Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 60796, Pages 1–11
DOI 10.1155/ASP/2006/60796

Subpixel Registration Directly from the Phase Difference

Murat Balci and Hassan Foroosh

School of Computer Science, University of Central Florida, Orlando, FL 32816, USA

Received 1 December 2004; Revised 31 May 2005; Accepted 21 June 2005

This paper proposes a new approach to subpixel registration, under local/global shifts or rotation, using the phase-difference matrix. We establish the exact relationship between the continuous and the discrete phase difference of two shifted images and show that their discrete phase difference is a 2-dimensional sawtooth signal. As a result, the exact shifts or rotations can be determined to subpixel or subangle accuracy by counting the number of cycles of the phase-difference matrix along the frequency axes. The subpixel portion is represented by a fraction of a cycle corresponding to the noninteger part of the shift or rotation. The rotation angle is estimated by applying our method using a polar coordinate system. The problem is formulated as an overdetermined system of equations and is solved by imposing a regularity constraint. The tradeoff for imposing the constraint is determined by exploiting the rank constraint leading to a closed-form expression for the optimal regularization parameter.

1. MOTIVATION

Registration is a crucial step in the analysis and fusion of information between multiple images. Examples can be found in remote sensing, robotics, and biomedical imaging, among others [1–3]. Some applications, such as coding and compression registration, need to be established locally, while other applications require only global registration. In this paper, we are particularly motivated by applications that require registration at subpixel accuracy. An important example of such applications is multi-frame super-resolution [4–20], which aims to combine several degraded low-resolution images into a single high-resolution image in order to approach the Nyquist rate.

The key to success in these multiframe super-resolution techniques is the accurate registration with subpixel precision. Essentially, subpixel registration is the step that allows merging the samples of the low-resolution data in a denser grid. In the absence of subpixel registration and fusion, the super-resolution problem would essentially reduce to that of classical deconvolution. Among existing subpixel registration techniques, Fourier domain methods [21–25] and also closely related spatial domain variations [26] are an important class of registration techniques that have gained popularity due to their remarkable accuracy. Most of these methods are in fact variations of the original phase-correlation method [27]. For instance, in [21], Foroosh et al. demonstrated how the method can be extended to subpixel accuracy, in [25] Stone et al. investigated the effect of aliasing error, and in [23] Hoge describes how the shift parameters can be decoupled in the dominant left and right eigenvectors of the phase-correlation matrix (not the phase-difference matrix) using its rank-one constraint.

This paper revisits the concepts underlying the phase-correlation method and shows how these methods are interrelated. By deriving an exact model for the phase difference of two discrete shifted images, we show that for shifted images (or for rotated images if we work with polar coordinates) the discrete phase-difference matrix is a 2-dimensional sawtooth signal. This in particular leads to a simple solution directly from the phase-difference matrix in the form of an overdetermined system of linear equations. The overdetermined nature of the formulation allows for handling random noise. In the next section, in order to describe the algorithms in the discrete domain, we will discuss the concepts and the theory that are commonly borrowed from the continuous domain. We then argue in Section 3 that the continuous domain representation is rather misleading, and that a proper representation can be derived directly in the discrete domain. Once an exact model is known, the formulation of the problem becomes straightforward as described in the subsequent sections.

Since by changing from Cartesian to polar coordinates both translation and rotation can be treated in the same manner, we discuss the background and our formulation in terms of translation.

2. BACKGROUND

Consider two continuous signals $f_1(x, y)$ and $f_2(x, y) = f_1(x - x_o, y - y_o)$, whose Fourier transforms exist (i.e., are square integrable). Their cross-power spectrum (also known as the phase correlation) is then given by

$$\hat{c}(u, v) = \frac{\hat{f}_1 \hat{f}_2^*}{|\hat{f}_1 \hat{f}_2^*|} = \exp(-ix_o u - iy_o v), \tag{1}$$

where the hat sign indicates the Fourier transform and the asterisk stands for the complex conjugate.

This result follows immediately from the Fourier shift property, which states that the translations in the spatial domain lead to linear phase differences between the two signals along each frequency axis, that is,[1]

$$\hat{\varphi}(u, v) = \angle \hat{c}(u, v) = x_o u + y_o v, \tag{2}$$

which is a planar surface through the origin. As a result, the spatial translations can be determined by inverse transforming $\hat{c}(u, v)$, which leads a Dirac delta function centered at (x_o, y_o).

A discrete interpretation of this result is used in practice for image registration [27, 28], which yields very good results. In the discrete case, the shift parameters are similarly determined by inverse transforming the discrete cross-power spectrum, which yields a unit impulse centered at (x_o, y_o). This, of course, is true only for pixel displacements. For shifts that include subpixel components, additional results are derived in [21].

When applying this approach locally, within a small window size, the estimation of local motion becomes inaccurate and dominated by noise. The main cause for this is due to the fact that the Fourier transform is an operator that cannot localize signals both in space and frequency. As a result, the noise process and the aliasing errors, which often are localized at the high-frequency components of the Fourier spectrum, become dispersed in the spatial domain upon inversion of $\hat{c}(u, v)$.

A possible approach to overcome this problem is therefore to avoid inverse transforming the cross-power spectrum and try to estimate the shifts directly in the Fourier domain, typically by also applying the method within a window of smaller size and smooth shape [23, 25]. A practical solution for estimating registration directly in the Fourier domain was first proposed by Hoge [23]. His method requires the following two steps prior to computing the shifts:

(i) a subspace approximation of the noisy phase-correlation matrix (not the phase difference) to impose a rank-1 constraint,

(ii) unwrapping of the dominant left and right eigenvectors.

He then estimates the shift parameters using the slopes of these unwrapped eigenvectors.

It is worth noting that 2-dimensional phase unwrapping is a notoriously ill-posed problem. This is perhaps why Hoge proposed to perform the unwrapping step on the 1-dimensional dominant eigenvectors of the phase-correlation matrix rather than directly on the phase-difference matrix itself. We will show below that due to the special shape of the phase-difference matrix, its unwrapped version has to be rank 2. As a result, the unwrapping process becomes separable along the two frequency axes. In other words, it reduces to two 1-dimensional unwrapping steps. Therefore, very good results may also be found without subspace approximation. We will also show that even phase unwrapping is an unnecessary step since we will determine the exact parametric shape of the discrete phase-difference matrix by examining the relationship between the discrete and the continuous cases. This exact relationship allows for determining the registration parameters by simply fitting a model directly to the phase-difference matrix.

