IEEE Int. Conference Neural Networks & Signal Processing
Zhenjiang, China, June 8~10, 2008

CONSTRAINED ONE-BIT TRANSFORM FOR RETINEX BASED MOTION ESTIMATION FOR SEQUENCES WITH BRIGHTNESS VARIATIONS

Hoi-Kok Cheung[+*], Wan-Chi Siu*, Dagan Feng*[+] and Zhiyong Wang[+]

[+]School of Information Technologies, J12
The University of Sydney
NSW 2006
Australia

*Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
enwcsiu@polyu.edu.hk

ABSTRACT

In this paper, we propose a DCT based binary matching approach for fast motion estimation. The proposed approach is suitable to be applied to our previously proposed retinex based coding system which is characterized with fast motion estimation and the capability of accurately estimating motions for sequences having inter-frame brightness variations. We propose to apply the DCT techniques to the transformation of images to the binary bit plane and to the scaled retinex domain, which is more computationally efficient compared to the conventional convolution based bit transformation approach. Experimental results show that our proposed DCT based bit transform has a very close prediction quality performance(less than 0.1dB drop) to that of the convolution based approach while our system can avoid the extra convolution procedure.

Index Terms – Motion estimation, video coding, two-bit transform(2BT), discrete cosine transform, H.264

1. INTRODUCTION

Conventional motion estimation(ME) methods[1] make use of the brightness constancy assumption, which assumes that object brightness remains constant between frames. However, for sequences with inter-frame brightness variations, the assumption is no longer valid and compression efficiency decreases. To handle brightness variations, the H.264 adopts the weighted prediction[2] approach. However, this feature is only effective for varying brightness of spatially uniform light source. The primary target application is the fade in/out and gradual scene change effects. To cope with spatially non-uniform brightness variations, we proposed previously a DCT based retinex system[3] for accurate motion estimation, which effectively removes the inter-frame de-correlation factor resulting from brightness variations. In this paper, we propose a fast motion estimation approach based on our previous work using an approach similar to the constrained one-bit transform[4] to facilitate Boolean Exclusive-OR(EX-OR) matching. Our proposed system is characterized with fast computation and the capability of accurately estimating motions under brightness variation environment. In[5], a multiplication free one-bit transform (1BT) based motion estimation approaches was proposed using the convolution techniques. Recently, two-bit transform (2BT)[4, 6] based approaches were introduced to enhance the accuracy in the motion estimation stage.

2. REVIEW OF PREVIOUS WORK

In this section, we briefly describe our previous work[3]. To model the influence of light source on image, Barrow and Tenenbaum[7] assumed the observed image I(**x**,t) being a product of two components: a reflectance image R(**x**,t) and an illumination image L(**x**,t).

$$I(\mathbf{x},t) = R(\mathbf{x},t)L(\mathbf{x},t) \qquad (1)$$

To decompose an image, an assumption has to be imposed. Land *et al.*[8] proposed the Retinex Theory which imposes the spatial smoothness assumption on the illumination image L(**x**,t). Instead of using the Gaussian filtering operations, we proposed to use the DCT techniques to estimate the illumination image L(**x**,t) to facilitate video coding.

$$L_d(\mathbf{x},t) = \mathrm{IDCT}(\ \mathrm{Clip}(\ \mathrm{DCT}(I(\mathbf{x},t)\ )\ ) \qquad (2)$$

