.61	-0.03	26.69	-0.84
SMS	26.29	-0.1	32.53	-0.72	37.54	-0.1	25.66	-1.87
DSS	26.36	-0.03	33.23	-0.02	37.59	-0.05	27.23	-0.3
MR-FS	28.20	0	33.61	0	37.80	0	27.99	0
MR-DSS	27.05	-1.15	33.41	-0.2	37.65	-0.15	27.44	-0.55
MS-DSS	27.36	-0.84	33.58	-0.03	37.71	-0.09	27.68	-0.31

Table 1. Computation speedup ratio R
(R = # Location MRFS / # Location MSDSS)

Sequence	Resolution	Speedup Ratio R
Mobile	QCIF, 240 frames	328.41
Coastguard	QCIF, 97 frames	299.22
News	QCIF, 240 frames	360.20
Stefan	QCIF, 300 frames	217.28

reference frame usually provides more accurate estimation results. However, it also takes more computational efforts. Here, an effective reference frame number selection algorithm is introduced to reduce unnecessary motion search based on spatial and temporal motion smoothness measures. Hence, it always provides better coding quality without increasing search numbers. Table 3 shows the number of search locations. Our proposed MSDSS algorithm has comparable search speed with original MRDSS while maintaining even better PSNR than that of MRDSS. In fact, our proposed method may compete with many single-reference block matching algorithms from the aspect of speed in some cases and it can achieve higher PSNR.

5. CONCLUSION

In this paper, a modified multi-reference downhill simplex search algorithm for fast motion estimation is proposed. Downhill Simplex Search is applied to each reference frame individually instead of a 3-D search version in MRDSS. Spatial-temporal smoothness of MVs is exploited as a criterion to avoid unnecessary motion search. First, the temporal smoothness factor is used to determine how many reference frames it will use. Secondly, spatial smoothness factor is computed to decide when to terminate motion search in multiple reference frames. Experimental results on different types of video sequences show the superior performance of the proposed modified downhill simplex search algorithm over the original MRDSS and some well-known fast motion estimation methods.

6. ACKNOWLEDGEMENTS

Table 3. Speed comparison of different motion estimation methods

BMA	Mobile	Coastguard	News	Stefan
	Location	Location	Location	Location
FS	82104.00	82104.00	82104.00	82104.00
FSS	1429.27	1507.57	1255.12	1952.76
DS	1368.41	1054.94	965.75	2268.09
SMS	1486.73	1478.89	1045.95	1501.34
DSS	536.46	548.96	515.96	638.95
MR-FS	387195	387195	387195	387195
MR-DSS	1377.76	1317.28	1035.48	1455.06
MS-DSS	1179.12	1294.79	1069.30	1782.86

This work was supported by MOEA research project under grant 95-EC-17-A-01-S1-034 and National Science Council, under grant 95-2220-E-007-028.

7. REFERENCES

[1] Y. Su and M.T. Sun, "Fast multiple reference frame motion estimation for H.264," IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, Vol.1, pp. 695-698, Jun. 2004.

[2] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "Simplex minimization for multiple-reference motion estimation," IEEE Trans. Circuits Syst. Video Techn, Vol.11, pp. 1209-1220, 2001.

[3] C. -K. Chiang, S.-H. Lai, "Fast Multi-Reference Frame Motion Estimation via Downhill Simplex Search," IEEE International Conference on Multimedia and Expo, Toronto, Canada, Jul. 2006.

[4] J.A. Nelder and R. Mead, "A simplex method for function minimization," The Comput. J., Vol. 7, pp. 308-313, 1965.

[5] H.-C. Fei, C.-J. Chen, S.-H. Lai, "Enhanced downhill simplex search for fast video motion estimation," Proc. Pacific Rim Conference on Multimedia, Jeju Island, Korea, pp. 512-523, 2005.

[6] Y. Liang, Ishfaq Ahmad, J. Luo, Y. Sun, and Viswanathan Swaminathan, "On Using Hierarchical Motion History For Motion Estimation in H.264/AVC," IEEE Trans. Circuits Syst. Video Techn, Vol.15, pp. 1594-1603, 2005.