High-Accuracy Sub-pixel Motion Estimation from Noisy Images in Fourier Domain

Jinchang Ren, Jianmin Jiang and Theodore Vlachos

Abstract—In this paper, we propose a new method for estimating sub-pixel motion via exploiting the principle of phase correlation in the Fourier domain. The method is based on linear weighting of the height of the main peak on the one hand and the difference between its two neighboring side-peaks on the other. Using both synthetic and real data we show that the proposed method outperforms many established approaches and achieves improved accuracy even in the presence of noisy samples.

Index Terms—phase correlation, motion estimation, sub-pixel registration, linear interpolation.

I. INTRODUCTION

ROBUST and accurate motion estimation is a critical component of many image analysis and computer vision applications, where pairs of images or sub-images need to be aligned and used for further processing. This can benefit a wide range of applications, including video coding, image stabilization, image registration, robotic/stereo vision, image mosaicking, super-resolution reconstruction, and object recognition [1-8, 13], and useful surveys can be found in [4, 7-8].

Among many existing motion estimation techniques, the phase correlation method has emerged as a particularly attractive solution and has been successfully applied to image registration, object recognition and other applications [9-14]. Three main advantages using frequency-domain approaches for motion estimation are: 1) robust to distortions caused by intensity and geometry changes [21]; 2) able to estimate shift, rotation and scale changes [10]; and 3) high efficiency due to the fast implementation of the Fourier transform.

Although integer-pixel accuracy can be adequate in some cases, fractional accuracy sub-pixel registration is generally beneficial to most applications [5, 18]. This is because that actual motion occurs in arbitrary increments and is oblivious of the discrete nature of the image registration grid. In other cases such as magnetic resonance imaging (MRI), data are usually sampled using non-integer offsets in the spatial Fourier domain before reconstruction and so sub-pixel registration by phase correlation is a natural approach in such a context [11]. Other application domains that have historically required the computation of sub-pixel estimates for registration include image down-sampling [18-20], spatial interpolation [21], interpolation-free [6] and pyramid-based approaches [2].

In phase correlation methods, sub-pixel registration is achieved using interpolation-based or interpolation-free approaches after determining the maximum peak of the correlation surface on the integer-accuracy grid. In [20], a least-squares fitting approach has been used with removal of outlier spectral components. In [15-17, 21], the maximum peak and its four surrounding points are used for interpolation-based sub-pixel registration via fitting of quadratic, Gaussian, *sinc* and the so-called *esinc* functions. Though good results have been reported, these methods are generally sensitive to noise.

Typical interpolation-free approaches can be found in [11, 18-19]. In [18], a closed-form solution is provided by modelling sub-pixel offsets as the result of down-sampling of images undergoing integer-accuracy offsets on a higher density grid and then analyzing the signal power distribution around the main peak. This is extended in

[12] to deal with noisy samples using higher order statistics. In [19], discrete phase difference is modeled as a 2-D saw-tooth signal and sub-pixel registration is performed by counting the number of cycles of the phase difference matrix along each frequency axis. In [11], subspace extension is presented for sub-pixel registration based on the fact that noise-free phase correlation matrix is a rank one, separable-variable matrix. In noisy cases, sub-pixel motion estimation is reduced to a problem of finding the rank one approximation to that matrix. All these methods require a high signal-to-noise ratio, otherwise performance degrades significantly [4, 12].

In this paper, we propose a high-accuracy sub-pixel registration method in the Fourier domain. The basics of phase correlation are introduced in Section II and the proposed method is presented in Section III. Experimental results and discussion are given in Section IV followed by a brief conclusion in Section V.

II. PHASE-CORRELATION ANALYSIS

Phase correlation is based on the Fourier shift theorem, which states that a shift between two functions will cause a phase shift in the Fourier domain. Let $f(x, y)$ and $g(x, y)$ be two $M \times N$ images and their 2-D discrete Fourier transforms are denoted as $F(u, v)$ and $G(u, v)$, where $x, u \in [0, M-1]$ and $y, v \in [0, N-1]$. If the above two images satisfy $f(x, y) = g(x - \Delta x, y - \Delta y)$ where $(\Delta x, \Delta y)$ is the 2-D offset, then, we have that

$$F(u, v) = G(u, v)e^{-j2\pi(\frac{u\Delta x}{M} + \frac{v\Delta y}{N})} \tag{1}$$

This can be re-written as

$$P(u, v) = \frac{F(u, v)G^*(u, v)}{|F(u, v)G^*(u, v)|} = e^{-j2\pi(\frac{u\Delta x}{M} + \frac{v\Delta y}{N})} \tag{2}$$

where * is the complex conjugate, $j = \sqrt{-1}$, and $P(u, v)$ is referred to as the cross power spectrum of the two signals.

If we apply the inverse Fourier transform to $P(u, v)$, the so-called phase correlation surface (PCS) can be obtained as $p(x, y)$. This is a 2-D impulse (Dirac) function located at $(\Delta x, \Delta y)$ if the two images under consideration are perfect replicas of each other. If not the surface is noisy but crucially still contains a dominant peak which yields an estimate of the shift parameters and can be recovered as

$$(\Delta \hat{x}, \Delta \hat{y}) = \arg \max_{x,y} | p(x, y) | \tag{3}$$

In this case, the peak value can be substantially less than unity (the expected value). To enhance the peak identification accuracy, pre-processing in the shape of filtering or windowing is often used. In our paper, however, such processing has not been considered in order to facilitate comparisons with competing methods so that the results are not conditional upon using a specific pre-processing regime.

