A Fast Direct Fourier-Based Algorithm for Subpixel Registration of Images

Harold S. Stone, *Fellow, IEEE*, Michael T. Orchard, *Fellow, IEEE*, Ee-Chien Chang, and
Stephen A. Martucci, *Senior Member, IEEE*

Abstract—This paper presents a new direct Fourier-based algorithm for performing image-to-image registration to subpixel accuracy, where the image differences are restricted to translations and uniform changes of illumination. The algorithm detects the Fourier components that have become unreliable estimators of shift due to aliasing, and removes them from the shift-estimate computation. In the presence of aliasing, the average precision of the registration is a few hundredths of a pixel.

Experimental data presented here show that the new algorithm yields superior registration precision in the presence of aliasing when compared to several earlier methods and has comparable precision to the iterative method of Thévenaz *et al.* [21].

Index Terms—Aliasing, Fourier transform, image registration, spline interpolation, subpixel.

I. INTRODUCTION

IMAGE registration is an important preprocessing operation that aligns the pixels of one image to corresponding pixels of a second image. Registration is the primary tool for comparing two or more images to discover the differences in the images or to fuse multiple modalities to create a composite that reveals information not easily accessible within individual images. It is used in the remote sensing community to study satellite images of the earth, and in the medical community to enhance the diagnostic capability of radiological imagery.

Registration algorithms typically assume that images differ by some transformation from a given family, and they find the transform within that family that optimizes a particular criterion. Transformation families include rigid transforms (translation, rotation, and rescaling), linear and affine (skewed and perspective transforms), and nonlinear warping. Optimization criteria include minimizing the sum of squares of pixel differences, maximizing the normalized correlation coefficient and maximizing the mutual information of the joint pixel-distributions of two images. For this paper, we assume that two observed sampled images represent the same scene sampled on identical grids but offset from each other by an unknown translational shift, as well as differing by a uniform change of intensity, perhaps also disturbed by independent additive noise.

Townshend *et al.* [1] make a case for subpixel accuracy of registration in their studies of vegetation changes because the changes that they are trying to measure are on the order of the errors introduced by misregistration by less than a pixel. The literature contains image-to-image registration algorithms of various precision that operate in the pixel (signal) domain [2]–[22] and in the Fourier domain [2], [5], [23]–[36].

Among the authors who report subpixel registration precision, Shekarforoush *et al.* [34] describes an algorithm that uses the cross-power spectrum of the Fourier transforms (FTs) of two ideally bandlimited images. The Fourier inverse of the cross-power spectrum is a sinc function displaced from the origin by the amount of the translation. Abdou [2] describes three algorithms that use various interpolation schemes to find the translational difference between images. Kim and Su [30] present a Fourier-based algorithm that estimates translation changes by modifying the phase of one FT to make it as similar as possible to a second FT. The phase change corresponds to the translational difference. To eliminate aliasing effects their algorithm relies on the low frequency components of the transforms. Thévenaz *et al.* [21] report an elegant pixel-based iterative algorithm that is able to register to high precision, and can deal with rotation, translation, changes of scale, and illumination changes. Of the schemes mentioned here, it is the most general.

Algorithms reported to have subpixel image-to-image registration precision for ideally bandlimited images typically have reduced precision in the presence of aliasing. Kim and Su's algorithm [30] treats aliasing explicitly, but [2], [21], [34] do not.

The main result of this paper is the development of a direct algorithm for image-to-image registration that achieves high precision in the presence of small amounts of aliasing. This paper models the aliased frequency components of the two images and predicts how this aliasing affects the phase relationships between their FTs. Aliasing causes some frequency components of the scene to be unreliable, and the new algorithm masks them out of the registration process. The new algorithm resembles [30] because it eliminates certain frequency components from the calculations, but it eliminates more than do Kim and Su. For our experimental data, the frequency-masking algorithm achieves a worst-case registration precision of a few hundredths of a pixel, and an average registration precision of less than a hundredth of a pixel for a broad range of nonideal prefilters. The iterative algorithm of [21] achieves comparable precision experimentally, even though the algorithm does not deal with aliasing explicitly. All other algorithms compared in this paper have poorer registration precision.

