RETINEX BASED MOTION ESTIMATION FOR SEQUENCES WITH BRIGHTNESS VARIATIONS AND ITS APPLICATION TO H.264

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

***Abstract*—Conventional motion estimation does not take inter-frame brightness variations into consideration, which causes inefficient video coding for sequences involving brightness variations. H.264 provides a specific mode called weighted prediction targeting to improve the coding efficiency for this case. In this paper, we propose a Retinex based motion estimation scheme which effectively removes the inter-frame de-correlation factor resulting from brightness variations. We also propose to use some DCT techniques to generate the Retinex images for both current and reference images and apply conventional motion estimation and compensation procedures for coding. We applied the scheme to the H.264 testing the efficiency in the multiple reference frame motion compensation environment. Experimental results show that our proposed scheme outperforms the H.264 system with weighted prediction enabled. It allows the system to use a smaller number of reference frames for coding, e.g. 2, to achieve a similar (or slightly better) compression efficiency of the H.264 system using 5 reference frames.**

Index Terms – Motion Estimation, video coding, motion compensation, discrete cosine transform, H.264.

I. Introduction

The latest video compression standard, H.264, has attracted much attention recently for its high compression efficiency. Block based motion estimation and compensation[1] play an indispensable role with extra features of variable block-size and multiple reference frame motion compensation (MRMC)[2,3]. The introduced rate distortion(RD) optimization[4] also gives an important role for its success. To handle video sequences having inter-frame brightness variations, weighted prediction technique is introduced to raise the prediction quality. However, the weighted prediction feature is only effective for spatially even brightness changes, primarily targeting the fade in/out effect. Barrow and Tenenbaum[5] proposed to model the brightness variations using a multiplicative model in which the observed image I($\mathbf{x}$,t) can be decomposed into two components: a reflectance image R($\mathbf{x}$,t) and an illumination image L($\mathbf{x}$,t), which describes the physical surface reflectance properties of the objects and the lighting condition respectively.

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

The decomposition is a classical ill-posed problem. Further assumptions have to be imposed to make it useful. Land *et al.*[6] proposed the Retinex Theory which imposes the spatial smoothness assumption on the illumination image L($\mathbf{x}$,t). In this paper, we propose a new Retinex based motion estimator which allows an efficient use of the conventional motion estimation and compensation scheme to handle inter-frame brightness variations

II. Proposed DCT Based Retinex for Brightness Compensation

A. Conventional Gaussian based Retinex

The primary goal of the Retinex is to decompose an image into a reflectance image and an illumination image to remove illumination effect. The core is the estimation of the illumination image L($\mathbf{x}$,t)[7,8]. One of the variations is the surround based model[9] where pixel values of L($\mathbf{x}$,t) are given by a weighted average of its surrounding. The Gaussian function is commonly used as the weighting function and the Retinex output R_g is defined as follows.

$$R_g(\mathbf{x},t) = \log I(\mathbf{x},t) - \log L_g(\mathbf{x},t) \qquad (2)$$

where $L_g(\mathbf{x},t) = F(\mathbf{x})*I(\mathbf{x},t)$ with "*" and F($\mathbf{x}$) denote the convolution operator and the Gaussian function respectively. $L_g(\mathbf{x},t)$ is essentially a low passed version of the image, I($\mathbf{x}$,t), which approximates the illumination function L($\mathbf{x}$,t).

1-4244-1484-9/08/$25.00 ©2008 IEEE

Substituting (1) into (2) and assuming $L(\mathbf{x},t)\approx\bar{L}(\mathbf{x},t)$, R_g can be rewritten as

$$R_g(\mathbf{x},t)=\log\frac{R(\mathbf{x},t)L(\mathbf{x},t)}{\bar{R}(\mathbf{x},t)\bar{L}(\mathbf{x},t)}\approx\log\frac{R(\mathbf{x},t)}{\bar{R}(\mathbf{x},t)} \qquad (3)$$

where the bar denotes the spatially weighted average value. Therefore, it is an approximation of the reflectance ratio which is characterized with illumination independence.