It is worth noting that sub-pixel offsets will also trigger lower peaks appearing in the same neighborhood of the PCS rather than a single peak. Adopting the assumption that sub-pixel offsets between images are correspond integer offsets on a higher-density grid followed by down-sampling, $p(x, y)$ is approximated as [18]

$$p(x, y) = \frac{\sin(\pi(x - \Delta x))}{\pi(x - \Delta x)} \frac{\sin(\pi(y - \Delta y))}{\pi(y - \Delta y)} + n(x, y) \tag{4}$$

J. Ren (on leave from School of Computers, Northwestern Polytechnical Univ., Xi'an, 710072, China) and J. Jiang are with Digital Media & Systems Research Institute, University of Bradford, U. K. (j.ren@bradford.ac.uk, j.jiang1@bradford.ac.uk).

T. Vlachos is with the Department of Audiovisual Arts, Ionian University, Corfu, Greece. (t.vlachos@ionio.gr).

where $n(x, y)$ refers to effect of interference terms including spatial-interpolation effects, non-overlapped regions, and noise.

If the interference term can be ignored when the SNR is high enough, the sub-pixel offset can be obtained as:

$$\Delta x_{For} = \frac{p(1,0)}{p(1,0) \pm p(0,0)}, \; \Delta y_{For} = \frac{p(0,1)}{p(0,1) \pm p(0,0)} \quad (5)$$

where the sub-pixel offsets are decided by linear weighting the main peak and one of the side-peaks. Although good results have been reported in [18], the above solution has the drawback that it doesn't take into account the interference term. When real data are involved lower accuracy is achieved as demonstrated below.

According to (4), $p(0,0)$ and $p(1,0)$ are obtained as:

$$p(0,0) = \frac{\sin(\pi\Delta x)}{\pi\Delta x} \frac{\sin(\pi\Delta y)}{\pi\Delta y} + n(0,0) \quad (6)$$

$$p(1,0) = \frac{\sin(\pi(1-\Delta x))}{\pi(1-\Delta x)} \frac{\sin(\pi\Delta y)}{\pi\Delta y} + n(1,0) \quad (7)$$

If we substitute (6) and (7) in (5) and let $\Delta x \to 0$, we have

$$\lim_{\Delta x \to 0} \Delta x_{For} = \frac{n(1,0)}{(\pi\Delta y)^{-1} \sin(\pi\Delta y) + n(1,0) + n(0,0)} \quad (8)$$

Obviously, due to the effect of the interference item $\Delta x_{For} \neq 0$ even when $\Delta x = 0$. Similarly, $\Delta y_{For} \neq 0$ even when $\Delta y = 0$. As a result, we present a new solution to overcome this ambiguity which is discussed in the next Section.

III. SUB-PIXEL MOTION ESTIMATION

To obtain sub-pixel accuracy, the location of an underlying peak defined on a higher density (in the limit continuous) PCS needs to be identified. To achieve this, some representative approaches using interpolation are summarized below.

A. Existing Methods

In general, the location of the highest peak (x_0, y_0) is taken as the central point, and one or more of its closest neighbors are then used for interpolation. For simplicity, we denote $C(k,l) = p(x_0 + k, y_0 + l)$ where $k, l \in [-1,0,1]$ as shown in Fig. 1. As mentioned above, in [18] linear weighting is used taking into account $C(0,0)$ and either $C(1,0)$ or $C(0,1)$ to respectively estimate Δx and Δy. In [21], two 1-D quadratic curves are fitted in x and y directions using three points including the main peak and its two neighboring peaks. Then, Δx and Δy are determined as follows:

$$\Delta x_{Qud} = \frac{C(1,0) - C(-1,0)}{2C(0,0) - C(1,0) - C(-1,0)}$$

$$\Delta y_{Qud} = \frac{C(0,1) - C(0,-1)}{2C(0,0) - C(0,1) - C(0,-1)} \quad (9)$$

In [16], two 1-D Gaussian functions are fitted using the same three points yielding:

$$\Delta x_{Gau} = \frac{\log|C(1,0)/C(-1,0)|}{\log|C^2(0,0)/C(1,0)/C(-1,0)|}$$

$$\Delta y_{Gau} = \frac{\log|C(0,1)/C(0,-1)|}{\log|C^2(0,0)/C(0,1)/C(0,-1)|} \quad (10)$$

Using the same triplets of points, two 1-D *sinc* functions are fitted in [15-17]. Further a modified *esinc* function is also employed in [15]. It is worth noting that the last two approaches do not lead to closed-form solutions and consequently optimality may not be demonstrable analytically.

$C(-1,-1)$	$C(0,-1)$	$C(1,-1)$
$C(-1,0)$	$C(0,0)$	$C(1,0)$
$C(-1,1)$	$C(0,1)$	$C(1,1)$

Figure 1. Correlation values of the main peak $C(0,0)$ and its direct neighbors for sub-pixel motion estimation.

B. Proposed Method

It is worth mentioning that one possible categorization of the above methods can be based upon whether they use (in addition to the maximum peak) neighborhoods of one-sided ([12], [18]) or two-sides ([15]-[17], 21). As implied by (6)-(8), using one-sided information is sensitive to noise. Consequently, in what follows we favor the use of two-sided neighborhoods.

