Extended Constraint Mask Based One-Bit Transform for Low-Complexity Fast Motion Estimation

Oğuzhan Urhan

***Abstract*—**In this paper, an improved motion estimation (ME) approach based on weighted constrained one-bit transform is proposed for block-based ME employed in video encoders. Binary ME approaches utilize low bit-depth representation of the original image frames with a Boolean exclusive-OR based hardware efficient matching criterion to decrease computational burden of the ME stage. Weighted constrained one-bit transform (WC-1BT) based approach improves the performance of conventional C-1BT based ME employing 2-bit depth constraint mask instead of a 1-bit depth mask. In this work, the range of constraint mask is further extended to increase ME performance of WC-1BT approach. Experiments reveal that the proposed method provides better ME accuracy compared existing similar ME methods in the literature.

***Keywords*—**Fast motion estimation, low-complexity motion estimation, video coding.

I. Introduction

ME is an unavoidable stage of the modern video coding methods because it enables efficient encoding of interframe redundancies existing between successive image frames. Typically, each image frame to be encoded is separated into non-overlapping blocks, and each block in the current frame is searched in the reference frame(s) within a fixed search range. If all possible positions in this range are controlled utilizing a matching criterion, this approach is called as full-search (FS) based ME. The most commonly employed matching criterion is sum of absolute difference (SAD). However, ME stage generally consumes most of the time in video compression [1]. Thus, it is required to speed-up ME process.

There are several groups of approaches to speed-up ME stage. The first group simply checks only a sub-set of all available positions in the search range. The second group limits the number of pixels that will be utilized in matching criterion computation. The third group aims to eliminate impossible candidate before computing full matching criterion for a given block. These groups of approaches are specifically designed for software based solutions. There is a last group of approach which modifies matching criterion to facilitate efficient hardware and parallel software implementations. The ME approaches belonging to this group generally requires binary input images.

In this work, we present a novel approach belonging to the last group which improves ME accuracy of the existing low bit-depth representation based ME algorithms by making use of an extended constraint mask.

O. Urhan is with the University of Kocaeli, Umuttepe Campus, Electronics & Telecom. Eng. Dept., 41380, Kocaeli, Turkey (hone: +90-262-303-33-52; fax: +90-262-303-30-03; e-mail: urhano@kocaeli.edu.tr).

II. Low Bit-Depth Representation Based ME

This group of ME approaches drawn a lot of attention recently, because of the potential efficient hardware and software implementations. One of the first method in this category is named as 1-bit (1BT) transform based ME [2]. In this approach, each image frame is firstly filtered by a multi band pass filter. Subsequently, the filtered image frame is compared against to the original image frame. Based on this comparison, a binary image frame is constructed. After obtaining binary images, a hardware efficient matching criterion named as number of non-matching points (NNMP) is utilized for ME process as:

$$NNMP_{1BT}(m,n)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left[B^{t}(i,j)\oplus B^{t-1}(i+m,j+n)\right], \quad (1)$$
$$-s\le m,n<s$$

where (m,n) is the candidate movement for $N\times N$ image block within a search range of s. B^t and B^{t-1} show the current and previous (reference) 1-bit depth images, whereas $\oplus$ shows the Boolean EX OR operation. The location providing the lowest NNMP value is assigned as the motion vector of the current block.

Since 1BT based ME approach in [2] utilizes only 1-bit depth images, its ME performance significantly decreases compared to ME approach which employs SAD based matching criterion. In [3], it is proposed to employ 2-bit planes to enhance ME accuracy of 1BT based ME. The first bit-plane is constructed by making use of block mean of a larger block covering the current block as a threshold, whereas the second bit-plane is obtained by considering mean and standard deviation of the larger block. In this case, matching criterion is defined as follows:

$$NNMP_{2BT}(m,n)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left\{\left[B_1^{t}(i,j)\oplus B_1^{t-1}(i+m,j+n)\right] \,\|\, \left[B_2^{t}(i,j)\oplus B_2^{t-1}(i+m,j+n)\right]\right\} \quad (2)$$

where B_1 and B_2 denote the first and second bit-planes, respectively, whereas || symbolizes Boolean OR operation. This approach is called as 2BT based ME and it provides better ME accuracy compared to 1BT based ME because of the additional bit-plane employed.

