Hardware Implementation of Local Binary Pattern Based Two-Bit Transform Motion Estimation

Seda Yavuz, Anıl Çelebi, Aysun Taşyapı Çelebi, Oğuzhan Urhan

Abstract—Nowadays, demand for using real-time video transmission capable devices is ever-increasing. So, high resolution videos have made efficient video compression techniques an essential component for capturing and transmitting video data. Motion estimation has a critical role in encoding raw video. Hence, various motion estimation methods are introduced to efficiently compress the video. Low bit-depth representation based motion estimation methods facilitate computation of matching criteria and thus, provide small hardware footprint. In this paper, a hardware implementation of a two-bit transformation based low-complexity motion estimation method using local binary pattern approach is proposed. Image frames are represented in two-bit depth instead of full-depth by making use of the local binary pattern as a binarization approach and the binarization part of the hardware architecture is explained in detail. Experimental results demonstrate the difference between the proposed hardware architecture and the architectures of well-known low-complexity motion estimation methods in terms of important aspects such as resource utilization, energy and power consumption.

Keywords—Binarization, hardware architecture, local binary pattern, motion estimation, two-bit transform.

I. Introduction

THANKS to recent advances in technology, digital video has started to find many applications in consumer electronic devices such as camcorders, smartphones, and mobile televisions. In video technology, ever-increasing resolution has made video compression techniques an essential requirement. By reducing the amount of required data, efficient encoding and decoding of videos have made it possible to capture and transmit real-time videos. Motion estimation (ME) is one of the most important parts of a typical video encoder. Various ME algorithms are introduced in the literature to serve for effectively encoding of raw video data.

In the block-based ME (BME), video frames are divided into non-overlapping blocks. Each block in the current frame is searched in the wider area, which is the corresponding location in the reference frame and named as the search window. Full search ME is an approach that a current block is searched in all candidate locations within the search window so as to find the location of the most similar block. It is important to note that many different ME approaches are presented to reduce the computational load and power

consumption which originate from the full search ME method.

One of these approaches is referred to as the bit plane matching (BPM) based technique, in which a lower bit-depth is used to represent image frames instead of 8-bit depth and Boolean operations are utilized to compute the matching criterion. Low bit-depth representation based ME methods are proposed to decrease computational complexity and accelerate computing matching criteria. In [1], one-bit transform (1BT) is proposed where image frames are converted into one-bit/pixel representations by making use of a multi band-pass filter. In order to obtain binary image, the original image frame is compared to its multi-band-pass filtered version. In [2], another approach is presented to further reduce the computational load of 1BT. This method is called as multiplication-free 1BT method (MF-1BT) which employs a diamond-shaped binarization kernel to facilitate 1BT to be carried out with integer arithmetic using addition and shifts only. In [3], two-bit transform (2BT) is proposed to achieve better ME accuracy compared to 1BT based ME methods by using two bit planes for representing an input image. The 2BT based ME method makes use of local mean and variances to construct representative images. In [4], constrained one-bit transform (C-1BT) method outperforms 1BT, MF-1BT and as well as, 2BT in terms of ME accuracy. This method also provides lower complexity than 2BT. In this approach, a multiplication-free filter is used to construct the first bit-plane; as well, it uses a constraint mask to obtain the second bit-plane by indicating the reliable pixels for the matching stage.

In the literature, several hardware implementations are presented for the aforementioned low-complexity ME methods. For instance, in [5]-[7], hardware architectures for the MF-1BT method are presented. In [8] and [9], the hardware implementation of the C-1BT method is presented. However, none of these architectures includes the binarization process of these methods. They only implement the matching process. Therefore in [10], the binarization cost is investigated and a novel hardware architecture is presented for the modified form of 1BT which is named as integer 1BT (I-1BT). In addition to the architecture for I-1BT, the hardware architecture of the MF-1BT method which includes both the binarization and matching part is also proposed in [10]. These methods are compared to each other in terms of binarization hardware cost and it is revealed that the binarization process should be taken into account for the whole ME architecture.

