MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

13—14

A Wavelet-Based Template locating Criterion for Electronic Digital Image Stabilizer Application

Hamid R. Pourreza
Amirkabir University of Technology
Tehran, Iran
pourreza @ce.aku.ac.ir

Farid Behazin
YMA College
Tehran, Iran
behazin @ieee.org

Mohammad Rahmati
Amirkabir University of Technology
Tehran, Iran
rahmati @ce.aku.ac.ir

Abstract

Template selection is an important step in a feature based electronic digital image stabilizer (EDIS). In this paper, we present a wavelet-based criterion for template locating that can be used in block-based electronic digital image stabilizers. This criterion computes the vertical and horizontal details of an image and employs them for template selection. The performance of this criterion is compared with an information-based algorithm in two ways. First it is used to evaluate registration of real and noisy images. For second evaluation, an electronic digital image stabilizer is designed and the stabilized sequences generated by using information-based algorithm and our algorithm have been compared. The evaluations show that our criterion has better performance than entropy based criterion.

1 Introduction

The problem of image stabilization is the process of generation of compensated video sequence in which any unwanted camera motion is removed from the original sequence [1]. Using mechanical stabilizers based on accelerometers, gyros, or mechanical dampers is one traditional solution to the image stabilization problem, but these techniques are typically not precise and even after mechanical stabilization there may be significant residual image motion [2]. This shortcoming has led to the use of electronic digital image stabilizers. These image stabilizers use digital image processing techniques [3-7].

Image stabilization can generally be obtained in two basic stages I) motion estimation and II) motion correction (warping), as shown in Fig. 1. In feature based stabilizers, several local motion vectors in different positions of an image are computed, then global motion is estimated from these vectors [7]. One method for local motion vector estimation is the use of block matching techniques [7,8]. In this technique, a motion vector is computed by finding best match between a block of current frame, i.e. feature point (FP), with a sub-image of previous frame. The proper selection of FP is very important in this method of motion estimation. The two important problems in proper template selection are template size and its location. Improper selection of FP can cause miss-registration and aperture problem. To avoid miss-registration and aperture problem, FP must contains enough information. In this paper, problem of locating the proper FP is addressed by introducing a new criterion. Then we compared its performance with an entropy-based criterion. We will show that the proposed criterion has better performance. For performance evaluation, we used two methods I) still images in registration and II) image sequences in stabilization.

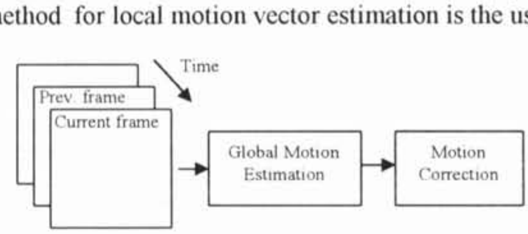

Fig. 1. Components of an EDIS.

This paper is organized as follows. The entropy-based criterion that called *Intensity Variation Number* (IVN) is introduced first section 2. In section 3 we explain our criterion that is the *Strong Vertical and Horizontal Edges* (SVHE). Next, in section 4 we design an electronic digital image stabilizer for evaluation of our method in FP selection. Some of the experimental results and conclusion are presented in section 5 and 6 consequently.

2 The IVN criterion

In this section we explain the IVN as a criterion for FP selection. This criterion is related with the entropy. Mustafa [9] has used the IVN criterion for evaluation of gray information by introducing:

$$IVN = \sum_{i=0}^{L-1} \Gamma(h_i), \tag{1}$$

where h denotes the normalized histogram of image intensity, L is the number of gray levels in the image and

$$\Gamma(u) = \begin{cases} 0, & if\ u = 0 \\ 1, & else \end{cases}. \tag{2}$$

He shows that IVN is highly correlated with entropy. Mustafa has calculated the local IVN in $n \times n$ image neighborhood for optimum selection of template location. The optimum template location is the one that maximized the local IVN.