Section II reviews theoretical issues related to aliasing and its effect on Fourier spectra, particularly on the phase of FTs. Section III describes the new algorithm, and how it deals with aliasing. The ground-truth model and experimental results appear in Section IV. A summary appears in the final section.

II. MATHEMATICAL BACKGROUND

The basic idea behind the new registration algorithm is that the phase of the Fourier spectra of an image pair contains sufficient information to determine the translation offset difference of the images. This section reviews how this can be exploited in the absence of aliasing, as described by [34], and then shows why aliasing causes problems that reduce the registration precision of this method.

Given a two-dimensional (2-D) image $f(x, y)$ and a translated version of the image $g(x, y) = f(x - x_0, y - y_0)$, we wish to find an efficient algorithm that gives the displacement vector (x_0, y_0). In the remainder of the discussion, we reduce the problem to one dimension and note that all the results generalize to two-dimensions straightforwardly.

Let $f_c(x)$ denote a continuous function with FT $F_c(\Omega)$. The FT of the shifted function $g_c(x) = f_c(x - x_0)$ is $G_c(\Omega) = F_c(\Omega)e^{-j\Omega x_0}$. This is the key relationship for all Fourier-based image-registration algorithms. In the ideal case of continuous transforms of noiseless images, for any Ω' the Fourier components $F_c(\Omega')$ and $G_c(\Omega')$ provide a perfect estimate of x_0, to within an integer multiple of $(2\pi/\Omega')$.

Now consider digital images, which are sampled versions of continuous images. Assume that $f_c(x)$ is continuous and periodic, and $F_c(\Omega) = 0$ for $|\Omega| > \Omega_0$. Sampled versions of this image are $f(n) = f_c(nT)$ and $g(n) = f_c(nT - x_0)$, where T is a sampling interval and x_0 is an arbitrary real-valued displacement with a magnitude less than $0.5T$. Let N be the number of samples per period of $f_c(x)$, so that $f_c(x) = f_c(x + NT)$.

In this formulation, the value of x_0 is the residual that remains after registering images to the nearest pixel by using any one of the several algorithms reviewed in [5]. Our goal is to find x_0. (Because $|x_0| \leq 0.5T$, we eliminate a phase ambiguity of $2\pi/\Omega_0$ in ratios of Fourier components in the algorithm developed later in this section.)

If the sampling interval T is less than π/Ω_0, then the Sampling Theorem [37] states that

$$f_c(x) = \sum_{n=-\infty}^{+\infty} f(n)\text{sinc}(x/T - n) \tag{1}$$

for any x and for integer values of n. Recall that $\text{sinc}(x) = \sin(\pi x)/(\pi x)$. The registration method of [34] uses discrete FTs of finite sampled images for which the reconstruction of (1) is approximated by the finite sum

$$g_c(x) = f_c(x - x_0) \approx$$
$$\sum_{n=0}^{N-1} f(n) \cdot \text{Dirichlet}_N\left((2\pi/N)((x - x_0)/T - n)\right). \tag{2}$$

The Dirichlet function is a periodic approximation to sinc, and is defined to be $\text{Dirichlet}_N(x) = \sin(Nx/2)/(N\sin(x/2))$ for x not a multiple of 2π, and is $\text{sign}(\cos(x))$ otherwise. It is the inverse discrete FT of a sampled window in the Fourier domain,
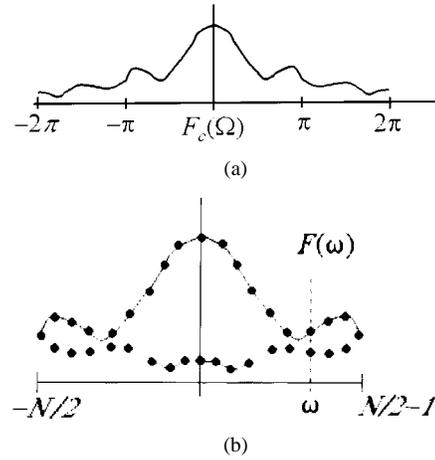