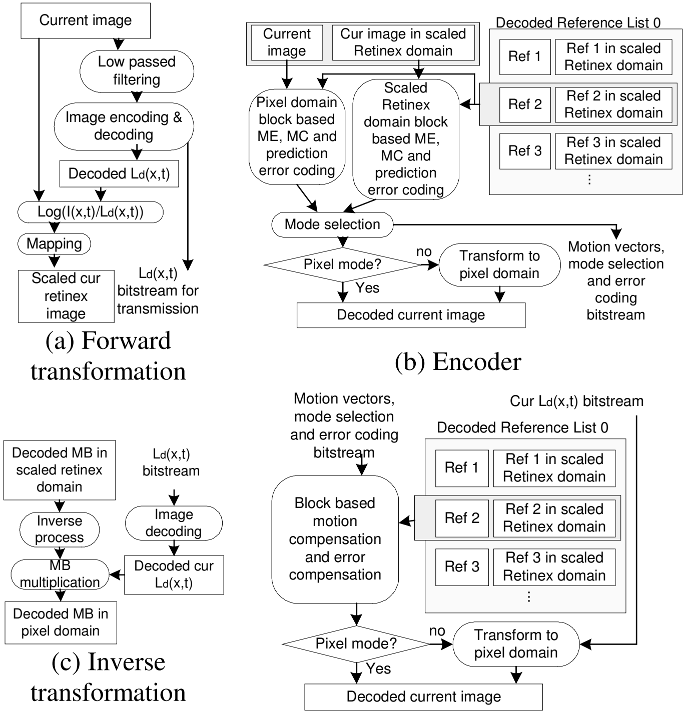
where *Clip()* is a function to preserve the first $(N_L+1)N_L/2$ low frequency coefficients in the zigzag scan order with N_L standing for the number of levels. The DCT coefficients are quantized and de-quantized with a quantization factor DCTQ before inverse DCT and the quantized coefficients are coded and sent to the decoder as the overhead bits for coding $L_d(\mathbf{x},t)$. The corresponding retinex output is

$$R_d(\mathbf{x},t) = \log I(\mathbf{x},t) - \log L_d(\mathbf{x},t) \qquad (3)$$

Subsequently, the floating point values of $R_d(\mathbf{x},t)$ from –K to K are then mapped (with upper and lower clipping) and quantized to integers ranging from 0 to 255. We refer the mapped image to the scaled retinex image(SRI). Fig. 1(a) and (c) show the block diagrams for transforming an image to and from the scaled retinex image respectively. We implemented the scheme and applied it to the multiple reference frame motion compensation environment of H.264. Figs. 1 (b) and (d) show the encoding and decoding system diagrams for P slices respectively. Each macroblock can be coded either in the conventional pixel domain or the scaled retinex domain with independent motion estimation. Mode

978-1-4244-2311-8/08/$25.00 ©2008 IEEE.

decision is made based on the mode giving the lowest rate-distortion cost. In summary, in addition to the conventional data, we need to transmit the bits for coding $L_d(\mathbf{x},t)$ for each image, and one pixel/retinex mode selection bit for each macroblock. The retinex based system is characterized with its high accuracy in estimated motion for sequences with brightness variations to improve coding efficiency.


Current image
Low passed filtering
Image encoding & decoding
Decoded Ld(x,t)
Log(I(x,t)/Ld(x,t))
Mapping
Scaled cur retinex image
Ld(x,t) bitstream for transmission
Current image
Cur image in scaled Retinex domain
Decoded Reference List 0
Ref 1
Ref 1 in scaled Retinex domain
Ref 2
Ref 2 in scaled Retinex domain
Ref 3
Ref 3 in scaled Retinex domain
Pixel domain block based ME, MC and prediction error coding
Scaled Retinex domain block based ME, MC and prediction error coding
Mode selection
Pixel mode?
no
Yes
Transform to pixel domain
Motion vectors, mode selection and error coding bitstream
Decoded current image
Decoded MB in scaled retinex domain
Ld(x,t) bitstream
Inverse process
Image decoding
MB multiplication
Decoded cur Ld(x,t)
Decoded MB in pixel domain
Motion vectors, mode selection and error coding bitstream
Cur Ld(x,t) bitstream
Decoded Reference List 0
Block based motion compensation and error compensation
Ref 1
Ref 1 in scaled Retinex domain
Ref 2
Ref 2 in scaled Retinex domain
Ref 3
Ref 3 in scaled Retinex domain
Pixel mode?
no
Yes
Transform to pixel domain
Decoded current image

(a) Forward transformation

(b) Encoder

(c) Inverse transformation

(d) Decoder

Fig. 1 Block diagrams of the image transformation and the overview video coding system

3. PROPOSED DCT BASED CONSTRAINED 1-BIT TRANSFORM

In this paper, we propose a fast approach for the ME using binary matching algorithm suitable for the DCT based retinex system. Originally, the motion vectors are estimated using the conventional sum-of-absolute-difference(SAD) approach in the scaled retinex domain.