The aperture problem can not be solved using IVN criterion, because the entropy and the IVN have no

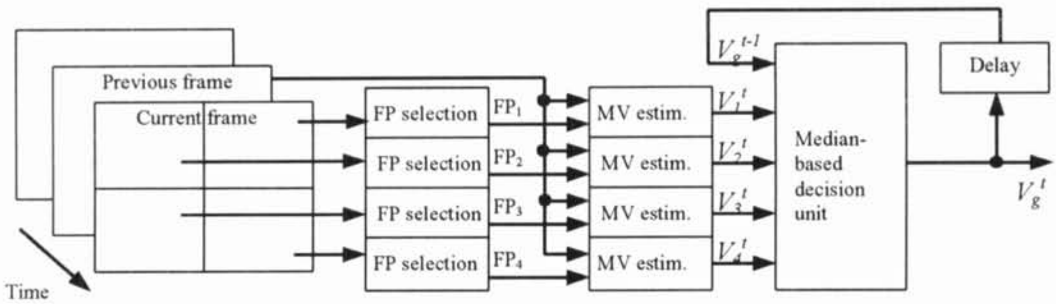

Fig. 2. Basic structure of global motion estimation stage.

information about arrangement of pixels. If selected FP contains any corner of an object or vertical and horizontal edges, then the aperture problem does not occured.

In the next section we present a new wavelet-based criterion for template selection. In this criterion, we search for a template that has both strong vertical and horizontal edges. We refer to this criterion the *Strong Vertical and Horizontal Edges* (SVHE) criterion.

3 The SVHE criterion

In this section we explain the SVHE as a criterion for selecting the FP based on the presence of strong vertical and horizontal edges. For evaluation of vertical and horizontal edges strength, the vertical and horizontal detail coefficients of discrete dyadic wavelet transform (DWT) are employed. For a continuous wavelet transform (CWT) of $f(x)$, we have the following function:

$$CWT(a,\tau) = \int_{-\infty}^{\infty} \psi_{a,\tau}(x) f(x)\,dx, \tag{3}$$

where $\psi_{a,\tau}(x)$, τ and a are respectively the wavelet function, the shift value of the original function and the scale change. Dyadic wavelet transform is a discrete wavelet transform that its parameters discretized as following:

$$a = 2^{v},\ \tau = u\,2^{v}, \tag{4}$$

($u, v \in Z$, the set of integers). Thus we have:

Dyadic Wavelet Transform:

$$d_{u,v} = \int_{-\infty}^{\infty} \psi_{u,v} f(x)\,dx. \tag{5}$$

An efficient approach to implement dyadic wavelet transform using filters was developed by Mallat [10]. The Mallat algorithm is, in fact, a classical well-known scheme in the signal processing community as a two-channel subband coder. This method first decomposes input signal into low-frequency (approximation content) and high-frequency (details content) components, then approximation and detail components are downsampled. The generated samples are called DWT coefficients. The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower-resolution components. This is called the *wavelet decomposition tree*. For an image or two-dimensional array, three detail components generated in any stage: vertical, horizontal and diagonal details.

For each template size, local vertical and horizontal coefficients energy can be measured as the vertical and horizontal edges strength, respectively. The local vertical and horizontal coefficient energy can be calculated as following:

$$E_v(x,y) = \sum_{i=-p}^{p} \sum_{j=-p}^{p} \left(c_v^k(x-i, y-j)\right)^2 \tag{6}$$

and

$$E_h(x,y) = \sum_{i=-p}^{p} \sum_{j=-p}^{p} \left(c_h^k(x-i, y-j)\right)^2, \tag{7}$$

where E_v and E_h denotes local energy of the vertical and horizontal wavelet coefficients (c_v^k and c_h^k) at level k and $p = \left\lfloor \dfrac{n}{2^k} \right\rfloor$ where n denotes template size.

To evaluate the presence of both vertical and horizontal edges, we use SVHE criterion as following:

$$SVHE(x,y) = Min\{E_v(x,y), E_h(x,y)\}. \tag{8}$$

The optimum template location corresponds to the location that the SVHE is maximized.

4 EDIS Design

Electronic Digital image stabilization is obtained in two basic stages I) motion estimation and II) motion correction (warping). We use several local motion vectors in different position of each frame for global motion estimation. This global motion is used for second stage. In second stage warping is doing by bilinear interpolation.

Global motion vector of a frame is estimated based on four local motion vectors obtained from four Cartesian regions and the past global motion vector, (see Fig. 2). For each of Cartesian regions, we first select a FP, then by using the template matching and previous frame, the motion vector is estimated for it in pixel accuracy. The motion vector is calculated with sub-pixel accuracy by using differential method [11].

