A NOVEL ILLUMINATION COMPENSATION SCHEME FOR SPRITE CODING

*Hoi-Kok Cheung**[+], *Wan-Chi Siu**, *Dagan Feng**[+] *and Kin-Wai Cho**

*Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

[+]School of Information Technologies, F09
The University of Sydney
NSW 2006, Australia

ABSTRACT

A novel illumination compensation scheme for sprite coding is presented in this paper. Inter-frame pixel discrepancy due to the effect of frustrating illumination conditions is one of the major challenges for most correlation-based global motion estimators. Our proposed system employs a lighting compensation technique, by means of some frame division operations, to eliminate the effect of illumination variation before motion estimation is conducted. The background information is then expressed in terms of our proposed mosaic based representation which can efficiently handle sequences involving frustrating illumination condition. This effectively increases the robustness of the global motion estimator. Meanwhile, this allows the normalization of the lighting conditions of each image in the sequence with respect to a particular base frame and gives rise to a sprite image characterized with global illumination homogeneity. Experimental results show that our proposed technique manages to increase the coding efficiency of video sequences which involve a strong variation of lighting conditions (up to 10 dB improvement), and is suitable for the sprite coding application in MPEG-4.

1. INTRODUCTION

Global motion estimation is one of the essential tasks in image processing and video compression fields. Sprite coding is an important core technique for video compression using global motion estimation and it is included as part of the coding standard, MPEG-4, for object-based coding. Sprite coding techniques involve generation of a high-resolution image called sprite. The image is composed of information belonging to objects visible throughout the video sequence. Usually the background objects are coded and the corresponding sprite is called "background mosaic".

The essential process involved in the generation of the background mosaic is the estimation of global motions. Irani, Anandan and Hsu[1] proposed a series of extensions to mosaic based representations of video sequences. Dufaux and Konrad[2] and Szeliski[3] proposed to estimate the global motion by an iterative minimization of the frame difference error. Smolic, Sikora and Ohm[4] proposed to employ a recursive closed-loop prediction scheme to reduce the error accumulation problem. Keller and Averbuch[5] proposed a fast gradient based approach successfully cutting computational complexity up to 33 folds. Lu, Gao and Wu[6] proposed a sprite coding scheme employing directional spatial-prediction technique.

In this paper, we propose a novel sprite generation technique which manages to handle video sequences with strong variation of atmospheric lighting conditions and express the background object information in terms of a single layered mosaic based representation. Generally speaking, one of the weaknesses for major global motion estimation techniques is the unreliability for handling sequences with background variations. One of the common background variation phenomena is the variation of atmospheric lighting condition. We propose an illumination compensation technique which can raise the robustness of global motion estimator by eliminating the de-correlation effect between frames due to a change of lighting environment and propose a mosaic based representation to code the background. With the illumination normalization effect, all the frames can be normalized with respect to a chosen frame in terms of the lighting condition and subsequently a sprite image with illumination homogeneity is constructed. The lighting information of individual frame is coded separately and the reconstructed frames can be de-normalized accordingly to restore the original sequence during decoding stage.

2. ILLUMINATION COMPENSATION

In conventional static sprite coding scheme[1], the background scene is represented by a static image. For high coding efficiency, there is a vital hypothesis that the intensity structures of a time-varying image are approximately constant under motion for at least a short duration. Therefore, conventional static sprite image cannot coherently represent the intensity variation over the time interval defined by the illumination-condition-varying video shot. In addition, conventional motion estimators[7] cannot accurately estimate the true motion activities between frames under different lighting conditions. To efficiently handle this problem, an image I(x,y) can be decomposed into two components: a reflectance image R(x,y) and an illumination image L(x,y). Barrow and Tenenbaum[8] introduced the term "intrinsic images" to refer to the decomposition. The reflectance image describes the physical surface reflectance properties of the objects in the image and the illumination image describes the light intensity distribution over the image at that instant of time. The three images are related by

$$I(x,y) = R(x,y)L(x,y) \tag{1}$$

Ideally, the reflectance images of a video sequence are extracted and are used to construct a static sprite image using conventional global motion estimation techniques[2,3,4,5]. Unfortunately, decomposition of a single image into two intrinsic images is a classic ill-posed problem as the number of unknowns is twice the number of equations. Assumptions have to be made for feasible decomposition. Weiss[9] proposed to employ the assumption that the distribution of the derivative filtered images is sparse. Land and McCann[10] proposed a Retinex theory and assumed that the reflectance image is piecewise constant. Kimmel, Elad, Shaked, Keshet and Sobel[11] proposed a variational framework for Retinex and a new model was proposed. Matsushita, Nishino, Ikeuchi and Sakauchi[12] proposed to use this technique to remove shadows for video surveillance purpose.

