Fast Full-Search Motion Estimation Method Based On Fast Fourier Transform Algorithm

Elena I. Zakharenko, Evgeniy A. Altman

Omsk State Transport University
{zaxarenko.elena, altmanea}@gmail.com

Abstract. Motion estimation (ME**)** is used extensively in video codecs based on MPEG-4 standards to remove interframe redundancy. Motion estimation is based on the block matching method which evaluates block mismatch by the sum of squared differences (SSD) measure. Winograd's Fourier transform is applied and the redundancy of the overlapped area computation among reference blocks is eliminated in order to reduce the computational amount of the ME. When the block size is $N \times N$ and the number of reference blocks in a search window is the same as the current block, this method reduces the computational amount (additions and multiplications) by 58 % of the straightforward approach for $N = 8$ and to 81 % for $N = 16$ without degrading motion tracking capability. The proposed fast full-search ME method enables more accurate motion estimation in comparison to conventional fast ME methods, thus it can be applied in video systems.

Keywords: motion estimation, fast Fourier transform, correlation, convolution theorem, full-search, video encoding.

1 Introduction

The popularity of video as a mean of data representation and transmission is increasing. Hence the requirements for a quality and size of video are growing. High visual quality of video is provided by coding. In 1960s the motion estimation (ME) and compensation were proposed to improve the efficiency of video coding [1].

The current frame is divided into non-overlapping blocks. For each block of the current frame the most similar block of the reference frame within the limited search area is found. The criterion of the similarity of the two blocks is called a metric comparison of the two blocks. The position of the block, for which an extremum of metric is founded, determines the coordinates of the motion vector of the current block.

The full search algorithm is the most accurate method of the block ME, i.e. the proportion of true motion vectors found is the highest [2]. The current block is compared to all candidate blocks within the restricted search area in order to find the best match.

This ME algorithm requires a lot of computing resources. Therefore, a lot of alternative fast motion estimation algorithms were developed. In 1981 T. Koga and other authors proposed a three-step search algorithm (TTS) [3].

The disadvantage of fast search methods is finding a local extremum of a function of the difference of two blocks. Consequently motion estimation degrades by half degradation in some sequences compared to brute-force and visual quality of video degrades as well [4].

2 The Criterion To Compare Blocks

The standards of video coding do not regulate the choice of criterion for matching two blocks (metric).

One of the most popular metrics is the sum of square difference (SSD):

$$SSD(i,j)=\sum_{y=0}^{N_h-1}\sum_{x=0}^{N_w-1}(B(x,y)-S(x+i,y+j))^2, \quad (1)$$

where i, j – the coordinates of the motion vector of the current block, $i \in (-V_w/2; V_w/2)$, $j \in (-V_h/2; V_h/2)$, where $V_w \times V_h$ – size of the area which can be is the upper left corner of the title block on the reference frame; x, y – coordinates of the current block B; $N_w \times N_h$ – block size B; S – reference area of size $S_w \times S_h$, where $S_w = N_w + V_w$, $S_h =$
$= N_h + V_h$; B and S – luminance images in color format YUV.

Inside the search area size $S_w \times S_h$ is the minimum value of SSD criterion for the current block B, which determines the coordinates of the motion vector in order.

SSD can be calculated through fewer number of operations by decomposition into three components [5]:

$$\sum_{y=0}^{N_h-1}\sum_{x=0}^{N_w-1}B^2(x,y)- \quad (2)$$

$$-\sum_{y=0}^{N_h-1}\sum_{x=0}^{N_w-1}B(x,y)S(x+i,y+j)+ \quad (3)$$

$$+\sum_{y=0}^{N_h-1}\sum_{x=0}^{N_w-1}S^2(x+i,y+j), \quad (4)$$

In [5, 6] for the (3) computation were used Fast Fourier Transform (FFT). We propose to replace this algorithm by other fast transforms: Winograd algorithm and the number-theoretic transform of Farm (NTT).

3 Research Results

A programming model for block motion estimation using SSD metrics was implemented in Matlab in order to analyze the effectiveness of the fast Fourier transform algorithm for computing the block matching criterion.

Expression (4) is calculated using the algorithm described in [2].

To analyze the effectiveness of the algorithms of fast Fourier transform (FFT) for motion estimation were considered Cooley-Tukey and Winograd algorithms. We selected *B* block 16 × 16 and 8 × 8 pixels, a reference area *S* is twice as much as *B* block, i.e. is 32 × 32 and 16 × 16, respectively.

