Phase Information and Space Filling Curves in Noisy Motion Estimation

V. Bruni, D. De Canditiis, D. Vitulano* *IEEE* Member

Istituto per le Applicazioni del Calcolo "M.Picone" - C.N.R.

Viale del Policlinico 137, 00161 Rome, Italy Tel. +39-6-49937771, Fax +39-6-4404306

E-mail: {bruni,d.decanditiis,vitulano}@iac.rm.cnr.it

Abstract

This paper presents a novel approach for translational motion estimation based on the phase of the Fourier Transform. It exploits the equality between the averaging of a group of successive frames and the convolution of the reference one with an impulse train function. The use of suitable space filling curves allows to reduce the error in motion estimation making the proposed approach robust under noise. Experimental results show that the proposed approach outperforms available techniques in terms of objective (PSNR) and subjective quality with a lower computational effort.

Edics: MDE-TRNS, MDE-OTHR

Index Terms: Motion estimation, FFT, Phase Correlation, Space Filling Curves, Complexity.

I. Introduction

Motion estimation is crucial in several research areas such as video coding, image registration etc. [1], [2]. The most famous approach is the Block Matching Algorithm (BMA) [3] because of its accuracy and robustness to noise. However, it suffers from a high complexity so that a lot of research effort has been devoted to its speed-up (see for instance [4]-[7]). More recently, transform-based approaches have received an increasing interest. Most of them exploit the Fast Fourier Transform (FFT) shifting property [8]-[15] and work on blocks of the image, instead of single pixels as the more recent pel recursive [16], [17] and optical flow approaches [18], [19]. More precisely, if $\{b_k(x,y)\}_{k=0,1}$

*Corresponding author

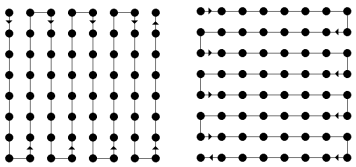

Fig. 1. The two adopted space-filling curves following the vertical (**Left**) and the horizontal (**Right**) direction.

are two aligned blocks, i.e. located at the same position in two successive frames (the analyzed Group Of Frames, GOF), for a translational motion it holds:

$$b_1(x,y) = b_0(x-d_x, y-d_y), \qquad 0 \le x,y \le M-1 \tag{1}$$

where $M \times M$ is the block size, $b_0(x,y)$ is the reference block, x is the row index, y is the column index and $\mathbf{d} \equiv (d_x, d_y)$ is the motion vector. Their normalized cross power spectrum is then:

$$e^{j(Arg(\hat{b}_0(\omega_x,\omega_y))-Arg(\hat{b}_1(\omega_x,\omega_y)))} = \frac{\hat{b}_0(\omega_x,\omega_y)\hat{b}_1^*(\omega_x,\omega_y)}{|\hat{b}_0(\omega_x,\omega_y)\hat{b}_1^*(\omega_x,\omega_y)|} \tag{2}$$

where $\hat{\cdot}$, $Arg(\cdot)$ and $\cdot^*$ respectively indicate the FFT, the phase and the complex conjugation of $\cdot$. The inverse FFT of the left member of eq. (2) combined with eq. (1) allow to achieve the phase correlation surface that contains a peak $\delta(x-d_x, y-d_y)$ in correspondence to the shift $\mathbf{d}$ (just one translational motion in each block is considered). That is why this procedure is called Phase Correlation Algorithm (PCA) [13]. The approaches oriented to reduce the computational effort of PCA can be split into two wide classes. The first one includes the Discrete Cosine Transform based methods: they are oriented to both avoid complex numbers and exploit the same domain used in video coding [14],[20]-[22]. Unfortunately, the resulting complexity is just slightly reduced. Approaches belonging to the second class estimate motion in the Fourier domain, avoiding the inversion of the transform. Among them, an efficient solution based on the sawtooth shape of the Fourier phase of two shifted blocks has been recently proposed by Balci and Foroosh [23], [24]. Using the phase separability [25], the number of phase cycles in each direction gives the motion:

$$d_x = \frac{M}{2\pi}\frac{\partial(Arg(\hat{b}_0) - Arg(\hat{b}_1))}{\partial\omega_x}, \quad d_y = \frac{M}{2\pi}\frac{\partial(Arg(\hat{b}_0) - Arg(\hat{b}_1))}{\partial\omega_y}. \tag{3}$$

Its robustness to noise can be increased by a regularization, but with a higher computational effort.

