Multiplication-Free One-Bit Transform for Low-Complexity Block-Based Motion Estimation

Sarp Ertürk, *Member, IEEE*

***Abstract*—A multiplication-free one-bit transform (1BT) for low-complexity block-based motion estimation is presented in this letter. A novel filter kernel is utilized to construct the 1BT of image frames using addition and shift operations only. It is shown that the proposed approach provides the same motion estimation accuracy at macro-block level and even better accuracy for smaller block sizes compared to previously proposed 1BT methods. Because the proposed 1BT approach does not require multiplication operations, it can be implemented in integer arithmetic using addition and shifts only, reducing the computational complexity, processing time, as well as power consumption.**

***Index Terms*—Block matching, motion estimation, one-bit transform (1BT), video coding.**

I. Introduction

DIGITAL video compression is required to reduce the necessary bandwidth for transmission or storage of video data in a wide range of applications. Video compression is, for instance, used in high definition television (HDTV) and standard definition television (SDTV) broadcasting equipment, video conferencing transmitters, video cellular phones, digital video camcorders, as well as multimedia services for networking applications. Motion estimation and compensation play key roles in video compression algorithms due to the ability of realizing high compression rates achieved by removing temporal redundancies between successive image frames.

The most popular technique used for motion estimation is the block matching algorithm (BMA), in which image frames are commonly divided into non-overlapping rectangular blocks. The best match to the current block of pixels is searched for in the previous frame of the sequence within a certain search area about the location of the current block. The optimal search strategy for BMA is the full search (FS) algorithm that exhaustively searches for the best matched block within all locations of the search window. The mean absolute difference (MAD) or mean square error (MSE) matching criteria are considered to be statistically optimal solutions to the matching process.

Motion estimation is usually remarked as the computationally most intensive part of the video compression system, performing up to 50% of the computations executed by the entire video coding system [1]. Various methods, proposed to reduce the high computational load of the FS-MAD block matching algorithm, can be divided into three main categories [2]: fast search techniques that select a subset of the possible search candidate locations; techniques based on various forms of pixel pattern or motion field decimation that employ a certain subsampling of the pixel pattern or motion field; and techniques that exploit different matching criteria instead of the classical MAD.

The multiplication-free one-bit transform (1BT)-based low-complexity motion estimation approach presented in this letter falls into the category of techniques that exploit different matching criteria to achieve reduction in computational complexity. A main literature review of similar approaches follows first.

Bit-plane matching as a preprocessing step to exhaustive search in order to eliminate unlikely locations has been proposed in [3], with the block mean being used as threshold to accomplish a 1BT. The resulting bit-plane of an image frame is obtained in the form of

$$B(i,j) = \begin{cases} 1, & \text{if } I(i,j) \geq t_{bm} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where t_{bm} represents the threshold value that is set equal to the block mean, and (i,j) is used as pixel index. Hierarchical feature matching-motion estimation (HFM-ME) that employs sign truncated feature (STF) matching has been proposed in [4]. Binary block matching on the binary edge maps of image frames has been proposed in [5]; however, it has been noted that the technique is inappropriate for blocks with inadequate edge information. In [6], motion estimation using the 1BT, where image frames are transformed into one-bit/pixel representations by comparing the original image frame against a multi band-pass filtered version is proposed. A 17 × 17 multiband-pass filter kernel in the form of

$$K(i,j) = \begin{cases} 1/25, & \text{if } i,j \in [0,4,8,12,16] \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

is used to filter the image frame, and the 1BT bit-plane of an image is constructed as

$$B(i,j) = \begin{cases} 1, & \text{if } I(i,j) \geq I_F(i,j) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $I_F(i,j)$ represents the filtered version of the image frame $I(i,j)$, obtained by applying the convolution kernel K to I. As the filtered version is compared against the original

Manuscript received March 16, 2006; revised July 1, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lina J. Karam.

