

COMBINE THE MULTIPLE CANDIDATES SEARCH WITH ONE-BIT TRANSFORM FOR MOTION ESTIMATION

Peter H. W. Wong *, Oscar C. Au **, A. Tourapis***

Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong.

Email: eepeter@ee.ust.hk *, eeau@ee.ust.hk**, alexis@ee.ust.hk***
Tel.: +852 2358-7053**

ABSTRACT

Motion estimation using one-bit information (1BT) to represent a pixel or a group of pixels becomes popular in recently years [1-2] as the distortion function between the reference block and the search block can be computed very efficiently. In [1], each frame is filtered and compared with the original frame to extract the edge information. However, its performance is not good enough as it degrades the predicted image by 1dB on the average comparing with full search. In this paper, we improve the performance using the results of one-bit transform as the initial estimation of the motion vector, and use re-examination strategies to find the final motion vector. Experiment results show that the proposed method improves the performance of the one-bit transform range up to 0.9 dB and it outperforms other motion estimation algorithms such as the Three-Step Search (3SS), New Three-Step Search (N3SS) and the Four-Step Search (4SS).

1. REVIEW OF MOTION ESTIMATION USING ONE-BIT TRANSFORM

For convenience, we use similar symbols in [1]. Each 8bit frame F is filter with a 17×17 kernel K which is a bandpass filter and given by (1). The filtered frame F' is then compared with the original frame F to create a 1-bit frame G . The creation of G is formulated in (2).

$$K_{i,j} = \begin{cases} \frac{1}{25}, & \text{if } i, j \in \{0, 4, 8, 12, 16\} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

$$G_{i,j} = \begin{cases} 1, & \text{if } F_{i,j} \neq F'_{i,j} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Sum of exclusive OR (SXOR) is used as the distance measure instead of Sum Absolute Difference (SAD). Let each block consists of $N \times N$ pixels and u_{ij} and v_{ij} denote the value of the pixel at row i and column j of block u and v respectively. The distance measure between two blocks is formulae as (3).

$$\|u, v\|_1 = \sum_{i,j} u_{i,j} \oplus v_{i,j} \quad (3)$$

where \oplus denotes the exclusive-or operation. For each reference block in a frame to be predicted, it search within a searching window to find a best match which has the least distortion from the reference block.

2. MULTIPLE CANDIDATES RE-EXAMINATION

We observe that the performance of 1BT can be significantly improved if we re-examine the best N results from the 1BT using other approach instead of using the best result. This can be explained that the 1BT limits the variation of small searching region. We propose the multiple candidates one-bit transform (M1BT) and the multiple candidates one-bit transform – full search (M1BTFS) approach to improve the performance of 1BT by re-examination process.

The M1BT is simply to re-examine the best N results obtained from the 1BT. The M1BTFS consists of two stages. In the first stage, the distortion for the best result obtained from by the 1BT is first compared with a threshold. If the distortion is greater than the threshold, we examine the second best result, if the distortion is still above the threshold, we examine the next best result and so on. This stage will stop when either the distortion is below the threshold or the best N motion vectors are examined. If the minimum distortion obtained after the best N candidates are examined is still above twice of the threshold, the process proceeds to the second stage at which Full Search is employed. The threshold value is the average distortion of the last frame.

2.1. Distortion Measure

To measure the distortion between the blocks, we compute the sub-sampled Sum Absolute Difference (SSAD) of the blocks. We used the checker board pattern to reduce the computational complexity by a factor of 2. Mathematically, for two blocks u and v , let u_{ij} and v_{ij} denote the intensity of a pixel in the i -th row and j -th column of u and v

respectively, the distortion D between two blocks is

$$D = \sum_{(i \bmod 2) \neq (j \bmod 2)} |u_{ij} - v_{ij}| \quad (4)$$

computed as following equation.

Where $(k \bmod 2)$ is the reminder when k is divided by 2.