B. Proposed DCT based scaled Retinex

Instead of using Gaussian filtering, we propose to use DCT techniques to estimate L(**x**,t) as shown below.

$$L_d(\mathbf{x},t) = \mathrm{IDCT}(\ \mathrm{Clip}(\ \mathrm{DCT}(I(\mathbf{x},t)\ )\ ) \qquad (4)$$

where *Clip()* is a function to preserve the first $(N_L+1)N_L/2$ low frequency coefficients in zigzag scan order with N_L standing for the number of levels. The corresponding DCT based Retinex output is

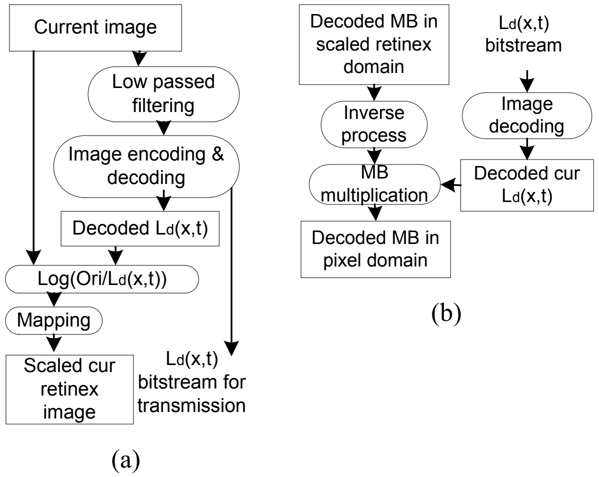
$$R_d(\mathbf{x},t) = \log I(\mathbf{x},t) - \log L_d(\mathbf{x},t) \qquad (5)$$

In the point of view of video coding, the major advantage of our proposed DCT based Retinex is its simplicity and efficiency in coding the illumination image L(**x**,t). All we need to do is to quantize and encode $(N_L+1)N_L/2$ DCT coefficients. The value of the quantization factor is denoted as DCTQ and we simply assume that each quantized coefficient is encoded to give a two-byte sized datum.

$R_d(\mathbf{x},t)$ is essentially a ratio image in log domain which is inefficient for computation and incompatible to the conventional video coders. We apply the mapping treatment of the Retinex output proposed by Jobson *et al.*[9]. We linearly map, with upper and lower bound clipping, $R_d(\mathbf{x},t)$ value ranging from –K to K to the range 0 to 255 and quantize it to a 1-byte-sized integer for storage. Jobson *et al.*[9] claimed that the gain/offset, in the mapping process, appears to be invariant from image to image. Thus, we experimentally determine the value of the mapping factor K and keep using the same value for different sequences. We refer the mapped Retinex image as scaled Retinex image in this paper.

C. Proposed video coding system

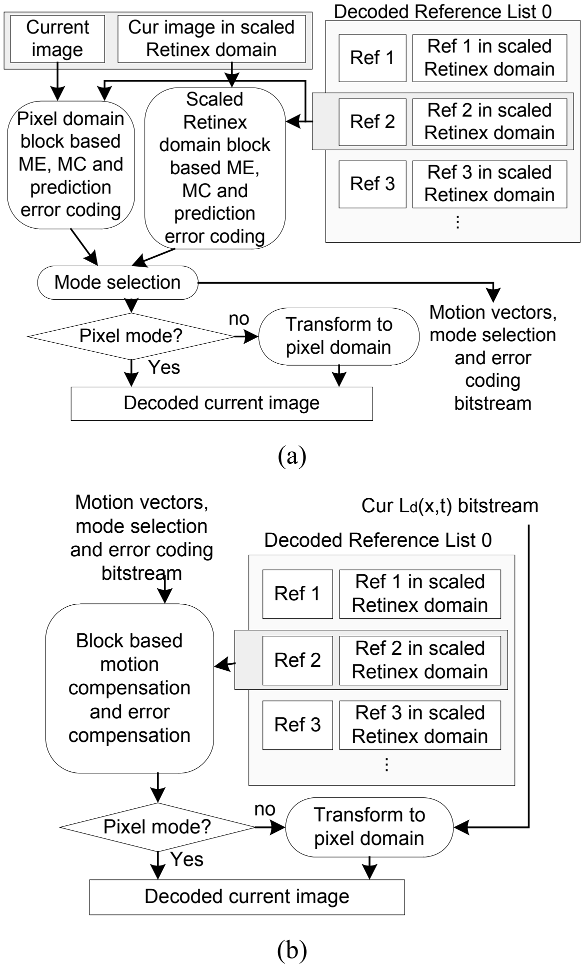
After transforming images into the scaled Retinex domain, the image contents are generally illumination free and are suitable for motion estimation and compensation using conventional motion estimator. We propose to transform the current image and all the reference images into the scaled Retinex domain and perform motion estimation, motion compensation and prediction error coding. Finally, the output image is transformed back to the pixel domain to form the decoded image. We implement our proposed scheme and apply it to the MRMC environment of H.264. Fig. 1(a) and (b) show the block diagrams for transforming the image to and from the scaled Retinex domain respectively.