Based on the same strategy, another 2-bit plane based method called as constraint 1-bit transform (C-1BT) is presented in [4] where a mask is generated to decide the reliable pixels for matching computation. This mask is called as constraint mask and is constructed as follows:

$$CM(i,j)=\begin{cases}1, & \text{if } \left|I(i,j)-I_F(i,j)\right|\geq D\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

In (3), I , I_F , and D correspond to the original frame, the multi band-pass filtered image frame, and a fixed threshold, respectively. The matching criterion (constrained NNMP) of C-1BT is computed as follows:

$$CNNMP\,(m,n)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left\{\left[CM^{t}(i,j)\,\|\,CM^{t-1}(i+m,j+n)\right]\bullet\left[B^{t}(i,j)\oplus B^{t-1}(i+m,j+n)\right]\right\} \tag{4}$$

where • denotes the Boolean AND operation. This approach eliminates unreliable pixels from matching computation and thus gives better ME accuracy compared to 1BT [2] and 2BT [3] based ME approaches.

The 1-bit depth constraint mask is modified to 2-bit resolution in [5] to further improve performance of C-1BT based ME. This approach employs the following matching criterion:

$$CNNMP_W\,(m,n)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left\{\left[B^{t}(i,j)\oplus B^{t-1}(i+m,j+n)\right]\times\max\left[2BCM^{t}(i,j),2BCM^{t-1}(i+m,j+n)\right]\right\} \tag{5}$$

where *2BCM* shows 2-bit depth constraint mask which takes values from 0 to 3. This approach is called as weighted C-1BT (WC-1BT) based ME and it performs better compared to other low bit-depth based ME methods since it utilizes three bits per pixel for matching.

In [6], it is proposed to extend dynamic range of the CNNMP matching criterion by making use of constraint mask and binary images in different combinations. Thus, the total dynamic range of the matching criterion is increased 5 times. This approach is called as enhanced C-1BT (EC-1BT) and it is shown that it is able to outperform previous low bit-depth ME methods. There are some efficient hardware implementations of these methods in the literature [7]-[10].

III. Proposed Approach

The proposed approach in this work aims to extend the range of constraint mask and thus improve the ME performance. As shown in WC-1BT based ME [5], utilizing two-bit depth constraint mask increases the ME accuracy of C-1BT based ME [4] where only one-bit depth constraint mask is employed. Following the same concept, we propose to utilize full bit-depth constraint mask for improved ME accuracy. We define the extended constraint mask (ECM) as

$$ECM(i,j)=\alpha\times\left|I(i,j)-I_F(i,j)\right| \tag{6}$$

where α is a factor controlling the dynamic range of the constraint mask. This factor does not require to be a fixed constant. The matching criterion in this case is defined as:

$$CNNMP_{EW}\,(m,n)=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left\{\left[B^{t}(i,j)\oplus B^{t-1}(i+m,j+n)\right]\times\max\left[ECM^{t}(i,j),ECM^{t-1}(i+m,j+n)\right]\right\} \tag{7}$$

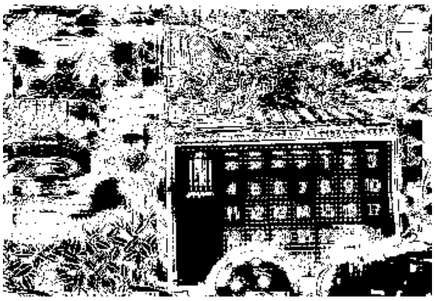
similar to the [5]. Since the proposed ECM is able to take full range allowed by the input image frame, the dynamic range of the proposed matching criterion is significantly improved compared to the methods in [4]-[6].

In order to show the core idea of the proposed approach, the constraint masks of the C-1BT, WC-1BT and the proposed method for an example image frame are depicted in Fig. 1. As seen from this figure, binary constraint mask in C-1BT based approach makes a strict decision about the pixel that will be included into the matching criterion computation. On the other hand, WC-1BT based approach employs 2-bit depth constraint mask and relaxes the strict classification between reliable and non-reliable pixels.

