Proceedings of ITC-CSCC 2000, Pusan, Korea

# Novel Image Stabilizing Techniques for Mobile Video Communications

Byoung-Su Kang, Jae-Won Kim, Jun-Suk Lee, Kang-Sun Choi, and Sung-Jea Ko
The Department of Electronics Engineering, Korea University, Seoul 136-701, Korea
Tel: +82-2-3290-3228, Fax: +82-2-928-0179
E-mail: sjko@dali.korea.ac.kr

**Abstract:** In this paper, we present two types of digital image stabilization (DIS) schemes for mobile video communications. In the first scheme, the DIS system, which is used as a preprocessor of the video encoder, compensates the camera's undesirable shakes before encoding. This method can reduce the bit rate of encoded video sequence by attenuating the prediction error to be encoded. In the second proposed scheme, the DIS system is coupled with the video decoder. The second scheme uses the K-means clustering algorithm to estimate the camera motion using motion vectors decoded from the received video stream. Simulation results show that the first scheme improves coding efficiency, while the second scheme is computationally efficient since it does not require motion estimation.

## 1. Introduction

Image stabilization, which compensates the image motion caused by the camera's undesirable shakes or jiggles, is a fundamental video sequence operation. Since mobile video communication systems such as mobile multimedia terminals generate the video data stream under motion, undesirable fluctuation of the camera degrades the picture quality. Thus, the DIS becomes more and more important in mobile video communications.

Various digital image stabilization systems have been developed for digital cameras or camcorders [1]-[3]. To combine the DIS algorithm with low bit-rate video codecs such as MPEG or H.263 codecs, two different coupling schemes can be considered:

*i*) The DIS is used as a preprocessor of a video encoder before low bit-rate video encoding. Although the jiggling analysis to obtain a global motion of the image frame is required additionally, the motion information as well as prediction errors can be reduced in a consequent motion compensated video. The conventional DIS techniques for video cameras or camcoders may be used for the preprocessor.

*ii*) The second coupling scheme is that the video decoder is combined with the DIS. This scheme is computationally efficient since the DIS system can directly obtain motion information from the video decoder.

In this paper, we present an efficient DIS system combined with the MPEG decoder for mobile communication. The proposed method is based on local motion information of MPEG encoded video stream. The motion field between two successive frames is constructed using local motion vectors obtained from the MPEG encoded video stream. We use the K-means clustering algorithm to classify the motion field into two clusters corresponding to background and foreground. This algorithm assumes that a dominant motion is only incurred by the camera motion in the entire frame. The global motion vector is determined by dominant motion vectors which are identified as the motion of background.

This paper is organized as follows. The proposed preprocessor DIS technique for a video encoder is presented in Section 2. In Section 3, the proposed DIS scheme coupling with MPEG decoder is presented. Simulation results are given in Section 4 and concluding remarks are in Section 5.

## 2. DIS Combined with Video Encoder

The configuration of the DIS system with combined with the video coder is shown in Fig. 1. The local motion estimation unit produces local motion vectors from subimages in the different position of the frame. The global motion decision unit determines the global motion vectors by processing these local motion vectors, and the integration unit using the accumulated global motion vector decides whether the motion of a frame is caused by undesirable fluctuation of the camera or intentional panning. Finally, the stabilized image is generated by reading out the proper block of fluctuated image in the frame memory.

Local motion estimation plays an important role in the

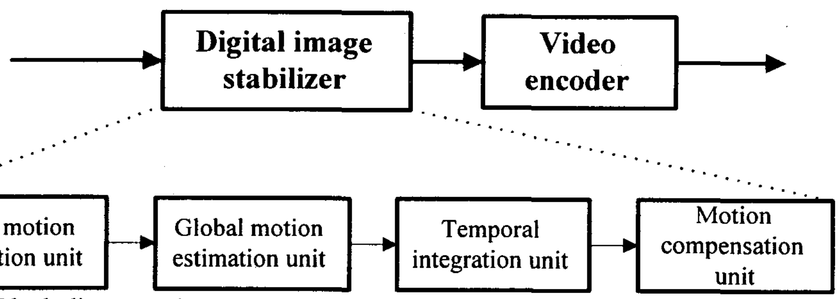



Fig. 1. Block diagram of the proposed DIS system combined with the video encoder.

DIS since local motion vectors are used for determining the global motion vector (hand-movement). Before describing the proposed method, we introduce the bit-plane and Gray-coded bit-plane decomposition of a gray-scale image.