Firstly, let us define the difference between the two-sided neighbors along the x and y directions as D_x and D_y, i.e.

$$D_x = C(1,0) - C(-1,0), \; D_y = C(0,1) - C(0,-1) \quad (11)$$

If interference terms can be ignored, i.e. under a high SNR, we can derive from (4) that $D_x = 2[\pi(1-\Delta x.\Delta x)]^{-1} \sin(\pi\Delta x)$ and $c(0,0) = (\pi\Delta x)^{-1} \sin(\pi\Delta x)$. When $\Delta x \in (0,1)$, our estimate in the x direction Δx_{Our} can be obtained as

$$\Delta x_{our} = \frac{D_x}{c(0,0) + D_x} = \frac{\Delta x}{1 - (1 - \Delta x)^2 / 2} \approx \Delta x \quad (12)$$

As seen from (12), more closer Δx to 1 is, more accurate the estimate becomes. Decreasing Δx to near zero may increase the relative error of estimate. However, the absolute error is still limited as Δx itself is small and the effect of interference terms becomes more important.

If we consider the interference terms and substituting (4) into (12), we obtain that

$$\lim_{\Delta x \to 0}(\Delta x_{Our}) = \frac{C(1,0) - C(-1,0)}{C(0,0) + C(1,0) - C(-1,0)}$$

$$= \frac{n(1,0) - n(-1,0)}{(\pi\Delta y)^{-1} \sin(\pi\Delta y) + n(0,0) + n(1,0) - n(-1,0)} \quad (13)$$

If the interference term cancels out within the boundaries of a compact neighborhood or it exhibits even-function characteristics, i.e. $n(1,0) \approx n(-1,0)$, we have $\Delta x_{Our} = 0$ when $\Delta x = 0$. This renders our estimate more robust than the one in [18] as illustrated in Eq. (8).

More generally, our estimate in the x direction is given by

$$\Delta x_{Our} = \frac{D_x}{C(0,0) + |D_x|} = \frac{sign(D_x)}{1 + C(0,0)/|D_x|} \quad (14)$$

As seen from (14), if $D_x > 0$ we have $\Delta x_{Our} > 0$. On the other hand, we have $\Delta x_{Our} < 0$ if $D_x < 0$. When the two side peaks are equal, i.e. $D_x = 0$, there is no such bias then $\Delta x_{Our} = 0$. This means that the obtained estimate is always biased towards the sign of D_x, the

difference of the two-side neighbors, rather than the direction of a higher side-peak. As a result, this may lead to better estimates in generating good results as explained below.

Consider the case when $\Delta x \in (0,1)$, in fact both Δx_{For} and Δx_{Our} can be written in the form of $[1 + C(0,0)/S_x]^{-1}$ where we have $S_x = C(1,0) - C(-1,0)$ but $S_x = C(1,0)$ in Foroosh [18]. Using the peak difference rather than the peak height helps to make a difference in the following cases as illustrated in Fig. 2.
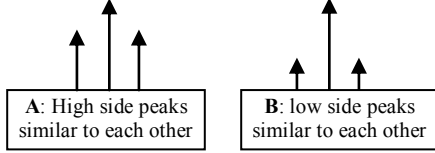
Figure 2. Two typical examples of peaks on the correlation surface, where A and B correspond to cases with more or less noise.

In Fig. 2, Case A is typical "noisy" motion estimation (not well-defined main peak and leakage of energy towards side peaks) and will force Δx_{For} to have given a significant bias away from the main peak. In contrast our scheme is insensitive to the actual heights of the side peaks and will still produce an estimate of high accuracy. For case B (typical of noise-less motion estimation i.e. low leakage) both schemes perform well in yielding a more reliable estimate.

Similarly, the estimate for Δy is given by

$$\Delta y_{Our} = \frac{D_y}{C(0,0)+|D_y|} = \frac{sign(D_y)}{1+C(0,0)/|D_y|} \quad (15)$$

C. Uncertainty of the Main Peak

When the sub-pixel shift is ±0.5, it can be found that there are two or more dominant peaks on the PCS. In the case $\Delta x = 0.5$, we have $C(1,0) - C(0,0) = n(1,0) - n(0,0)$ if assuming $\Delta y = 0$ for simplicity. If the interference item can be ignored subject to a high SNR, then the difference becomes zero, which renders the heights of the two peaks involved equal. Similarly if $\Delta x = \Delta y = 0.5$, it can be shown that the following four correlation values involved i.e. $C(0,0)$, $C(1,0)$, $C(0,1)$ and $C(1,1)$ are all equal. Due to the effect of interference terms, any of them can be extracted as the highest peak and leads to ambiguity. To overcome this drawback, an additional condition is introduced below, where $\delta \in [0.85,1)$ is used to measure how close the two correlation values are.

$$\Delta x_{Our} = \begin{cases} 0.5, & if \quad C(1,0)/C(0,0) > \delta \\ -0.5, & if \quad C(-1,0)/C(0,0) > \delta \end{cases} \quad (16)$$

$$\Delta y_{Our} = \begin{cases} 0.5, & if \quad C(0,1)/C(0,0) > \delta \\ -0.5, & if \quad C(0,-1)/C(0,0) > \delta \end{cases} \quad (17)$$

IV. RESULTS AND DISCUSSION

Both synthetic and real data have been used for evaluating the performance. Synthetic data is generated by displacing the original test images with known sub-pixel shifts (considered as ground truth) using linear interpolation. For experiments using real data, a set of MRI images are employed which have undergone sub-pixel displacements.