Another approach presented in [11] uses local binary pattern (LBP) method to represent image frames in lower bit-depth for video stabilization. Recently, in [12], an LBP based binarization approach called as LBP-1BT is presented to

A. Taşyapı Çelebi, S. Yavuz, and A. Çelebi are with Kocaeli University Integrated Systems Laboratory (KUTSAL), Electronics and Telecom. Eng. Dept., Umuttepe Campus, 41380, İzmit/Kocaeli, Turkey (e-mail: aysun.tasyapi@kocaeli.edu.tr, seda.yavuz@kocaeli.edu.tr, anilcelebi@kocaeli.edu.tr).

transform an 8-bit plane image frame into a single bit-plane. Besides, the hardware architecture of LBP-1BT is proposed in [12]. Because of ME accuracy reduction compared to other lower bit-depth representations based on ME methods, an LBP based binarization approach is presented in [13]. It is named as LBP-2BT and its main motivation is to provide better ME accuracy and lower complexity.

In this paper, a hardware implementation for the LBP-2BT method is proposed. ME performance of the LBP-2BT method is investigated in detail for different aspects and its hardware efficiency is compared to other low-complexity ME methods in the experimental results section.
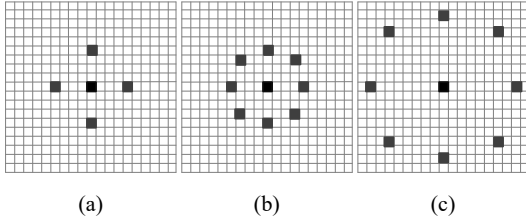

Fig. 1 Different LBP configurations LBP(P,R) (a) LBP(4,4), (b) LBP(8,4), (c) LBP(8,8)

II. LBP BASED BINARIZATION APPROACH

Local Binary Pattern is an efficient texture operator which has become popular in many image processing applications thanks to its discriminative power and computational simplicity. LBP approach is applied by thresholding the neighboring of each pixel with the value of the center pixel. As a result of this thresholding, it represents each pixel as a binary number. Due to its provided advantages in extracting useful features for classification, LBP is used in many applications such as face detection and recognition [14], age estimation [15], texture classification [16], colonoscopy image classification [17]. As mentioned in the previous section, recently in [11] and [12], LBP is utilized as a binarization approach for ME. Eventually, in [13], LBP-2BT method is presented as a two-bit transformation-based binarization approach to provide better ME performance. In [13], 8-bit depth input images are converted into 2-bit depth images by making use of LBP-2BT binarization approach.

In the LBP approach, the center pixel is compared with the P neighbor pixels which are equally spaced and forming a circle of radius R. Fig. 1 illustrates different configurations of LBP. In the binarization approach in [11], the binary representation of a given image is computed as:

$$B(i,j) = \begin{cases} 1 & , \quad \sum_{p=0}^{P-1} \text{sgn}(I_p - I(i,j)) >= P/2 \\ 0 & , \quad \quad \quad otherwise \end{cases} \quad (1)$$

$$sgn(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases}$$

where P and $I(i,j)$ denote the number of neighbor pixels, and corresponding pixel in the image. $B(i,j)$ shows the binary

image pixel at location (i,j). The LBP-1BT method presented in [12] uses this motivation and obtains 1-bit images as proposed method in [11].

Unlike the LBP-1BT method presented in [11] and [12], LBP-2BT computation is carried out by adding a fixed-threshold value which provides robustness for noise.