(a)

(b)

Fig. 1.   (a) Spectrum of image after prefiltering. (b) Downsampling and resulting aliasing of prefiltered image.

and is the function obtained from the algorithm of [34]. Equations (1) and (2) differ in their summation limits and in the substitution of $\text{Dirichlet}_N(2\pi x/N)$ for $\text{sinc}(x)$. [A reviewer noted that (1) converges with equality to (2) for N odd, but not for N even.]

The periodicity of $f(n)$ implies that its FT $F(\Omega)$ is discrete, which we represent as $F(\omega)$ for integers ω, $-N/2 \leq \omega \leq N/2 - 1$. Let $F(\omega)$, $G(\omega)$, and $D(\omega)$ be the respective discrete FTs of $f(n)$, $g(n)$, and $\text{Dirichlet}_N(2\pi n/N)$. Then the convolution in (2) corresponds to

$$G(\omega) = F(\omega)D(\omega). \tag{3}$$

This explains why the inverse FT of the spectral ratio $D(\omega) = G(\omega)/F(\omega)$, which is called the *cross-power spectrum*, is approximately a sinc function displaced from the origin by an amount x_0[34]. If $f_c(x)$ is sampled below its Nyquist rate then its continuous transform has energy in frequencies higher than Ω_0, in which case (2) is not guaranteed to be a good approximation because the perfect reconstruction of (1) no longer holds.

To estimate the phase of an aliased cross-power spectrum when $f_c(x)$ is sampled below its Nyquist rate, we assume that the optical prefilter is not an ideal low-pass filter, and that the spectral leakage is most likely to be in the frequencies just above the sampling cutoff frequency. Also, the attenuation increases strongly with frequencies higher than sampling cutoff, with essentially no energy at frequencies above twice the sampling cutoff. Fig. 1(a) illustrates this situation. The figure shows a Fourier spectrum plotted against normalized continuous frequency Ω, $-2\pi \leq \Omega < 2\pi$, which represents the spectrum in a one-dimensional image after prefiltering according to our assumptions. Sampling the image at one half the Nyquist rate with an interval of $2T$ produces the N spectral values shown in Fig. 1(b). The spectra for $F(\omega)$ and $G(\omega)$ in Fig. 1(b) are sums of phase-shifted components of frequencies $F_c(\Omega)$ in Fig. 1(a). For integer ω in the interval $-N/2 \leq \omega < 0$ we have

$$F(\omega) = F_c(2\pi\omega/N) + F_c(2\pi(\omega + N)/N)$$
$$G(\omega) = F_c(2\pi\omega/N)e^{-2\pi j\omega x_0/N}$$
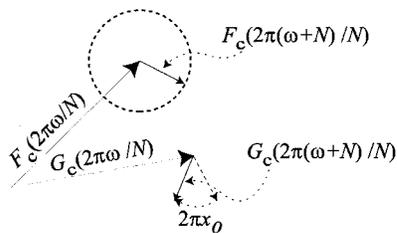$$\quad + F_c(2\pi(\omega + N)/N)e^{-2\pi j((\omega+N)x_0/N)}. \tag{4}$$

Fig. 2. Vector illustration of aliasing.

For ω in the interval $0 \leq \omega \leq N/2 - 1$, replace $(\omega + N)$ by $(\omega - N)$ in the second term of (4).