If we consider the ground truth data as a reference, error vectors for the estimates obtained by each competing method can be computed. Let d_x and d_y denote the corresponding vectors of absolute error, i.e. $d_x(i) = |x_i - \hat{x}_i|$ and $d_y(i) = |y_i - \hat{y}_i|$, where (x_i, y_i) are the ith real displacements and $(\hat{x}_i, \hat{y}_i)$ their corresponding estimates. The mean μ and standard deviation σ of error vectors are computed to assess accuracy. Additionally, the mean squared error (MSE) between the estimates and the ground truth is also used as a performance measure. This is consistent with work reported elsewhere (i.e. [4]). These three measures are defined as follows.

$$\mu(z) = n^{-1}\sum_{i=1}^{n} d_z(i), \quad z = x, y$$

$$\sigma(z) = \frac{1}{n-1}\sqrt{\sum_{i=1}^{n}[d_z(i) - \mu(z)]^2}, \quad z = x, y \quad (18)$$

$$MSE(z) = \sqrt{n^{-1}\sum_{i=1}^{n}[d_z(i)]^2}, \quad z = x, y$$

Consequently, a good quality estimate is expected to minimize such measures, where MSE and μ provides accuracy of the estimate and σ reflects how robust or consistent the estimate can be under different conditions. In addition, we also require that the values obtained for each of the three measures above are close to each other. This is important as it will show if the estimate is sensible to image contents such as edge effects under interpolation.

A. Synthetic Data

To generate sub-pixel shifts and evaluate the corresponding algorithms, five well-known images are employed which include "Airfield", "Barbara", "Image043", "Paris" and "Pentagon" as shown in Fig. 3. These are monochrome 8bpp images, and four of them are of 512×512 pixels except "Paris" which is of 200×200. They are shifted by linear interpolation to obtain sub-pixel displacements. This is different from that followed in [18] and [19], in which sub-pixel shifts are obtained by low-pass filtering and down-sampling of a real high resolution.

To fully test the effectiveness of the proposed method, in each direction 17 shifts are generated within the interval of [-1,1] which
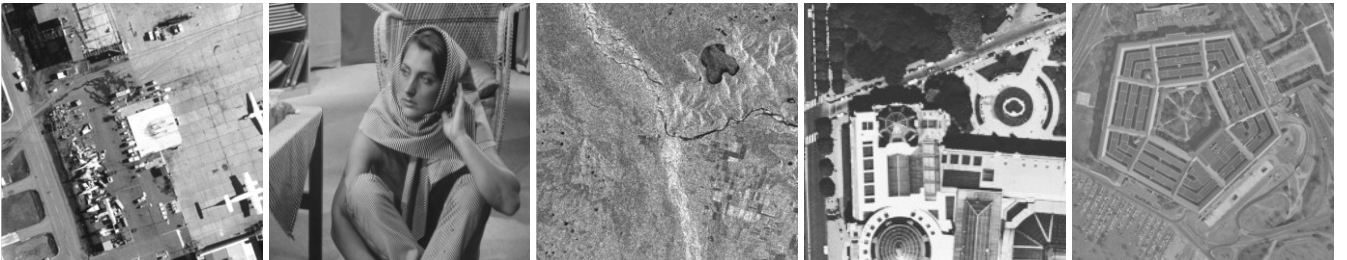
Fig. 3. Five test images used to generate sub-pixel shifts namely "Airfield", "Barbara", "Image043", "Paris" and "Pentagon", respectively.

Table I. Measurement of μ values along x- and y- axes for the five test images in Fig. 3 using 289 shifts between [-1,1] for each image.

Test images	Hoge [11]	Foroosh [18]	Stone [20]	Gaussian [16]	Quadratic [21]	Sinc [17]	ESinc [15]	Ours
Airfield	(0.1334,0.1328)	(0.0831,0.1327)	(0.7306,0.7066)	(0.0559,0.0479)	(0.1118,0.1106)	(0.0786,0.0786)	(0.0686,0.0919)	(0.0228,0.0132)
Barbara	(0.1937,0.1330)	(0.1634,0.1403)	(0.7458,0.6913)	(0.0643,0.1195)	(0.1167,0.1212)	(0.0881,0.1125)	(0.0998,0.1194)	(0.0429,0.0954)
Image043	(0.1331,0.1330)	(0.0828,0.0902)	(0.7428,0.6705)	(0.0515,0.0498)	(0.1108,0.1090)	(0.0837,0.0854)	(0.0342,0.0288)	(0.0066,0.0052)
Paris	(0.1336,0.1337)	(0.1631,0.1671)	(0.6990,0.7057)	(0.1100,0.0560)	(0.1052,0.1152)	(0.0897,0.0777)	(0.0996,0.0885)	(0.0831,0.0269)
Pentagon	(0.1330,0.1330)	(0.0800,0.1416)	(0.6805,0.6852)	(0.0576,0.0485)	(0.1123,0.1098)	(0.0768,0.0766)	(0.0777,0.0960)	(0.0276,0.0208)
Mean (x, y)	(0.1454,0.1331)	(0.1145,0.1344)	(0.7197, 0.6919)	(0.0679,0.0644)	(0.1114,0.1132)	(0.0833,0.0862)	(0.0760,0.0849)	(0.0366,0.0379)

Table II. Measurement of σ values along x- and y- axes for the five test images in Fig. 3 using 289 shifts between [-1,1] for each image.