The regions including any moving objects, can produce motion vectors which are significant different from the other motion vectors. These vectors must be excluded from the global motion estimation process. Moreover, for a camera motion that is relatively slower than the frame rate, the global movements for two successive frames are approximately the same. Based on these properties, use of a median filter, as shown in Fig. 2 (median-based decision unit), can be produced a robust estimation for global motion by using (V_1^t, V_2^t, V_3^t, V_4^t, V_g^{t-1}) [7]. Thus global motion vector is:

$$V_g^t = median\left\{V_1^t, V_2^t, V_3^t, V_4^t, V_g^{t-1}\right\}. \qquad (9)$$

For comparing the performance of EDIS systems, different evaluation measures have presented in [11]. One of these measures is fidelity. The fidelity measure is PSNR between stabilized frames of a sequence. For fully motion compensated images, the same frame should be obtained repeatedly. Then the difference between the two stabilized images should be zero. Many factors such as noise, estimation error distortion caused by departures from the motion model and by interpolating during warping, contribute to this difference being non-zero. PSNR between images I_1 and I_0 of size N×M pixels and 8 bits/pixel is given by:

$$PSNR(I_1, I_0) = 10\log\frac{255^2}{MSE(I_1, I_0)}, \qquad (10)$$

where MSE denotes the mean squared error:

$$MSE(I_1, I_0) = = \frac{1}{N\times M}\sum_{i=1}^{N}\sum_{j=1}^{M}(I_1(i,j) - I_0(i,j))^2 \qquad (11)$$

The PSNR gives a relation between the desired output and the residual image. The higher the PSNR between two stabilized frames means the better the fidelity of the system.

5 Experimental results

The experiments are performed in two ways. The first experiment was conducted on ten 256x256 grayscale still image scenes as shown in Fig. 3 in the presence of Gaussian noise with zero mean and various variances. To select the template location, for each size $n\times n$, the SVHE is calculated as described earlier. The selected template corresponds to the location that the SVHE is maximized. Our experiments evaluation is performed for 10 different template sizes, n={5, 7, 9, 11, 13, 15, 19, 23, 27, 31}, and 10 different signal to noise ratio (SNR), SNR={7, 8, 10, 12, 15, 18, 21, 24, 27, 30} dB. The variance of Gaussian noise is calculated by:

$$\sigma_n^2 = \frac{P_s}{10^{\frac{SNR}{10}}}, \qquad (12)$$

where P_s denotes the signal power, that is:

$$P_s = \frac{1}{N\times M}\sum_x\sum_y f^2(x,y), \qquad (13)$$

where $f(x,y)$ and N×M denote the original image and its size respectively.

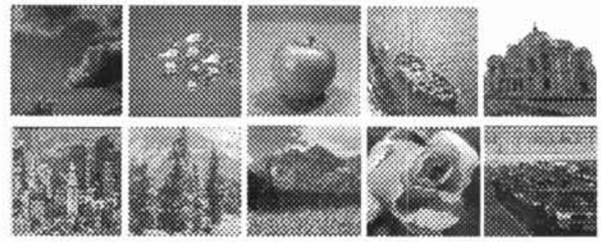

Fig. 3. Test images.

When the FP selection is carried-on, then the evaluation of best match is done by full-search and sum of absolute difference (SAD) measure for 'best fit' [12,13]. Template selection and template matching are evaluated based on a single image with specific SNR but at two different seeds. In other word, at each SNR, two different noisy images are generated from original images. The FP is selected from one noisy image and the template matching is done with second noisy image. This process is repeated 10 times and correct registrations are counted. From this simulation several observation can be stated about the overall matching results:

1. For low SNRs and small template sizes, the results of SVHE are much better than IVN, as shown in Fig.4.
2. Based on Ref. [9] and our experiments, at a template size of n=5 and using the IVN criterion the template did not match correctly, while using the SVHE at SNR>27dB, the FP matched correctly (Fig. 4).

The matching performance for the SNRs 7, 8 and 10dB and various template sizes are shown in Fig. 4, top row. For the template sizes 5,7 and 9 and various SNRs, matching performance are shown at the bottom row of this figure.

Second experiment is performed on our electronic digital image stabilizer. The IVN and SVHE criteria are used for feature point selection in EDIS, and stabilization results for five real sequences (Fig. 5) are compared. Fidelity measure is used to compare the performance of two stabilizers. Table 1 shows the result of our experiments and it is seen that the SVHE has a higher fidelity from IVN criterion.