This paper focuses on a novel approach for estimating motion directly in the FFT domain with a computational effort saving. It exploits a very simple but general result in the time domain: the average B of K aligned blocks in K successive frames is equivalent to the convolution $b_0 * \delta_{comb}$ between the reference block b_0 (i.e. lying in the reference frame) and a suitable impulse train δ_{comb}

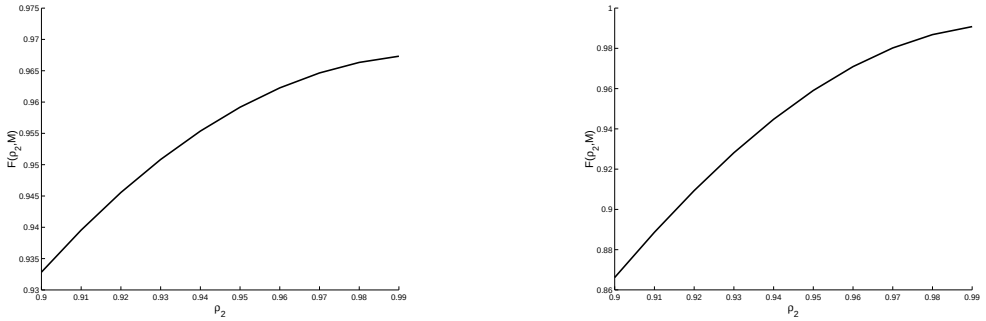

Fig. 2. $F(\rho_2, M)$, as in eq. (13), versus $\rho_2 \in [.9, .99]$ with $M = 8$ (**Left**) and $M = 16$ (**Right**).

Fig. 3. Considered test sequences and images. From left to right: Mobile, Coastguard, Flower, Tennis and Pentagon.

(also known as *Dirac comb* or *sampling function*). This result allows us to estimate the displacement $\mathbf{d}$ by comparing the FFT phases of B and $b_0 * \delta_{comb}$ in a non zero frequency. This approach, called MEPI (Motion Estimation based on Phase Information), has the following advantages: *i)* motion is automatically estimated without search — as it happens using PCA (search of the best peak) and BMA (search of the best block); *ii)* a drastic reduction of the computational effort (with respect to PCA) is provided by avoiding the inverse FFT and by using two only values of the forward FFT; *iii)* GOF can contain an arbitrary number of frames further reducing its complexity. However, MEPI may fail for noisy sequences because it is based on two only phase values. A theoretical result that allows us to evaluate the precision of the estimated motion in case of noise is presented and then exploited by embedding two suitable space-filling curves in the MEPI's scheme [26]. These curves allow to increase the signal-to-noise ratio by following the main orientation of eventual edges in the considered blocks. Experimental results show that MEPI outperforms available techniques in terms of quality for not severe noise with a lower complexity. The paper is organized as follows. Next section presents MEPI's theoretical model, the use of space filling curves in the noisy case (II.A), the proposed algorithm (II.B) and its complexity (II.C). Section III includes experimental results, discussions and comparisons with the most performing motion estimators.

II. The proposed Model

It can be proved that the average B of K aligned blocks subjected to a translational motion is equivalent to the convolution of the reference block b_0 with a suitable function δ_{comb}.

Proposition 1: If

$$B(x,y) = \frac{1}{K}\sum_{k=0}^{K-1} b_k(x,y), \qquad 0 \le x,y \le M-1 \tag{4}$$

is the average of $K \in \mathbf{N}$ aligned blocks $\{b_k(x,y)\}_{k=0,1,..,K-1}$ subjected to a translational motion whose shift is $(d_x, d_y) \in \mathbf{Z}^2$, b_0 is the reference block and δ_{comb} is an impulse train function:

$$\delta_{comb}(x,y) = \frac{1}{K}\sum_{k=0}^{K-1} \delta(x-kd_x, y-kd_y), \tag{5}$$

then

$$B(x,y) = (b_0 * \delta_{comb})(x,y). \tag{6}$$

Proof By putting eq. (1) into eq. (4) we get the equality: $B(x,y) = \frac{1}{K}\sum_{h=0}^{K-1} b_0(x-hd_x, y-hd_y)$. Each element in $B(x,y)$ is the sum of K suitably sampled elements of b_0. From eq. (5), we get:

$$B(x,y) = \frac{1}{K}\sum_{h_1=0}^{(K-1)d_x}\sum_{h_2=0}^{(K-1)d_y}\left(b_0(x-h_1, y-h_2)\sum_{k=0}^{K-1}\delta(h_1-kd_x, h_2-kd_y)\right) =$$

$$= \sum_{h_1=0}^{(K-1)d_x}\sum_{h_2=0}^{(K-1)d_y} b_0(x-h_1, y-h_2)\delta_{comb}(h_1,h_2) = (b_0 * \delta_{comb})(x,y). \qquad \bullet$$