3. RELATION BETWEEN DISCRETE AND CONTINUOUS PHASE DIFFERENCES

From (2), the phase difference of the underlying continuous signals has the following representation in the spatial domain:

$$\varphi(x, y) = \iint_{-\infty}^{\infty} (x_0 u + y_o v) \exp(iux + ivy) du\, dv \tag{3}$$

$$= -ix_o \frac{d\delta(x)}{dx} - iy_o \frac{d\delta(y)}{dy}, \tag{4}$$

where the derivatives are understood in the distributional sense [29].

Although inverse transforming the phase-difference function in this way may not be customary, it facilitates our understanding of the relationship between the continuous and the discrete phase difference. In particular, this will lead to an exact parametric model of the discrete phase-difference matrix, as shown below.

From bandlimited sampling theory and (4), it follows that the spatial domain representation of a component of the discrete phase-difference matrix is given by

$$\varphi_{kl} = -i\frac{x_o}{\pi k}\left(2\,\mathrm{sinc}\,\frac{k}{x_o} - 2\cos\frac{\pi k}{x_o}\right) - i\frac{y_o}{\pi l}\left(2\,\mathrm{sinc}\,\frac{l}{y_o} - 2\cos\frac{\pi l}{y_o}\right). \tag{5}$$

[1] Note that, throughout the paper, we call $\exp(-ix_o u - iy_o v)$ the phase-correlation function and $\hat{\varphi}(u, v) = \angle \hat{c}(u, v)$ the phase-difference function. Similarly, their discrete counterparts are referred to as the discrete phase-correlation matrix and the discrete phase-difference matrix, respectively.

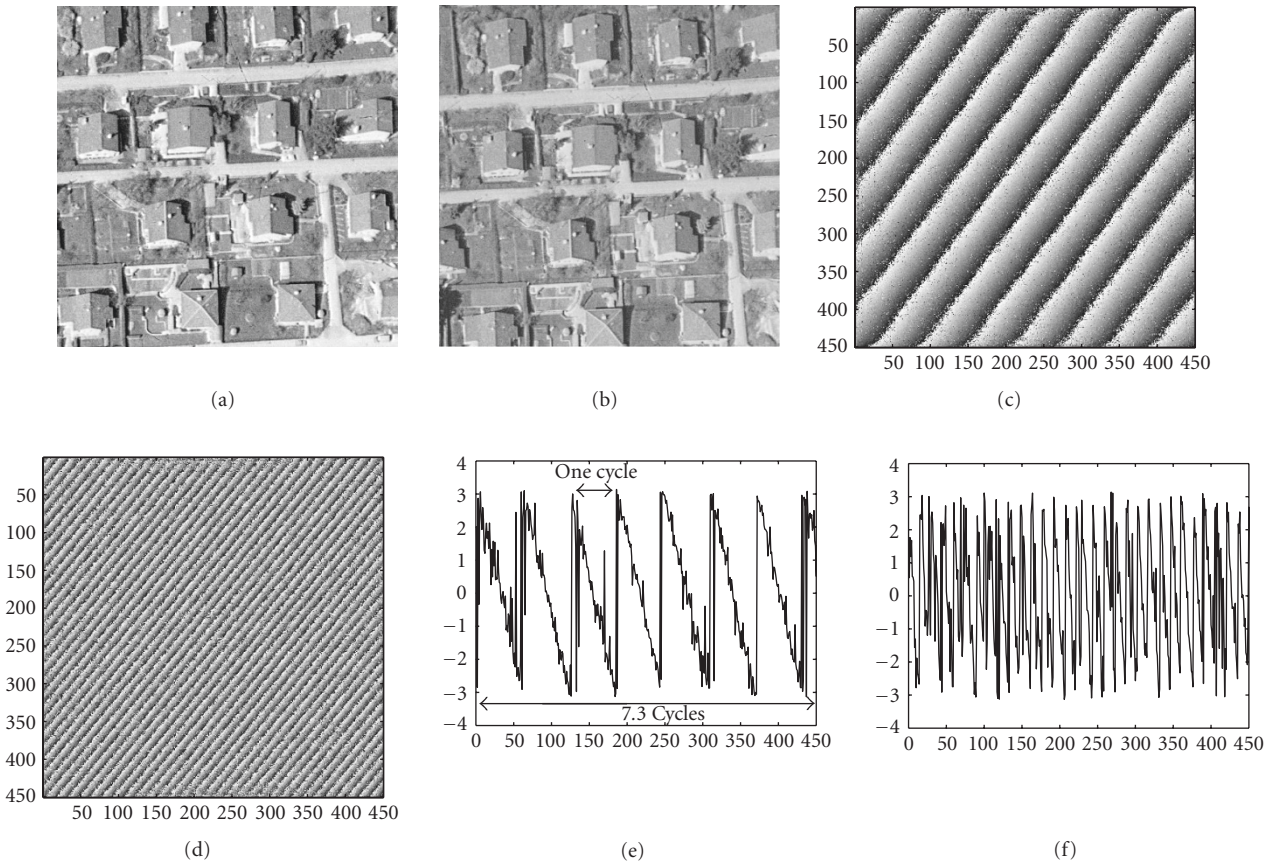

Figure 1: (a), (b) Two aerial images with some shifts, (c) noisy sawtooth phase-difference matrix corresponding to shifts of (7.3,5.6) pixels, (d) noisy sawtooth phase-difference matrix corresponding to shifts of (30.5,25.4) pixels, and (e), (f) one row of the phase-difference matrices shown in (c) and (d), respectively.

On the other hand, note that

$$\begin{aligned} i\varphi_{kl} &= \frac{x_o^2}{\pi^2 k^2}\left(2\sin\frac{\pi k}{x_o} - 2\frac{\pi k}{x_o}\cos\frac{\pi k}{x_o}\right) \\ &\quad + \pi\frac{y_o^2}{\pi^2 l^2}\left(2\sin\frac{\pi l}{y_o} - 2\frac{\pi l}{y_o}\cos\frac{\pi l}{y_o}\right) \end{aligned} \tag{6}$$

$$\begin{aligned} &= \frac{x_o}{\pi}\frac{2}{2\pi/x_o}\int_{-\pi/x_o}^{\pi/x_o} u\sin ku\,du \\ &\quad + \frac{y_o}{\pi}\frac{2}{2\pi/y_o}\int_{-\pi/y_o}^{\pi/y_o} v\sin lv\,dv. \end{aligned} \tag{7}$$