Table 1. Stabilizer fidelity for SVHE and IVN based stabilizers for five real sequences

Sequence #	PSNR (dB)	
	SVHE	IVN
1	31.20	30.02
2	19.47	17.69
3	35.62	35.09
4	26.19	26.47
5	31.68	31.58
Average	28.83	28.17

For our experiments we employed the *Haar* wavelet transform in the second level of decomposition tree. The selection of this level is a tradeoff between computational load, noise immunity, and locating

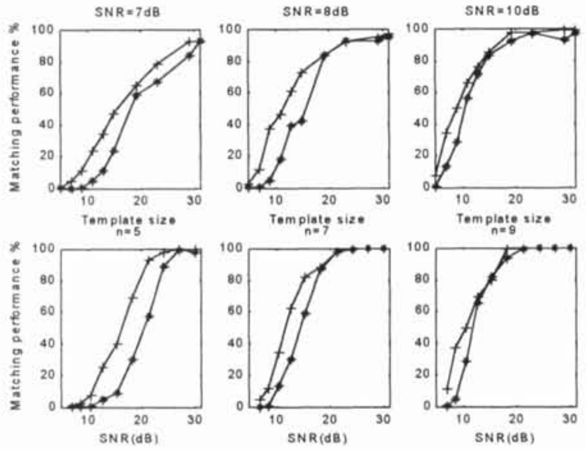

Fig. 4. Matching performance. Top: curves for constant SNR. Bottom: curves for constant Template size (+: SVHE , *:IVN).

Fig. 5. 1st, 15th and 30th frames of sequence #5.

accuracy. The results verified by using other wavelet functions with larger sizes, but increasing the complexity of the function does not yield any considerable effects on the results.

6 Conclusions

The SVHE is used as a criterion for proper template selection in template matching. The SVHE is minimum local energy of vertical and horizontal details in wavelet representation, for grayscale images. The selected template corresponds to the location on the image that maximize the amount of SVHE for a given $n \times n$ template size. The advantage of SVHE is verified by conducting tests on 10000 images from ten real images. The experimental results indicate that SVHE presents better performance than IVN. Then by using five real sequences and an EDIS, it is shown that the SVHE is better than IVN for EDIS applications.

References

[1] C. Morimoto; R. Chellappa; "Evaluation of Image Stabilization Algorithms"; *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*; pp. 2789-2792; May 1998.

[2] M. Hansen; P.Anandan; K. Dana; G. van der Wal; P. Burt; "Real-time Scene Stabilization and Mosaic Construction"; *Proc. of the Second IEEE Workshop on Application of Computer Vision*; pp. 54-62; Dec. 1994

[3] K. Uomori, A. Morimora, 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.

[4] J.-K. Paik; Y.-C. Park; and D.-W. Kim; "An Adaptive Motion Decision System for Digital Image Stabilizer Based on Edge Pattern Matching"; *IEEE Trans. On Consumer Electronics*; Vol. 38; No. 3; pp. 607-615; Aug. 1992.

[5] T. Kinugasa; N. Yamamoto; H. Komatsu; S. Takase; and T. Imaide; "Electronic Image Stabilizer for Video Camera Use"; *IEEE Trans. On Consumer Electronics*; Vol. 36; No. 3; pp. 520-525; Aug. 1990.

[6] C. Guestrin; F. Cozman; M.G. Simoes; "Industrial Application of Image Mosaicing and Stabilization"; *Proc. Second International Conf. On Knowledge-Based Intelligent Electronic Systems*; pp. 174-183; Apr. 1998.

[7] 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.

[8] R.D. Horton; "A Target Cueing and Tracking System (TCATS) for Smart Video Processing"; *IEEE Aerospace and Electronic Systems*; Vol. 6; Issue 3; pp. 8-13; Mar. 1991.

[9] A.A. Mustafa; "Optimum Template Selection for Image Registration Using ICMM"; *Proc. On British Machine Vision Conf. 98*; pp. 811-821; 1998.

[10] S. Mallat; "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation"; *IEEE Trans. on Pattern Analysis and Machine Vision*; Vol. 11; No. 7; pp. 674-693; 1989.

[11] Morimoto, C.; "Electronic Digital Image Stabilization: Design and Evaluation, with Applications"; PHD Thesis; *University of Maryland; Collage Park*; 1997.

[12] V. Bhaskaran; K. Konstantinides; "**Image and Video Compression Standards**"; *Kluwer Academic Publishers*; 1997.

[13] H.R. Pourreza; M. Rahmati; F. Behazin; "Simple and Efficient Bit-Plane Matching Algorithms for Video Compression"; *Proc. of Real-Time Image Sequence Analysis Workshop* (RISA2000); Aug. 2000.