Hindawi Publishing Corporation
Research Letters in Signal Processing
Volume 2008, Article ID 417915, 5 pages
doi:10.1155/2008/417915

Research Letter

Estimation of Subpixel Motion Using Bispectrum

El Mehdi Ismaili Aalaoui[1, 2] and Elhassane Ibn-Elhaj[2]

[1] *Faculté des Sciences de Rabat, Université Mohammed V Agdal, 4 Avenue Ibn Battouta, B.P. 1014 RP, Rabat, Morocco*
[2] *Institut National des Postes et Télécommunication, Madinat Al Irfane, Rabat, Morocco*

Correspondence should be addressed to El Mehdi Ismaili Aalaoui, ismailimehdi@gmail.com

Received 19 September 2007; Accepted 1 February 2008

Recommended by Liang-Gee Chen

Motion estimation techniques are widely used in today's video processing systems. The frequently used techniques are frequency-domain motion estimation methods, most notably phase correlation (PC). If the image frames are corrupted by Gaussian noises, then cross-correlation and related techniques do not work well. In this paper, however, we have studied this topic from a viewpoint different from the above. Our scheme is based on the bispectrum method for sub-pixel motion estimation of noisy image sequences. Experimental results show that our proposed method performs significantly better than PC technique.

1. INTRODUCTION

Image frames are generated by scanning a scene several times a second where each frame, generally, consists of two regions. One region, referred to as stationary background,is virtually the same as the previous frame. The other region, referred to as moving data, has moved with respect to the previous frame. Estimating the motion between image frames has long been a problem of interest in areas such as video compression, robot vision, and biomedical engineering [1].

Many motion estimation schemes have been developed. They can be classified into spatial-domain and frequency-domain approaches. The spatial-domain algorithms consist of matching algorithms and gradient-based algorithms. The frequency domain algorithms consist of phase correlation algorithms, wavelet transform-based algorithms, and DCT-based algorithms [2]. The vast majority of these algorithms consider noise-free data, although in [3] the displacement vector is estimated from noisy data using the generalized maximum likelihood criterion. If the image frames are corrupted by Gaussian noises, then cross-correlation and related techniques do not work well. In this circumstance, higher-order spectra (HOS) in general and the bispectrum in particular have recently been widely used as an important tool for signal processing. The classical methods based on the power spectrum are now being effectively superseded by the bispectral ones due to some definite disadvantages of the former. These include the inability to identify systems fed by non-Gaussian noise (NGN) inputs and nonminimum phase (NMP) systems, and by identification of system nonlinearity. In these cases, the autocorrelation-based methods offer no answer. Out of all these, the identifiability of NMP systems has received the maximum attention from researchers.

HOS-based methods have already been proposed to estimate motion between image frames [4–9]. In [6], the displacement vector is obtained by maximizing a third-order statistics criterion. In [8], the global motion parameters obtained by a new region recursive algorithm. In [4, 5], several algorithms are developed based on a parametric cumulant method, a cumulant-matching method, and a mean-kurtosis error criterion. In this correspondence, a novel algorithm for the detection of motion vectors in video sequences is proposed. The algorithm uses bispectrum method for subpixel motion estimation of noisy image sequences to obtain a measure of content similarity for temporally adjacent frames and responds very well to scene motion vectors.

2. BISPECTRUM-BASED IMAGE MOTION ESTIMATION

The problem of motion estimation can be stated as follows: "given an image sequence, compute a representation of the motion field that best aligns pixels in one frame of the sequence with those in the next" [9]. This is formulated as

$$\begin{aligned} g_{k-1}(x,y) &= f_{k-1}(x,y) + n_{k-1}(x,y), \\ g_k(x,y) &= f_k(x,y) + n_k(x,y) \\ &= f_{k-1}(x - d_k x, y - d_k y) + n_k(x,y) \\ &= g_{k-1}(x - d_k x, y - d_k y) + w_k(x,y), \end{aligned} \tag{1}$$

where $w_k(x,y) = n_k(x,y) - n_{k-1}(x - d_k x, y - d_k y)$; (x,y) denotes spatial image position of a point; $g_k(x,y)$ and $g_{k-1}(x,y)$ are observed image intensities at instant k and $k-1$, respectively; $f_k(x,y)$ and $f_{k-1}(x,y)$ are the noise-free frames; $n_k(x,y)$ and $n_{k-1}(x,y)$ are assumed to be spatially and temporally stationary, zero-mean image Gaussian noise sequences with unknown covariance; $(d_k x, d_k y)$ is the displacement vector of the object during the time interval $[k-1,k]$.