Test images	Hoge [11]	Foroosh [18]	Stone [20]	Gaussian [16]	Quadratic [21]	Sinc [17]	ESinc [15]	Ours
Airfield	(0.0813,0.0806)	(0.0943,0.1502)	(0.6899,0.6640)	(0.0370,0.0328)	(0.0696,0.0629)	(0.0629,0.0633)	(0.1437,0.2431)	(0.0171,0.0128)
Barbara	(0.4120,0.0870)	(0.1811,0.0596)	(0.5047,0.4671)	(0.0394,0.0581)	(0.0676,0.0953)	(0.0811,0.1122)	(0.1985,0.1190)	(0.0317,0.0624)
Image043	(0.0811,0.0809)	(0.0777,0.0820)	(0.7776,0.5169)	(0.0391,0.0419)	(0.0730,0.0770)	(0.0620,0.0622)	(0.0766,0.0372)	(0.0067,0.0047)
Paris	(0.0831,0.0821)	(0.1111,0.1812)	(0.5602,0.5876)	(0.0539,0.0351)	(0.0891,0.0642)	(0.0753,0.0660)	(0.0942,0.1672)	(0.0588,0.0220)
Pentagon	(0.0808,0.0809)	(0.0935,0.1628)	(0.5277,0.5497)	(0.0415,0.0315)	(0.0621,0.0672)	(0.0638,0.0640)	(0.1397,0.2515)	(0.0211,0.0177)
Mean (x, y)	(0.1477,0.0823)	(0.1116,0.1271)	(0.6120, 0.5571)	(0.0422,0.0399)	(0.0711,0.0747)	(0.0692,0.0735)	(0.1305,0.1636)	(0.0271,0.0239)

include -1, -0.875, -0.75, -0.667, -0.5, -0.333, -0.25, -0.125, 0, 0.125, 0.25, 0.333, 0.5, 0.625, 0.75, 0.875 and 1. Consequently, for each test image in total a group of 289 2-D offsets are formed to generate shifted samples in the test, and results from each group are then evaluated in terms of μ, MSE and σ measurements for evaluations.

Here, our results are compared with those obtained from Hoge [11], Foroosh [18], Stone [20] and interpolation based approaches using Gaussian [16], Quadratic [21], Sinc [17] and ESinc [15] models. The corresponding results in terms of MSE and σ measurements are illustrated in Table I and Table II, respectively for quantitative evaluations. Please note that the standard Blackman window is employed to all the methods except Hoge in the spatial domain towards removing image-boundary effects for robustness. As can be seen, our approach consistently yields the best results in both x and y directions achieving the minimum values in terms of MSE and σ measurements. In addition, the two measures of our results along the x and y directions are very close to each other, which also show that our result is insensitive to image contents.

Regarding MSE measurements in Table I, Stone's approach almost fail in these tests, and its average errors are over half pixels in both the x and y directions. The results of Hoge, Foroosh and quadratic interpolation are very comparative with an average error between 0.11 and 0.15 pixels, though Hoge's method appears slightly worse. Interpolation using Gaussian, Sinc and ESinc produces similar results, whose average error is between 0.064 and 0.085 pixels, the

second best in the group. Finally, our proposed method yields the most accurate results with an average error of less than 0.038 pixels.

Regarding the measurement of σ in Table II, it is useful in measuring how robust and consistent the results are for one image under different sub-pixel shifts. Not surprisingly, it is found that Stone [20] still generates worst results. However, the 2nd worst results are from ESinc interpolation, followed by the methods from Hoge and Foroosh, which shows that these methods seem lack of robustness and may produce inaccurate estimates under certain circumstances. Interpolation based methods using Gaussian/quadratic and our scheme generate much less σ values, where our results are again the best.

1) Error Distribution Analysis

Although the σ above can provide a rough measurement of the robustness of these methods, it fails to show how accurate the estimates are under various shifts. To further evaluate the error distribution of these methods, we further analyze the MSE and σ measurements in the x and y directions as follows. Firstly, the shift in the y direction is fixed. For all the estimates of horizontal shifts among the five test images in Fig. 3, their MSE and σ measurements are attained. When the fixed shift varies among its 17 possible values, for each method two curves of MSE and σ vs. the fixed vertical shift are obtained. Similarly, two curves of MSE and σ values over estimates in the y direction vs. the fixed horizontal shift are also obtained for each method. Curves obtained from different methods are
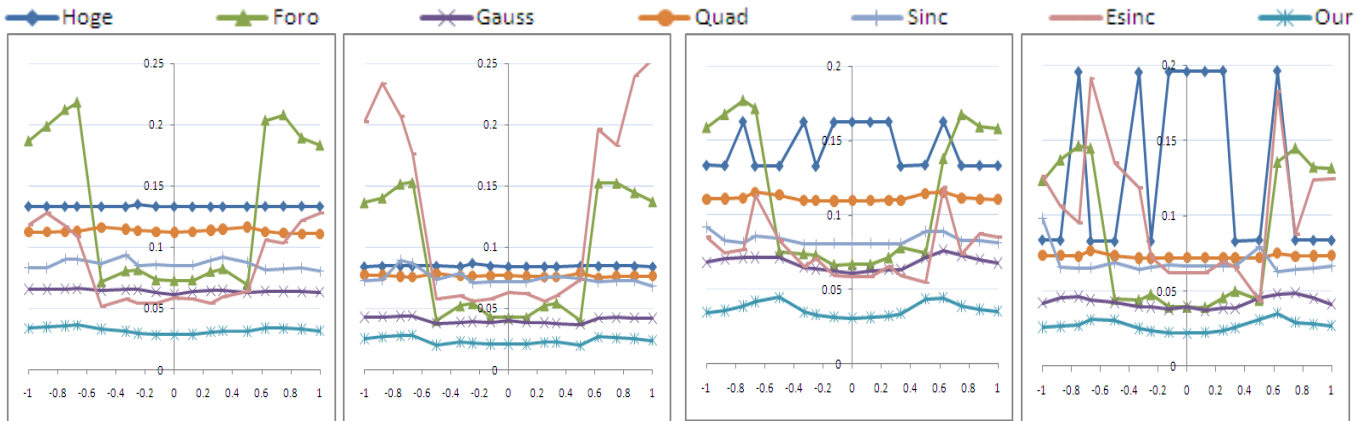
Figure 4. Measurements of μ and σ (y-axis) under various sub-pixel shifts (x-axis) over the five test images in Fig. 3: The left two are obtained respectively as average μ and σ values over all vertical shifts with fixed sub-pixel shifts in the horizontal direction, and the right two are for horizontal shifts under fixed sub-pixel shifts in the vertical direction.