It can also be verified that

$$\begin{aligned} &\frac{x_o}{\pi}\frac{2}{2\pi/x_o}\int_{-\pi/x_o}^{\pi/x_o} u\cos ku\,du \\ &\quad + \frac{y_o}{\pi}\frac{2}{2\pi/y_o}\int_{-\pi/y_o}^{\pi/y_o} v\cos lv\,dv = 0 \end{aligned} \tag{8}$$

and similarly,

$$\frac{x_o}{\pi}\frac{2}{2\pi/x_o}\int_{-\pi/x_o}^{\pi/x_o} u\,du + \frac{y_o}{\pi}\frac{2}{2\pi/y_o}\int_{-\pi/y_o}^{\pi/y_o} v\,dv = 0. \tag{9}$$

From (7), (8), and (9), and using the definition of the discrete Fourier transform (DFT) based on Fourier series [30], it follows immediately upon substituting $u = n(2\pi/N)$ and $v = m(2\pi/M)$ that φ_{kl} is a DFT coefficient of the following discrete periodic signal:

$$\phi_{mn} = \frac{2\pi}{N}\left(x_o n + J\frac{N}{x_o}\right) + \frac{2\pi}{M}\left(m y_o + K\frac{M}{y_o}\right), \tag{10}$$

where J and K are arbitrary integers.

In other words, the discrete phase-difference matrix for a pair of shifted images is given by

$$\mathbf{\Phi} = [\phi_{mn}], \tag{11}$$

where $m = 0,\ldots,M-1$ and $n = 0,\ldots,N-1$.

This is a discrete 2D periodic sawtooth signal as opposed to the continuous phase-difference function in (2), which is a plane through the origin. Figure 1 shows examples of noisy discrete phase-difference matrices. The underlying 2-dimensional sawtooth signals are clearly visible. Figures 1(e) and 1(f) show one row of Figures 1(c) and 1(d). The first observation that can be made from this result is that the unwrapping of a 2-dimensional sawtooth signal is separable, since its unwrapped matrix has to be rank-2. This also correlates to the fact that the phase-correlation matrix has to be rank-1 [23]. In particular, it implies that a subspace

approximation can be disregarded. But even a more interesting conclusion that can be drawn from the above result is that unwrapping is also an unnecessary step for registration.

For this purpose, note how the period of the sawtooth signal along each axis determines the shifts along corresponding axis: the period along the u-axis is N/x_o, and hence there are x_o repeated cycles along each row of the phase difference, where x_o may or may not be an integer. When x_o is not an integer, the number of repeated cycles in a row is given by the integer part of x_o plus a fraction of a cycle defined by the noninteger portion of x_o. A similar argument applies to the columns of $\mathbf{\Phi}$. This process of counting the number of cycles along the rows and columns of the phase-difference matrix is essentially all that is required to determine the local or the global shifts. The challenge of course is to determine the exact fractional portion of the repeated cycles. In the next two sections, we will design a linear estimator for this problem.

4. PROBLEM FORMULATION

As indicated above, the key to solve the problem is to find how many cycles of the discrete sawtooth phase difference fit in the range $[0, 2\pi]$ along each frequency axis. The number of cycles, that is, x_o and y_o, may or may not be integer values and are given by

$$\begin{aligned} x_o &= \frac{\text{cycles}}{2\pi} = \frac{N}{2\pi}\frac{d\mathbf{\Phi}(m,n)}{dn} \\ y_o &= \frac{\text{cycles}}{2\pi} = \frac{M}{2\pi}\frac{d\mathbf{\Phi}(m,n)}{dm}. \end{aligned} \tag{12}$$

Due to noise and other sources of error, however, counting the number of cycles per 2π using (12) may lead to inaccurate results. To overcome this problem, we exploit the fact that a total of $M \times N$ data points are available for regression. Therefore, an accurate solution can be obtained by minimizing an appropriate error function, that is, by solving a largely overdetermined problem of fitting the parameters to the data set.

For this purpose note that the 2D sawtooth signal $\mathbf{\Phi}$ has constant slopes for the vast majority of frequencies along each row or each column, except for a small number of frequencies, where discontinuities occur. As depicted in Figure 1(e), visually we use a landmark point of the cycle to count the number of cycles along each row or each column. This idea can be implemented as follows. A particular useful landmark point is the zero-crossing of the phase-difference matrix, where $\phi_{mn} = 0$. Since the same set of zero-crossings can be independently obtained by scanning along the rows or columns, we can fix $J = 0$ in (10), and after some algebraic manipulations obtain

$$n\cos\theta + m\sin\theta + \rho = 0, \tag{13}$$

where

$$\tan\theta = \frac{Ny_o}{Mx_o}, \quad \rho = K\frac{M}{y_o}\sin\theta. \tag{14}$$

This shows that the zero-crossings are represented by a family of lines in the phase space parameterized by K—that is, each integer value of K would give a different line along which the phase difference is zero. Each line itself is parameterized by the angle θ and its distance ρ from the center frequency (i.e., the origin of the frequency domain). This set of zero-crossing lines form a function that is ideal for detection using Hough transform. The Hough transform basically maps these lines to a parameter space of (θ, ρ). As can be verified from the above derivations, θ remains invariant among all lines and ρ varies as integer multiples of some other invariant parameter, that is, $\rho = K\rho_o$. Therefore in the Hough-transform domain (i.e., in the Hough matrix), we expect to see a set of peak values situated at equal distances from each other, and parallel to the ρ-axis. Figure 2 shows an example of the Hough transform of the zero-crossing of the phase-difference matrix of two shifted images, where the peaks can be clearly identified by a simple thresholding process. As is customary in Hough transform, we used the local maxima for finding a suitable threshold value. In our case, since all the peaks are known to be aligned parallel to the ρ-axis, we took the maximum of the Hough matrix for each ρ as the local maximum. This yields a curve similar to the one shown in Figure 2(b). We then used the average of the local maxima curve as a threshold.

The problem now reduces to estimating (x_o, y_o) from these peak values in the Hough transform domain. For this purpose note that by combining (14), we can obtain the following linear constraint on x_o and y_o:

$$\frac{\rho}{\cos\theta}x_o + \frac{\rho}{\sin\theta}y_o = K(M+N). \tag{15}$$

Each peak point in the Hough-transform domain provides one such linear constraint on x_o and y_o. Given a total of t such peak values, we can construct a system of linear equations of the form

$$\mathbf{Ar} = \mathbf{b}, \tag{16}$$

where

$$\mathbf{A} = \begin{bmatrix} \dfrac{\rho_1}{\cos\theta_1} & \dfrac{\rho_1}{\sin\theta_1} \\ \vdots & \vdots \\ \dfrac{\rho_t}{\cos\theta_t} & \dfrac{\rho_t}{\sin\theta_t} \end{bmatrix}, \tag{17}$$

$\mathbf{r} = [x_o \;\; y_o]^T$, and $\mathbf{b} = (M+N)[K_1 \;\; \cdots \;\; K_t]^T$.