The third-order autocumulants and cross-cumulants of a zero-mean from 2D random field $g_k(x,y)$ are defined, respectively, as follows:

$$\begin{aligned} &C_3^{g_k g_k g_k}(r_1, r_2; s_1, s_2) \\ &\quad = E\big[g_k(x,y) g_k(x + r_1, y + r_2) \cdot g_k(x + s_1, y + s_2)\big], \\ &C_3^{g_k g_{k-1} g_k}(r_1, r_2; s_1, s_2) \\ &\quad = E\big[g_k(x,y) g_{k-1}(x + r_1, y + r_2) \cdot g_k(x + s_1, y + s_2)\big], \end{aligned} \tag{2}$$

where $E\{\cdot\}$ represents the expectation operator.

The discrete bispectrum of the frame $g_k(x,y)$ is defined as follows:

$$\begin{aligned} &B_3^{g_k g_k g_k}(u_1, u_2; v_1, v_2) \\ &\quad = \mathcal{F}\big[C_3^{g_k g_k g_k}(r_1, r_2; s_1, s_2)\big] \\ &\quad = G_{g_k}(u_1, u_2) G_{g_k}(v_1, v_2) G_{g_k}^*(u_1 + v_1, u_2 + v_2), \end{aligned} \tag{3}$$

where $\mathcal{F}$ denotes the Fourier transform operation; $G_{g_k}(u)$ corresponds to the DFT of the frame $g_k(x,y)$; $*$ indicates the complex conjugate; (u_1,u_2) and (v_1,v_2) are the frequency coordinates for the 2D Fourier transform.

Due to the shift-invariance of the bispectrum

$$\begin{aligned} &B_3^{g_k g_k g_k}(u_1, u_2; v_1, v_2) \\ &\quad = B_3^{g_{k-1} g_{k-1} g_{k-1}}(u_1, u_2; v_1, v_2) + B_3^{w_k w_k w_k}(u_1, u_2; v_1, v_2), \end{aligned} \tag{4}$$

where $B_3^{g_{k-1} g_{k-1} g_{k-1}}(u_1, u_2; v_1, v_2)$ and $B_3^{w_k w_k w_k}(u_1, u_2; v_1, v_2)$ denote bispectrum of the frame $g_{k-1}(x,y)$ and noise, respectively. If the probability density function of the noise is symmetrical, that is, $p(n) = p(-n)$, or at least not skewed, that is, $\int n^3 p(n) dn = 0$, then the term $B_3^{w_k w_k w_k}(u_1, u_2; v_1, v_2)$ is negligible which renders the triple-correlation very effective in detecting a signal embedded in noise [10]. Then

$$B_3^{g_k g_k g_k}(u_1, u_2; v_1, v_2) = B_3^{g_{k-1} g_{k-1} g_{k-1}}(u_1, u_2; v_1, v_2). \tag{5}$$

The cross-bispectrum is defined as follows:

$$\begin{aligned} &B_3^{g_k g_{k-1} g_k}(u_1, u_2; v_1, v_2) \\ &\quad = \mathcal{F}\big[C_3^{g_k g_{k-1} g_k}(r_1, r_2; s_1, s_2)\big] \\ &\quad = G_{g_k}(u_1, u_2) G_{g_{k-1}}(v_1, v_2) G_{g_k}^*(u_1 + v_1, u_2 + v_2). \end{aligned} \tag{6}$$

Then,

$$\begin{aligned} &B_3^{g_k g_{k-1} g_k}(u_1, u_2; v_1, v_2) \\ &\quad = B_3^{g_{k-1} g_{k-1} g_{k-1}}(u_1, u_2; v_1, v_2) \cdot e^{-j2\pi(v_1 d_k x + v_2 d_k y)}. \end{aligned} \tag{7}$$