It turns out that the temporal averaging of corresponding blocks is equivalent to the spatial convolution of the reference block b_0 with an impulse train, whose sampling period depends on the motion vector $\mathbf{d} = (d_x, d_y)$. Using the convolution property, in the Fourier domain we get:

$$\hat{B}(\omega_x, \omega_y) = \hat{b}_0(\omega_x, \omega_y)\hat{\delta}_{comb}(\omega_x, \omega_y), \tag{7}$$

where $\hat{\delta}_{comb}(\omega_x, \omega_y) = \sum_{k=0}^{K-1} e^{-i\frac{2\pi}{2M-1}k(\omega_x d_x + \omega_y d_y)} = |\hat{\delta}_{comb}(\omega_x, \omega_y)| e^{-i\frac{\pi}{2M-1}(d_x\omega_x + d_y\omega_y)(K-1)}$. A comparison of the phases of both members of eq. (7) at frequencies $(\omega_x, 0)$ and $(0, \omega_y)$ gives:

$$d_x = -(Arg(\hat{B}(\omega_x,0)) - Arg(\hat{b}_0(\omega_x,0)))\ \ P(\omega_x, M), \qquad \forall\ (\omega_x, 0), \tag{8}$$

$$d_y = -(Arg(\hat{B}(0,\omega_y)) - Arg(\hat{b}_0(0,\omega_y)))\ \ P(\omega_y, M), \qquad \forall\ (0, \omega_y), \tag{9}$$

where $P(\omega_x, M) = \frac{(2M-1)}{(K-1)\pi\omega_x}$ and $P(\omega_y, M) = \frac{(2M-1)}{(K-1)\pi\omega_y}$. The choice of two independent frequencies allows a direct solution of the linear system giving the motion vector $\mathbf{d}$ without any search.

A. The Noisy Case

In presence of noise, eq. (6) becomes:

$$C(x,y) = (b_0 * \delta_{comb})(x,y) + \frac{1}{K}\sum_{k=0}^{K-1}\epsilon_k(x,y), \tag{10}$$

where $C(x,y) = \frac{1}{K}\sum_{k=0}^{K-1} c_k(x,y)$ and $c_k(x,y) = b_k(x,y) + \epsilon_k(x,y)$ are the original (clean) blocks $b_k(x,y)$ with an additive i.i.d. Gaussian noise $\epsilon_k(x,y) \sim N(0,\sigma^2) \quad \forall k$. Eq. (10) is invariant to any linear preprocessing on the noisy frames c_k. On the other hand non linear denoising operators are usually expensive and less performing than Singular Value Decomposition (SVD) based approaches [25]. Therefore, our goal is not to remove noise, but to derive an upperbound for the error on motion estimation when eqs. (8) and (9) are evaluated on noisy blocks c_k:

Proposition 2: *The upperbound of the error Δ_x for d_x in eq. (8) with noisy blocks c_k is:*

$$\Delta_x^2 \leq \sigma^2 \left(\frac{1}{K}\frac{1}{|\hat{C}(\omega_x,0)|^2} + \frac{1}{|\hat{c}_0(\omega_x,0)|^2} + \frac{2}{K}\right) P^2(\omega_x, M) \qquad \forall\,(\omega_x,0). \tag{11}$$

The upperbound of Δ_y for d_y in eq. (9) has the same form, with $(\omega_x,0)$ replaced by $(0,\omega_y)$.

Proof Eq. (8) evaluated on noisy frames c_k can be rewritten as:

$$d_x = P(\omega_x, M)\left(\arctan\left(\frac{m_4}{m_3}\right) - \arctan\left(\frac{m_2}{m_1}\right)\right)\Big|_{m_i=\overline{m}_i \quad i=1,2,3,4}$$

where $\overline{m}_1 = Re(\hat{C}(\omega_x,0)) = Re(\hat{B}(\omega_x,0)) + \frac{1}{K}\sum_{k=0}^{K-1} Re(\hat{\epsilon}_k(\omega_x,0)) = m_1 + \eta_1$, $\overline{m}_2 = Im(\hat{C}(\omega_x,0)) = m_2 + \eta_2$, $\overline{m}_3 = Re(\hat{c}_0(\omega_x,0)) = m_3 + \eta_3$, $\overline{m}_4 = Im(\hat{c}_0(\omega_x,0)) = m_4 + \eta_4$. Hence, d_x depends on the variables m_1, m_2, m_3, m_4 that are affected by the errors $\eta_1, \eta_2, \eta_3, \eta_4$. Exploiting the property of an i.i.d. normal random array with zero mean and finite variance in the Fourier domain, as well as the orthogonality property of the cosine and sine bases, the errors η_j, $j = 1, ..., 4$ are still normal distributed with zero mean and covariance matrix with non zero entries: $Cov(\eta_1,\eta_3) = Cov(\eta_2,\eta_4) = \sigma^2/K$ and $diag(Cov) = \{\sigma^2/K, \sigma^2/K, 1, 1\}$. Therefore, the error Δ_x^2 for d_x can be estimated by applying the error propagation law for derived measures (see chap. V of [27]): $\Delta_x^2 = \sum_{i=1}^4\sum_{j=1}^4 \frac{\partial d_x}{\partial m_i}\frac{\partial d_x}{\partial m_j}\Big|_{\substack{m_i=\overline{m}_i,\\ m_j=\overline{m}_j}} Cov(\eta_i,\eta_j)$. After a simple algebra, we get: $\Delta_x^2 = \sigma^2 P^2(\omega_x,M)\left(\frac{1}{K}\frac{1}{|\hat{C}(\omega_x,0)|^2} + \frac{1}{|\hat{c}_0(\omega_x,0)|^2} - \frac{2}{K}\frac{Re\left(\hat{C}(\omega_x,0)\hat{c}_0^*(\omega_x,0)\right)}{|\hat{C}(\omega_x,0)|^2|\hat{c}_0(\omega_x,0)|^2}\right)$. Considering $|Re(z)|/|z|^2 \leq 1$ for any complex number z, we get eq. (11). A similar result can be provided for Δ_y^2. •