The author is with the Kocaeli University Laboratory of Image and Signal Processing (KULIS), Electronics and Telecommunication Engineering Department, University of Kocaeli, Kocaeli 41040, Turkey (e-mail: sertur@kou.edu.tr; sarp@ieee.org).

Digital Object Identifier 10.1109/LSP.2006.882088

image to construct the representative bit-plane, the filtered image is mainly used as a sort of pixel-wise threshold, and the 1BT is constructed by comparing the original image against the threshold pattern obtained by applying the image to a multi-band-pass filter. The addition of conditional local searches has been proposed in the modified 1BT to improve the predicted image at the expense of increased computational complexity in [7]; however, in this case, the binary only matching characteristics are destroyed as additional local MAD matches have to be executed. A two-bit transform (2BT) that makes use of local mean and variances to obtain a two-bit representation has been proposed in [8], and a DCT-based adaptive thresholding approach for binary motion estimation has been proposed in [9] to improve motion estimation accuracy at the cost of higher computational complexity.

While the 1BT proposed in [6] and shown in (2) and (3) provides a simple and fairly effective approach (particularly for block motion estimation at macro-block level), an important drawback is the requirement of floating number multiplication for normalizing the kernel shown in (2). Floating point multiplications are slow in both hardware and software implementations [10]. A novel 1BT is proposed in this letter that is multiplication-free and can be carried out using addition and shift operations only, for low-complexity binary motion estimation. It is shown that the proposed approach gives the same performance as the 1BT presented in [6] at macro-block level and even provides a superior performance for lower block sizes, with reduced computational complexity.

II. Multiplication-Free One-Bit Transform

The main reason why the multiplication operation in the 1BT proposed in [6] cannot be avoided is that the normalization coefficient of the kernel shown in (2) is not a power of two. If the number of nonzero elements in the kernel is changed to a power of 2, without changing the structure of the kernel, the motion estimation accuracy is degraded. In this letter, it is proposed to use a kernel that has a power of 2 nonzero elements, and the kernel structure is changed to avoid any loss in accuracy. Instead of utilizing a rectangular-structured kernel as in (2), the utilized kernel has a diamond-shaped structure. The kernel utilized for this purpose is shown in (4) at the bottom of the page. A sample 1BT result obtained with the proposed kernel using (3) is shown in Fig. 1.

Compared to the rectangular-shaped kernel utilized in [6], the proposed kernel can be considered as a similar shaped kernel but rotated by 45° around the center to obtain the diamond-shaped structure observed in (4). As the nonzero elements mainly define a multiband-pass filter, and a rotation in spatial domain has only the effect of a similar rotation in the frequency domain, this kernel also defines a multiband-pass filter. The frequency response of this kernel is shown in Fig. 2. Because the normalization coefficient is now a power of 2, it becomes possible to carry out this filtering by simply adding the corresponding 16 pixels and shifting the final result. Hence, the complete filtering operation can be carried out using integer arithmetic, and the computational load is 16 addition and 1 shift operation per pixel.

The single bit-plane representation (i.e., 1BT) of each image frame is constructed as shown in (3), by comparing the filtered version against the original image frame. The motion vector of a block is then decided based on the number of non-matching points (NNMP) measure

$$\text{NNMP}(m,n) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left\{B^t(i,j) \oplus B^{t-1}(i+m, j+n)\right\} \qquad -s \le m,n \le s-1 \quad (5)$$

$$K = \frac{1}{16}\begin{bmatrix}
0&0&0&0&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&\mathbf{1}&0&0&0&0&0&0&0&0&0
\end{bmatrix} \quad (4)$$

(a)

(b)

Fig. 1. (a) Sample frame of the *Football* sequence. (b) Corresponding 1BT obtained with the proposed approach.

where (m, n) shows the candidate displacement, s determines the search range, and $\oplus$ shows the exclusive-OR (XOR) operation. The candidate displacement that gives the lowest NNMP is designated to be the block motion vector. If two candidate displacements result in the same NNMP value, the one with the lowest distance to the block position (i.e., the smallest motion vector) is selected.

