AN IMPROVED FAST FEATURE-BASED BLOCK MOTION ESTIMATION

Yiu-Hung Fok* and Oscar C. Au**

Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
Email: eeyhfok@uxmail.ust.hk*
eeau@uxmail.ust.hk**

ABSTRACT

A fast block matching algorithm in the feature domain was proposed by Fok and Au which achieves a computation reduction factor of $N/2$ with search block size of $N \times N$. This paper presents an improved fast block matching algorithm in the integral projections feature domain. This algorithm is based on the fast block matching algorithm in the feature domain and a new motion field sub-sampling scheme. By utilizing the properties of the features and the motion field sub-sampling scheme, with a search block size of $N \times N$, the improved algorithm can achieve a computation reduction factor of N while retaining close-to-optimal performance in the mean absolute difference (MAD) sense.

1. INTRODUCTION

Video compression is vital and necessary for transmission and storage of video sequences because of the huge data rate of the raw video data. There are many video compression standards, such as CCITT H.261 and MPEG, which use block-based motion estimation and compensation to reduce the temporal redundancy in video sequences. Conventionally, motion estimation is done by performing block matching exhaustively in the search area which is optimal with respect to the mean absolute difference (MAD) but it requires extremely large amount of computation. Therefore, different fast block matching algorithms had been proposed[2]-[7]. A fast block matching algorithm in feature domain (FBMA) was proposed in [1] in which there was a reduction of computation complexity by a factor of $N/2$ when compared with the conventional exhaustive search (CES) algorithm. Here the basic search blocks are of size $N \times N$. This reduction factor of $N/2$ is significant.

Another fast block matching algorithm called Subsampled Motion-Field Estimation with Alternating Pixel-Decimation Patterns (SMAP) was proposed in [3]. It first uses Alternating Pixel-Decimation Patterns[3] to achieve a computation reduction by a factor of four. It then applies Subsampled Motion-Field Estimation [3] to achieve another computation reduction by a factor of two. Therefore, SMAP can achieve a computation reduction factor of 8, independent of the block size.

FBMA first calculates the Horizontal Integral

Projections (HIP) and the Vertical Integral Projections (VIP) of the current frame k . The HIP and the VIP at the (x,y) -th location of the k -th frame are defined in equation (1) and (2) respectively. They are:

$$H_k(x, y) = \sum_{n=0}^{N-1} f_k(x+n, y) \quad (1)$$

$$V_k(x, y) = \sum_{n=0}^{N-1} f_k(x, y+n) \quad (2)$$

where $f_k(x,y)$ is the pixel value at the (x,y) -th location of frame k .

The best R search locations are found according to a distance measure which was defined as MAD in the HIP and the VIP feature domains defined in equation (3).

$$MAD_f(i, j) = \sum_{n=0}^{N-1} |H_k(x, y+n) - H_{k-1}(i, j+n)| + |V_k(x+n, y) - V_{k-1}(i+n, j)| \quad (3)$$

where (i,j) is the search location in the search area of the previous frame and (x,y) is the location of the block in the current frame.

In order to improve the matching result, these R locations are re-examined using another distance measure, the MAD in the pixel domain given in equation (4). The one with the least MAD is declared the best match. This step is called Re-examination[1].

$$MAD_p(i, j) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} |f_k(x+m, y+n) - f_{k-1}(i+m, j+n)| \quad (4)$$

In this paper, a new fast block matching algorithm (NFBMA) is presented. This algorithm makes use of FBMA and a different scheme of subsampled motion-field estimation. With a block of size $N \times N$, this algorithm can achieve a computation reduction factor of

N instead of $N/2$ of both FBMA and SMAP while having close-to-optimal performance in the MAD sense.

2. ALGORITHM

The new improved fast block matching algorithm, NFBMA, is based on FBMA and a new subsampled motion-field estimation scheme. This is described below:

Step 1. We use FBMA to estimate the motion vectors for half of the blocks in a frame as shown in Figure 1(a).

Figure 1: Pattern of Motion Vector Estimation.

Step 2. The motion vectors for the other half of the blocks are estimated using the motion vectors of the neighbouring blocks. For instance, as shown in Figure 1(b), there are five initial estimated motion vectors for block C: the motion vectors for block A, B, D, E and zero (assuming block C is stationary). In addition, we allow a displacement of ± 1 pixel either in the horizontal or vertical directions of these initial estimates. So, we estimate a motion vector for block C by calculating the MAD in the pixel domain of these 25 motion vectors.