*A. Bit-plane and Gray-coded Bit-plane Decomposition of a Grayscale Image*

Let the graylevel of the pixel at location (x,y) in the $t^{th}$ image frame with $2^K$ graylevels be represented as

$$f^t(x,y) = a_{K-1}2^{K-1} + a_{K-2}2^{K-2} + \cdots + a_1 2^1 + a_0 2^0, \quad (1)$$

where $a_k$, $0 \le k \le K-1$, is either 0 or 1. Let the $k^{th}$ order bit-plane image be denoted by $b^t_k(x,y)$. This plane contains all the $k^{th}$ order ($a_k$) bits. For the case of the 8-bit image, an image is composed of eight 1-bit planes $b^t_0(x,y) \sim b^t_7(x,y)$, ranging from plane 0 to plane 7. $b^t_0(x,y)$ contains all the least significant (lowest order) bits comprising pixels in the image and $b^t_7(x,y)$ contains all the most significant (highest-order) bits.

The $t^{th}$ image frame $f^t(x,y)$ can be also represented by $k$-bit gray-code. The $k$-bit gray code $g_{K-1} \cdots g_2 g_1 g_0$ can be computed from

$$g_{K-1} = a_{K-1}$$
$$g_k = a_k \oplus a_{k+1}, \quad 0 \le k \le K-2, \quad (2)$$

where, $\oplus$ is the exclusive OR operation and $a_k$ is the $k^{th}$ bit of the base 2 representation given by (1).

*B. Motion Estimation Based on Gray-coded Bit-plane Matching (GC-BPM)*

In the proposed scheme, three local motion vectors are estimated from subimages (regions 1, 2, and 3) placed in appropriate positions in the Gray-coded bit-plane as shown in Fig. 2. In the proposed method, the camera motion is initially estimated using regions 1 and 2. In the case of detecting movements due to persons in the background or other moving objects in the background, the estimated motion of region 3 becomes important and the stabilizing mechanism of the image is switched to reduce the jiggling with respect to the foreground [4] (see Fig. 3).

Fig 2: Regions for subimage motion estimation.


(a) (b)

Fig. 3. Motion field obtained by local motion estimation with 16X16 macroblock. (a) A frame of video sequence. (b) Motion field.

Next we present the GC-BPM for local motion estimation. Let the size of each block be $M \times N$ and a search window be $(M+2p) \times (N+2p)$. For the proposed GC-BPM, we define the correlation measure given by

$$C_j(m,n) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g_k{}^t(x,y) \oplus g_k{}^{t-1}(x+m, y+n),$$
$$-p \le m,n \le p \quad (3)$$

where $g_k{}^t(x,y)$ and $g_k{}^{t-1}(x,y)$, respectively, are the current and previous $k^{th}$ order Gray-coded bit-planes, and $p$ is the maximum displacement in the search window.

At each $(m,n)$ within the search range, the proposed matching method calculates $C_j(m,n)$ which is the number of unmatched bits between the reference subimage in the current Gray-coded bit-plane and the compared subimage in the previous Gray-coded bit-plane. The smallest $C_j(m,n)$ yields the best matching for each subimage, and thus motion vector $V_j$ from $j^{th}$ subimage is obtained as

$$V_j = \arg\min\{C_j(m,n), -p \le m,n \le p\} \quad (4)$$

This motion estimation technique can replace the arithmetic calculation of BMA's based on conventional MAD and MSE criteria with simple Boolean exclusive-OR operations, and thus has significantly reduced computational complexity. Fig. 4 shows the motion estimation system using the proposed algorithms. In Fig. 4, the GC-BPM performs motion estimation using a single bit-plane, it is important to select an appropriate Gray-coded bit-plane for GC-BPM. We experimentally determine the most desirable Gray-coded bit-plane for GC-BPM.

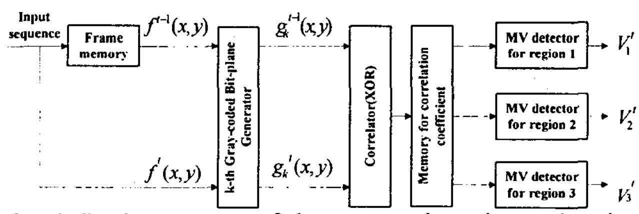



Fig. 4. Basic structure of the proposed motion estimation system.

## 3. Image Stabilization Technique Coupled with MPEG Decoder

MPEG frames are coded in one of three modes: I frames are coded using intraframe coding. P frames are encoded using interframe prediction from the previous P (or I) frames. B frames are coded using bi-directional interframe prediction from the previous and the next P (or I) frames.