The aliasing terms in (4) create $F(\omega)$ and $G(\omega)$ such that the vector ratio $G(\omega)/F(\omega)$ is not equal to $e^{-2\pi j\omega x_0/N}$, as required for (1) and (2). A graphic illustration of this situation appears in Fig. 2. The complex vector $F(\omega)$ is shown to be the sum of two complex components of $F_c(\Omega)$. The complex vector $G(\omega)$ is the complex sum of those same components, one of which is rotated by $2\pi\omega x_0/N$, and the other is rotated by $2\pi(\omega + N)x_0/N$. The vector $G(\omega)$ is not a rotation of $F(\omega)$ by $2\pi\omega x_0/N$ because of the excess phase of $2\pi x_0$ in the second component of the sum. In general, the aliasing caused by the excess rotational phase leaves $G(\omega)$ with an amplitude different from that of $F(\omega)$ whereas bandlimited images experience no change of amplitude in the frequency domain after translation.

Our problem is to estimate the phase $2\pi\omega x_0/N$ as a function of ω from observations of the vector sums shown in Fig. 2. The components of F_c are not directly observable. The next section shows how we can eliminate badly aliased frequency components from further consideration, and thereby estimate x_0 from frequency components whose phases yield accurate estimates of x_0.

III. FREQUENCY MASKING SUBPIXEL SHIFT ESTIMATION

Examination of Fig. 2 shows that the observed relative phase of $G(\omega)/F(\omega)$ for a specific value of ω is likely to be a good estimate of x_0 if the magnitude of the alias component $F_c(2\pi(\omega + N)x_0/N)$ is small compared to the magnitude of the in-band component $F_c(2\pi\omega x_0/N)$. Fig. 1(b) shows that this is likely to occur under our assumptions at frequencies near the origin because of the attenuation of aliasing magnitude with increasing frequency by the prefilter. Hence, we should limit the frequency range to frequencies near the origin [30].

But this is not sufficient to attain high precision in registration, as indicated by the experimental data later in the paper. There usually exist highly aliased frequency components near the origin. Using these frequencies greatly reduces the precision of the estimate of x_0. The frequencies that are most likely to be corrupted are those for which the spectral magnitude is small. Therefore, the algorithm masks out contributions from spectral components whose magnitudes are small relative to the rest of the magnitudes, regardless of whether they occur at low or high frequencies.

The full algorithm for 2-D data is very simple and consists of these major steps.

1) Use any image registration algorithm to find a translation that registers the two images to the nearest integral pixel coordinates.

2) Apply a Blackman or Blackman–Harris window in the pixel domain to eliminate image-boundary effects in the Fourier domain [38].
3) Calculate the discrete FTs of $f(x, y)$ and $g(x, y)$.
4) Mask out spectral components that lie outside a radius of R from the central peak. A suitable value of R is $0.6N/2$ where N is the minimum of the number of samples in the x and y dimensions.
5) Mask out spectral components for which either $F(\omega, \nu)$ or $G(\omega, \nu)$ have magnitudes less than a specified threshold α.
6) Using the frequencies that remain after masking, find a least-squares estimate of (x_0, y_0).

The windowing operation is well known and eliminates the spurious introduction of high-frequency spectral energy due to edge effects. We found that a separable Blackman window (as well as a separable Blackman–Harris window) worked quite well [38]. We also tested a radially symmetric Blackman window and several other windows that are flatter than the Blackman window in the middle of the image. The radially symmetric window gave results comparable to the separable window, but is more complex to create. The flatter windows use more information from the center of the image, but they tend to be less effective in eliminating spurious high frequency energy from the edges of the images. The separable Blackman and Blackman–Harris windows yielded the best results for the least computation.

The use of radius $R = 0.6N/2$ constrains the frequencies to be close to the origin. Under the assumption that the underlying spectrum has most of its energy in the low frequencies as shown in Fig. 1(a), aliasing effects are smaller in this region as indicated in Fig. 1(b). We found empirically that the constant factor 0.6 can be as small as 0.5 or as large as 0.7 without materially affecting the algorithm.