The error Δ_x depends on the inverse signal-to-noise ratio (SNR): $\frac{\sigma^2}{|\hat{C}(\omega_x,0)|^2}$ and $\frac{\sigma^2}{|\hat{c}_0(\omega_x,0)|^2}$. Hence, the selection of the two lowest frequency values could be the best choice, since the spectrum of many natural images has a decreasing decay [26], [28]. A way for reducing the error in eq. (11) may

be to increase the size of the GOF. However, the hypothesis of translational motion may become false. A simple but effective tool can be then adopted: the space-filling curves [26]. A space-filling curve passes through every point of a 2D region of the image. Its aim is to increase the clean signal correlation, while leaving the noise unchanged. Each block can be then modeled through a curve whose preferential direction is parallel to the main orientation of the edges inside it. Bearing in mind the motion separability [25], the most suitable curves are the ones shown in Fig. 1, that account for horizontal and vertical edges. Curves like the Hilbert (sometimes called Peano) one [26] are not suitable for our purposes because of their anisotropy and the lack of both a closed form and a linear dependence on the horizontal and vertical motion. MEPI can be then extended as follows. The errors Δ_x and Δ_y are compared with a threshold τ, that depends on the required precision for the motion vector. If they are less than τ, the estimated motion vector (d_x, d_y) is adopted. If one of the two errors, say Δ_y, over-exceeds τ, then d_y is estimated as follows. The block is reorganized using the horizontal scan for further correlating the clean information. Then, d_y is estimated using eq. (9) where $P(1, M) = (2M^2 - 1)/(K - 1)\pi$ while $\hat{C}(2\pi/M^2)$ and $\hat{c}_0(2\pi/M^2)$ are the second coefficient of the 1-D FFT of the resulting noisy 1D signals. Since the motion for this signal is $Md_x + d_y$, d_y can be derived by simply subtracting the known quantity Md_x from it. Finally, if both Δ_x and Δ_y over-exceed τ, motion vector cannot be estimated and then it is supposed equal to $(0, 0)$. Next proposition proves the existence of cases where the reorganization of the block via one of the two proposed space filling curves improves the accuracy of motion estimation. To this aim, each block is modeled as a first order discrete stationary random field with a separable covariance matrix (page 36 of [28]).

Proposition 3: Let c be a generic block of size $M \times M$ with separable and stationary covariance matrix $r(u, v) = \sigma^2 \rho_1^{|u|} \rho_2^{|v|}, \quad \forall\ u, v$, where $\rho_1 = r(1, 0)/\sigma^2$ and $\rho_2 = r(0, 1)/\sigma^2$ are the one-step correlations in the u and v directions. If $\rho_1 < \rho_2$, let $\hat{c}(0, 2\pi/M)$ and $\hat{c}(2\pi/M^2)$ be respectively be the second Fourier coefficient of the block and its 1D horizontal raster scan. Hence,

$$|\hat{c}(0, 2\pi/M)|^2 < |\hat{c}(2\pi/M^2)|^2 \tag{12}$$

is equivalent to:

$$\rho_1 < F(\rho_2, M) \quad \text{with} \quad F(\rho_2, M) = \frac{\rho_2(cos(2\pi/M^2) - cos(2\pi/M))}{1 + \rho_2^2 - \rho_2(cos(2\pi/M^2) - cos(2\pi/M))}. \tag{13}$$

A similar result holds if $\rho_2 < \rho_1$ and the vertical raster scan is considered.

Proof In case of an horizontal raster scan, the correlation matrix for the resulting one-dimensional signal is $r(u) = \sigma^2 \rho_2^{|u|}, \quad \forall u$. It is the correlation matrix of a stationary first-order

		Mobile		*Coastguard*		*Flower*	
Method	**No. of op. per Pixel** (16×16)	**PSNR** (16×16)	**PSNR** (8×8)	**PSNR** (16×16)	**PSNR** (8×8)	**PSNR** (16×16)	**PSNR** (8×8)
MEPI	4	26.72 db	28.52 db	33.85 db	35.20 db	24.63 db	25.69 db
PDM	44	25.64 db	26.98 db	32.90 db	33.20 db	24.23 db	25.43 db
PCA	45	24.78 db	25.44 db	32.26 db	32.53 db	24.07 db	25.37 db
FS-BMA	589	22.99 db	23.88 db	30.47 db	31.58 db	23.79 db	25.22 db
2BT	10	22.72 db	22.99 db	29.93 db	30.50 db	23.43 db	24.55 db

TABLE I

RESULTS OF MEPI, PDM, PCA, FS-BMA AND 2-BT ON MOBILE, COASTGUARD AND FLOWER SEQUENCES.