In the next section, we devise an approach to solve this overdetermined problem.

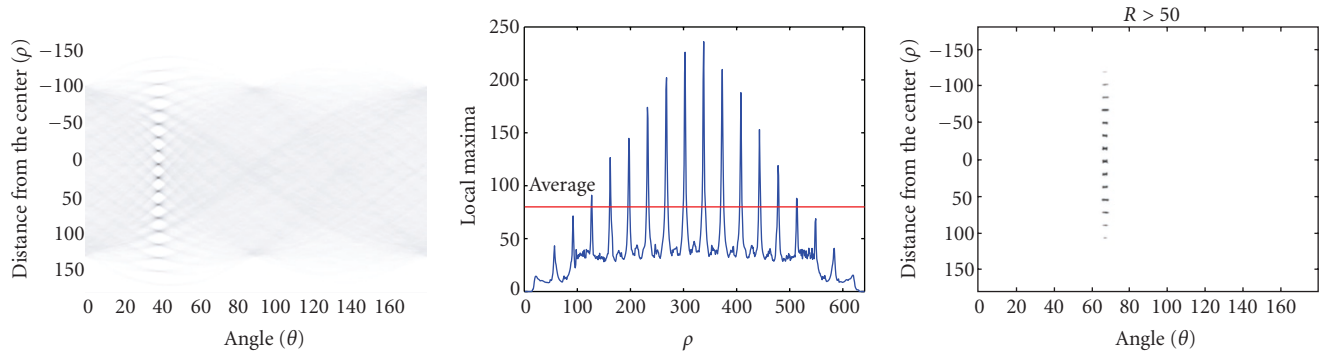

FIGURE 2: (a) The Hough transform of the phase discontinuities, (b) the local maxima and the threshold, and (c) the peaks detected by thresholding.

5. PROPOSED SOLUTION

In order to solve the overdetermined system of equations in (16), we formulate it as the following constrained optimization problem

$$\mathbf{r}_{\text{opt}} = \arg\min \|\mathbf{Ar} - \mathbf{b}\| + \lambda \|\mathbf{Lr}\|, \tag{18}$$

where λ is the regularization parameter, and $\mathbf{L}$ is such that

$$\mathbf{L}^T\mathbf{L} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}. \tag{19}$$

This choice of $\mathbf{L}$ [31] implies that our a priori belief is that our solution should be constant over all equations in the system, that is, the equations should be consistent with each other. The formulation in (18) is basically a Tikhonov-Arsenin norm-regularization of the least-squares solution to the problem in (16). The first term imposes the faithfulness to data and the second term the regularity. The solution is given by

$$\mathbf{r}_{\text{opt}} = \left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{A}^T\mathbf{b}. \tag{20}$$

We now have the solution given by (20) up to an unknown regularization parameter. The optimal value of this parameter is given by the method of generalized cross validation (GCV), which amounts to minimizing

$$\text{GCV}(\lambda) = \frac{\left\|\left(\mathbf{I} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{A}^T\right)\mathbf{b}\right\|^2}{\left(\text{tr}\left(\mathbf{I} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{A}^T\right)\right)^2} \tag{21}$$

with respect to λ, where $\text{tr}(\cdot)$ is the trace of a matrix.

The minimizer of (21) is usually obtained using numerical techniques by making simplifying assumptions, for example, circulant $\mathbf{A}$ and $\mathbf{L}$, or by using numerical techniques such as quadrature rules and Lanczos algorithm [32]. However, in our case, due to the rank constraint of $\mathbf{A}$, we can find a simplified closed-form solution. For this purpose, let $\mathbf{P} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{A}^T$. The GCV function can then be written as

$$\text{GCV}(\lambda) = \frac{\|\mathbf{Pb}\|^2}{\left(\text{tr}(\mathbf{P})\right)^2}. \tag{22}$$

Upon rearranging $\mathbf{P}$ as follows:

$$\begin{aligned} \mathbf{P} &= \mathbf{I} - \mathbf{A}\mathbf{L}^{-1}\mathbf{L}\left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{L}^T\mathbf{L}^{-T}\mathbf{A}^T \\ &= \mathbf{I} - \mathbf{A}\mathbf{L}^{-1}\left(\mathbf{L}^{-T}\mathbf{A}^T\mathbf{A}\mathbf{L}^{-1} + \lambda\mathbf{L}^{-T}\mathbf{L}^T\mathbf{L}\mathbf{L}^{-1}\right)^{-1}\mathbf{L}^{-T}\mathbf{A}^T \\ &= \mathbf{I} - \mathbf{K}\left(\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I}\right)^{-1}\mathbf{K}^T \end{aligned} \tag{23}$$

and applying the matrix inversion lemma, we find

$$\mathbf{P} = \left(\mathbf{I} + \frac{1}{\lambda}\mathbf{K}\mathbf{K}^T\right)^{-1}, \tag{24}$$

where $\mathbf{K} = \mathbf{A}\mathbf{L}^{-1}$.

Now, let

$$\mathbf{K}\mathbf{K}^T = \mathbf{V}\Sigma\mathbf{V}^T = \sum_{j=1}^{t}\sigma_j\mathbf{v}_j\mathbf{v}_j^T \tag{25}$$

be the eigen decomposition of $\mathbf{K}\mathbf{K}^T$, where $\mathbf{v}_j$'s are the columns of $\mathbf{V}$ that form a set of orthonormal basis, and σ_j's are the corresponding eigenvalues. $\mathbf{P}$ can then be written as

$$\begin{aligned} \mathbf{P} &= \sum_{j=1}^{t}\frac{\lambda}{\lambda+\sigma_j}\mathbf{v}_j\mathbf{v}_j^T \\ &= \sum_{j=1}^{2}\frac{\lambda}{\lambda+\sigma_j}\mathbf{v}_j\mathbf{v}_j^T + \sum_{j=3}^{t}\mathbf{v}_j\mathbf{v}_j^T, \end{aligned} \tag{26}$$

where the last equality follows from the fact that $\mathbf{A}$ is rank 2.