The choice of threshold α warrants a brief discussion. The algorithm sorts the frequencies by magnitude and retains the K largest in the spectrum, for some value of K. In the absence of other information, a good way to choose K is to vary K over a range of values, and observe the estimated translation as a function of K. Experimental data in the next section show that there is a region where the estimated translation is virtually independent of K. The displacement estimate produced by K in this region is the one to use. This approach works very well, so we did not pursue other possible approaches.

To estimate translation displacements from the Fourier spectra, let $\mathcal{B}$ denote the set of frequency coordinate pairs (ω, ν) that survive the masking operations. Let $phase(\omega, \nu)$ be the phase of the complex ratio $G(\omega, \nu)/F(\omega, \nu)$ at point $(\omega, \nu) \in \mathcal{B}$. In the absence of aliasing, $phase(\omega, \nu)$ has x and y slopes equal to $2\pi x_0/N$ and $2\pi y_0/N$, respectively. The least squares estimate of the slope of a plane that passes through the origin is

$$x_0 = \left(\frac{N}{2\pi}\right) \left(\frac{\overline{\omega\nu}\,\overline{\nu\ \text{phase}} - \overline{\nu^2}\,\overline{\omega\ \text{phase}}}{\overline{\omega\nu}^2 - \overline{\omega^2}\,\overline{\nu^2}}\right)$$

$$y_0 = \left(\frac{N}{2\pi}\right) \left(\frac{\overline{\omega\nu}\,\overline{\omega\ \text{phase}} - \overline{\omega^2}\,\overline{\nu\ \text{phase}}}{\overline{\omega\nu}^2 - \overline{\omega^2}\,\overline{\nu^2}}\right)$$

where the quantities with overbars are the means of the respective products taken over all frequency pairs that survive the masking operation. Note that any uniform change of intensity of the form $p' = Ap + B$ for constants A and B has no effect on the registration process because A changes the magnitudes of the FT by a factor of A, leaving phase unchanged, whereas B changes only the spectral coefficient at the origin, leaving its phase unchanged.

The insensitivity to uniform changes of intensity is quite powerful. Artigas *et al.* [39] showed that by factoring out uniform changes of illumination, a set of two cloudfree AVHRR images of Long Island can be registered to all other cloudfree images of Long Island, (98 images) from the same five-month period, and thereby can be used as an image-search key for Long Island.

This completes the description of the registration algorithm.

IV. EXPERIMENTAL DATA

A. Ground Truth

In order to measure the precision of the registration algorithm, we prepared ground truth using a scheme described in [34]. The idea is to use a single high-resolution image to represent the actual scene, and to create an image pair from this scene by filtering and downsampling the high-resolution image in two ways. The downsampled images are shifted with respect to each other by integer amounts x_0 and y_0 in the high-resolution grid. After downsampling by a factor of M in each dimension, the relative shifts become fractional shifts of size x_0/M and y_0/M, respectively. One downsampled image optionally has its intensity values rescaled to new values by means of the formula $p' = Ap + B$ for fixed constants A and B. The rescaled intensities allow us to investigate the effect of such rescaling on the registration precision.

For the detailed analysis of the frequency-masking algorithm, we used 52 aerial photos of various urban and agricultural landscapes for this study. Each of the images is 1024×1024 pixels, and was downsampled by a factor of $M = 8$ in each dimension. Note that there are 64 distinct phases for a downsampled image depending on how it is shifted in x and y with respect to the origin before downsampling. Shifts of greater than eight pixels in either coordinate direction produce the same phase as a shift of 8 fewer pixels in the same direction. (The phases are equal, but the images are displaced by an integral number of low-resolution pixels.) Each of the two images was shifted by integer amounts ranging from -4 to $+3$ pixels in each dimension, thereby creating all 64 distinct phases of each image. Registering the two sets of 64 image-phases in all possible ways created 4096 registration tests for each image and over 200 000 registrations overall. The Blackman window was positioned at a fixed position, and the registration pairs varied over all possible phase-pairs relative to that fixed position. Our results show that the algorithm exhibits virtually no bias with respect to the position of the Blackman window relative to the images for our data. For the comparisons of the frequency-masking algorithm with other algorithms, we used a single aerial photo in 4096 relative pairs of shifts.