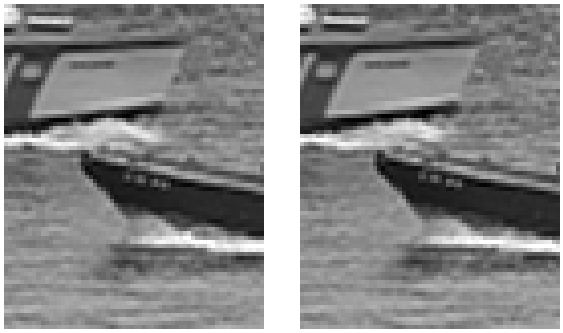

Fig. 4. Detail of Frame no. 31 of Coastguard sequence: original (**Left**) and motion compensated via MEPI (**Right**).

Markov sequence. From the Wiener-Khinchin Theorem it holds: $|\hat{c}(\omega)|^2 = \hat{r}(\omega) \quad \forall\, \omega$, where $\hat{r}(\omega)$ is the spectral density function [28]. Then, eq. (12) is equivalent to: $\hat{r}(0, 2\pi/M) < \hat{r}(2\pi/M^2)$, i.e.

$$\sigma^2 \frac{(1-\rho_1^2)(1-\rho_2^2)}{(1+\rho_1^2-2\rho_1\cos(0))\,(1+\rho_2^2-2\rho_2 cos(2\pi/M))} < \sigma^2 \frac{1-\rho_2^2}{1+\rho_2^2-2\rho_2 cos(2\pi/M^2)}.$$

Since $(1+\rho_2^2-2\rho_2 cos(\bullet)) = |1-\rho_2 e^{-i\bullet}|^2 > 0$, a simple algebra gives eq. (13). •

Eq. (13) is graphically proved in Fig. 2. It shows the behavior of $F(\rho_2, M)$ for $\rho_2 > .9$, that is the usual range in real images (page 37 of [28]). Therefore, when eq. (13) holds, space filling curves are convenient since they increase SNR that regulates the error Δ_y.

B. Algorithm

For a GOF composed of K frames f_k of size $\overline{M} \times \overline{M}$, split each f_k into non overlapping $M \times M$ blocks $c_{j,k}$. $c_{j,k}$ is the j^{th} block in the k^{th} frame f_k with $j = 1, \ldots, \left\lfloor \frac{\overline{M}\times\overline{M}}{M\times M} \right\rfloor$. For each j:

1) Compute the average C_j of the K corresponding blocks in K successive frames.

2) Evaluate the phase of $\hat{C}_j$ at the frequency values $(\omega_x, \omega_y) = (0, 1)$ and $(\omega_x, \omega_y) = (1, 0)$.
3) Compute the phase of $\hat{c}_{j,0}$ at the same frequency values.

   If $\sigma^2 = 0$, then

   - Compute the motion $\mathbf{d} = (d_x, d_y)$ using eqs. (8) and (9).

   else

   - Evaluate Δ_x and Δ_y using eq. (11). Let $3|\Delta_x|$ be the maximum error for d_x, according to the 3σ law [27] (similarly for Δ_y), and let τ be the largest admissible error value, then:
     a) if $3|\Delta_x| < \tau$ and $3|\Delta_y| < \tau$, then estimate d_x and d_y using eqs. (8) and (9);
     b) if $3|\Delta_x| < \tau$ and $3|\Delta_y| > \tau$, then estimate d_x using eq. (8). Compute the second coefficient of the FFT of the 1D signal achieved through the curve in Fig. 1.right. Its motion is $(d_y + Md_x)$. Then subtract Md_x (known) from it to get d_y;
     c) if $3|\Delta_x| > \tau$ and $3|\Delta_y| < \tau$, then apply operations in 3.b), but inverting the roles of x and y and using the curve in Fig. 1.left;
     d) if $3|\Delta_x| > \tau$ and $3|\Delta_y| > \tau$, then set $d_x = 0$ and $d_y = 0$.
4) Assign the motion vector $\mathbf{d} = (kd_x, kd_y)$ to each block $c_{j,k}$, $\quad k = 1, 2, \ldots, K-1$.