plotted in Fig. 4 for comparisons, where the curves from Stone's method are omitted for better visualization purpose as they suffer large data ranges due to significant estimate errors produced.

As can be seen, these curves appear nearly bilateral symmetry as even functions, where the reason is that the shift values that we adopted are almost symmetry. Again, it has clearly demonstrated that our proposed method help to produce least MSE and σ values, i.e. the most accurate and most robust results, followed by the results using Gaussian fitting. Although Foroosh's method and ESinc fitting occasionally generate good results, they suffer large estimate errors when the absolute values of sub-pixel shifts are more than 0.6. Fitting using Sinc and quadratic functions yield consistent estimates, which makes the corresponding curves appear like lines. In addition, Sinc fitting outperforms quadratic fitting, and they both defeat Hoge's method. Although Hoge's method generates consistent measurement in the vertical direction (see the left two images in Fig. 4), the results in the horizontal direction are less consistent. This is due to occasional incorrect estimates from the test set, which has inevitably shown lack of robustness of this method.

B. Synthetic Data with Additive Noise

In this section, performance assessment of competing sub-pixel estimation methods is carried out using data with additive Gaussian and uniform noise. For simplicity, the six best performing methods from the previous test (i.e. Tables I and II) are selected. Before adding noise, the intensity level of the original images is normalized within $[0,1]$. Then, zero-mean noise is generated with different variance values to be added to the test images. In total, five levels of noise are used for both Gaussian and uniform noise, and the corresponding variance values are given in Table III. Example images with additive Gaussian / uniform noise are shown in Fig. 5 where it can be visually appreciated the extent to which the images have been degraded.


Figure 5. Examples of two groups of noisy test samples with additive Gaussian noise (the left two) and uniform noise (the right two) embedded to the "Barbara" image, where the noise levels in each group are 2 and 5, respectively.

Table III. Variance values for five levels of Gaussian and uniform noises used to generate noisy samples.

Variance	Level-1	Level-2	Level-3	Level-4	Level-5
Gaussian	0.005	0.015	0.025	0.035	0.045
uniform	0.05	0.10	0.15	0.20	0.25

For each of the five noise levels, 289 noisy samples are generated with embedded additive noise. Displacement estimates are computed between each pair of the 289 noisy samples for quantitative analysis. The MSE measures along x-axis and y-axis are then obtained and plotted in Fig. 6. Firstly, we can see that curve fitting using *sinc* and *esinc* is very sensitive to noise, which makes these two methods weak performers. Secondly, Foroosh's method yields third worst results, followed by quadratic interpolation, where the latter produces much better results than the former when samples with Gaussian noise are used. Thirdly, the results from Gaussian fitting and our approach are the best and robust over the entire set of test conditions. Moreover our approach exhibits lower MSE values slightly outperforming Gaussian fitting and is shown to be the overall best option in this comparison.
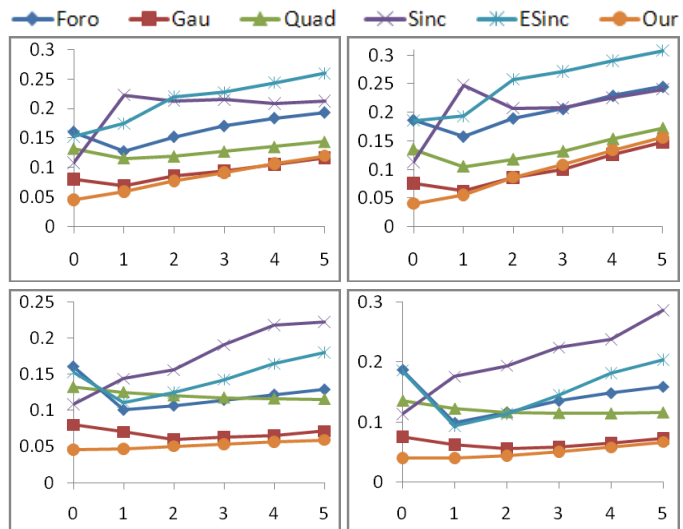
Figure 6. Comparison of MSE results (y-axis) vs. noise levels (x-axis) along the x- and y- directions (as the left and the right columns) for Gaussian (top row) and uniform noise (bottom row), respectively.

C. Real Data

In this section, real MRI data are used for evaluations. The data set is from Hoge and consists of five MRI images of a grapefruit [11]. The true offsets between each pair of images are known so they can be used as ground truth for performance evaluation. The first MRI image and its two noisy samples are shown in Fig. 7.