The modified LBP based binarization is given in (2), where T denotes number fixed-threshold value. $B_1(i,j)$ and $B_2(i,j)$ are the first and second bit plane images.

$$B_1(i,j) = \begin{cases} 1, & if \sum_{p=0}^{P-1} sign (I(p) - I(i,j) - T) \geq P/2 \\ 0, & otherwise \end{cases}$$

$$B_2(i,j) = \begin{cases} 0, & if \sum_{p=0}^{P-1} sign (I(p) - I(i,j) - T) = \{0 \ or \ P\} \\ 1, & otherwise \end{cases} \quad (2)$$

$$sgn(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases}$$

After two-bit planes are obtained, Modified Number of Non-Matching Points (*MNNMP*) criterion is used at the matching stage as in the following:

$$MNNMP(c_x, c_y) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \{B_1^t(i,j) \oplus B_1^{t-1}(i + c_x, j + c_y)\}$$
$$+ \{B_2^t(i,j) \oplus B_2^{t-1}(i + c_x, j + c_y)\} \quad (3)$$

where $-s \leq c_x, c_y \leq s-1$, (c_x, c_y) determine the candidate motion vector, $\oplus$, s and N represent EX-OR operation, search range and block size, respectively. $B_1^t(i,j)$, $B_2^t(i,j)$ and $B_1^{t-1}(i,j)$, $B_2^{t-1}(i,j)$ show the first and second bit planes of the current frame and reference frame, respectively.

In [13], different LBP configurations and fixed-threshold values are investigated for best ME performance. Experimental results of this work reveal that when the LBP configuration, in which P and R are selected as 8 and 12, is preferred and the T value is set to 16, ME accuracy become the best. Hence, in this paper this configuration is selected and its hardware implementation is proposed accordingly. The hardware architecture of method presented in [13] is examined in detail in the next section.

III. PROPOSED HARDWARE ARCHITECTURE

In the matching process, utilizing lower bit-depth images instead of 8-bit pixel intensity values reduces the computational load and accelerates the calculation of the matching criterion. It is proved several times in the literature that making use of simple EX-OR operation as a matching criterion is advantageous in terms of hardware implementation. As previously mentioned in the introduction section, most of the presented hardware architectures for low bit depth representation-based ME methods in the literature only include the matching process. In [10]-[12], in addition to the matching process, the binarization process is also taken into account and experimental results reveal that ignoring the binarization cost is not an appropriate assumption for the whole video encoder. Thus, in this paper, the hardware implementation of both binarization and the matching part for the LBP-2BT method presented in [13] is proposed.

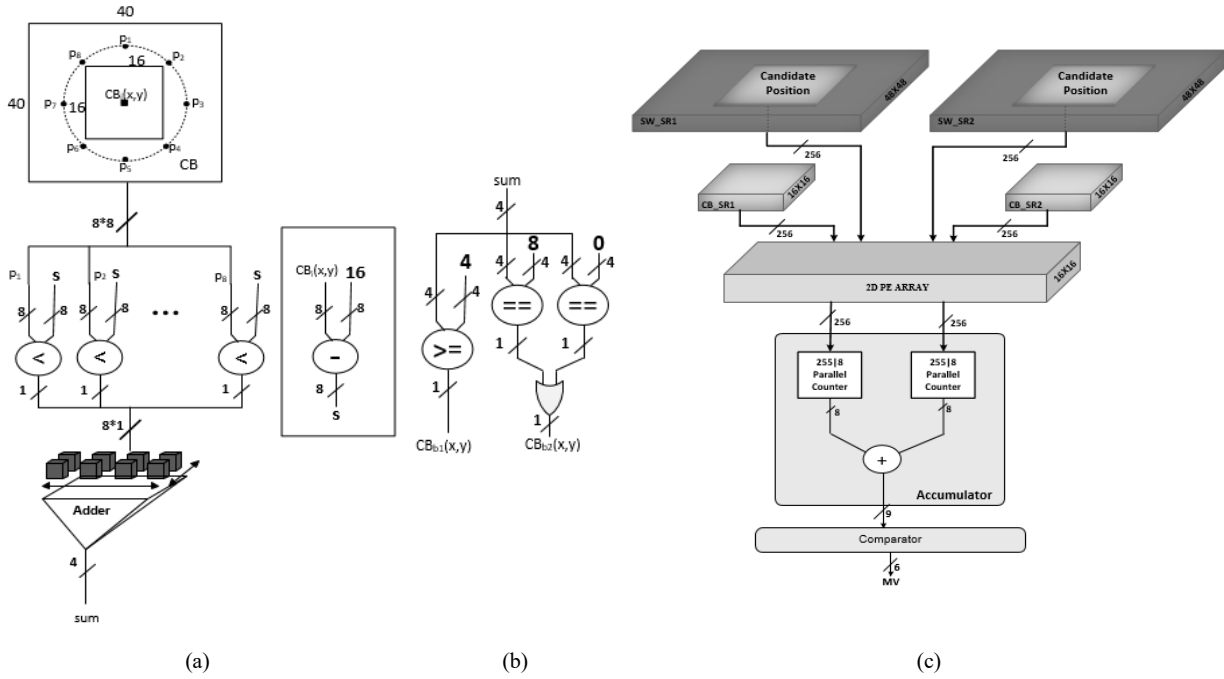(a)   (b)   (c)