To control the amount of aliasing, we used a Gaussian filter prior to downsampling. The filter was characterized by a support area and a filter half-width σ. Since downsampling by eight
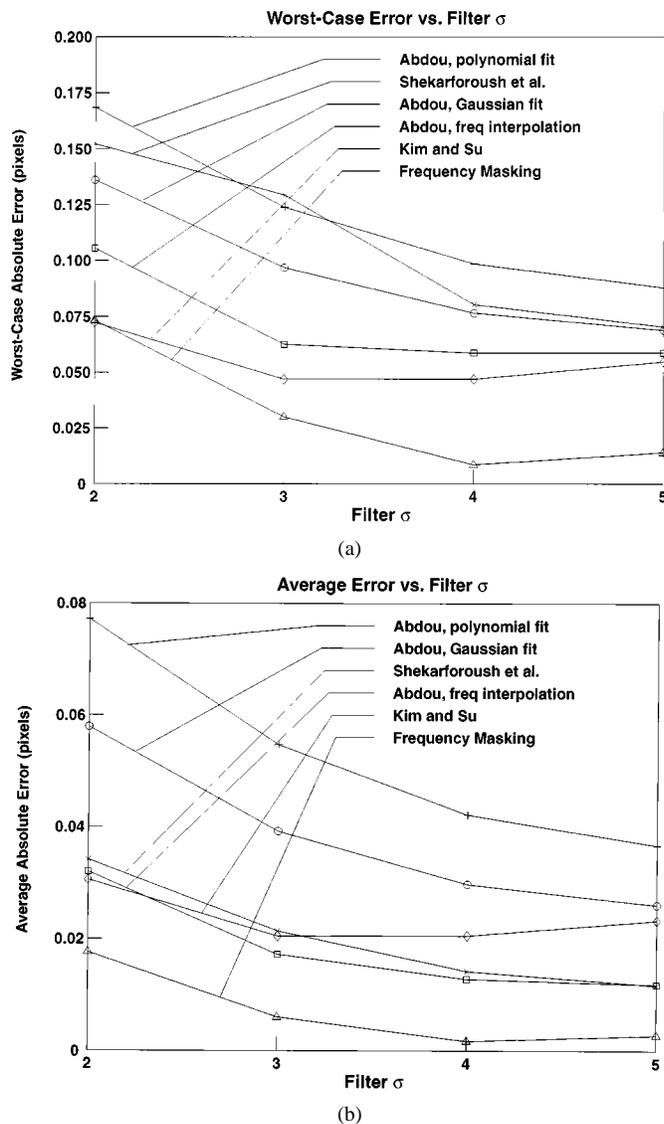

Fig. 3. (a) Comparison of worst-case error per coordinate for six direct subpixel registration algorithms. (b) Comparison of average error per coordinate for the same six algorithms.

compresses an 8×8 region of pixels into a single pixel, a bandlimiting filter must have a central peak at least 8×8 to keep aliasing small. Our experiments changed σ from 2 to 5 in steps of 1. The width of the central peak is approximately 2σ, so that $\sigma = 2$ produces substantial aliasing and $\sigma = 5$ produces a very small amount of aliasing. The support of the filter was 17×17 for the data reported here. We also explored other ranges of support and σ to confirm that the registration algorithm behaves as expected as we move outside the parameter region studied in detail, and found that the results were consistent with the data reported here.

B. Experimental Results

The first experiment is a comparison of the precision of our algorithm and other algorithms as a function of σ. For these experiments, we used a single image, and no change of illumination. These results appear in Fig. 3 for worst-case errors over both individual coordinates and average errors per coordinate direction. The mean-square error in two-dimensions is approximately $\sqrt{2}$

(a)

(b)

Fig. 4. Aerial photograph used in the comparisons. (a) Original figure. (b) Figure after applying frequency masking. (Courtesy of Positive Systems, Inc.)

times the average error per coordinate. Fig. 4 shows the image used to compare the algorithms. Fig. 4(b) is the same image after applying frequency masking to illustrate the low-pass behavior of the operation.