Step 3. We use the estimated motion vector found in step 2, to calculate the MAD of the estimated block C. If it is larger than the MAD of the estimated block A, B, D and E, an exhaustive search in pixel domain is applied to find the motion vector for block C. Otherwise, we use the motion vector found in step 2 as the motion vector for block C.

3. COMPUTATIONAL COMPLEXITY

For an $Q \times P$ frame, with a block size of $N \times N$ and maximum displacement of $\pm W$ pixels in both the horizontal and the vertical directions, the number of additions needed for motion estimation of one frame using CES is:

$$N_{CES} = \frac{PQ}{N^2} (2W+1)^2 (N^2 - 1) \quad (5)$$

Let R be the number of locations used in re-

examination in FBMA and NFBMA. From [1], the additions needed for motion estimation of one frame using FBMA is:

$$N_{FBMA} = \frac{PQ}{N^2} [(2W+1)^2 (2N-1) + R(N^2 - 1)] \quad (6)$$

Since the calculation of integral projections is insignificant compared to that of motion estimation, for $W \gg N \gg R > 1$, the computation reduction factor of FBMA is:

$$\frac{N_{CES}}{N_{FBMA}} \approx \frac{N^2 - 1}{2N - 1} \approx \frac{N}{2} \quad (7)$$

For SMAP, at different search location, it uses different 4:1 sub-sampled block to calculate the block distance[3]. Moreover, the motion vectors of half of the blocks in the frame are interpolated using the motion vectors of the neighbouring blocks. The computation for interpolation is insignificant compared to that of motion estimation. Therefore, the number of additions needed for motion estimation using SMAP is:

$$N_{SMAP} = \frac{PQ}{2N^2} (2W+1)^2 \left(\frac{N^2}{4} - 1 \right) \quad (8)$$

The computation reduction factor for SMAP is:

$$\frac{N_{CES}}{N_{SMAP}} = \frac{N^2 - 1}{\frac{1}{2} \left(\frac{N^2}{4} - 1 \right)} \approx 8 \quad (9)$$

Notice that, it is independent of the size of the search block.

For the case of NFBMA, half of the blocks are estimated using FBMA and the motion vectors of the remaining blocks are estimated using those of the neighbouring estimated blocks. If step 3 of NFBMA is rarely fired, the computation for motion-field interpolation is insignificant compared to that of motion estimation. So, the number of addition needed is:

$$N_{NFBMA} = \frac{PQ}{2N^2} [(2W+1)^2 (2N-1) + R(N^2 - 1)] \quad (10)$$

Therefore, the computation reduction factor of NFBMA is:

$$\frac{N_{CES}}{N_{NFBMA}} \approx \frac{N^2 - 1}{\frac{1}{2} (2N - 1)} \approx N \quad (11)$$

The following table summarizes the computation reduction factor of FBMA, SMAP and NFBMA.

Algorithm	Computation Reduction Factor
FBMA	$N/2$
SMAP	8
NFBMA	N

Table 1: Computational reduction factors.

4. SIMULATION RESULTS

Our algorithm was simulated using the luminance component of 200 frames of the "Football" and the "Tennis" sequences. Each frame contained 352×240 pixels quantized uniformly to 8 bits. The size of the block was 16×16. The maximum displacement in the search space was ±16 pixels in both the horizontal and the vertical directions. For re-examination in each algorithm, we used five best candidates in the feature domains.

Figure 2 and 3 show the MAD and Peak Signal-To-Noise Ratio (PSNR) between 31 estimated frames and 31 original frames of the "Football" sequence started from the 80th frame. While Figure 4 and 5 show the MAD and the PSNR between the estimated frames and the original frames of "Tennis" sequence from the 50th to the 80th frame. They show the general profiles of different algorithms. The PSNR of the estimated frame is defined as:

$$PSNR = -10 \log_{10} \frac{\sum_{y=0}^{P-1} \sum_{x=0}^{Q-1} [f(x,y) - \tilde{f}(x,y)]^2}{Q \times P \times 255^2} \quad (12)$$

where $f(x,y)$ and $\tilde{f}(x,y)$ are the original frame and the estimated frame of size $Q \times P$ respectively.