In order to simplify the GCV function in (22), we make a first-order approximation of the rank-2 matrix $\mathbf{K}\mathbf{K}^T$ using its largest eigenvalue $\mathbf{K}\mathbf{K}^T \simeq \sigma_1\mathbf{v}_1\mathbf{v}_1^T$. In practice, we found that the dominant eigenvalue is usually orders of magnitude larger than the second eigenvalue (see the appendix for more formal discussions). Therefore

$$\text{GCV}(\lambda) \simeq \frac{\left(\lambda/(\lambda+\sigma_1)\right)^2 s_1^2 + \sum_{j=2}^{t} s_j^2}{\left(t - 1 + \lambda/(\lambda+\sigma_1)\right)^2}, \tag{27}$$

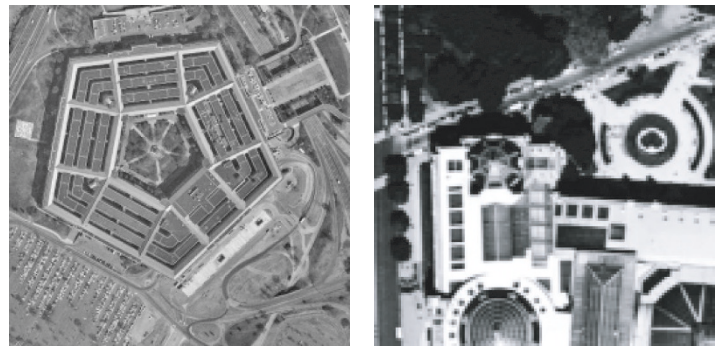

Figure 3: Some of the images used for simulation.

where s_j are the components of the vector $\mathbf{V}^T\mathbf{b} = [s_1 \ s_2 \ \cdots \ s_t]^T$.

Differentiating this equation with respect to λ and setting it equal to zero, we find after simplification that the optimal regularization parameter is given by

$$\lambda^* = \frac{\sigma_1 \sum_{j=2}^{t} s_j^2}{(t-1)s_1^2 - \sum_{j=2}^{t} s_j^2}. \tag{28}$$

Using (20) and (28), we can compute the optimal solutions for x_o and y_o.

6. EXPERIMENTAL RESULTS

We applied the technique to both the global subpixel registration problem and the local motion estimation. We also applied our approach to rotated images using polar coordinate transformation. The experimentations included both synthetically generated shifts and real data with and without ground truth. In all cases, very good results were obtained. We used the synthetic simulations to evaluate our approach under various conditions.

For global registration, we synthetically generated the shifts by filtering and downsampling shifted versions of a high-resolution image. Using appropriate combinations of these operations, shifts with subpixel contents were produced. Figure 3 shows some of the images to which the technique was applied. Results are shown in Table 1 and are compared to those reported in [21]. The accuracy was predominantly higher than [21], which we attribute to the largely overdetermined nature of the problem in the proposed approach. This of course introduces resiliency to random noise (assuming that random noise is zero-mean and has a well-behaved distribution).

As for random noise, it is worth noting that in many practical problems, dealing with data obtained from real imaging instrumentation, one may typically have the high-frequency portion of the phase-difference matrix cluttered and highly corrupted with noise. Figure 4 shows an example of such phase-difference matrix. This in practice does not introduce a major difficulty in our algorithm, since similar to [23, 25], we can apply our algorithm within a window in the low-frequency portion by cropping out the lower frequencies of the phase-difference matrix. The important remark to realize is that, we would then be counting the number of cycles within less than 2π. So for instance (12) (or other equations thereafter) would be adjusted accordingly.

Table 1: Results for global shifts of the images in Figure 3.

Image	True shifts	Foroosh et al. [21]	Proposed method
Pentagon	(0.50, −0.50)	(0.48, −0.51)	(0.496, −0.493)
	(0.25, 0.50)	(0.28, 0.49)	(0.255, 0.498)
	(−0.25, −0.50)	(−0.25, −0.52)	(−0.25, −0.52)
	(0.0, 0.75)	(0.0, 0.80)	(0.0, 0.744)
Paris	(0.167, −0.5)	(0.152, −0.49)	(0.163, −0.51)
	(0.67, 0.25)	(0.69, 0.33)	(0.679, 0.242)
	(−0.33, −0.167)	(−0.32, −0.15)	(−0.331, −0.159)
	(0.33, 0.33)	(0.325, 0.32)	(0.333, 0.329)

When the data is contaminated with random noise or aliasing, the high-frequency portion of the phase-difference matrix is not useful, and in fact can introduce error in the results. In order to avoid this problem, similar to [23, 25], we windowed the Fourier spectra of the images to their low-frequency portion. Therefore, in order to evaluate the performance of our approach under noise, we introduced additive Gaussian random noise with standard deviation in the range $[0, 5]$. We then computed the peak signal-to-noise ratio (PSNR) over 100 independent trials as we varied the size of the lowpass window, where we took the true shift value as the peak value of the signal. The results are shown in Figure 4 for $\sigma = 5$, and lowpass windows of width 10–100. It can be seen that the method is fairly stable over a wide range of window sizes. The SNR for the y coordinate was better in this experimentation, because the shift along that direction was an integer value.

We then applied the technique to real data in a framework using short-length Fourier transform. The images used are stereo pairs from the CMU data set [33] and the Tsukuba pair. Results are shown in Figures 5, 6, and 7 for the Pentagon, baseball, and the Tsukuba image pairs, respectively. Of course, one challenge in the case of real data is that due to lack of ground truth, the performance evaluation is

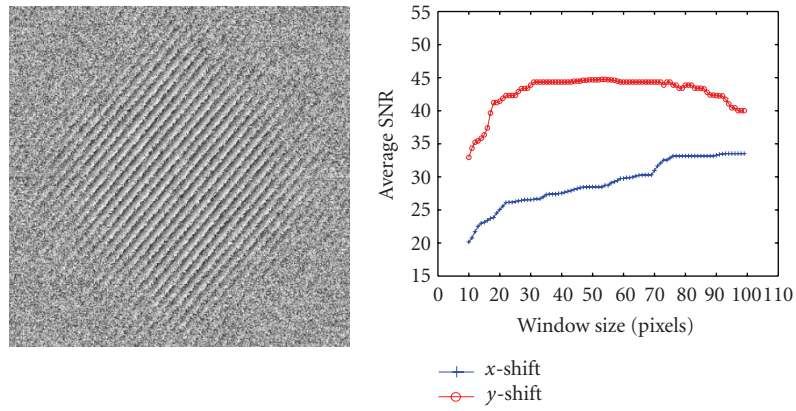

FIGURE 4: (a) A phase-difference matrix corrupted by aliasing and high-frequency noise, (b) average PSNR over 100 independent trials with varying window sizes and the standard deviation of noise $\sigma = 5$.