The algorithm of Kim and Su is most like ours because it restricts its attention to low frequencies in the Fourier domain. It seeks a phase for which the sum of squares differences between the spectrum of the first image and the phase-shifted spectrum of the second image is minimal. Our implementation used a Blackman window to eliminate spectral leakage, and affirmed that windowing made the registration robust with respect to boundary effects. The data in [30] were taken without a window,

although the paper did indicate that windowing should be considered. When we tried this algorithm without windowing, the results tended to be more accurate than with windowing, but were highly dependent on how the boundaries of the ground truth were prepared for the experiment. Windowing removes this dependence and removes artifacts of spectral leakage. The frequency region to which we restricted the algorithm had a size 89×89 for a frequency domain of size 128×128. We found experimentally that this size yields the most precise results for the experiment. Our implementation of the algorithm uses iterative hill-climbing to locate the optimal value of the Fourier phase. All of the algorithms except Kim and Su's produce higher registration precision as aliasing decreases.

The Shekarforoush algorithm estimates the displacements by fitting points to a sinc function, but when the points are corrupted by aliasing, the algorithm produces poor estimates. We found a slight instability in the implementation when the ratio $G(\omega)/F(\omega)$ became large because of a small value of $F(\omega)$, which we removed by artificially setting $F(\omega)$ to unity at this frequency.

The plot shows three different algorithms proposed by Abdou [2]. The first fits a polynomial curve through points that lie on a correlation peak. The second fits a Gaussian through those points, and the third linearly interpolates the spectra at integer offsets to find the closest approximation to the observed spectrum. The first two algorithms are direct algorithms. The third algorithm iterates a search over the interpolation coefficients. Because the third algorithm operates in the Fourier domain, we used a Blackman window on the images prior to registration. We did not window the data when applying the first two algorithms.

Abdou's [2] spectral interpolation algorithm required special treatment and some modifications. This algorithm computes the spectra of four copies of one image, each displaced relative to the other by displacements that lie on the corners of a unit square in the pixel-domain grid. The algorithm interpolates the cross products of each of these with the spectrum of the second image and finds the interpolation closest to the cross product of the spectra of the two images. Since the central peak dominates the spectral magnitudes, virtually the entire estimate of displacement is due to interpolation of the central peaks. To avoid this dependency, we zeroed out the central peaks in the spectra, which is equivalent to normalizing the images to the same average intensities. Nevertheless, the spectral magnitudes near the central peak are very large compared to magnitudes elsewhere in the frequency plane, so that the interpolation depends on relatively few of the frequency coefficients.

The iterative algorithm of [21] was studied somewhat differently from the other algorithms. The important aspect of this algorithm is that it drives toward a minimum sum of squared pixel differences between two images by performing a sequence of spline interpolations in the image domain. The spline interpolations are very close to sinc interpolations, and therefore they tend to affect the phase of the corresponding Fourier spectra without changing the magnitudes. Hence, the interpolations do not remove aliasing artifacts that may exist in one or both images as they interpolate one image into the other. For this reason, the iterations are unlikely to be able to drive the differences in the images to zero. In the absence of aliasing, it is clear that the

iterations can reduce the sum of squares to near zero, and the point at which this occurs corresponds to the subpixel translation difference of the images. In the presence of aliasing, it is not clear that the subpixel coordinates of the minimum sum of squared pixel differences are the same as those that minimize the subpixel translational difference of the images.