Using the relation in (5), we can transform (7) as

$$\begin{aligned} &B_3^{g_k g_{k-1} g_k}(u_1, u_2; v_1, v_2) \\ &\quad = B_3^{g_k g_k g_k}(u_1, u_2; v_1, v_2) \cdot e^{-j2\pi(v_1 d_k x + v_2 d_k y)}. \end{aligned} \tag{8}$$

As we can see from (3), the bispectrum has two vector arguments containing totally four scalar frequency variables. Assuming that $G_{g_k}(u_1,u_2)$ is an N-by-N discrete Fourier transform of $g_k(x,y)$, the bispectrum becomes a four-dimensional N-by-N-by-N-by-N matrix. It is therefore not practical to evaluate the whole bispectrum. A better solution is to take 2D slices of the 4D spectrum. There are basically various ways of defining these slices, but we will only consider the case where

$$\begin{aligned} B_{3,l}^{g_k g_k g_k}(u_1, u_2) &= B_3^{g_k g_k g_k}(u_1, u_2; l u_1, l u_2) \quad \forall l \in R, \\ B_{3,l}^{g_k g_{k-1} g_k}(u_1, u_2) &= B_3^{g_k g_{k-1} g_k}(u_1, u_2; l u_1, l u_2) \quad \forall l \in R. \end{aligned} \tag{9}$$

Although we have now taken only a small portion of the whole spectrum [11], it can be shown that the motion vector is still possible and no essential information has been lost.

Thus, the third-order hologram, $h_l(r_1,r_2)$, is defined by

$$h_l(r_1, r_2) = \mathcal{F}^{-1}\left[\frac{B_{3,l}^{g_k g_{k-1} g_k}(u_1, u_2)}{\left|B_{3,l}^{g_k g_k g_k}(u_1, u_2)\right|}\right] = \delta(r_1 - d_k x, r_2 - d_k y). \tag{10}$$

As a result, by finding the location of the pulse in (10) we are able to tell the displacement, which is the motion vector. Since third-order statistics are used, the method is insensitive (in theory) to both spatially and temporally corrupted by noise which is symmetrically distributed (e.g., Gaussian). In practice, the motion vector is not an impulse; hence, we estimate $(d_k x, d_k y)$ as the index (r_1, r_2), which maximizes $|h_l(r_1,r_2)|$.

The co-ordinates (r_{1m}, r_{2m}) of the maximum of the real-valued array $h_l(r_1,r_2)$ can be used as an estimate of the horizontal and vertical components of motion between $g_k(x,y)$ and $g_{k-1}(x,y)$ as follows:

$$(r_{1m}, r_{2m}) = \arg\max\left(h_l(r_1, r_2)\right). \tag{11}$$

3. SUBPIXEL ACCURACY

Subpixel performance is a critical element of the proposed algorithm. With reference to our previously published work

[12, 13], we are introducing a number of important new features, which improve the accuracy of the motion estimates.

Subpixel accuracy of motion measurements is obtained dy variable-separable fitting performed in the neighborhood of the maximum using one-dimensional quadratic function. Using the notation in (11) above, prototype functions are fitted to the triplets

$$\{h_l(r_{1m}-1, r_{2m}), h_l(r_{1m}, r_{2m}), h_l(r_{1m}+1, r_{2m})\}, \tag{12}$$

$$\{h_l(r_{1m}, r_{2m}-1), h_l(r_{1m}, r_{2m}), h_l(r_{1m}, r_{2m}+1)\}, \tag{13}$$

that is, the maximum peak of the phase correlation surface and its two neighboring values on either side, vertically and horizontally.

The location of the maximum of the fitted function provides the required subpixel motion estimate $(\hat{d}_k x, \hat{d}_k y)$. Fitting a parabolic function horizontally to the data triplet (12) yields a closed-from solution for the horizontal component of the motion estimate $\hat{d}_k x$ as follows:

$$\hat{d}_k x = \frac{h_l(r_{1m}+1, r_{2m}) - h_l(r_{1m}-1, r_{2m})}{2H}, \tag{14}$$

where $H = [h_l(r_{1m}+1, r_{2m}) - 2h_l(r_{1m}, r_{2m}) + h_l(r_{1m}-1, r_{2m})]$.