III. Experimental Results

In order to evaluate the performance of the proposed approach, motion estimation has been performed for various video sequences using exhaustive full search. For the matching criteria, MAD matching, 2BT as proposed in [8], 1BT as proposed in [6], and the proposed multiplication-free 1BT (MF-1BT) are evaluated. Note that other 1BTs proposed are not evaluated for comparison as the procedure of [6] is mainly the most accurate and simple 1BT process. In order to evaluate the motion estimation accuracy, peak signal-to-noise ratio (PSNR) values obtained between the original frames and frames reconstructed from the previous frame using motion vectors calculated from one of the methods are used. The motion estimation accuracy is assessed for a block size of 16 $\times$ 16 pixels with a search range of 16 pixels and a block size of 8 $\times$ 8 pixels with a search range of 8 pixels. The PSNR results for various test sequences are given in Tables I and II, respectively.

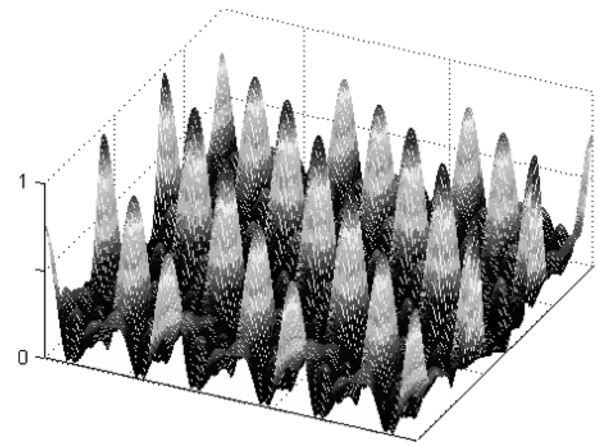

Fig. 2. Frequency response of the filter used in the proposed 1BT.

Table I shows that for a block size of 16 $\times$ 16 pixels with a search range of 16 pixels, the proposed MF-1BT provides the same motion estimation accuracy as the 1BT presented in [6]. In this case, 1BT results are usually within 1 dB of MAD matching; hence, the 1BT performance is reasonably close to MAD. Therefore, 1BT-based motion estimation is particularly suitable for video coding approaches that utilize motion estimation at macro-block level such as MPEG1 and MPEG2. It is also seen from Table I that the performance improvement of 2BT over 1BT is small for a block size of 16 $\times$ 16 pixels as 1BT is already close to MAD. The main reason for this is that for larger block sizes, sufficient information is already captured within the blocks in the 1BT process so that a reasonable match can be achieved.

Table II shows that for a block size of 8 $\times$ 8 pixels with a search range of 8 pixels, the proposed MF-1BT provides an average PSNR performance of about 0.2 dB better than the 1BT proposed in [6]. Hence, the proposed MF-1BT not only reduces the computational complexity but also improves motion estimation accuracy for block sizes below macro-block level. It is observed that in this case, 2BT can result in more than 1 dB better than the 1BT; however, this improvement comes at the cost of increased complexity particularly compared against the proposed MF-1BT approach.

The computation load of the 2BT is 2.8125 addition, 1.0625 multiplication, 0.03125 subtraction, three comparison operations, and one Boolean OR operation per pixel to construct the final 2BT [8]. The computational complexity of the 1BT presented in [6] is comprised of 25 addition operations, one multiplication operation, and one comparison per pixel. The computational complexity of the proposed MF-1BT, on the other hand, is 16 addition operations, one shift, and one comparison per pixel. As the proposed MF-1BT does not require a multiplication operation, it can be implemented using integer arithmetic addition and shift operations, therefore reducing the computational complexity of the system, the processing time, as well as power consumption.