Firstly, 10-pairs of combinations are extracted from the five noise-free MRI images and used for motion estimation, and the estimated shifts as well as their corresponding average MSE measurements are given in Table IV. In general, the estimate along the x-axis has much larger MSE value than that of the y-axis, and the reasons are analyzed in detail in [11]. Among all the test methods, again it was found that the results generated by our method as well as

Table IV. Estimated shifts and their average MSE measurements for 10 image pairs of MRI data.

Image pairs	Real-shifts	Hoge [11]	Foroosh [18]	Balci [19]	Gaussian [16]	Quadratic [21]	Sinc [17]	ESinc [15]	Ours
[1,2]	(-2.40, 4.00)	(-2.03, 4.01)	(-2.22, 4.23)	(-2.11, 4.10)	(-2.07, 4.02)	(-2.03, 4.01)	(-2.0, 4.00)	(-2.0, 4.00)	(-2.09, 4.02)
[1,3]	(-4.80, 8.00)	(-4.13, 8.01)	(-4.36, 8.24)	(-3.9, 8.05)	(-4.33, 8.01)	(-4.18, 8.00)	(-4.12, 8.00)	(-4.25, 8.0)	(-4.34, 8.01)
[1,4]	(-7.20, 4.32)	(-6.81, 4.17)	(-6.59, 4.41)	(-6.22, 4.34)	(-6.57, 4.37)	(-6.73, 4.25)	(-6.72, 4.12)	(-6.54, 4.31)	(-6.58, 4.38)
[1,5]	(-7.20, 12.0)	(-6.82, 12.02)	(-6.59, 12.26)	(-6.39, 12.15)	(-6.57, 12.06)	(-6.74, 12.03)	(-6.73, 12.0)	(-6.54, 12.04)	(-6.59, 12.08)
[2,3]	(-2.40, 4.00)	(-2.1, 3.99)	(-2.32, 3.75)	(-2.18, 3.86)	(-2.26, 3.97)	(-2.13, 3.98)	(-2.09, 4.0)	(-2.19, 4.00)	(-2.27, 3.96)
[2,4]	(-4.80, 0.32)	(-4.28, 0.15)	(-4.55, 0.39)	(-4.16, 0.30)	(-4.55, 0.35)	(-4.65, 0.22)	(-4.72, 0.11)	(-4.59, 0.28)	(-4.54, 0.36)
[2,5]	(-4.80, 8.00)	(-4.78, 8.00)	(-4.55, 8.24)	(-4.13, 7.92)	(-4.56, 8.01)	(-4.65, 8.00)	(-4.71, 8.0)	(-4.60, 8.00)	(-4.54, 8.01)
[3,4]	(-2.40, -3.68)	(-2.17, -3.84)	(-2.4, -3.61)	(-2.34, -3.62)	(-2.433, -3.66)	(-2.25, -3.78)	(-2.27, -3.89)	(-2.46, -3.72)	(-2.40, -3.65)
[3,5]	(-2.40, 4.00)	(-2.18, 4.51)	(-2.41, 3.76)	(-2.49, 4.07)	(-2.44, 4.00)	(-2.27, 4.00)	(-2.27, 4.0)	(-2.47, 4.0)	(-2.41, 4.00)
[4,5]	(0.00, 7.68)	(0.01, 7.85)	(-0.18, 7.61)	(-0.03, 7.66)	(-0.013, 7.64)	(-0.01, 7.78)	(0.00, 7.87)	(0.00, 7.54)	(-0.02, 7.64)
Mean MSE (x,y)		(0.367, 0.191)	(0.337, 0.195)	(0.586, 0.084)	(0.355, 0.031)	(0.333, 0.061)	(0.349, 0.130)	(0.385, 0.051)	(0.349, 0.040)

Gaussian fitting are the best. The corresponding MSE values are (0.349, 0.040) and (0.355, 0.031), respectively, which are slightly better than those using quadratic, *sinc* and *esinc* fitting. The methods by Hoge [11], Foroosh [18] and Balci [19] yield significant errors, and their MSE values were found respectively as (0.367, 0.191), (0.337, 0.195) and (0.586, 0.084).