The fractional part $\hat{d}_k y$ of the vertical component can be obtained in a similar way using (13) instead of (12).

Finally, the horizontal and vertical components of the subpixel accurate motion estimate are obtained by computing the location of the maxima of each of the above fitted quadratics.

In [14], it is shown that half-pixel accuracy motion vectors leads to a very significant improvement when compared to one-pixel accuracy, where as a higher precision results in negligible changes. Therefore, a half-pixel accuracy was chosen in our simulations.

4. EXPERIMENTAL RESULTS

To prove the feasibility of the proposed method, we compared it to a PC technique implemented in a similar manner as our approach. In this section, we examine a few examples and compare the performance, efficiency, and complexity of the two methods. In our experiments, we used the well-known test sequences: foreman (176 pixels by 144 lines), mother-daughter and Stefan (352 pixels by 288 lines), table tennis (352 pixels by 240 lines). Although the original sequences are in color, only the luminance (brightness) component is used to estimate the motion vectors.

To assess the performances of the different motion estimation techniques, the following comparisons were made. First, the subjective quality of the estimated motion field was evaluated, showing the capability of the algorithm to estimate the true motion in the scene. Second, the PSNR of motion compensated was measured, giving insight about the quality of the prediction. Results obtained using the foreman, mother-daughter and Stefan sequences are shown in Figure 1. All image sequences are degraded with additive zero-mean Gaussian noise to a signal-to-noise ratio (SNR) of 10 dB. Our results demonstrate that the proposed method

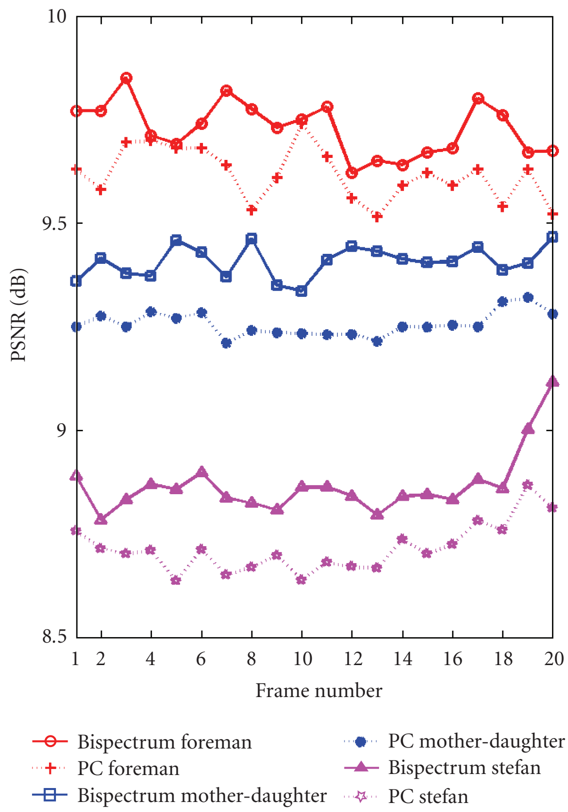
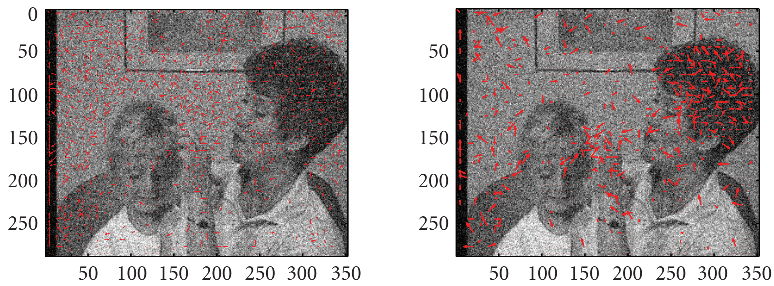


FIGURE 1: PSNR obtained for noisy sequences (SNR = 10 dB).

outperforms PC, achieving higher precision and a significantly smaller corresponding measurement error. This confirms the motion that the proposed technique of an image is a superior feature selector utilizing the portions of the image spectrum most likely to contribute to reliable motion estimation.