In this paper, we assume that the reflectance of a scene is constant over time while the illumination image varies from time to time to account for the lighting conditions. The underlying concern of our proposed representation is to separate the factor of the varying illumination from the images so as to facilitate the coding

0-7803-8406-7/04/$20.00

efficiency and the global motion estimation process. Instead of decomposing the reflectance images from a video sequence, the illumination ratio of each frame and a reference frame is computed and encoded. We arbitrary choose the first frame of the sequence to be the reference for simplicity. As the reflectance of the scene is constant, the ratio is identical to the ratio of the illumination images as stated below.

$$R_m(x,y) = \frac{MC(I_m(x,y), A_{m,0})}{I_0(x,y)} = \frac{MC(L_m(x,y), A_{m,0})}{L_0(x,y)} \qquad (2)$$

where I_0, I_m, L_0, L_m are the first frame, m-th frame, first illumination image and m-th illumination images, respectively. $A_{m,0}$ is the registration matrix denoting the global motions between frame m and frame 0 and $MC(I_m, A_{m,0})$ denotes the global motion compensated frame of I_m using $A_{m,0}$. The ratio $R_m(x,y)$ denotes the illumination ratio between frame m and frame 0. By computing the ratio $R_m(x,y)$, the illumination variation factor is extracted. Therefore, frame m can be normalized, in the sense of global illumination, with respect to the first frame by the following simple division operation.

$$I_m'(x,y) = \frac{I_m(x,y)}{R_m(x,y)} \qquad (3)$$

where $I'_m(x,y)$ is frame m under the illumination condition as frame 0. However, it is a paradox that an image can be normalized once the registration matrix $A_{m,0}$ is estimated while accurate estimation of global motions requires that one of the images be normalized with respect to the other. To resolve this problem, we assume that the illumination is uniform over the images. $R_m(x,y)$ is approximated by making use of the registration matrix of the previous frame A_{m-1} and the mean of $R_m(x,y)$ is used to normalize the current image $I_m(x,y)$.

$$\hat{R}_m(x,y) = \frac{MC(I_m(x,y), A_{m-1,0})}{I_0(x,y)} = \frac{MC(L_m(x,y), A_{m-1,0})}{L_0(x,y)} \qquad (4)$$

$$\hat{I}_m'(x,y) = \frac{I_m(x,y)}{mean(\hat{R}_m(x,y))} \qquad (5)$$

where $\hat{R}_m(x,y)$ and $\hat{I}_m'(x,y)$ are respectively the approximated ratio image and normalized image m with respect to image 0. Subsequently, $A_{m,0}$ is estimated using $\hat{I}_m'(x,y)$, $I_0(x,y)$ and A_{m-1}. Finally, $I'_m(x,y)$ is computed using (2) and (3). With all the frames normalized and registered, a sprite image having homogeneous lighting condition of the scene can then be constructed using the conventional method[2,4].

In the decoding stage, the reconstructed image is warped from the sprite, which is under the lighting condition of the first frame, and is de-normalized to its original illumination condition.

$$\hat{I}_m(x,y) = R_m(x,y) \cdot W(sprite, A_{m,0}) \qquad (6)$$

where *sprite* and $\hat{I}_m(x,y)$, are the sprite image and the reconstructed frame m respectively. $W(sprite, A_{m,0})$ is the warped image from the sprite using registration matrix $A_{m,0}$. With the assumption that the lighting effect is globally uniform and the energy of $R_m(x,y)$ mainly concentrates on the low frequency part. We propose to transform $R_m(x,y)$ using the 2D DCT and encode only the first six low frequency coefficients. These six coefficients are referred to as illumination compensation coefficients(ICC) of frame m.

3 THE PROPOSED SYSTEM

The proposed system employs our earlier robust global motion estimation scheme[13] as the basis of the encoder. Fig. 1(a) shows the structure of the system proposed in this paper. To reduce error accumulation, a particular frame in the history of time, frame k, is chosen to be the reference for motion estimation[13]. Registration of the current frame, frame m+1, is accomplished by projecting frame m+1 into the space of frame k using registration matrix A_{mk} to generate frame z. Subsequently, the global motion between frame z and frame k is estimated. Initially, the reference frame is chosen to be the first frame of the sequence. The current frame, frame m+1, would replace the reference frame if the relative displacement between the current frame and the reference frame is large. This can be detected if either the overlapping area between the registered current frame and the reference frame is smaller than a threshold T1 or the non-overlap area is larger than a threshold T2.

$$T_1 = w \times h - T_2 \text{ and } T_2 = w \times h \times Nr \qquad (7)$$

where w and h are the width and height of the frame respectively. Nr is a parameter defined between 0 and 1 controlling the value of the two thresholds.