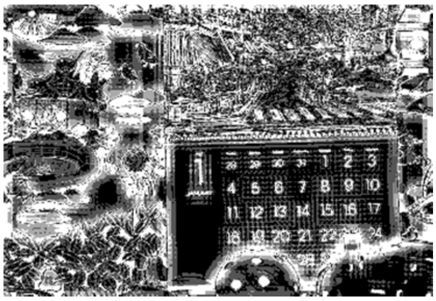
The constraint mask in the proposed approach extends the bit-depth of the constraint mask in previous methods further in order to improve ME performance. The extended constraint mask presented in the proposed method enables determination of the contribution level of each pixel in higher accuracy compared to C-1BT and WC 1BT based approaches.

Tables I and II show the number of operations required for the transform and matching criteria computation stages of the low-complexity ME approaches per pixel, respectively. As seen from Table I, the proposed approach has the lowest binarization cost similar to the C-1BT based ME [4], whereas the EC-1BT based ME approach in [6] has the highest computational cost at this stage. When the matching stage computational complexity is assessed, the proposed approach has similar complexity with the W-C1BT based ME and it has considerably lower complexity than the method in [6]. It is clear that the multiplication operation for the suggested matching criterion could be implemented as a simple fetch from a look up table (LUT).

(a) Original (Mobile sequence, frame #1)



(b) C-1BT constraint mask (1-bit depth)



(c) WC-1BT constraint mask (2-bit depth)

(d) Proposed approach constraint mask (8-bit depth)

Fig. 1 Constraint masks utilized in the C-1BT, WC-1BT and proposed method

TABLE I
NUMBER OF OPERATIONS REQUIRED FOR TRANSFORM STAGE OF THE LOW COMPLEXITY ME APPROACHES (PER PIXEL)

Operation / Method	C-1BT [4]	W-C1BT [5]	EC-1BT [6]	Proposed
Boolean Op.	-	-	-	-
Comparison	2	8	2	2
Add./Sub.	17	17	25	17
Shift	1	1	-	1
F.P. Mult.	-	-	1	-

TABLE II
NUMBER OF OPERATIONS REQUIRED FOR MATCHING CRITERIA COMPUTATION OF THE LOW-COMPLEXITY ME APPROACHES (PER PIXEL)

Operation / Method	C-1BT [4]	W-C1BT [5]	EC-1BT [6]	Proposed
Boolean Op.	3	1	6	1
Comparison	-	1	-	1
Addition	-	-	5	-
Shift	-	-	1	-
Mult./LUT	-	1	-	1

The dynamic range control factor α can be generated using a non-linear function for each possible absolute difference value in (6) and in this case, the computational complexity of the proposed method does not change since it is possible to implement this multiplication as a LUT operation.

IV. Experimental Results

For evaluating the performance of different ME approaches, the peak signal to noise ratio (PSNR) measure is often utilized in the literature. We will use same method to assess the ME performance of the presented method by estimating the current frame using the previous frame and related motion vectors. Next, PSNR between the current and estimated frame is computed frame by frame.

Table III shows the average PSNR values for seven different video sequences displaying different motion characteristics for N=16, s=16, α=1. The best and second best among the low bit-depth based approaches are shown using bold and italic fonts, respectively.

TABLE III
PSNR VALUES FOR DIFFERENT TEST SEQUENCES

Sequence / Method	Football (125 frame)	Coastguard (300 frame)	Mobile (300 frame)	Tennis (150 frame)
FS-SAD	22.88	30.48	23.94	29.45
1BT [2]	21.83	29.83	23.61	28.11
2BT [3]	22.06	29.94	23.66	28.46
C-1BT [4]	22.10	29.98	23.69	28.71
WC-1BT [5]	*22.37*	*30.19*	*23.81*	*29.00*
EC-1BT [6]	22.35	30.17	23.76	28.97
Proposed	**22.41**	**30.28**	**23.84**	**29.06**
Sequence / Method	Garden (115 frame)	Foreman (300 frame)	Bus (150 frame)	Average
SAD	23.79	32.09	25.02	27.58
1BT [2]	23.31	30.32	23.83	26.62
2BT [3]	23.43	30.70	24.59	26.87
C-1BT [4]	23.38	30.86	24.34	26.92
WC-1BT [5]	*23.56*	31.03	*24.58*	27.12
EC-1BT [6]	23.52	*31.24*	24.52	*27.13*
Proposed	**23.61**	**31.25**	**24.65**	**27.21**

As seen from this table, the proposed method mostly outperforms the existing approaches. The results given in Table III are obtained when α is fixed to 1. If an appropriate non-linear function can be found for each possible absolute difference value in (6), then it might be possible to further improve the accuracy of the presented approach.