Fig. 2 Proposed hardware architecture (a) Filter structure, (b) The component of the final step in binarization process, (c) The structure for matching process

Fig. 2 shows the proposed hardware architecture for the method presented in [13]. Binarization approach of LBP-2BT is similar to LBP-1BT, but as seen from (2), an extra threshold parameter, the value which is set to 16, is added to comparison with each neighbor pixel of center pixel. In this paper, a filter structure is implemented for LBP (8, 12) configuration. Therefore, 8 neighbor pixels of equal spaced on the circle of radius 12 are included in the comparison. Because of selecting 12 for R value, the memory area for binarization is bigger than the 16×16 block size. Hence, additional pixels are needed. This situation is illustrated in Fig. 2 (a). Filter structure is shown for the pixel in the center of the 16×16 current block. As seen from this figure, the size of memory area becomes $(16 + (2 \times R))$. The structure consists of 1 subtraction, 8-comparison and 8 adder blocks. For each pixel in the current block, firstly, the thresholding value is subtracted from the pixel value and then the result is compared with each neighbor pixel value. Comparison results are summed by 8 adders and summation is compared to P□ 2, P and 0, as explained in (2). Fig. 2 (b) illustrates the final operations which are carried out to obtain the first and second bit plane of the current block.

After obtaining 2-bit depth bit planes for search window and current block, the matching process is performed and its hardware architecture is illustrated in Fig. 2 (c). Both search window and current block consist of shift register arrays which are composed of flip flops with three and four direction shifting capabilities, as presented in [9]. The 2D processing element (PE) array, parallel counter and comparator are similar to the architecture presented in [18], however, two units are utilized for 2D PE array and parallel counter because

these blocks carry out the same functionality for 2-bit planes.

IV. EXPERIMENTAL RESULTS

In order to evaluate ME performance of many low bit-depth based ME methods introduced in the literature, the open loop approach is used. The reason of performing this approach is to investigate only ME accuracy regardless of other effects in the whole encoder. The current frame is estimated from the previous frame then peak signal to noise (PSNR) values are obtained between the original and estimated frame to assess ME accuracy. In our experiments, block size and search range are selected as 16 and [-16,16], respectively. Table I demonstrates the PSNR results of low bit-depth based ME methods for various test sequences. As seen from this table, the LBP-2BT method generally outperforms the other methods. It is clear that instead of using one-bit plane as in the LBP-1BT, using an LBP method-based two-bit transformation is advantageous in terms of ME accuracy. Moreover, it is revealed in [13] that the LBP-2BT method reduces the computational load of the binarization stage compared to other methods while achieving better ME performance.

The hardware resource utilization of methods is given in Table II. All methods in the table are implemented using the same FPGA device, which is 28 nm (Xilinx Zynq 7000 - Package Code: xc7vx690tffg1761-2) FPGA, so as to make a fair comparison. Because of bigger kernel size and using two-bit transformation for the matching process, the proposed hardware architecture utilizes more resources compared to others.