Fig. 1(b) shows the details of the block "Motion estimation and ICC estimation" in Fig. 1(a). Initially, a rough illumination compensation operation is performed for frame z with respect to frame k. Subsequently, global motions between frame k and frame z is estimated generating registration matrix $A_{(m+1)k}$. Finally, ICC of frame m+1 is estimated using frame m+1 and illumination compensated frame k, with respect to the first frame.

Fig. 1(c) shows the details of the block "Initial illumination compensation" in Fig. 1(b). Firstly, pixels of frame k are divided by pixels of frame z to generate the ratio image, $\hat{R}(x,y)$. As frame z is not properly registered to frame k, only the DC component of $\hat{R}(x,y)$ is considered. Frame z is then illumination compensated with respect to frame k by multiplying each pixel by the mean value of $\hat{R}(x,y)$ generating illumination compensated frame z.

Fig. 1(d) shows the details of the block "Global motion estimation" in Fig. 1(b) which attempts to estimate the global motion between frame k and illumination compensated frame z in a coarse-to-fine manner. The estimator firstly builds two three-level pyramids for frame k and the illumination compensated frame z using a low-pass filter. Subsequently, the initial translation is estimated at the top level of the pyramids using a three-step search matching. The illumination compensated frame z is then motion compensated using the estimated global motion again for subsequent GME. At the medium-level, the partial distortion search algorithm is employed to estimate motion vectors of some selected pixel blocks with a search range of ± 2. The block selection technique employed is our earlier work on global motion estimation[14]. The aim is to exclude blocks of pixels, which mainly consist of highly texture area or homogeneous area, to avoid aperture problem and to reduce the chance of trapping into a local minimum for fast gradient method in the later stage. Registration matrix A_{zk}, in the form of affine motion model, is obtained in the least square sense. The block-based global motion estimation is performed again in the base-level with a search range of ±4 and the resultant registration matrix A_{zk} is refined iteratively using the fast gradient method proposed by Keller and Averbuch[5]. Finally, the estimated A_{zk} is then concatenated with A_{mk} by simple matrix multiplication to obtain $A_{(m+1)k}$.

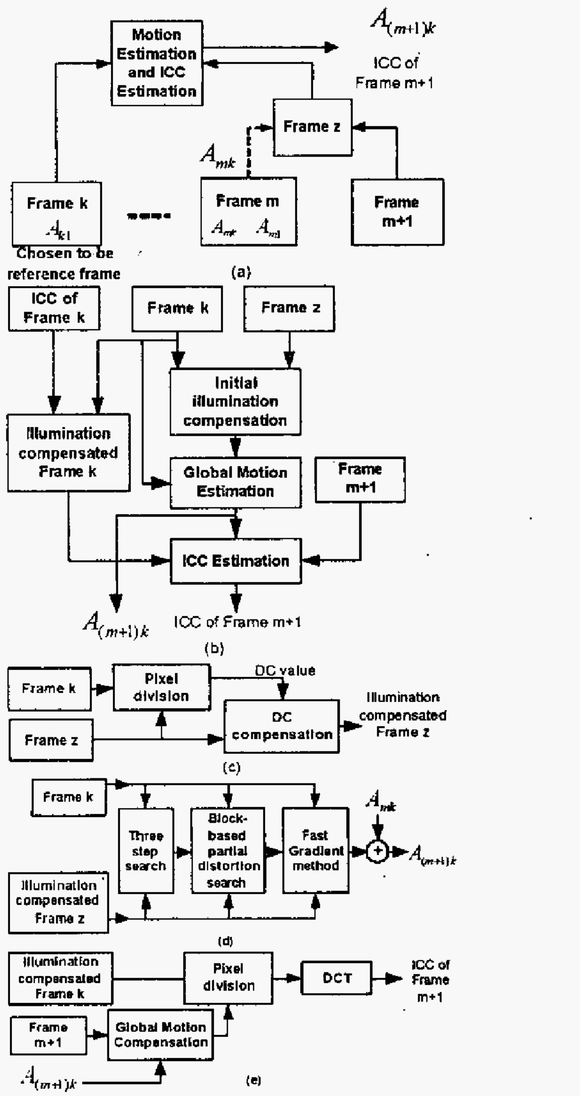


Fig. 1 (a)Block diagram of the global motion estimation algorithm. Frame m+1 is projected to form frame z using registration matrix A(mk). (b) Details of he motion estimation and ICC estimation block in (a). (c) Details of the initial illumination compensation block in (b). (d) Details of the global motion estimation block in (b). (e) Details of the ICC estimation block in (b).