Table 2 shows the percentage of MAD of different algorithms on the average larger than that of the CES and Table 3 shows the percentage of PSNR of different algorithms on the average smaller than that of the CES.

Algorithm	Football sequence	Tennis sequence
FBMA	2.5%	1.99%
SMAP	6.95%	3.59%
NFBMA	2.99%	1.61%

Table 2: MAD larger than CES.

Algorithm	Football sequence	Tennis sequence
FBMA	0.74%	0.53%
SMAP	2.85%	2.04%
NFBMA	0.97%	0.60%

Table 3: PSNR smaller than CES.

As expected, the CES is optimal in the MAD sense. Simulation showed that, with a block of size $N \times N$, NFBMA can achieve close-to-optimal MAD performance with a computation reduction factor of N approximately.

Notice that, in 23 out of 200 frames of the "Football" sequence, the PSNR of NFBMA can even be larger than or equal to those of the CES. Moreover, in 41 out of 200 frames of the "Tennis" sequence, the PSNR of NFBMA are larger than or equal to those of the CES.

For the "Football" sequence, on the average 8.6% of the blocks fire step 3 of NFBMA while there are 7.03% of the blocks, on the average, fire step 3 of NFBMA for the "Tennis" sequence. However, step 3 of NFBMA can generate better results for, on the average, 3.85% and 1.18% of the blocks of the "Football" and "Tennis" sequences respectively. In these cases, the computation reduction factor of NFBMA is about $0.9N$ instead of N .

Figure 6 shows the estimated 82th frames of CES, FBMA, SMAP and NFBMA of the "Football" sequence. In addition, Figure 7 shows the estimated 82th frames of CES, FBMA, SMAP and NFBMA of the "Tennis" sequence. These frames are of lots of motions. In terms of subjective image quality, NFBMA performs better than SMAP. Also, the subjective image quality of NFBMA are very closed to that of FBMA while both are as good as that of CES.

5. CONCLUSION

In this paper, we present an improved fast block matching algorithm in the feature domain for motion estimation. This algorithm utilizes the properties of the integral projections features and motion field sub-sampling. With block size of $N \times N$, this algorithm can achieve a computation reduction factor of N which is significant. Simulations show that this algorithm can achieve close-to-optimal MAD performance with good subjective image quality. However, it needs more computation when step 3 of the improved algorithm fires. There is a trade-off between the computation needed and the accuracy of the estimation.

6. ACKNOWLEDGMENTS

This project is supported in part by the HKTIIT grant #HKTIIT 92/93.001 and in part by the UPGC grant #RI92/93.EG02.

7. REFERENCES

- [1] Y. H. Fok and O. Au, "A Fast Block Matching Algorithm in Feature Domain," Proc. of IEEE Workshop on Visual Signal Processing and Communications '93, Melbourne, 21-22 September 1993, pp. 199-202.
- [2] J. S. Kim and R. H. Park, "A fast feature-based block matching algorithm using integral projections," IEEE Journal on Selected Areas in Communications, vol. 10, no. 5, June 1992, pp. 968-979.
- [3] B. Liu and A. Zaccarin, "New Fast Algorithms for the Estimation of Block Motion Vectors," IEEE Trans. Circuits and Systems for Video Technology, vol. 3, No. 2, April 1993, pp. 148-157.
- [4] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," IEEE Trans. Commun., vol. COM-29, December 1981, pp. 1799-1808.
- [5] S. Kappagantula and K. R. Rao, "Motion Compensated Interframe Image Prediction," IEEE Trans. Commun., vol. COM-33, no. 9, September 1985, pp. 1011-1015.
- [6] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing," in Proc. Nat. Telecommun. Conf., New Orleans, L. A., Nov. 29-Dec. 3, 1981, pp. G5.3.1-5.3.5.
- [7] Q. Wang and R. J. Clarke, "Motion estimation and compensation for image sequence coding," Signal Processing: Image Communication, vol. 4, 1992, pp. 167-174.

Figure 2: MAD for Football sequence.

Figure 3: PSNR for Football sequence.

Figure 4: MAD for Tennis sequence

Figure 5: PSNR for Tennis sequence.

Figure 6: The Estimated 82th frame for Football sequence using (a)CES, (b)FBMA, (c)SMAP and (d)NFBMA.

Figure 7: The Estimated 82th frame for Tennis sequence using (a)CES, (b)FBMA, (c)SMAP and (d)NFBMA.