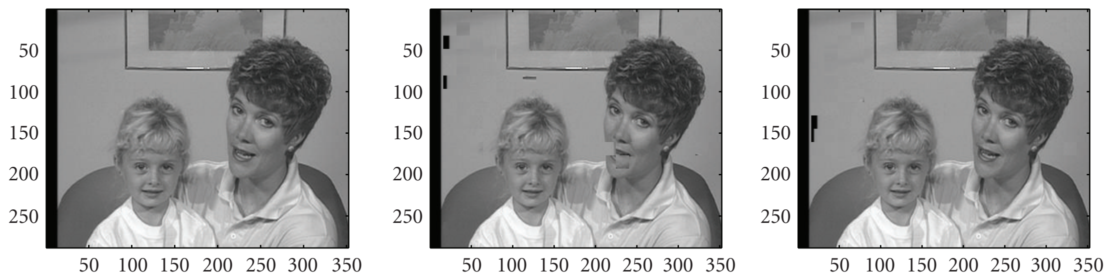
The ability of the bispectrum method to accurately estimate the displacement vector field from a degraded sequence is demonstrated in Figure 2. In this Figure, the estimated motion vector fields for the mother-daughter sequence using the two aforementioned motion estimation methods. The motion vectors estimated between the frames 126 and 127 are shown for the mother-daughter sequence. We can see that the estimates from the PC seem very random, but the bispectrum technique gives better results, producing the same motion vectors. Thus, the motion fields estimated by the “our approach” tend to be very smooth due to the smoothness constraint. Because of the noise-resistant property of the bispectrum, it produces more reliable estimates. Therefore, the proposed method motion estimation results globally in motion fields more representative of the true motion in the scene.

In terms of motion compensated images, from mother-daughter sequence, we observe better compensated images by the proposed method. We also observe that the motion compensated images for the “our method” are much closer to the original images. Thus, the “our scheme” is able to measure the motion vector more accurately and is more robust in



(a) PC (b) Bispectrum

Figure 2: Motion field for the mother-daughter sequence in the presence of noise.



(a) Original image (b) PC (c) Bispectrum

Figure 3: Prediction for frame 3 of the mother-daughter.

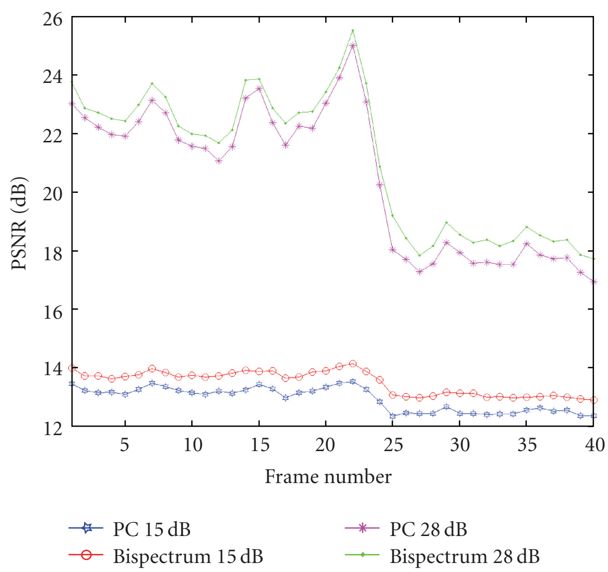


Figure 4: PSNR versus frame number for motion compensated prediction of the table tennis sequence.

general. Overall, the bispectrum typically offers better visual quality images than the PC method. Figure 3 gives examples of motion compensated images.

Comparisons of the PC and bispectrum methods indicate that the bispectrum is a robust technique for motion vector. Results of these comparisons are shown for different noise levels and video sequences. Additive Gaussian noise (AGN) was added to image sequences with an input SNR varying from 15 dB to 28 dB. Figure 4 shows the PSNR of the motion compensated prediction error for noise power of 15 dB for table tennis sequence, confirming that our scheme is consistently more immune to noise. Experiments with other levels of noise power demonstrated that similar performance gains are achievable. Similar experiments were performed for the other sequences. The results further confirm that the "our method" consistently outperforms PC. In terms of complexity, this is measured by the computation time. All the computations are performed on Intel(R) Pentium(R)D CPU 3.4 GHz with Windows XP. The two algorithms have been implemented using a prototype written in MATLAB 6.5 R13. The comparison between the "our method" and the PC confirmed that the two methods have the complexity on the same order. This is shown in Table 1.