(a)

(b)

Figure 1: Block diagrams for transforming image (a) to and (b) from scaled Retinex domain.


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
Motion vectors, mode selection and error coding bitstream
Cur Ld(x,t) bitstream
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
Block based motion compensation and error compensation
Pixel mode?
no
Yes
Transform to pixel domain
Decoded current image

(a)

(b)

Figure 2: Overview of the proposed video coding system, (a) encoding and (b) decoding.

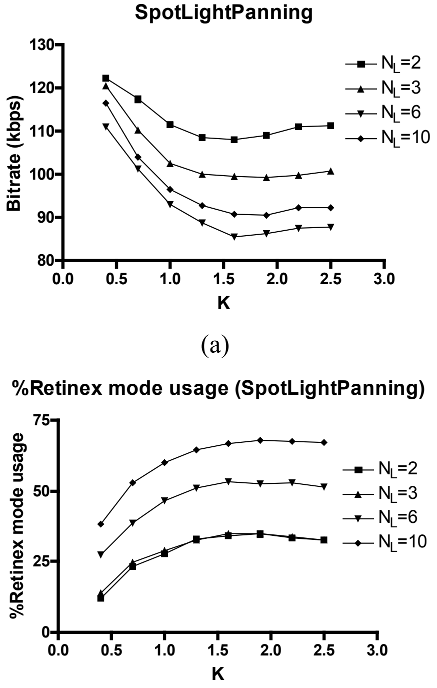
Mapping block in Fig. 1(a) stands for the linear mapping using the parameter K as explained in the previous session and the inverse process block in Fig. 1(b) stands for the corresponding reverse process. Each macroblock(MB) is allowed to choose from the coding procedures to be done in the conventional pixel domain and our proposed scaled Retinex domain. The mode decision is made based on the mode giving the lowest rate-distortion cost (the extra bits for coding $L_d(\mathbf{x},t)$ is not considered). Fig. 2(a) and (b) show the encoding and decoding system diagrams of the proposed system for P slices.

In summary, in addition to the conventional data including the motion vectors, the error compensation data and the coding mode and reference frame selection bits, we need to transmit the pixel/Retinex mode selection bits for each macroblock and the bits for coding $L_d(\mathbf{x},t)$.

III. Experimental Results

To test the performance of our proposed scheme in H.264, we used a number of real and synthetic sequences having various forms of inter-frame brightness variations. Because of page constraint, we just quote two typical results using two real sequences, CameraFlash and SpotLightPanning, in the SIF format. CameraFlash was created by a static camera taking video of static objects which were exposed to a sequence of consecutive flashes. To increase the relative portion of frames (to a total number of frames) involving significant inter-frame brightness variations, we manually picked up 6 frames, which have the flash captured, together with the immediate preceding and following frames to form a sequence of 18 frames (since there is no inter-frame motions). For SpotLightPanning, it is a sequence of 43 frames involving the lighting effect of a moving spotlight which is characterized by significant brightness variations in both spatial and time domains.