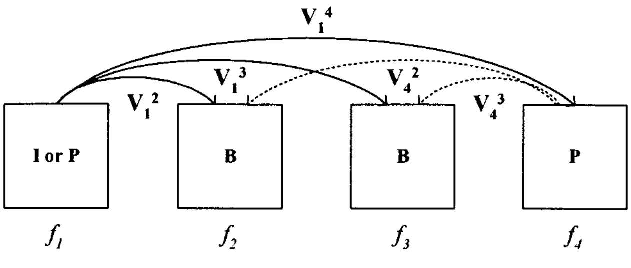
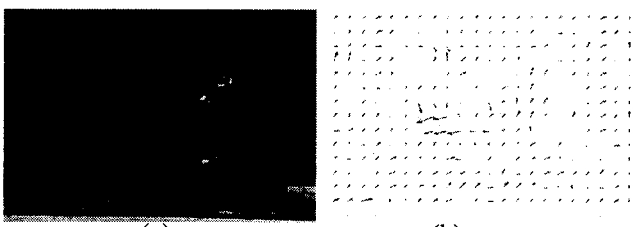



Fig. 5. Temporal prediction between frames of MPEG.

Let M denote the period between two adjacent reference frames for motion prediction. In Fig. 5, for example, $f_1$ and $f_4$ are the reference frames and M = 3.

In order to estimate the GMV $\mathbf{V}_i^j$, we utilize the MV's from blocks of 16x16 pixels called macroblocks (MB's), each of which is represented as $\mathbf{v} = [v_h \quad v_v]^T$ where $v_h$ and $v_v$ are the horizontal and vertical components of $\mathbf{v}$, respectively.

In the proposed method, we first construct a motion field using the MB motion vectors of $f_j$ as shown in Fig. 6. Since the MV's obtained from the image background tend to be very similar in both magnitude and direction, we use the K-means clustering algorithm [5] to classify the motion field into two clusters corresponding to the background and foreground.



(a) (b)

Fig. 6. Motion field obtained from an MPEG encoded video stream. (a) A frame of a video sequence. (b) Motion field.

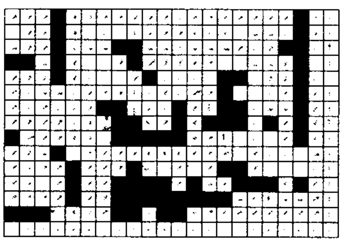


Fig. 7. Motion field segmentation.

Fig. 7 illustrates the result of the clustering algorithm where the shaded region is the MB in the foreground.

Next, we estimate the GMV $\mathbf{V}_i^j$ using the clustered motion field as follows:

$$\mathbf{V}_i^j = \mathbf{v}_k, \quad |\mathbf{v}_k| = median\left\{|\mathbf{v}_1|, |\mathbf{v}_2|, \cdots, |\mathbf{v}_L|\right\}, \qquad (5)$$

where $\mathbf{v}_i$'s are $L$ MV's in the cluster corresponding to the background and $|\mathbf{v}| = \sqrt{v_h^2 + v_v^2}$ .

Under the assumption that hand-movement is uniform for the period $M$, the GMV between two successive frames, $\mathbf{V}_{i-1}^i$, can be obtained using GMV's, $\{\mathbf{V}_1^2, \mathbf{V}_1^3, \mathbf{V}_1^4, \mathbf{V}_4^2, \mathbf{V}_4^3\}$, as follows:

$$\mathbf{V}_{i-1}^i = \frac{\mathbf{G}_k}{M}, \quad |\mathbf{G}_k| = median\left\{|\mathbf{G}_1|, |\mathbf{G}_2|, |\mathbf{G}_3|\right\}, \qquad (6)$$

where $\mathbf{G}_1 = \mathbf{V}_1^4$, $\mathbf{G}_2 = \mathbf{V}_1^2 - \mathbf{V}_4^2$, and $\mathbf{G}_3 = \mathbf{V}_1^3 - \mathbf{V}_4^3$.

To decide whether the motion of a frame is caused by hand-movement or intentional panning, the global motion vector of a frame is accumulated with a damping coefficient as

$$\mathbf{A}_i = \alpha \ \mathbf{A}_{i-1} + \mathbf{V}_{i-1}^i, \qquad (7)$$

where $\mathbf{A}_i$ represents the accumulated motion vector at the $i^{th}$ frame and $\alpha$ $(0 < \alpha < 1)$ is a damping coefficient. The accumulated motion vector becomes the final motion vector of the frame for motion compensation.