C. Complexity

We will denote addition, multiplication, division and comparison respectively with **a**, **m**, **d** and **c**. For the component x, setting $K = 2$ and omitting arguments, eq. (8) can be rewritten as: $d_x = -P\ (atan(Im(\hat{B}\,\hat{b}_0^*)/Re(\hat{B}\,\hat{b}_0^*))$. Hence, *atan* argument computation, in terms of $\hat{b}_0$ and $\hat{b}_1$, requires 6 **m**s, 4 **a**s and 1 **d**. For $\omega_x = 1$, $-P$ requires 1 **m**, 1 **a** and 1 **d**. *atan* computation requires $log_2(M/2)+1$ **c**s: one for the sign, the remaining $log_2(M/2)$ to determine its absolute value through the search in a binary tree (built once for all). In fact, the (precomputed) values, corresponding to all possible integer shifts are $M/2$, which is the largest integer shift that guarantees a comparison among aligned blocks. With regard to $\hat{b}_0(\omega_x, 0)$, $M\ (M-1)$ **a**s are required for the sum along columns (or rows) plus the 1D FFT of the resulting signal. The latter exploits the symmetry of *cos* function in the unitary circle: *Re* requires $M-3$ **a**s and $M/4-1$ **m**s, while $M/2-1$ **a**s and $M/4-1$ **m**s are necessary for *Im* by exploiting the complementary role of *cos* and *sin*. Finally, 2 **d**s are required for the normalization of FFT. The same amount of operations is required for $\hat{b}_1(\omega_x, 0)$ (or for $\sum_{k=1}^{K-1} \hat{b}_k$ in case of $K > 2$). Considering both x and y direction, the total amount of operations per pixel is 4.3281 for $M = 16$, while its complexity is $O((K+2)M^2)$.

True shifts	(0.167, -0.5)	(0.67, 0.25)	(-0.33, -0.167)	(0.33, 0.33)
PDM	(0.161, -0.502)	(0.68, 0.244)	(-0.34, -0.161)	(0.333, 0.328)
MEPI	(0.163, -0.502)	(0.672, 0.272)	(-0.333, -0.177)	(0.334, 0.338)

TABLE II

PENTAGON IMAGE. SUBPIXEL COMPARISON AMONG MEPI AND PDM [24].

In case of noise, the error bound in eq. (11) requires 6 **a**s, 3 **d**s and 8 **m**s. The same is required for Δ_y^2 plus the two **c**s in step 3) of the algorithm. The computation of *atan* is still obtained by searching among all its possible (precomputed) values, corresponding to all possible shifts, $[-M/2+Md_x, M/2+Md_x]$ in step 3.b) (or $[-M/2+Md_y, M/2+Md_y]$ in step 3.c)); the number of **c**s is still $log_2(M/2)+1$. The two (for $\hat{c}_0$ and $\hat{c}_1$) 1D FFTs on M^2 samples organized along one of the 2 space filling curves again need $2(2M^2-4)$ operations. The amount of operations per pixels is 4.4687 in case of step 3.a) (as well as in 3.d except for atan computation) of the algorithm and 8.4414 in case of step 3.b) or 3.c). The complexity is $O((K+6)M^2)$.

III. EXPERIMENTAL RESULTS AND CONCLUSIONS

MEPI has been tested on three well-known sequences (see Fig. 3): Mobile (140 frames of size 240×352), Coastguard (300 frames of size 352×288) and Flower (60 frames of size 352×240). Table I shows its PSNR (Peak Signal-to-Noise Ratio averaged on the whole motion compensated sequence) values for $\sigma = 0$, using GOFs of size two ($K = 2$) and blocks of size 8×8 and 16×16. MEPI has been compared with the classical PCA (Phase Correlation Algorithm) [2], Balci and Foorosh approach in [23], denoted with PDM (Phase Difference Method), the classical FS-BMA (Full Search - Block Matching Algorithm) [3] and the recent 2BT (Two Bit Transform) [30] based on bit-planes. We note that MEPI performs better than the other methods, improving both the more sophisticated PDM and the classical PCA. In particular, this latter may give false peaks in case of multiple motion or a not precise equivalence between two blocks, i.e. when the actual block is not the precise shifted copy of the reference one. On the contrary, MEPI is more robust thanks to the selection of the lowest frequency in step 2) of the Algorithm and the use of the sum of subsequent frames, that is able to suppress or reduce eventual differences between two aligned blocks. In order to show that this choice does not produce artifacts, in Fig. 4 there is a detail of frame no. 31 of Coastguard that contains the aforementioned effects along with its

motion compensated version. This region is composed of 80×64 pixels with topmost and leftmost corner located at pixel (97,81). For this test, a GOF of size two with 16×16 blocks has been used. It can be noted that no evident distortion appears. But, the most attractive aspect of MEPI stems from its complexity, as shown in Section II.C. The second column of Table I compares the (rounded) number of operations (where the same weight has been assigned to **a**, **m**, **d** and **c**) required by MEPI, PDM [23], 2BT [30], FS-BMA [3] and PCA [2]. It can be noted that there is an evident saving of the computational effort with respect to the most comparable approach 2-BT that requires 9.9025 operations per pixel and also with respect to a possible speed-up of PCA via the Discrete Cosine Transform [21], that requires 20 operations per pixel. In a non optimized matlab code, the CPU time for Mobile is 47.2 secs (Pentium 4, 3.06 GHz, ram 512 Mb). Another test sequence has been selected: Tennis, shown in Fig. 3. In particular, the first 25 frames have been considered, where there is just pure translational motion, using 16×16 blocks. The results, in terms of PSNR, of the three Fourier based approaches MEPI, PCA and PDM respectively are 33.98, 33.61 and 33.41 db. They are very similar because the common hypothesis of pure global translation is completely fulfilled. However, MEPI requires about 1/10 of the operations that are necessary for PCA and PDM. The same observation holds for Flower sequence (in Table I).