To determine the coding parameters K and N_L, a series of tests were done. We fixed the encoder configuration to high profile, IBP coding pattern, search range=32, DCTQ=3, Q=35 for I, P and B slices, and using 3 reference images for MRMC (#Ref=3). Fig. 3(a) shows the influence of K and N_L on the bitrate (PSNR is roughly 31.8dB for all cases). For all values of N_L, the bitrate decreases as K increases from 0.4 to 1.9 and increases slightly for larger K values. For small value of K, information loss due to clipping is severe while a large value of K might merge adjacent grey levels causing distortion. For $N_L< 6$, the bitrate generally decreases as the value increases and reaches a minimum at $N_L=6$. This can be explained that small value of N_L cannot effectively remove the inter-frame lighting variations affecting motion estimation while large value increases the overhead bits of coding $L_d(\mathbf{x},t)$ and outweighs the advantage. Fig. 3(b) shows the percentage of MBs selecting the new Retinex mode. The results also show that increasing the value N_L manages to increase the Retinex mode usage as the lighting effect is more effectively removed. Thus, we arbitrarily set K=1.8 and $N_L=6$ for the rest of the experiments which gives approximately the best coding efficiency. Fig. 4(a)(b) show the original images #17 and #19 of SpotLightPanning respectively. They are characterized with Brightness variation in both spatial and temporal domains. Fig. 4(c)(d) show the corresponding images in the scaled Retinex domain for K=1.8 and $N_L=6$. Note that the difference between the two images is effectively removed.



(a)

(b)

Figure 3: Relationship between K, N_L and the coding performance for sequence SpotLightPanning, with Q=35, #Ref=3.

Fig. 5(a) and (b) show the rate distortion graphs for coding the two sequences with different systems including the proposed system and the H.264 system with and without the weighted prediction(WP) enabled (explicit mode). RX denotes the number of reference frames, X, for MRMC (e.g. R5 = 5 reference frames). The encoder configuration is the same as the previous test except the value of Q and the number of reference frames for MRMC are varying. From the results, our proposed coding system outperforms the other two, especially at low bitrates. We can use a smaller number of reference frames, e.g. 2, for MRMC using our proposed scheme to achieve similar (or slightly better) coding performance to the H.264 with weighted prediction enabled using 5 reference frames. The major reason is that images in the scaled Retinex domain manage to decrease the lighting influence and increase the average usage rate of the adjacent images for reference purpose. For high bitrate, the information loss due to the quantization step in transformation process and the overhead bit for coding $L_d(\mathbf{x},t)$ might outweigh the advantage. However, the performance is comparable to the H.264 with weighted prediction enabled.

IV. CONCLUSION

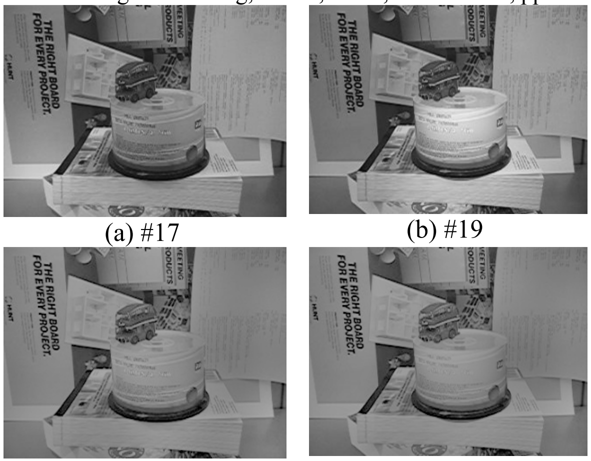
To handle sequences having inter-frame brightness variations, we propose a Retinex based motion estimation and compensation scheme which manages to remove the lighting effect. Considering the low frequency property of the scene illumination, we propose to approximate the illumination image of each frame using some DCT techniques. The Retinex image characterized with illumination independence can be produced and the conventional motion estimation and compensation techniques can be efficiently applied. Experimental results show that our proposed scheme applying to H.264 outperforms the original H.264 with weighted prediction enabled, for sequences involving brightness variations. It manages to effectively reduce the number of reference frames needed to achieve a similar coding performance

V. ACKNOWLEDGMENT

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.