Regarding motion estimation from noisy samples, among all methods, as shown in Fig. 8, our proposed method together with curve fitting using Gaussian and quadratic functions were shown to be the ones less sensitive to noise. Fitting using *sinc* and *esinc* functions are found very sensitive to noise. Although good results are generated in the horizontal direction, Foroosh's method [18] yields the worst results in the vertical direction, which makes it less robust in this test.
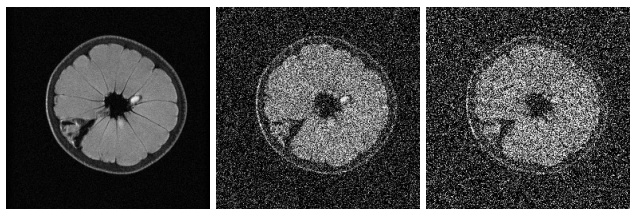
Figure 7. One of the five original MRI images (left, Courtesy of W. S. Hoge [11]) and its two noisy versions (the right two) embedded with the Gaussian and uniform noise at variance level 3, respectively.
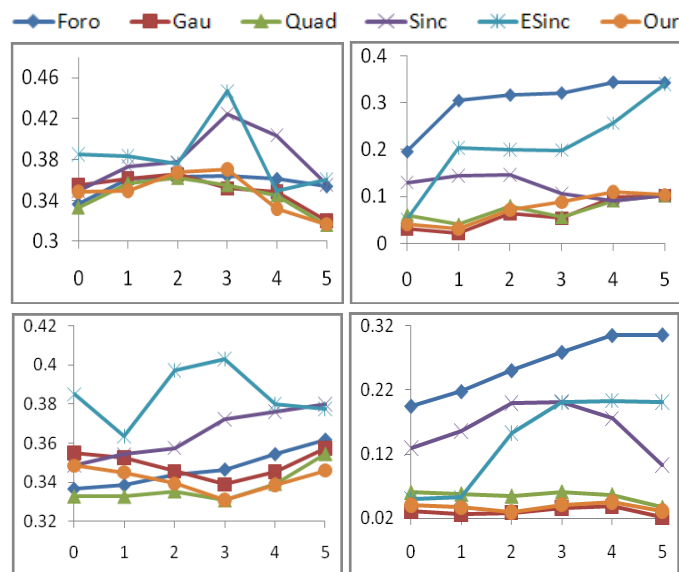
Figure 8. Comparison of MSE results (y-axis) vs. noise levels (x-axis) for the 10 pairs of noisy MRI data, where the top and bottom rows respectively refer to results from noisy samples with Gaussian and uniform noise. The left and the right columns are for shifts along the horizontal and vertical directions.

V. CONCLUSIONS

We have proposed a new scheme for the identification of sub-pixel motion estimates on phase correlation surfaces. The robustness of the proposed method in the presence of the noise and other forms of interference such as non-overlapping image material has been demonstrated both theoretically and empirically. It was found that taking into account the difference between the two secondary side-peaks as opposed to a one-sided peak improves considerably the sub-pixel accuracy performance. Results obtained using a variety of test material, including manually displaced digital imagery and MRI data, under a variety of test conditions have confirmed the superiority of our approach.

REFERENCES

[1]. S.-D. Wei and S.-Hi, Lai, "Robust and efficient image alignment based on relative gradient matching," *IEEE Trans. Image. Proc.*, vol. 15, no. 10, pp. 2936-2943, 2006.

[2]. P. Thevenaz, U. E. Ruttimann, and M. Unser, "A pyramidal approach to sub-pixel registration based on intensity," *IEEE Trans. Image Proc.*, vol. 7, no. 1, pp. 27-41, 1998.

[3]. Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronbin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, 43(9), pp. 2127-2137, 2005.

[4]. D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Trans. Image Proc.*, vol. 13, no. 9, pp. 1185-1199, 2004.

[5]. R. J. Althof, M. G. J. Wind, and J. T. Dobbins, "A rapid and automatic image registration algorithm with sub-pixel accuracy," *IEEE Trans. Medical Imaging*, vol. 16, no. 3, pp. 308-316, 1997.

[6]. U. Koc and K. J. R. Liu, "Interpolation-free sub-pixel motion estimation techniques in DCT domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 4, pp. 460-487, 1998.

[7]. L. Zagorchey and A. Goshtasby, "A comparative study of transformation function for nonrigid image registration," *IEEE Trans. Image. Proc.*, vol. 15, pp. 3, pp. 529-538, 2006.

[8]. B. Zitova and J. Flusser, "Image registration methods, a survey," *Image and Vision Comput.*, vol. 21, no. 11, pp. 977-1000, 2003.

[9]. E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms". *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 9, no. 5, pp. 700-703, 1987.

[10]. Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition". *IEEE Trans. PAMI*, 16(12), pp. 1156-1168, 1994.

[11]. W. S. Hoge, "Subspace identification extension to the phase correlation method". *IEEE Trans. Medical Imaging*, vol. 22, no. 2, pp.277-280, 2003.

[12]. L. Chen and K.-H. Yap, "An effective technique for subpixel image registration under noisy conditions," *IEEE Trans. System, Man and Cybernetics (A)*, vol. 38, no. 4, pp. 881-887, 2008.

[13]. P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP J. Applied Signal Process.*, vol. 2006, pp. 1-14, 2006.

[14]. B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration". *IEEE Trans. Image Proc.*, vol. 5, no. 8, pp. 1266–1271, 1996.

[15]. V. Argyriou and T. Vlachos, "A study of sub-pixel motion estimation using phase correlation". In *Proc. BMVC*, pp. 387-396, 2006.

[16]. I. E. Abdou, "Practical approach to the registration of multiple frames of video images," in *Proc. SPIE Conf. Vis. Comm. Image Proc.*, vol. 3653, pp. 371-382, 1999.

[17]. H. Shekarforoush,, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of cross power spectrum," in *Proc. CVPR*, pp. 532-537, 1996.

[18]. H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to sub-pixel registration". *IEEE Trans. Image Proc.*, vol. 11, no. 3, pp. 188–200, 2002.

[19]. M. Balci and H. Foroosh, "Sub-pixel estimation of shifts directly in the Fourier domain". *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1965-1972, 2006.

[20]. H. S. Stone, M. Orchard, E.-C. Chang, and S. Martucci, "A fast direct Fourier-based algorithm for sub-pixel registration of images". *IEEE Trans. Geosci. Remote Sens.*, 39(10), pp. 2235–2243, 2001.

[21]. Q. Tian and M. N. Huhns, "Algorithms for sub-pixel registration". *Computer Vision, Graphics and Image Proc.*, vol. 35, no. 2, pp. 220-233, August 1986.