$$SAD(\mathbf{mv_i}) = \Sigma|SRI(\mathbf{x}, t) - SRI(\mathbf{x}+\mathbf{mv_i}, t-1)| \qquad (4)$$

where $\mathbf{mv_i}$ is a candidate motion vector and $SRI(\mathbf{x},t)$ is the image $I(\mathbf{x},t)$ in the scaled retinex domain. To increase the computational efficiency, we propose to perform the binary block matching using a two-bit transform. The advantage of bit transform is that it allows an 8-bit pixel being represented by 2 bits such that only 2 bytes are required to store 8 pixels(points). Thus, matching operations for the 8 pixels can be done concurrently by simple binary operations, instead of 8 subtraction operations in the conventional matching process. The applied 2BT is a modification of the constrained 1BT proposed in [4]. Instead of using a multi-band-pass filtered image of $I(\mathbf{x},t)$, we make use of the computed illumination image $L_d(\mathbf{x},t)$ in (2) which is essentially a low pass image. The image $L_d(\mathbf{x},t)$ is compared against the original image $I(\mathbf{x},t)$ to construct the binary bit plane $B(\mathbf{x},t)$ and it is mainly used as a sort of pixel-wise threshold.

$$B(\mathbf{x},t) = 1 \text{ , if } I(\mathbf{x},t) \geq L_d(\mathbf{x},t) \qquad (5)$$
$$= 0, \text{ otherwise}$$

Generally, $B(\mathbf{x},t)$ represents the high frequency components of the image which is suitable for ME. In order to increase the accuracy in 1BT matching process, a constrain mask $CM(\mathbf{x},t)$ is computed so as to avoid including pixels that have values close to the transform threshold.

$$CM(\mathbf{x},t) = 1 \text{ , if } |I(\mathbf{x},t) - L_d(\mathbf{x},t)| \geq D \qquad (6)$$
$$= 0, \text{ otherwise}$$

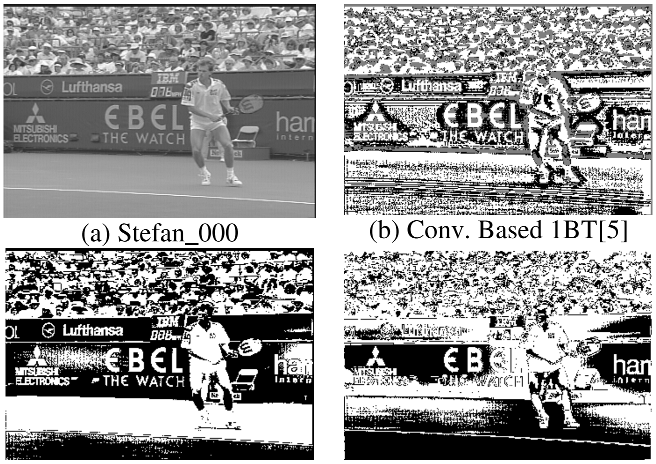
where D is a user defined value. $CM(\mathbf{x},t)$ value of 1 indicates that this pixel is reliable for the 1BT matching process. This mask can effectively increase the accuracy of motion estimation by excluding pixels from homogeneous area.

In the search process of the motion estimation, instead of using (4), the constrained number of non-matching points (CNNMP) measure is used.

$$CNNMP(\mathbf{mv_i}) = \Sigma\{[CM(\mathbf{x}, t) \,|\, CM(\mathbf{x}+\mathbf{mv_i})] \,\&\, \qquad (7)$$
$$[\, B(\mathbf{x},t) \oplus B(\mathbf{x}+\mathbf{mv_i}, t-1) \,]\}$$

where |, & and ⊕ denote Boolean OR, AND and Exclusive-OR operations respectively. Pixels that have values close to the transform threshold will be counted as a match regardless of their 1BT value. The candidate motion vector $\mathbf{mv_i}$ scoring the lowest CNNMP value is designated to be the block motion vector. If two candidate motion vectors result the same CNNMP, the one having the smaller motion vector magnitude is chosen. With the selected motion vector, motion compensation is done in the scaled retinex domain followed by the prediction error coding as shown in Fig. 1. In summary, the binary bit plane $B(\mathbf{x},t)$ and the constrain mask $CM(\mathbf{x},t)$ are generated and stored only in the encoder for fast motion estimation purpose. No extra bit is sent to the decoder and the decoder is identical to the one shown in Fig. 1(d). The computational saving should be similar to the constrained 1BT proposed in [4].