TABLE I
PSNR PERFORMANCE (IN DB) OF DIFFERENT LOW COMPLEXITY ME METHODS IN THE OPEN LOOP SCHEME

ME Method	Video Sequences (Frame Size, Sequence Length)						
	Football (352×240) (125 frames)	Foreman (352×288) (300 frames)	Flowergarden (352×240) (115 frames)	Coastguard (352×288) (300 frames)	Tennis (352×240) (150 frames)	Mobile (352×240) (300 frames)	Average of six video sequence
SAD (8-bit depth)	22.88	32.09	23.79	30.48	29.45	23.94	27.10
1BT [1]	21.83	30.32	23.31	29.83	28.11	23.61	26.17
MF-1BT [2]	21.81	30.38	23.26	29.88	28.18	23.63	26.19
2BT [3]	22.06	30.70	*23.43*	29.94	28.46	23.66	26.38
C-1BT [4]	*22.10*	**30.86**	23.38	*29.98*	*28.71*	*23.69*	*26.45*
LBP [11]	21.40	29.20	23.07	29.72	27.43	23.43	25.59
LBP-1BT [12]	21.40	29.33	23.08	29.73	27.45	22.59	25.59
LBP-2BT[13]	**22.18**	*30.80*	**23.52**	**30.11**	**28.73**	**23.72**	**26.51**

TABLE II
HARDWARE RESOURCE UTILIZATION OF DIFFERENT LOW COMPLEXITY ME METHODS SCHEME

	ME Method	LBP-2BT (Proposed)	LBP1BT [12]	I-1BT [10]	MF-1BT [10]
Binarization	Kernel Size	**25×25**	17×17	17×17	19×19
	# of LUTs	**55089**	41628	42140	45006
	# of DFFs	**55019**	41565	42636	45468
Matching	# of LUTs	**8070**	4030	4047	4036
	# of DFFs	**5856**	2928	2928	2928
Total	# of LUTs	**63159**	45658	46187	49042
	# of DFFs	**60875**	44493	45564	48396

Power consumption is one of most important limitations of consumer electronic devices, especially handheld devices in which energy budget is constrained by the batteries. In order to assess power consumption performance, power analysis is performed for proposed hardware implementation and it is compared to other methods, as seen in Table III. According to the results given in this table, the proposed hardware consumes more power and energy. Nevertheless, it is not fair to assess only energy consumed. Normally energy/power consumption of the ME approach should be evaluated together with its ME accuracy. For this purpose, the PSNR vs. Energy plot of different ME approaches is given Fig. 3. As seen from this figure, even though the proposed approach has the highest energy consumption, it also has best ME accuracy. Thus, the proposed architecture provides a good-balance between energy consumption and ME accuracy for the LBP-2BT approach.

TABLE III
POWER AND ENERGY CONSUMPTION OF DIFFERENT LOW COMPLEXITY ME METHODS

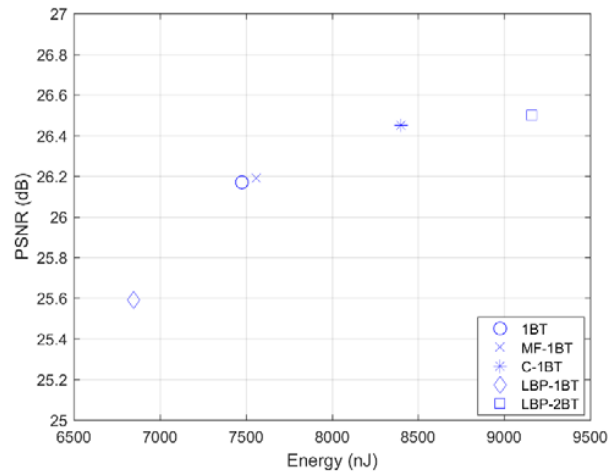| ME Method	Signal Power (mW)	Logic Power (mW)	Time (μs)	Energy (nJ)
MF-1BT [2]	101	79	42	7560
I-1BT [10]	103	75	42	7476
LBP-1BT [12]	90	73	42	6846
LBP-2BT(Proposed)	120	96	42.4	9158.4

Fig. 3 Relationship between PSNR values and amount of energy consumption for different ME approaches

V. CONCLUSIONS

In this work, a hardware architecture of the LBP-2BT-based ME method is proposed. Additionally, the binarization part is included the hardware implementation, whereas it is ignored by most of works presented in the literature. Experimental results reveal that the LBP-2BT method provides reasonable energy and PSNR trade-off compared to other low-complexity ME methods at the same category.