Fig. 1(e) shows the details of the block "ICC estimation" of Fig. 1(b). After the estimation of registration matrix $A_{(m+1)k}$, the ICC of frame m+1, with respect to the first frame, can be estimated. Frame m+1 is firstly global motion compensated using $A_{(m+1)k}$ and the result is divided by the illumination compensated frame k, with respect to the first frame, generating the ratio frame $R_{m+1}(x,y)$. Finally, 2D DCT of $R_{m+1}(x,y)$ is performed and the first six low frequency coefficients, $(c_0, c_1, c_2, c_3, c_4, c_5)$ in the zigzag scan order as shown in Fig. 2, are taken to form the illumination compensation coefficients (ICC) of frame m+1. Finally, the estimated registration matrix $A_{(m+1)k}$ is concatenated to $A_{k,0}$ to give the long-term registration matrix $A_{(m+1)0}$.

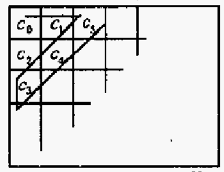


Fig. 2 Six illumination compensation coefficients

4. CONSTRUCTION OF SPRITE

In the construction of the sprite, we employ our earlier technique[13] by choosing the frame having the highest resolution of the scene to be the base frame. Before projecting and blending the frames on to the space of the chosen base frame, each of the frame is normalized with respect to the illumination condition of the first frame according to the estimated ICC as stated in equation (3). This operation allows a construction of a sprite having homogeneous lighting condition over the scene.

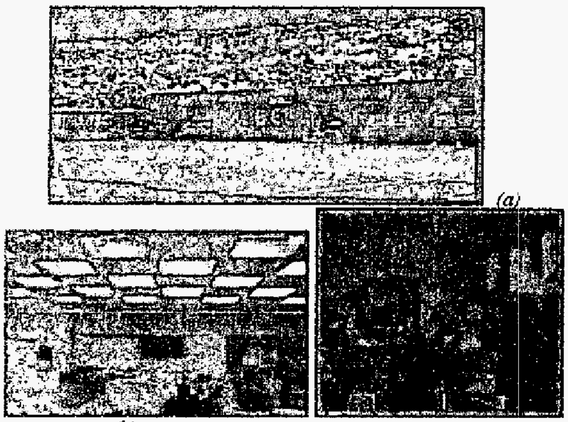
In the decoding stage, images are formed by the projection of the sprite according to the estimated registration matrixes. As images produced are all normalized with respect to the first image, it is necessary to de-normalize the lighting conditions according to their respective ICC to complete the frame reconstruction operations.

5. EXPERIMENTAL RESULTS

In this section, we will show some experimental results using our proposed system. Four test sequences with CIF format including 150 frames were tested, namely "Stefan" (352x240), "Coast Guard" (352x288), and two of our home-made sequences. "ZoomIn" (352x240) and "Brighter" (352x288). The first two sequences are provided with segmentation mask while the last two sequences do not involve foreground object. Sequence "Brighter" involves a significant amount of variation in lighting conditions while sequence "ZoomIn" involves a considerable amount of zoom operations in which the automatic gain control feature of the camera causes the atmospheric brightness variation in the frames.

In our experiments, we chose to use perspective motion model to describe the global motions (with Nr = 0.1). Fig. 3(a) depicts the sprite image generated with our system for sequence "Stefan". Fig. 3(b) and (c) show the sprite images having homogeneous illumination condition as the first frame for sequences "ZoomIn" and "Brighter". Fig. 4(a) to (c) show some of the decoded and de-normalized frames from sequence "Brighter" which involves strong variations of lighting conditions. This shows that our proposed system can effectively handle sequences with a strong variation of lighting conditions.

Table 1 depicts the quality of the reconstructed frames generated by different coding systems. These two systems make use of the same global motion estimator as described in section 3 and the reconstructed frames are compared with the original sequences. Generally speaking, our system with illumination compensation(IC) feature outperforms the system without IC feature except for sequence "CoastGuard". For sequences "Stefan" and "CoastGuard", variations of lighting condition is insignificant and the effect of IC feature is unimportant. The variation of seawater pattern in sequence "CoastGuard" might account for the quality degradation in the reconstructed frames using the IC feature. For sequences "ZoomIn" and "Brighter", the advantage of the proposed IC feature is revealed especially for sequence "Brighter" which involves a strong variation of lighting conditions. The quality improvement is mainly brought by the increase in robustness of the global motion estimator, as the illumination is compensated before processing, and the capability of our proposed sprite coding scheme which manages to dynamically vary the lighting condition accordingly for frame reconstruction. Table 1 also shows the performance of the proposed system.