FIGURE 5: Pentagon stereo pair and the estimated local motion.

somewhat subjective. For our real data, however, it is important to note that we used small windows (e.g., of width 15–25) to compute the short-length Fourier transform. This has a very important implication, since unlike global registration, we now have a very small set of linear equations to estimate the parameters.

Essentially, in global registration, we estimate two parameters using a huge number of pixels (i.e., for a 256×256 image, t is typically a very large number). In the case of local motion estimation, if we for instance use window sizes of 15×15, t is typically reduced by a factor of 300. As a result, it is quite natural to assume that the accuracy will drop. It should be however noted that this comment equally applies to all existing methods in the literature. In order to be able to evaluate how the method performs for varying lengths of short-length Fourier transform, we synthetically generated a stereo pair with known ground truth (see Figure 8), where the parallax for the house in the image is artificial. We then computed the PSNR for varying range of window sizes of short-length Fourier transform. The plot in Figure 8 shows the result of this experimentation. One can notice that the PSNR reaches the highest value for an optimal window size. Our interpretation is that for local motion, if the window size is too small, then the number of data points will be insufficient to get the optimal result, and if it is too large it may get dominated by the motion in the neighboring pixels. This is similar to what is known to as the aperture problem in optical flow. There is in fact an optimal window size for which we get the best tradeoff between the two constraints.

Finally, we applied our method to estimate rotation in polar coordinates. For this purpose we used also a real stereo pair from the CMU data collection [33] for which the ground truth was known. Figure 9 shows the stereo pair, their polar coordinate representations, and the resulting phase-difference matrix. We estimated the rotation as 4.737 degrees, which nicely matched the ground truth. Note that the number of the sawtooth cycles can be used to determine the rotation even if only a small portion is useful (i.e., not badly contaminated with noise). Also, note that we can virtually get the correct solution regardless of number of steps used to sample the angular axis in the polar domain. For instance, in Figure 9 the steps used (as can be seen from the horizontal axis) are every one degree. But our solution is not limited to one degree accuracy due to subdegree (subpixel) accuracy provided by the proposed method. Therefore, for instance, we could sample the angular axis at every 2 degrees and still get sub-angle results as long as the aliasing can be avoided by lowpass windowing as discussed above.

Figure 6: Baseball stereo pair and the estimated local motion.

Figure 7: Tsukuba stereo pair and the estimated local motion.

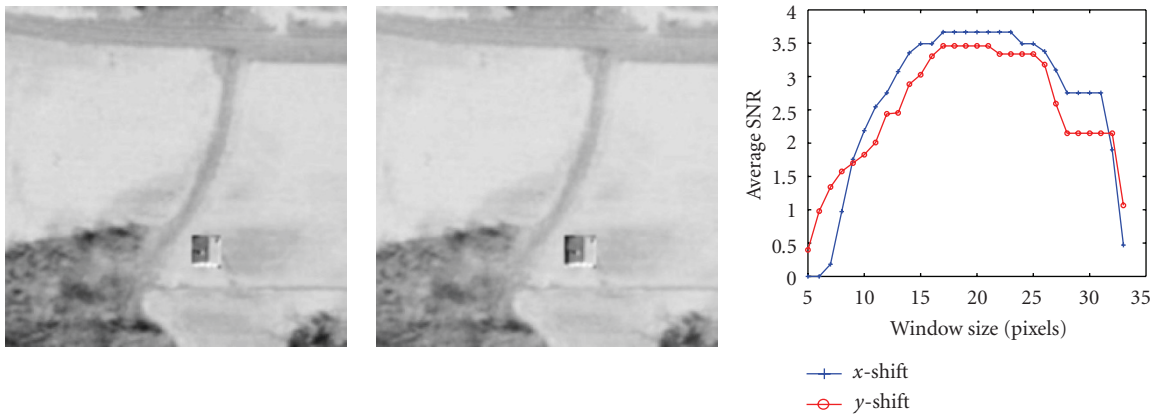

Figure 8: A synthetic stereo pair and the PSNR as a function of the window size used for short-length Fourier transform in local motion estimation.

7. CONCLUSION

The analyses and the experimentations presented in this paper show that accurate results can be obtained for subpixel registration directly in the Fourier domain by counting the number of cycles of the phase-difference matrix. Since no inverse transforming is required, the computational complexity is mostly dominated by the forward Fourier transform, which is $N \log N$. There are of course small overheads associated with computing the phase-difference matrix and its Hough transform. In the context of super-resolution, subpixel registration directly in the Fourier domain is particularly advantageous and of interest for the super-resolution algorithms that work directly in the Fourier domain. It is also worth mentioning that some sensor modalities naturally provide the Fourier spectrum of the field of view. Examples of such imaging modalities are magnetic resonance imaging (MRI) [34] and synthetic aperture radar (SAR) [35].

APPENDIX

In this appendix we show that the largest singular value of $\mathbf{K}\mathbf{K}^T$ is at least three times the second singular value. For this purpose we use the following results from linear matrix algebra.