Fig. 2 shows the first image of sequence Stefan, the generated mask image CM(x,t) with D=18 and the corresponding 1BT generated using our proposed DCT based 1BT and the multiplication free 1BT proposed in [5]. It is seen from the mask image CM(x,t) that the pixels in the homogeneous area are classified as unreliable and are excluded from 1BT matching process.



(a) Stefan_000 (b) Conv. Based 1BT[5]

(c) DCT based 1BT (d) Mask CM(**x**,t)

Fig. 2 (a) Sample frame of the Stefan sequence (b) 1BT obtained using convolution based approach[5] (c) 1BT obtained using our proposed approach (d) corresponding mask CM(**x**,t) with D=18

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed approach, motion estimation has been tested using various sequences with and without inter-frame brightness variations. Six matching approaches for motion estimation with full search scheme are tested: the conventional SAD in pixel domain(SAD-pixel), SAD in the scaled retinex image domain(SAD-SR), binary matching with 1BT[5] and constrained 1BT[4] using convolution based approach(Conv-1BT-SR and Conv-2BT-SR respectively) and our proposed binary matching with constrained 1BT and its variant 1BT (without CM(**x**,t)) using DCT based approach(DCT-2BT-SR and DCT-1BT-SR respectively). Note, the abbreviation name with "SR" indicates the retinex based coding system as described in section III. PSNR values obtained between the original frames and frames reconstructed from previous frames using the computed motion vectors are used to evaluate the motion estimation accuracy. The motion estimation accuracy is assessed for a block size of 16x16, 8x8 and 4x4 with a search range of 16. A number of real and synthetic sequences having various forms of inter-frame brightness variations are used for testing. In this paper, we just quote some typical results using two sequences without brightness variation ("Stefan" and "Foreman") and 3 sequences with brightness variations. "StefanFading" (originally sequence "Stefan") involves synthetic linear fade in and fade out effect. "CameraFlash" and "SpotLightPanning" involve static objects exposed to various types of lighting effects: successive camera flashes and a moving spotlight respectively.

The average prediction quality using block sizes of 16x16, 8x8 and 4x4 is summarized in Tables 1, 2 and 3 respectively with D value set to 18 (determined experimentally) for the two variants of constrained 1BT(DCT-2BT-SR and Conv-2BT-SR). For the two sequences without brightness variation ("Stefan" and "Foreman"), conventional SAD matching in pixel domain outperforms other approaches and the prediction quality increases as the block size decreases. However, for sequences having brightness variations, most of the matching approaches using the retinex based system outperform the conventional SAD matching approach as more accurate motion estimation is achieved. In addition, the prediction quality of all the binary matching approaches generally decrease as the block size decreases indicating that binary matching is more appropriate to be applied to large block sizes. Generally, the prediction quality of the binary matching approaches is slightly lower than that of the SAD-SR approach, and the constrained 1BT approaches outperform the 1BT approaches. The performance of our proposed DCT based constrained 1BT is very close to that of the convolution based approach (with a drop of less than 0.1dB on average). However, our proposed approach is more suitable to be applied to the retinex based system as shown in Fig. 1. It is because the image used for threshold purpose, L_d(**x**,t), has already been generated during the transformation process to the scaled retinex domain while the convolution based constrained 1BT[4] needs to generate another image for threshold purpose which implies extra computational load. Therefore, the retinex based coding system with our proposed DCT based constrained 1BT method is very efficient in terms of the high computational efficiency and the capability of accurately estimating the motions under brightness variation environment.

5. CONCLUSION

In this paper, we propose a DCT based binary matching approach for fast motion estimation to be applied in our previously proposed retinex based video coding system which is characterized with the capability of accurately estimating motions under brightness variations environment. Instead of using the convolution based bit transformation techniques, we propose to use the DCT based bit transform which facilitates both the transformation of image into the binary bit plane and the scaled retinex image domain. Experimental results show that our proposed DCT based bit transform has a very close prediction quality performance to that of the convolution based approach while our system can avoid the extra convolution procedure.

6. ACKNOWLEDGEMENT

This research is supported by the NICTA (via The University of Sydney), ARC grants, and the Centre for Multimedia Signal Processing, (via project RGC-PolyU5268/06E), Hong Kong Polytechnic University.