When the proposed approach is compared with WC-1BT based ME [5], it is clear that the proposed method displays better accuracy for all sequences thanks to extended constraint mask approach presented in this work. The overall ME accuracy of the presented approach is slightly better than the proposed method in [6]. Additionally, note that the proposed method has considerably lower computational complexity compared to [6] as shown in Tables I and II. Thus, the proposed method not only have computational benefits but also higher ME accuracy compared to existing methods in the literature falling into same category.

The proposed method can be combined with an adaptive search range based method as in [11] to further reduce the computational complexity. On the other hand, hybrid methods which perform an additional local search using the SAD criterion around the best ME result obtained by low bit-depth representation methods as in [12] can be integrated into the proposed approach to further improve the ME accuracy.

V. CONCLUSION

In this paper, extended constraint mask based low-complexity ME approach is presented. The proposed approach employs the full potential of the constrain- mask by making use of full bit-depth constraint-mask without introducing any additional computational load. Experiments show that the proposed ME method is able to outperform existing low bit-depth representation based approaches in terms of ME accuracy.

ACKNOWLEDGMENT

This work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant No. EEEAG/115E921.

REFERENCES

[1] T. C. Chen, Y. H. Chen, S. F. Tsai, S. Y. Chien, and L.G. Chen, “Fast algorithm and architecture design of low-power integer motion estimation,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, pp. 568-577, 2007.

[2] B. Natarajan, V. Bhaskaran, and K. Konstantinides, “Low-complexity block-based motion estimation via one-bit transforms,” IEEE Trans. Circuit Syst. Video Technol., vol. 7, pp. 702-706, 1999.

[3] A. Ertürk, S. Ertürk, “Two-bit transform for binary block motion estimation,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 15, pp. 938-946, 2005.

[4] O. Urhan, S. Ertürk, “Constrained one-bit transform for low-complexity block motion estimation,” IEEE Trans. Circuits and Syst. Video Technol., vol. 14, pp. 478–482, 2007.

[5] M. K. Güllü, “Weighted constrained one-bit transform based fast block motion estimation,” *IEEE Trans. Consum. Electron.*, vol. 57 pp. 751-755, 2011.

[6] S. Lee, G. Jeon, and J. Jeong, “Fast motion estimation based on enhanced constrained one-bit transform,” *Electron. Lett.*, vol. 50, pp. 746-748, 2014.

[7] A. Çelebi, O. Urhan, I. Hamzaoğlu, S. Ertürk, “Efficient hardware implementations of low bit depth motion estimation algorithms,” *IEEE Signal Process. Letts.*, vol. 16, pp. 513-516, 2009.

[8] A. Çelebi, O. Urhan, “High performance hardware architecture for constrained one-bit transform based motion estimation,” *in Proc 19th European Signal Processing Conference (EUSIPCO 2011)*, 2011, pp. 2151-2155.

[9] S. K. Chatterjee, “Implementation of weighted constrained one-bit transformation based fast motion estimation,” *IEEE Trans. Consum. Electron.*, vol. 58, pp. 646-653, 2012.

[10] A. Akın, I. Hamzaoğlu, “A high performance hardware for early terminated C-1BT based motion estimation,” *in Proc Design and Technology of Integrated Systems in Nanoscale Era (DTIS) Conference*, 2016.

[11] O. Urhan, “Truncated gray coding–based fast block motion estimation,” *J. Electron. Imaging*, vol. 22, Article ID: 023018, 2013.

[12] S. Lim, J. Kim, S. Yu, “Modified one-bit transform method with dynamic search range,” *J. Electron. Imaging*, vol. 22, Article ID: 023001, 2013.