Even if MEPI has not been specifically designed for subpixel estimation, in principle it may be adopted for this purpose too. MEPI has been tested on Pentagon image in Fig. 3, using the method in [11] for producing two shifted images. Table II contains a comparison with PDM. MEPI's performance is promising and very similar to PDM one, except for some cases where it may give worse results — see d_y in the second test. It can be explained by considering that PDM is based on a very powerful regularization approach that exploits all available frequencies, while MEPI uses just one phase value that might result not accurate for precise non integer shift estimation.

In presence of noise, MEPI has been tested on Mobile, Coastguard and Flower sequences using 16×16 blocks, GOFs of size two, with and without the use of space filling curves, as shown in Fig. 5. PSNR is the average on 30 independent trials. The threshold τ in step 3) of the algorithm has been set to .5 in order to achieve a precision of 1 pixel for (integer) motion estimation. MEPI outperforms alternative approaches for $\sigma < 30$ with space filling curves and $\sigma < 15$ without, confirming the advantage in using space filling curves in case of noise. MEPI has also been compared in Fig. 5 using GOFs of size 3 and 4 on Mobile sequence highlighting a further advantage of this approach. MEPI's experiments and theoretical model assume i.i.d. Gaussian noise, since it embeds both the (additive) thermal photodetector noise and the (multiplicative) film grain noise [31].

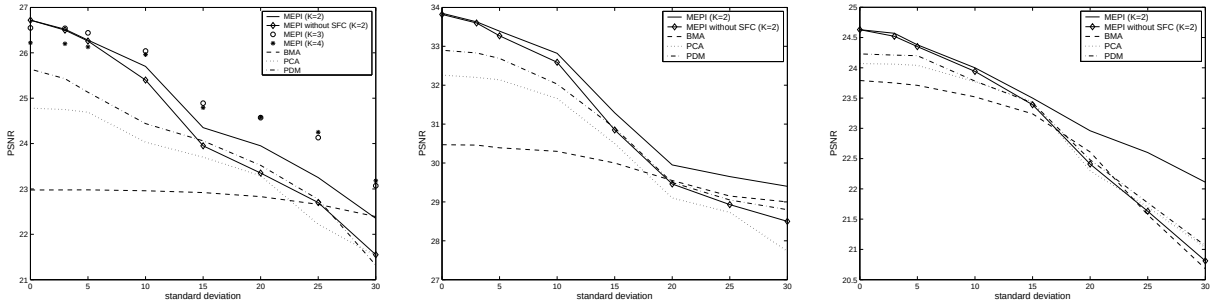

Fig. 5. Mobile (**Left**), Coastguard (**Middle**) and Flower (**Right**) sequences: MEPI with $K = 2$ (with and without space filling curves), PCA, PDM and BMA results. Mobile figure also includes MEPI with $K = 3, 4$.

Hence, in real applications, the robust median estimator in [29] can be applied if the noise variance σ^2 is unknown. For more complicated noise (additive signal dependent) a simple algorithm like MEPI is not effective, since more operations or, at least, more frequencies should be accounted for. However, MEPI's primary objective was to limit the influence of noise in motion estimation without increasing the computational effort, as it is required, for instance, in ancient film restoration. In fact, MEPI's computational effort is still competitive under noise, as shown in Section II.C. Nonetheless, MEPI's performance can be further improved in step 3.d) of the algorithm, by embedding more sophisticated schemes like the one in [24]. Future research will be oriented to both make MEPI more robust for subpixel estimation and extend it to more general motion and noise models.

References

[1] J. K. Aggarwal and N. Nandhakumar *On the computation of motion from sequences of images: A review*, Proc. of IEEE, Vol. 76, pp. 917-935, August 1988.

[2] Y. Q. Shi, H. Sun, Image and Video Compression for Multimedia Engineering, CRC Press, New York, 2000.

[3] J. M. Jou, P.-Y. Chen and J.-M. Sun, *The Gray Prediction Search Algorithm for Block Motion Estimation*, IEEE Trans. Circuits Syst. Video Technol., vol 9, no. 6, pp. 843-848, September 1999.

[4] H.-W. Cheng and L.-R. Dung, *EFBLA: A two phase matching algorithm for fast motion estimation*, Electronic Letters, Vol. 40, No. 11, pp. 660-661, May, 2004.

[5] C. J. Kuo, C. H. Yeh, and S. F. Odeh, *Polynomial Search Algorithm for Motion Estimation*, IEEE Trans. Circuits Syst. Video Technology, vol. 10, no.5, pp. 813-818, August 2000.

[6] K. N. Nam, J.-S. Kim, R.-H. Park and Y. S. Shim, *A Fast Hierarchical Motion Vector Estimation Algorithm using Mean Pyramid*, IEEE Trans. Circuits Syst. Video Technology, vol. 5, no. 4, pp. 344-351, August 1995.