7. REFERENCES

[1] K.-C. Hui, W.-C. Siu, and Y.-L. Chan, "New adaptive partial distortion search using clustered pixel matching error Characteristic," *IEEE Transactions on Image Processing,* vol. 14, no. 5, pp. 597-607, May 2005.

[2] P.Yin, A.M.Tourapis, and J.Boyce, “Localized weighted prediction for video coding,” *Proceedings, IEEE International Symposium on Circuits and Systems, ISCAS 2005,* vol. 5, pp. 4365-4368, 2005.

[3] H.-K. Cheung, W.-C. Siu, D. Feng *et al.*, “Retinex based motion estimation for sequences with brightness variations and its application to H.264,” *Proceedings, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008*, pp. 1161-1164, 30 March - 4 April 2008.

[4] O.Urhan, and S.Erturk, “Constrained one-bit transform for low complexity block motion estimation,” *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 17, no. 4, pp. 478-482, April 2007.

[5] S.Erturk, “Multiplication-Free One-Bit Transform for Low-Complexity Block-Based Motion Estimation,” *IEEE Signal Processing Letters,* vol. 14, no. 2, pp. 109-112, Feb 2007.

[6] A.Erturk, and S.Erturk, “Two-bit transform for binary block motion estimation,” *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 15, no. 7, pp. 938-946, July 2005.

[7] H.G.Barrow, and J.M.Tenenbaum, “Recovering intrinsic scene characteristics from images,” *Computer Vision Systems, Academic Press*, 1978.

[8] E.H.Land, and J.J.McCann, “Lightness and retinex theory,” *Journal of the Optical Society of America,* vol. 61, no. 1, pp. 1-11, 1971.

Table 1

Average PSNR(dB) of sequences reconstructed by various ME techniques (full search, block size = **16x16**, search range = 16 pixels)

	Video sequences (frame size, sequence length)				
	Stefan (352x240, 100)	Foreman (352x288, 100)	Stefen-Fading (352x240, 100)	Camera-Flash (320x240, 24)	SpotLight-Panning (320x240, 40)
SAD-pixel	25.09	32.92	29.1	16.19	26.76
SAD-SR	25.07	32.83	29.32	26.81	28.63
DCT-1BT-SR	24.23	30.08	28.42	25.32	27.7
Conv-1BT-SR	24.55	31.39	28.72	26.17	26.9
DCT-2BT-SR*	24.37	30.37	28.48	26.36	28.35
Conv-2BT-SR	24.55	31.25	28.46	26.37	28.39

*Proposed algorithm

Table 2

Average PSNR(dB) of sequences reconstructed by various ME techniques (full search, block size = **8x8**, search range = 16 pixels)

	Video sequences (frame size, sequence length)				
	Stefan (352x240, 100)	Foreman (352x288, 100)	Stefen-Fading (352x240, 100)	Camera-Flash (320x240, 24)	SpotLight-Panning (320x240, 40)
SAD-pixel	26.27	34.35	30.25	17.61	27.47
SAD-SR	26.21	34.08	30.46	27.75	28.96
DCT-1BT-SR	23.83	29.32	27.98	25.22	26.6
Conv-1BT-SR	23.92	29.05	28.18	25.27	25.59
DCT-2BT-SR*	24.43	30.34	28.29	26.17	28.15
Conv-2BT-SR	24.58	30.85	28.35	26.23	28.27

*Proposed algorithm

Table 3

Average PSNR(dB) of sequences reconstructed by various ME techniques (full search, block size = **4x4**, search range = 16 pixels)

	Video sequences (frame size, sequence length)				
	Stefan (352x240, 100)	Foreman (352x288, 100)	Stefen-Fading (352x240, 100)	Camera-Flash (320x240, 24)	SpotLight-Panning (320x240, 40)
SAD-pixel	28.15	36.21	32.00	18.81	29.20
SAD-SR	27.98	35.80	32.14	29.74	30.02
DCT-1BT-SR	22.04	29.34	26.07	25.31	26.28
Conv-1BT-SR	21.49	27.87	25.58	25.06	24.97
DCT-2BT-SR*	23.27	30.20	26.82	25.95	28.20
Conv-2BT-SR	23.35	30.47	27.02	25.97	27.90

*Proposed algorithm