REFERENCES

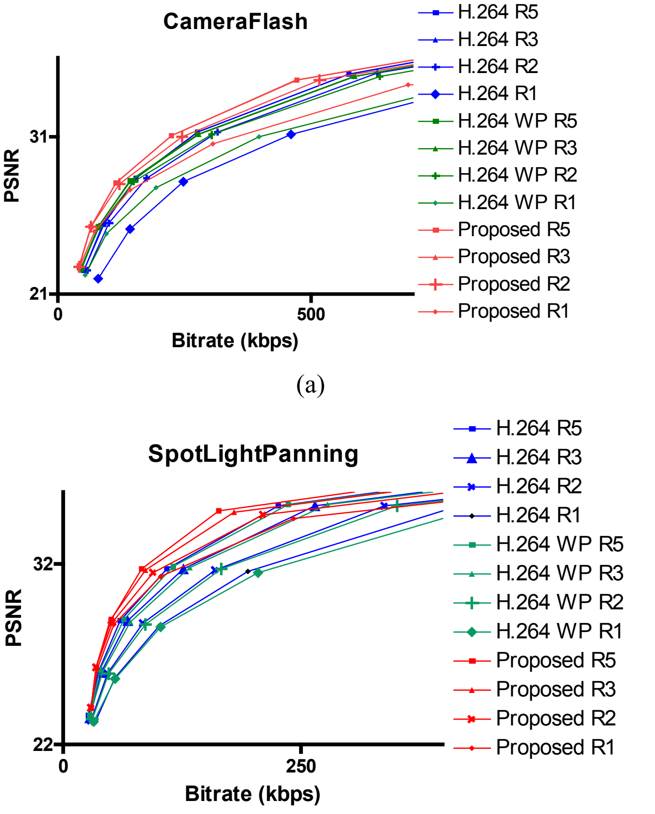
1. K.C.Hui, W.C.Siu and Y.L.Chan, “New adaptive partial distortion search using clustered pixel matching error characteristic”, IEEE Transactions on Image Processing, Vol. 14, No. 5, May 2005, pp.597-607.
2. T.Wiegand, E.Steinbach nad B.Girod, “Affine multipicture motion-compensated prediction”, IEEE Transactions On Circuits and Systems for Video Technology, Vol. 15, No. 2, Feb 2005, pp.197-209.
3. S-E.Kim, J-K.Han and J-G.Kim, “An efficient scheme for motion estimation using multireference frames in H.264/AVC”, IEEE Transactions on Multimedia, Vol. 8, No.3, June 2006, pp.457-466.
4. E-H.Yang and X.Yu, “Rate distortion optimization for H.264 interframe coding: a general framework and algorithms”, IEEE Transactions on Image Processing, Vol. 16, No. 7, July 2007, pp.1774-1784.
5. H.G.Barrow and J.M.Tenenbaum, “Recovering intrinsic scene characteristics from images”, Computer Vision Systems, Academic Press, 1978.
6. E.H.Land and J.J.McCann, “Lightness and retinex theory”, J. Opt. Soc. Am., 61(1), 1971, pp.1-11.
7. L.Meylan and S.Susstrunk, “High dynamic range image rendering with a retinex-based adaptive filter”, IEEE Transactions on Image Processing, Vol.15, No.9, Sep 2006, pp.2820-2830.
8. S.Lee, “An efficient content-based image enhancement in the compressed domain using retinex theory”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.17, No.2, Feb 2007, pp.199-213.
9. D.J.Jobson, Z.Rahman and G.A.Woodell, “Properties and performance of a center/surround retinex”, IEEE Transactions on Image Processing, Vol. 6, No. 3, March 1997, pp.451-462.



(a) #17 (b) #19

(c) #17 (d) #19

Figure 4: Images 17 and 19 of SpotLightPanning; (a) amd (b) are in pixel domain, whereas (c) and (d) are in the scaled Retinex domain, respectively.



(a)

(b)

Figure 5: Rate distortion graphs for (a)CameraFlash and (b)SpotLightPanning.