3. SIMULATIONS AND RESULTS

We applied the M1BT and M1BTFS on two testing sequences 'Tennis' and 'Football'. Each frame consists of 360x240 pixels. The first 88 frames of 'Tennis' are considered while the first 99 frames of 'Football' are considered. The search window has the size of 7x7 pixels relative to the location of the reference block. The average Mean Square Error (MSE) against N is shown in Fig. 1. The M1BT using the full resolution is labeled as M1BT1 while for the checker board sub-sampling pattern is labeled as M1BT2. For the M1BTFS, a checker board sub-sampling is used to reduce the computation by a factor of two. We choose $N=6$ for M1BT and M1BTFS. Table 1 and 2 list the average (MSE) and the difference in PSNR from Full Search (FS) against different algorithm. The relative computation complexities for different algorithm are also listed in the tables. Note that, if the relative computation complexity of FS is 225, the relative complexity of 1BT is approximately 1 according to [1]. This computation is included in the values for M1BT1, M1BT2 and M1BTFS.

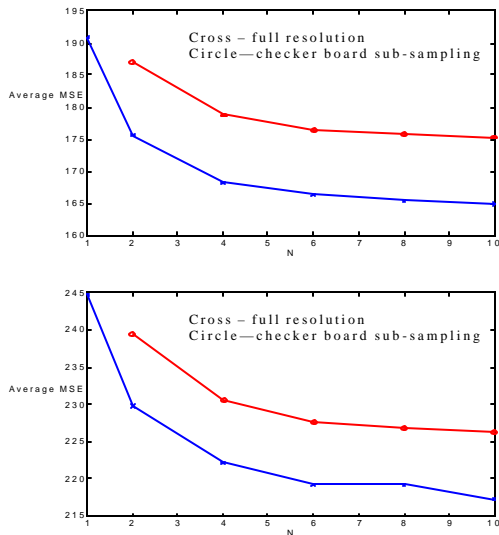


Figure 1 Average MSE against the number of candidates (Top: Tennis, bottom: Football)

4. CONCLUSIONS

We proposed a hybrid approach for motion estimation using the 1-bit transform. Initially, the motion vector is estimated using the 1-bit transform and the motion vector is then refined in the re-examination process. Experimental results show that the proposed M1BT can achieve good estimation while maintaining very low complexity and the M1BTFS can achieve close to optimal estimation with some increasing in complexity. However, the complexity is still low compared with other well known algorithms.

Method	MSE	Relative complexity	PSNR to FS (dB)
FS	148.47	225	0
1BT	190.85	1 (approx.)	1.09
3SS	208.28	25	1.47
N3SS	183.57	22.53	0.92
4SS	177.32	21.38	0.77
M1BT1	166.38	7	0.49
M1BT2	176.47	4	0.75
M1BTFS	154.62	12.43	0.18

Table 1 Comparison between different methods for the sequence 'Tennis'

Method	MSE	Relative complexity	PSNR to FS (dB)
FS	201.91	225	0
1BT	244.77	1 (approx.)	0.84
3SS	225.75	25	0.48
N3SS	215.63	20.79	0.29
4SS	222.11	19.38	0.41
M1BT1	219.15	7	0.31
M1BT2	227.70	4	0.52
M1BTFS	211.15	14.29	0.19

Table 2 Comparison between different methods for the sequence 'Football'

REFERENCE

- [1] B. Natarajan, V. Bhaskaran, and K. Konstantinides, "Low-Complexity Block-Based Motion Estimation via One-Bit Transforms," *IEEE Trans. on Circuits System for Video Technology*, vol. 7, No 4, pp. 702-706, August 1997
- [2] Xiaobing Lee, and Ya-Qin Zhang, "A Fast Hierarchical Motion-Compensation Scheme for Video Coding Using Block Feature Matching" *IEEE Trans. on Circuits System for Video Technology*, vol. 6, No. 6, pp. 627-635, Dec 1996.