Operating time of algorithm is significant for processing and analysis of video, it depends on its computational complexity. The number of arithmetic operations (multiplications (×) and additions (+)) is significant when the complexity of the method for motion estimation is measured. Therefore, this criterion was chosen as the metric calculation efficiency of SSD.

Table 1 shows the results for a block size of 16 × 16 pixels and the reference area *S* 32 × 32 dots.

Table 1. The computational complexity of algorithms SSD when $S = 2B = 32$ for one of the current block

Algorithms	Real arithmetic operations								In total	The complexity, %
	expression (2)		expression (3)		expression (4)		in total			
	×	+	×	+	×	+	×	+		
SSD according to (1), FS	–	–	–	–	–	–	65 536	130 816	196 352	100
SSD according to (1), TTS	–	–	–	–	–	–	2 560	5 110	7 670	4
SSD according to (2), (3) and (4)	256	255	65 536	65 280	65 536	65 280	131 328	130 815	262 143	134
Winograd FFT	0	0	7 203	25 923	1 426	2 115	8 629	28 038	36 667	19
Cooley-Tukey FFT	0	0	23 552	33 792	1 426	2 115	24 978	35 907	60 885	31

The results for block 8×8 and the search area 16×16 similar to presented in Table 1. Winograd algorithms reduces the computational complexity of the ME algorithm by 58 % for block size 8×8.

Having analyzed the results shown in Tables 1, it is obvious that the choice of convolution algorithm has a significant impact on the computational complexity of motion estimation.

For the hardware implementation of a significant impact on the area occupied on the chip, providing multiplication. As can be seen from Table 1 application of the Winograd algorithms for calculating SSD reduces number of the real multiplications.

The algorithms examined operate with complex numbers in order to get the result. For identifying more efficient ways to compute the two-dimensional convolution the number-theoretic transform of Farm (NTT) was considered. This algorithm is based on modular arithmetic and uses only integer real numbers.

It is hard to assess the computational efficiency by the number of arithmetic operations of NTT relative to considered algorithms, because besides addition and multiplication there are operations modulo Farm. In this case the criterion of comparison of these methods is the estimated performance. This criterion evaluates the time required to calculate the coordinates of the moving blocks with each of the FFT algorithms considered in the developed model. The results of Matlab simulation are presented in Table 2.

Table 2. The estimated performance of the algorithms SSD when S=2B=16 for one of the current block

Algorithms	Time, ms
Winograd FFT	5,5
Cooley-Tukey FFT	6,7
NTT	6,1

As shown in Table 2 motion estimation method based on block matching using Winograd algorithm has the lowest complexity.

4 Conclusions

A new fast full search algorithm for block motion estimation based on convolution theorem and Winograd's Fourier transform is presented. Proposed method reduces the computational complexity of the ME algorithm by 58 % the size of block 8 × 8, and – 81 % the size of block 16 × 16 pixels for the search area, exceeding twice the size of the current block. The developed motion estimation algorithm of full search can be used in the environment with a high demand for the quality of video, and the computational complexity of the algorithm is not critical. The further direction of research in this area will be the analysis of fast algorithm for computing two-dimensional convolution through its decomposition into several convolutions shorter length.

References

1. Haskell B.G., Limb J.O.: Predictive video encoding using measured subjective velocity. U.S. Patent No. 3,632,865 (1972)
2. Lee C.-H., Chen L.-H.: A Fast Motion Estimation Algorithm Based on the Block Sum Pyramid. J. IEEE Trans. Image Processing, vol. 6, pp. 1587–1591 (1997)
3. Koga T., Iinuma K., Hirano A., Iijima Y., Ishiguro T.: Motion compensated interframe coding for video conferencing. In: Proc. Nat. Telecommun. Conf., New Orleans, LA, pp. G5.3.1–5.3.5 (1981)
4. Cheung C. H., Po L. M.: A novel cross-diamond search algorithm for fast block motion estimation. J. IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 12, pp. 1168–1177 (2002)
5. Kilthau S. L., Drew M. S., Moller T. Full search content independent block matching based on the fast Fourier transform. IEEE 1CIP, I, pp. 669–672 (2002)
6. Zubarev Iu.B., Dvorkovich V.P., Nechepaev V.V., Sokolov A.Iu.: Metody analiza i kompensatsii dvizheniia v dinamicheskikh izobrazheniiakh. J. Elektrosviaz', no. 11, pp. 15–21 (1998)