[7] S. Zhu and K.-K. Ma, *A new Diamond Search Algorithm for Fast Block Matching Estimation*, IEEE Trans. Circuits Syst. Video Technology, vol. 9, no.2, pp. 287-290, February 2000.

[8] L. H. Chien and T. Aoki *Robust motion estimation for video sequences based on phase-only correlation*, Proc. of IASTED Int. Conf. on Signal and Image Proc., pp. 441-446, August, 2004.

[9] Y.M. Chou, H.M. Hang, *A New Motion Estimation Method using Frequency Components*, Journal of Visual Communication and Image Representation, Vol. 8, No. 1, pp. 83-96, March 1997.

[10] S. Erturk *Digital Image Stabilization with sub-image phase correlation based global motion estimation*, IEEE Trans. on Consumer Electronics, Vol. 49, No. 4, pp. 1320-1325, November, 2003.

[11] H. Foroosh, J. Zeroubia, M. Berthold *Extension of phase correlation to sub-pixel registration*, IEEE Trans. on Image Processing, Vol. 11, No. 3, pp. 188-200, 2002.

[12] L. Hill and T. Vlachos *Motion measurement using shape adaptive phase correlation*, Electronic Letters, Vol. 37, Issue 25, pp. 1512-1513, December, 2001.

[13] C.D. Kughlin, D.C. Hines, *The phase correlation image alignment method*, Proceedings of IEEE Int. Conf. Systems, Man. and Cybernetics, pp. 163-165, September 1975.

[14] M. Biswas, T. Nguyen, *A novel motion estimation algorithm using phase plane correlation for frame rate conversion*, Proc. of IEEE Conf. on Signals, Systems and Computers, Vol. 1, pp. 492-496, November, 2006.

[15] A. Molino, F. Vacca, G. Masera *Design and implementation of phase correlation based motion estimator*, Proc. of IEEE International SOC Conf., pp. 291-294, September, 2005.

[16] A. Netravali, J. Robbins, *Motion-Compensated Television Coding: Part I*, Bell Syst.Tech.J., 58(3), pp. 631-670, 1979.

[17] J. D. Robbins, A. N. Netravali, *Recursive Motion Compensation: A Review, in Image Sequence Processing and Dynamic Scene Analysis*, T. S. Huang, Ed. Berlin, Germany: Springer-Verlag, pp. 73-103, 1983

[18] G. Le Besnerais, F. Champagnat, *Dense Optical Flow by Iterative Local Window Registration*, Proc. of IEEE International Conference on Image Processing, pp. 137-140, 2005.

[19] M. Ye, R.M. Haralick, L.G. Shapiro *Estimating Piecewise - Smooth Optical Flow with Global Matching and Graduated Optimization* , IEEE PAMI, Vol. 25, No. 12, pp. 1625-1630, December 2003.

[20] M.Li, M. Biswas, S. Kumar, T. Nguyen, *DCT- based phase correlation motion estimation*, Proc. of IEEE Int. Conf. on Image Processing, pp. 445-448, 2004.

[21] U.-V. Koc, K. J. R. Liu, *DCT-Based Motion Estimation*, IEEE Trans on Image Proc., vol. 7, no. 7, 1998.

[22] M. Song, A. Cai, J.-A. Sun, *Motion estimation in DCT domain*, Proc. of ICCT, Vol. 2, pp. 670-674, 1996.

[23] M. Balci, H. Foroosh, *Inferring motion from the rank constraint of the phase matrix*, Proc. of ICASSP, Vol. 2, pp. 925-928, March, 2005.

[24] M. Balci, H. Foroosh, *Subpixel Estimation of Shifts Directly in the Fourier Domain*, IEEE Trans. on Image Proc. Vol. 15, No. 7, pp. 1965-1972, July 2006.

[25] W. S. Hoge, *Subspace identification extension to the phase correlation method*, IEEE Trans. on Medical Imaging, Vol. 22, No. 2, pp. 277-280, 2003.

[26] D. Salomon, Data Compression: The complete reference, Springer, New York, 2004.

[27] J. K. Taylor, C. Cihon, Statistical Techniques for Data Analysis, CRC Press, New York, 2004.

[28] A. K. Jain, Fundamentals of digital image processing, Prentice Hall, 1989.

[29] D. L. Donoho, Denoising by soft-thresholding, IEEE Trans. on Inf. Th., Vol. 41, No.3, 1995, pp. 613-627.

[30] A. Erturk, S. Erturk, *Two-Bit Transform for Binary Block Motion Estimation*, IEEE Trans. on Circ. and Syst for Video Tech., Vol. 15, No. 7, pp. 938-946, July 2005.

[31] K. J. Boo, N. K. Bose, *A Motion-Compensated Spatial-Temporal Filter for Image Sequences with Signal-Dependent Noise*, IEEE Trans. on Circ. and Syst for Video Tech., Vol. 8, No. 3, pp. 287-298, June 1998.