## 4. Experimental Results

To evaluate the performance of the first proposed DIS system, preprocessor scheme, we compare the bit rate plot for the fluctuated original sequence with that for the stabilized sequence. As shown in Fig. 8, the bit rate of stabilized sequence is lower than that of fluctuated original sequence. This result means that motion information is reduced in consequent motion compensated video as well as prediction errors.

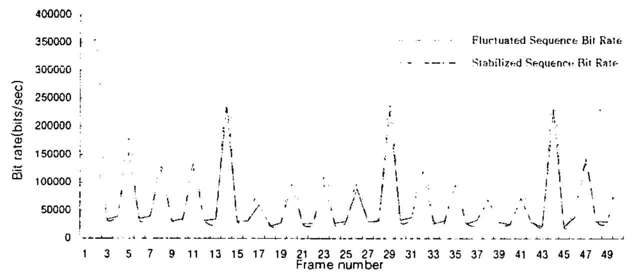



Fig. 8. The comparison of MPEG encoded bit rate.

In the second proposed scheme where DIS is coupled with MPEG decoder, we compare the two measures given by

$$R_1 = PSNR(f_i, f_{i-1}), \quad R_2 = PSNR(f_i^g, f_{i-1}^g), \qquad (8)$$

where $f_i^g$ is the $i^{th}$ frame stabilized by the proposed method. $R_1$ and $R_2$ measure the similarity between two successive frames of fluctuated sequence and stabilized sequence, respectively. The proposed method is applied to a test sequence in Fig. 9(a). Fig. 9(b) shows that the proposed coupling with MPEG decoder scheme improves the PSNR performance by reducing the difference between consecutive frames.

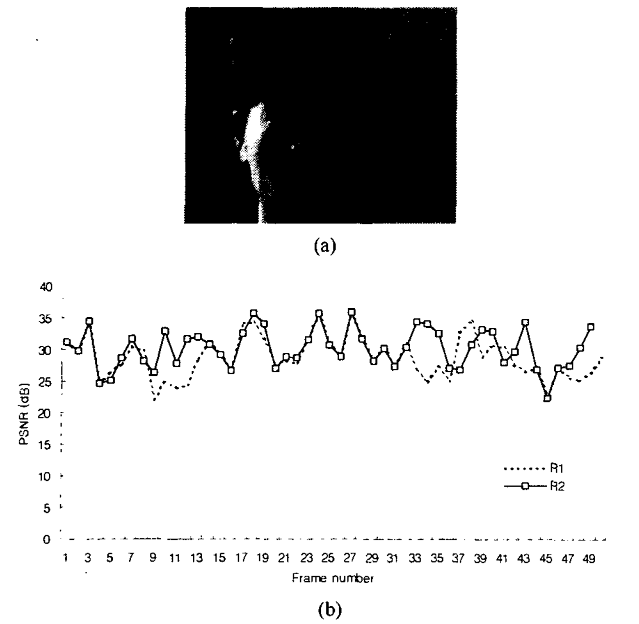



Fig. 9. A fluctuated MPEG video sequence. (a) A frame of the sequence. (b) Comparison of fluctuated and stabilized sequences.

## 5. Conclusions

In this paper, we proposed two types of DIS schemes for mobile video communications. The first DIS scheme used as a preprocessor of video encoder can improve the coding efficiency of MPEG encoder. The second scheme coupled with MPEG decoder uses motion information obtained from an MPEG video stream within an MPEG decoder. Therefore, additional motion estimation process is not required. By exploiting the characteristics of hand-movement, the proposed DIS technique effectively estimates camera motion with the K-means clustering algorithm. Simulation results show that both schemes improve the overall visual quality.

## References


[1] K. Sato, S. Ishizuka, A. Nikami, and M. Sato, "Control techniques for optical image stabilizing system," *IEEE Trans. on Consumer Electronics*, vol. 39, no. 3, pp. 461-466, Aug. 1993.

[2] K. Uomori, A. Morimura, H. Ishii, T. Sakaguchi, and Y. Kitamura, "Automatic image stabilizing system by full-digital signal processing," *IEEE Trans. on Consumer Electronics*, vol. 36, no. 3, pp. 510-519, Aug. 1990.

[3] S. –J. Ko, S. –H. Lee, and K. –H. Lee, "Digital image stabilizing algorithms based on bit-plane matching," *IEEE Trans. on Consumer Electronics*, vol. 44, no. 3, pp. 617-622, Aug. 1998.

[4] A. Engelsberg, and G. Schmidt, "A comparative review of digital image stabilizing algorithms for mobile video communications," *IEEE Trans. on Consumer Electronics*, vol. 45, no. 3, pp. 591-597, Aug 1999.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, London: Academic Press, 1990.