Table 1: The comparison between two methods for the computation time.

Sequences	Methods	Motion estimation computation time (sec)
Foreman	Bispectrum	0.72
	PC	0.61
Mother-daughter	Bispectrum	0.61
	PC	0.57
Stefan	Bispectrum	0.67
	PC	0.59
Table tennis	Bispectrum	0.57
	PC	0.48

5. CONCLUSION

In this paper, the bispectrum method for subpixel motion estimation of noisy image sequences in frequency-domain was presented. The "our" proposed method provides an advantage over the PC algorithm in the presence of AGN. With "our method," the displacement vector field is smoother, providing a more accurate measure of object motion. At relatively low noise levels, the bispectrum performance is comparable to its performance in the noise-free environment. At high noise levels SNR around 10 dB, the PC fails, yet even under these extreme conditions, the bispectrum provides improvement in performance over the PC algorithm. In addition to its PSNR performance, the bispectrum also yields smooth motion fields.

REFERENCES

[1] F. Kelly and A. Kokaram, "Graphics hardware for gradient based motion estimation," in *Embedded Processors for Multimedia and Communications*, vol. 5309 of *Proceedings of SPIE*, pp. 92–103, San Jose, Calif, USA, January 2004.

[2] L. Jooheung, N. Vijaykrishnan, M. J. Irwin, and W. Wolf, "An efficient architecture for motion estimation and compensation in the transform domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 191–201, 2006.

[3] N. M. Namazi and C. H. Lee, "Nonuniform image motion estimation from noisy data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 364–366, 1990.

[4] J. M. Anderson and G. B. Giannakis, "Noise insensitive image motion estimation using cumulants," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '91)*, vol. 4, pp. 2721–2724, Toronto, Ontario, Canada, April 1991.

[5] J. M. Anderson and G. B. Giannakis, "Image motion estimation algorithms using cumulants," *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 346–357, 1995.

[6] R. P. Kleihorst, R. L. Lagendijk, and J. Biemond, "Noise reduction of severely corrupted image sequences," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '93)*, vol. 5, pp. 293–296, Minneapolis, Minn, USA, April 1993.

[7] E. Sayrol, T. Gasull, and J. R. Fonollosa, "Cost function for motion estimation based on higher order statistics," in *Proceedings of the 7th European Signal Processing Conference (EUSIPCO '94)*, pp. 1117–1120, Edinbourgh, Scotland, UK, September 1994.

[8] E. Ibn-Elhaj, D. Aboutajdine, S. Pâteux, and L. Morin, "HOS-based method of global motion estimation for noisy image sequences," *Electronics Letters*, vol. 35, no. 16, pp. 1320–1322, 1999.

[9] E. Sayrol, A. Gasull, and J. R. Fonollosa, "Motion estimation using higher order statistics," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 1077–1084, 1996.

[10] C. L. Nikias and A. P. Petrpulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[11] A. P. Petropulu and H. Pozidis, "Phase reconstruction from bispectrum slices," *IEEE Transactions on Signal Processing*, vol. 46, no. 2, pp. 527–530, 1998.

[12] E. M. Ismaili Aalaoui and E. Ibn-Elhaj, "Estimation of motion fields from noisy image sequences: using generalized cross-correlation methods," in *Proceedings of the IEEE International Conference on Signal Processing and Communications (ICSPC '07)*, Dubai, United Arab Emirates, November 2007.

[13] E. M. Ismaili Aalaoui and E. Ibn-Elhaj, "Estimation of displacement vector field from noisy data using maximum likelihood estimator," in *Proceedings of the 14th IEEE International Conference on Electronics, Circuits and Systems (ICECS '07)*, Marrakech, Morocco, December 2007.

[14] G. Madec, "Half pixel accuracy in block matching," in *Proceedings on the Picture Coding Symposium (PCS '90)*, Cambridge, Mass, USA, March 1990.