IV. Conclusion

An MF-IBT-based low-complexity motion estimation approach has been presented in this letter. The proposed approach

TABLE I
AVERAGE PSNR (dB) OF SEVERAL SEQUENCES RECONSTRUCTED BY VARIOUS MOTION ESTIMATION TECHNIQUES, WITH FULL SEARCH AND A BLOCK SIZE OF 16 × 16 PIXELS WITH A MOTION VECTOR SEARCH RANGE OF 16 PIXELS

Method	Video Sequences (Frame Size, Sequence Length)					
	Football (352×240) (125 frames)	Foreman (352×288) (299 frames)	Tennis (352×240) (112 frames)	Flowergarden (352×240) (115 frames)	Mobile (352×240) (140 frames)	Coastguard (352×288) (299 frames)
MAD	22.88	32.10	29.87	23.79	22.99	30.47
2BT [8]	22.08	30.71	28.89	23.43	22.72	29.93
1BT [6]	21.83	30.36	28.77	23.32	22.71	29.84
MF-1BT	21.81	30.39	28.78	23.29	22.73	29.88

TABLE II
AVERAGE PSNR (dB) OF SEVERAL SEQUENCES RECONSTRUCTED BY VARIOUS MOTION ESTIMATION TECHNIQUES, WITH FULL SEARCH AND A BLOCK SIZE OF 8 × 8 PIXELS WITH A MOTION VECTOR SEARCH RANGE OF 8 PIXELS

Method	Video Sequences (Frame Size, Sequence Length)					
	Football (352×240) (125 frames)	Foreman (352×288) (299 frames)	Tennis (352×240) (112 frames)	Flowergarden (352×240) (115 frames)	Mobile (352×240) (140 frames)	Coastguard (352×288) (299 frames)
MAD	24.73	32.89	31.25	25.22	23.88	31.58
2BT [8]	23.36	30.64	29.91	24.55	22.99	30.50
1BT [6]	22.72	29.82	29.23	24.10	22.71	29.20
MF-1BT	22.75	30.01	29.40	24.25	22.84	29.50

has a lower computational complexity compared to previous approaches and provides the same motion estimation accuracy at macro-block level, while the motion estimation accuracy is even superior compared to previous 1BT methods for lower block sizes. The proposed MF-1BT does not require multiplication operations and can therefore be implemented using integer arithmetic addition and shifts, reducing the computational complexity of the system, the processing time, as well as power consumption.

REFERENCES

[1] Z.-L. He, C.-Y. Tsui, K.-K. Chan, and M. L. Liou, "Low-power VLSI design for motion estimation using adaptive pixel truncation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 5, pp. 669–678, Aug. 2000.

[2] M. Mattavelli and G. Zoia, "Vector-tracing algorithm for motion estimation in large search windows," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1426–1437, Dec. 2000.

[3] J. Feng, K.-T. Lo, H. Mehrpour, and A. E. Karbowiak, "Adaptive block matching motion estimation algorithm using bit-plane matching," in *Proc. ICIP*, 1995, pp. 496–499.

[4] X. Lee and Y. Zhang, "A fast hierarchical motion-compensation scheme for video coding using block feature matching," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 6, no. 8, pp. 627–635, Dec. 1996.

[5] M. M. Mizuki, U. Y. Desai, I. Masaki, and A. Chandrakasan, "A binary block matching architecture with reduced power consumption and silicon area requirement," in *Proc. IEEE ICASSP*, Atlanta, GA, 1996, vol. 6, pp. 3248–3251.

[6] 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. 5, pp. 702–706, Aug. 1997.

[7] P. H. W. Wong and O. C. Au, "Modified one-bit transform for motion estimation," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 9, no. 7, pp. 1020–1024, Oct. 1999.

[8] 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, Jul. 2005.

[9] C.-B. Wu, C.-Y. Yao, B.-D. Liu, and J.-F. Yang, "DCT-based adaptive thresholding algorithm for binary motion estimation," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 15, no. 5, pp. 694–703, May 2005.

[10] J. Liang and T. D. Tran, "Fast multiplierless approximations of the DCT with the lifting scheme," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 49, no. 12, pp. 3032–3044, Dec. 2001.