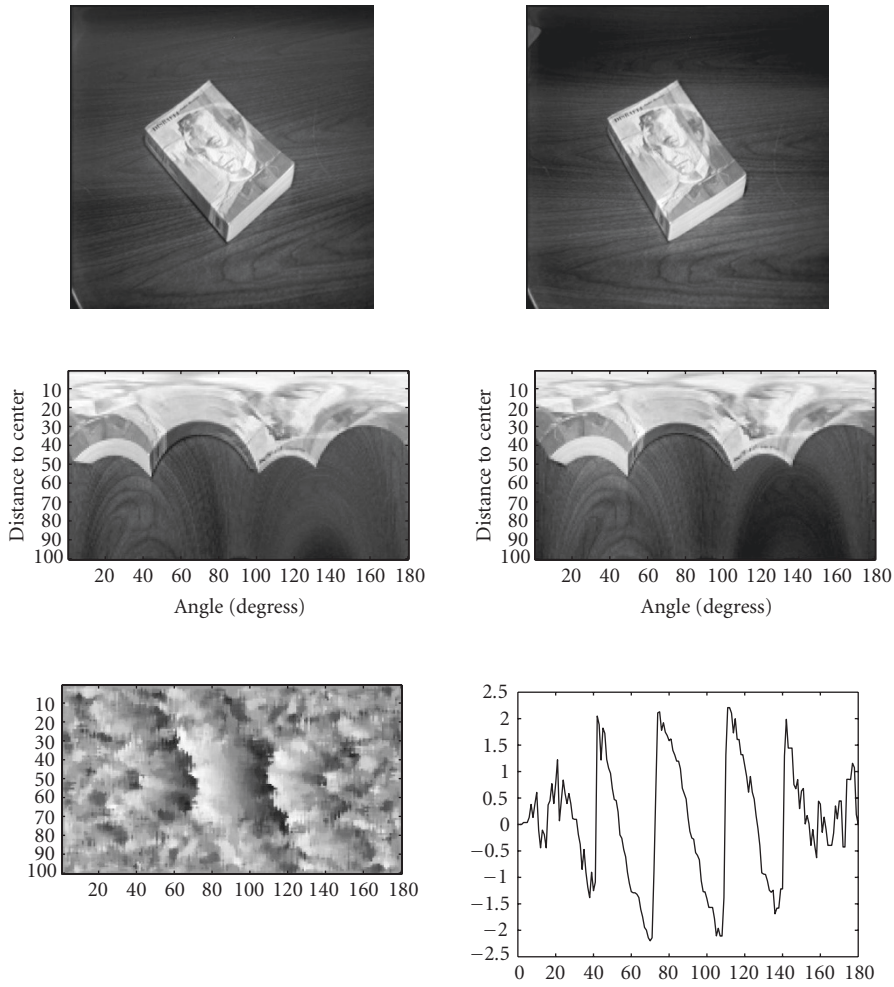

Figure 9: Top and middle: stereo pair and their corresponding polar representations; bottom: their phase-difference matrix and one row of the matrix.

(i) The nonzero singular values of $\mathbf{K}\mathbf{K}^T$ are equal to the singular values of $\mathbf{K}^T\mathbf{K}$.
(ii) The singular values of a real-square symmetric matrix are always real, see [36, Theorem 6.6.3].

It therefore follows that the two nonzero singular values of $\mathbf{K}\mathbf{K}^T$ (which are also the two largest singular values) are given by the solution of the characteristic polynomial of the 2×2 real-square symmetric matrix $\mathbf{K}^T\mathbf{K}$, that is,

$$\lambda^2 - \operatorname{tr}(\mathbf{K}^T\mathbf{K})\lambda + \det(\mathbf{K}^T\mathbf{K}) = 0. \tag{A.1}$$

The two solutions are given by

$$\begin{aligned}\lambda_1 &= \frac{1}{2}\operatorname{tr}(\mathbf{K}^T\mathbf{K})\left(1 + \sqrt{\left[\operatorname{tr}(\mathbf{K}^T\mathbf{K})\right]^2 - 4\det(\mathbf{K}^T\mathbf{K})}\right),\\ \lambda_2 &= \frac{1}{2}\operatorname{tr}(\mathbf{K}^T\mathbf{K})\left(1 - \sqrt{\left[\operatorname{tr}(\mathbf{K}^T\mathbf{K})\right]^2 - 4\det(\mathbf{K}^T\mathbf{K})}\right).\end{aligned} \tag{A.2}$$

Since $\mathbf{K}^T\mathbf{K}$ is a real-square symmetric matrix, both singular values have to be real. This implies that

$$\frac{\operatorname{tr}(\mathbf{K}^T\mathbf{K})^2}{\det(\mathbf{K}^T\mathbf{K})} \geq 4. \tag{A.3}$$

First-order Taylor series approximation of these singular values are given by

$$\begin{aligned}\lambda_1 &= \frac{\left[\operatorname{tr}(\mathbf{K}^T\mathbf{K})\right]^2 - \det(\mathbf{K}^T\mathbf{K})}{\operatorname{tr}(\mathbf{K}^T\mathbf{K})} + \mathcal{O}^2,\\ \lambda_2 &= \frac{\det(\mathbf{K}^T\mathbf{K})}{\operatorname{tr}(\mathbf{K}^T\mathbf{K})} + \mathcal{O}^2.\end{aligned} \tag{A.4}$$

The ratio of these two singular values is therefore given by

$$\frac{\lambda_1}{\lambda_2} = \frac{\left[\operatorname{tr}(\mathbf{K}^T\mathbf{K})\right]^2}{\det(\mathbf{K}^T\mathbf{K})} - 1. \tag{A.5}$$

Combining (A.3) and (A.5), we conclude that

$$\frac{\lambda_1}{\lambda_2} \geq 3. \tag{A.6}$$

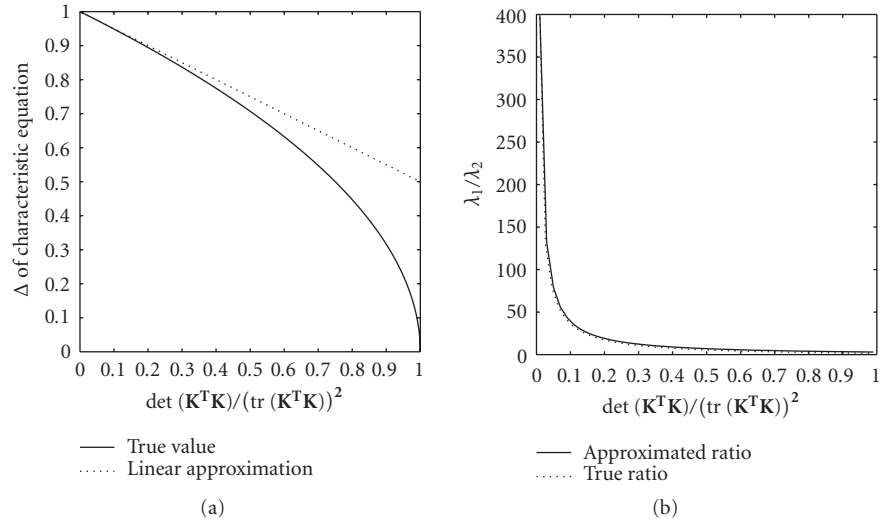

Figure 10: (a) Discriminant of the characteristic equation and its linear approximation, (b) ratios of the singular values.

A question that may be raised with regards to this demonstration is the error involved in the Taylor series truncations and its effect on the results in (A.6). These truncations are basically based on expanding the discriminant of the quadratic characteristic equation of the matrix $\mathbf{K}^T\mathbf{K}$. We now show that this truncation is practically unconsequential. Basically, although (A.4) have truncation errors, the error in their ratio is extremely negligible.