REFERENCES

[1] 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, no. 4, pp. 702-706, Aug. 1997.
[2] S. Ertürk, "Multiplication-free one-bit transform for low-complexity block-based motion estimation," *IEEE Signal Process. Lett.*, vol. 14, no. 2, pp. 109-112, Feb. 2007.
[3] A. Ertürk, and S. Ertürk, "Two-bit transform for binary block motion estimation," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 15, no. 7, pp. 938-946, July 2005.
[4] O. Urhan, and S. Ertürk, "Constrained one-bit transform for low-complexity block motion estimation," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 17, no.4, pp. 478-482, Apr. 2007.
[5] A. Akin, G. Sayilar, and I. Hamzaoglu, "High performance hardware architectures for one bit transform based single and multiple reference frame motion estimation," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 1144–1152, July 2010.
[6] S. Chatterjee, and I. Chakrabarti, "Low power vlsi architectures for one bit transformation based fast motion estimation," *IEEE Trans. Consum.*

Electron., vol. 56, no.4, pp. 2652–2660, January 2011.

[7] A. Celebi, H. J. Lee, and S Erturk, "Bit plane matching based variable block size motion estimation method and its hardware architecture," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1625-1633, 2010.

[8] A. Çelebi, O. Urhan, İ. Hamzaoğlu, and S. Ertürk, "Efficient hardware implementations of low bit depth motion estimation algorithms," *IEEE Signal Process. Letters*, vol. 16, no. 6, June 2009.

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

[10] S. Yavuz, A. Taşyapı Çelebi, A. Çelebi, and O. Urhan, "Integer 1-bit transform method and its hardware architecture for low-complexity block-based motion estimation," in Proc. of *25th Signal Processing and Communication Applications Conf. (SIU)*, Antalya, Turkey. DOI: 10.1109/SIU.2017.7960295, June 2017.

[11] B. Kır, M. Kurt, and O. Urhan, "Local binary pattern based fast digital image stabilization," IEEE Signal Processing Letters, vol. 22, no. 3, pp. 341-345, 2015.

[12] S. Yavuz, A. Taşyapı Çelebi, A. Çelebi, and O. Urhan, "Local binary pattern method and its hardware architecture for low-complexity motion estimation," in Proc. of *25th Signal Processing and Communication Applications Conf. (SIU)*, Antalya, Turkey. DOI: 10.1109/SIU.2017.7960443, June 2017.

[13] A. Taşyapı Çelebi, "Two-bit transform using local binary pattern method for low complexity block motion estimation," revised version submitted for publication in Turkish Journal of Electrical Engineering and Computer Sciences.

[14] X. M. Zhao, and S. Q. Zhang, "Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding," EURASIP Journal of Advances in Signal Processing, pp. 1–9, 2012.

[15] Ylioinas J., Hadid A., Hong X., and Pietikainen M., "Age estimation using local binary pattern kernel density estimate," Lecture Notes in Computer Science, vol. 8156, pp. 141–150, 2013.

[16] L. Liu, L. Zhao, G. Kuang, and P. Fieguth, "Extended local binary patterns for texture classification," Image and Vision Computing, vol. 39, no. 2, pp. 86-99, 2012.

[17] S. Manivannan, R. Wang, and E. Trucco, "Extended gaussian-filtered local binary patterns for colonoscopy image classification," in Int.Conf. on Computer Vision (ICCV-2013).

[18] S. Yavuz, A. Çelebi, M. Aslan, and O. Urhan, "Selective gray-coded bit-plane based low complexity motion estimation and its hardware architecture," IEEE Trans. Consum. Electron, vol. 62, no. 1, pp. 76-84, Feb. 2016.