(a) (b) (c)

Fig.3 Sprite generated by our proposed system (a)Stefan (b)ZoomIn (c)Brighter

(a) decoded frame 0 (b) decoded frame 48 (c)decoded frame 149

Fig.4 Decoded frames for sequence Brighter

Sequence	Stefan	Coast-Guard	Brighter	ZoomIn
System w/o IC*	22.65dB	23.45dB	20.63dB	27.87dB
Proposed system	**22.77dB**	**22.99dB**	**31.41dB**	**28.75dB**

** - without illumination compensation*

Table 1 Comparison of the quality of the reconstructed frames generated with different systems

6. CONCLUSION

In this paper, we have presented a novel illumination compensation scheme for sprite coding application. The proposed technique manages to estimate accurately the global motion activities for video sequences involving a strong variation of lighting conditions and to represent the illumination-varying scene of the sequence using a single layered sprite image. In most of the conventional correlation-based global motion estimators, the lighting effect severely weakens the reliability of the global motion estimation results. With our proposed illumination compensation technique, the lighting conditions of frames can be normalized with respect to the first frame. It also facilitates the improvement of accuracy in global motion estimation. In addition, the sprite image can be constructed within a homogeneous illumination environment which greatly improves the visual quality of the decoded sequence with a strong lighting variation effect.

7. REFERENCES

[1]M.Irani, P.Anandan and S.Hsu, "Mosaic based representations of video sequences and their applications", Fifth International Conference on Computer Vision, June 1995, pp.605-611

[2]F. Dufaux and J.Konrad, "Efficient, robust, and fast global motion estimation for video coding", Image Processing, IEEE Transactions on Image Processing, Vol. 9, Issue 3, March 2000, pp.497-501

[3]R. Szeliski, "Video mosaics for virtual environments", IEEE Computer Graphics and Applications, Vol. 16, Issue 2, March 1996, pp.22-30

[4]A. Smolic, T. Sikora and J. R. Ohm, "Longterm global motion estimation and its application for sprite coding, content description, and segmentation", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9 Issue 8, Dec 1999, pp.1227-1242

[5]Y.Keller and A.Averbuch, "Fast gradient methods based on global motion estimation for video compression", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No.4, April 2003, pp.300-309

[6]Y.Lu, W.Gao and F.Wu, "Efficient background video coding with static sprite generation and arbitrary-shape spatial prediction techniques", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No.5, May 2003, pp.394-405

[7]Y.L.Chan and W.C.Siu, "Edge oriented block motion estimation for video coding", IEE Preceedings Vision, Image and Signal Processing, Vol. 144, Issue 3, June 1997

[8]H.G.Barrow and J.M.Tenenbaum, "Recovering intrinsic scene characteristics from images", Computer Vision Systems, Academic Press, 1978

[9]Y.Weiss, "Deriving intrinsic images from image sequences", IEEE International Conference on Computer Vision, Vol. 2, July 2001, pp.68-75

[10]E.H.Land and J.J.McCann, "Lightness and retinex theory", Journal of the optical Society of America, 61(1):1-11, 1971

[11] R.Kimmel, M.Elad, D.Shaked, R.Keshet and I.Sobel, "A variational framework for retinex", International Journal of Computer Vision, Vol.52, Number 1, April 2003, pp.7-23

[12]Y.Matsushita, K.Nishino, K.Ikeuchi and M.Sakauchi, "Illumination normalization with time-dependent intrinsic images for video surveillance", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 2003, pp.I-3-I-10

[13]H.K.Cheung and W.C.Siu, "A robust global motion estimation scheme for sprite coding", IEEE International Symposium on Circuits and Systems, Vol. 2, 2003, pp.II-632-II-635

[14]H.K.Cheung and W.C.Siu, "Fast global motion estimation for sprite generation", IEEE International Symposium on Circuits and Systems, Vol 3, 2002, pp.1-4

[15]Y.He, B.Feng, S.Yang and Y.Zhong, "Fast global motion estimation for global motion compensation coding", IEEE International Symposium on Circuits and Systems, Vol. 2, 2001, pp. 233-236