Instead of giving a formal proof of this claim, we give a graphical proof. Figure 10(a) shows the possible values of the discriminant of the characteristic equation and its linear approximation using first-order Taylor series, where the discriminant of the characteristic equation is plotted against $\det(\mathbf{K}^T\mathbf{K})/[\mathrm{tr}(\mathbf{K}^T\mathbf{K})]^2$ (i.e., the independent variable in the Taylor series expansion). Although the errors grow for small values of the discriminant, the error in the ratio of the two singular values is practically negligible as shown in Figure 10(b).

REFERENCES

[1] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[2] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[3] B. A. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[4] J. B. Abbiss, B. J. Brames, and M. A. Fiddy, "Super-resolution algorithms for a modified Hopfield neural network," *IEEE Transactions on Signal Processing*, vol. 39, no. 7, pp. 1516–1523, 1991.

[5] K. Aizawa, T. Komatsu, and T. Saito, "A scheme for acquiring very high resolution images using multiple cameras," in *Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, vol. 3, pp. 289–292, San Francisco, Calif, USA, March 1992.

[6] T. E. Boult and G. Wolberg, "Local image reconstruction and subpixel restoration algorithms," *CVGIP: Graphical Models and Image Processing*, vol. 55, no. 1, pp. 63–77, 1993.

[7] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, 1997.

[8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super-resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[9] H. (Shekarforoush) Foroosh, M. Berthod, J. Zerubia, and M. Werman, "Sub-pixel Bayesian estimation of albedo and height," *International Journal of Computer Vision*, vol. 19, no. 3, pp. 289–300, 1996.

[10] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

[11] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '88)*, pp. 742–746, Ann Arbor, Mich, USA, June 1988.

[12] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Transactions on Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.

[13] W. B. Lim, M. K. Park, and M. G. Kang, "Spatially adaptive regularized iterative high-resolution image reconstruction algorithm," in *Visual Communications and Image Processing (VCIP '01)*, vol. 4310 of *Proceedings of SPIE*, pp. 10–20, San Jose, Calif, USA, January 2001.

[14] M. K. Ng and A. M. Yip, "A fast MAP algorithm for high-resolution image reconstruction with multisensors," *Multidimensional Systems and Signal Processing*, vol. 12, no. 2, pp. 143–164, 2001.

[15] N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient super-resolution image reconstruction algorithm," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 573–583, 2001.

[16] A. J. Patti, M. I. Sezan, and A. Murat Tekalp, "Super-resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1064–1076, 1997.

[17] S. Peleg, D. Keren, and L. Schweitzer, "Improving image resolution using subpixel motion," *Pattern Recognition Letters*, vol. 5, no. 3, pp. 223–226, 1987.

[18] R. R. Schultz and R. L. Stevenson, "Improved definition video frame enhancement," in *Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95)*, vol. 4, pp. 2169–2172, Detroit, Mich, USA, May 1995.

[19] H. (Shekarforoush) Foroosh and R. Chellappa, "Data-driven multichannel super-resolution with application to video sequences," *Journal of Optical Society of America A*, vol. 16, no. 3, pp. 481–492, 1999.

[20] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1, chapter 7, pp. 317–339, JAI Press, Greenwich, Conn, USA, 1984.

[21] H. (Shekarforoush) Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.

[22] H. (Shekarforoush) Foroosh, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of cross power spectrum," in *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '96)*, pp. 532–537, San Francisco, Calif, USA, June 1996.

[23] W. S. Hoge, "A subspace identification extension to the phase correlation method [MRI application]," *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 277–280, 2003.

[24] S. P. Kim and W. Y. Su, "Subpixel accuracy image registration by spectrum cancellation," in *Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '93)*, vol. 5, pp. 153–156, Minneapolis, Minn, USA, April 1993.

[25] H. S. Stone, M. T. Orchard, E.-C. Chang, and S. A. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2235–2243, 2001.

[26] G. Wolberg and S. Zokai, "Robust image registration using log-polar transform," in *Proceeding of International Conference on Image Processing (ICIP '00)*, vol. 1, pp. 493–496, Vancouver, British Columbia, Canada, September 2000.

[27] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proceeding of IEEE International Conference on Cybernetics and Society*, pp. 163–165, New York, NY, USA, September 1975.

[28] E. De. Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 700–703, 1987.

[29] A. H. Zemanian, *Distribution Theory and Transform Analysis*, Dover, New York, NY, USA, 1965.

[30] A. V. Oppenheim and R. W. Schafer, *Discrete Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1st edition, 1989.

[31] A. Tikhonov and A. Arsenin, *Solutions of Ill-Posed Problems*, Winston & Sons, Washington, DC, USA, 1977.

[32] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[33] CMU, Vision and autonomous systems center's image database. http://vasc.ri.cmu.edu/idb/.

[34] G. A. Wright, "Magnetic resonance imaging," *IEEE Signal Processing Magazine*, vol. 14, no. 1, pp. 56–66, 1997.

[35] M. I. Skolnik, *Radar Handbook*, McGraw-Hill, New York, NY, USA, 2nd edition, 1990.

[36] M. Stein, *Introduction to Matrices and Determinants*, Wadsworth, Belmont, Calif, USA, 1967.

Murat Balci received the B.S. and M.S. degrees in computer engineering from METU, Turkey. He is currently a doctoral student in the School of Computer Science, University of Central Florida. His research interests include image processing, computer vision, computer graphics, and pattern recognition. He is an ACM Student Member and a Member of Upsilon Pi Epsilon Computing Sciences Honor Society.

Hassan Foroosh is an Assistant Professor of computer science at the University of Central Florida (UCF). Prior to joining the UCF, he was a Senior Research Scientist at the University of California, Berkeley (2000–2002), and prior to that an Assistant Research Scientist at the Center for Automation Research, University of Maryland, College Park (1997–2000). He received the M.S. and Ph.D. degrees in computer science, specializing in computer vision and image processing, from INRIA Sophia Antipolis in France in 1993 and 1996, respectively. He has authored and coauthored over 40 peer-reviewed journal and conference papers, and has been in the organizing and the technical committees of several international colloquia. He is a Senior Member of the IEEE and an Associate Editor of the IEEE Transactions on Image Processing. He has served on the committees of the IEEE International Conference on Image Processing, the IEEE Workshop on Image and Video Registration, the IEEE Workshop on Motion and Video Computing, the IEEE Workshop on Applications of Computer Vision, and the IEEE Workshop on Stereo and Multi-baseline Vision. In 2004, he was a recipient of an Academic Excellence Award from Sun MicroSystems and the Pierro Zamperoni Best Paper Award in the International Conference on Pattern Recognition.