We studied this question both analytically and experimentally. The analysis shows that both the frequency masking and sum-of-squares of pixel differences lead to approximately equal solutions. To see this observe that in the pixel domain one wants to find Δx that minimizes $\sum_x (f(x) - g(x + \Delta x))^2$. By Parseval's Theorem, this is the same as finding Δx, which minimizes $\sum_\omega |F(\omega) - G(\omega)e^{j\omega\Delta x}|^2 = \sum_\omega |F(\omega)|^2 + \sum_\omega |G(\omega)|^2 - 2\sum |F_\omega| |G_\omega| \cos(\theta_\omega - \omega\Delta x)$ where θ_ω is the phase angle between $F(\omega)$ and $G(\omega)$. Hence, the Δx that minimizes the sum of squares of pixel differences maximizes $2\sum |F_\omega| |G_\omega| \cos(\theta_\omega - \omega\Delta x)$. Our algorithm selects a set of ω for which both $|F_\omega|$ and $|G_\omega|$ are large relative to other components. Because real images tend to have their energy concentrated in the lower frequencies, the algorithm tends to choose a set of frequencies that dominate the full sum. By selecting a Δx that maximizes $|F_\omega| |G_\omega| \cos(\theta_\omega - \omega\Delta x)$ for these components, the algorithm tends to maximize the sum over all components, and thus finds a Δx that tends to be one that minimizes the sum of squares of pixel differences.

Experiments confirmed that the minimum occurs at a registration point consistent with our ground truth. Hence, the iterative algorithm can achieve comparable precision if it can drive interpolations to the minimum sum of squares value. This was tested and confirmed by Thévenaz et al. [21], who graciously ran their code on a sample image pair supplied by the authors. The observed precision for an image pair filtered with σ of 3.0 was approximately 0.005 pixels, which is approximately the same as the frequency masking algorithm on the same image pair.

We also attempted to do registration by blind matching of the centers-of-gravity of an image pair. The process failed badly on satellite images. The center of gravity for such images is essentially unchanged by shifts when the distribution of pixel values is similar throughout an image. The center-of-gravity algorithm has been used successfully in the literature in contexts in which the center of gravity of specific regions or features drive the registration [40]–[43]. This requires region or feature identification, which is not required by "blind" algorithms like ours.

To check the robustness of our algorithm, we tested it on a suite of 52 images with a $\sigma = 3$. In this study, two of the 52 images were outliers, and had insufficient detail in their respective centers to give good worst-case registrations. The Blackman window weighs the center of the images very strongly, and for these images, not enough detail was left after windowing to yield very precise results at the subpixel level. The worst case registrations for these two images had errors that lay between 0.1 and 0.2 pixels in at least one dimension. The remaining 50 had worst-case errors that did not exceed 0.067 pixels in either dimension. The average error in each dimension was on the order of 0.0055 pixels. We repeated the experiment with uniform changes of intensity applied to one of the images, and ob-
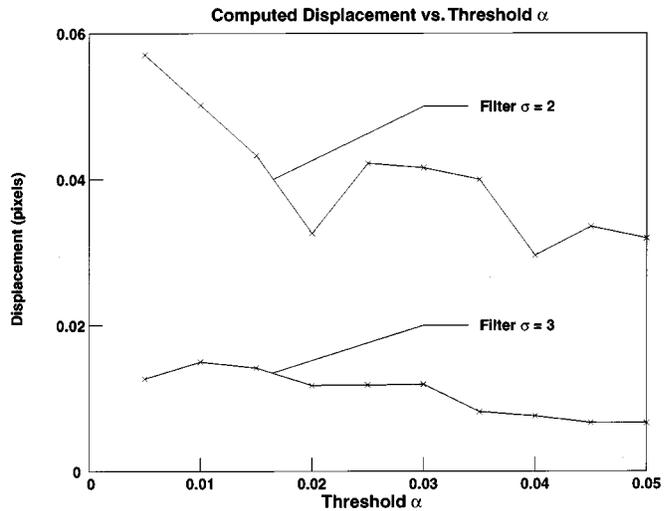

Fig. 5. Plot of displacement as a function of threshold variable α.
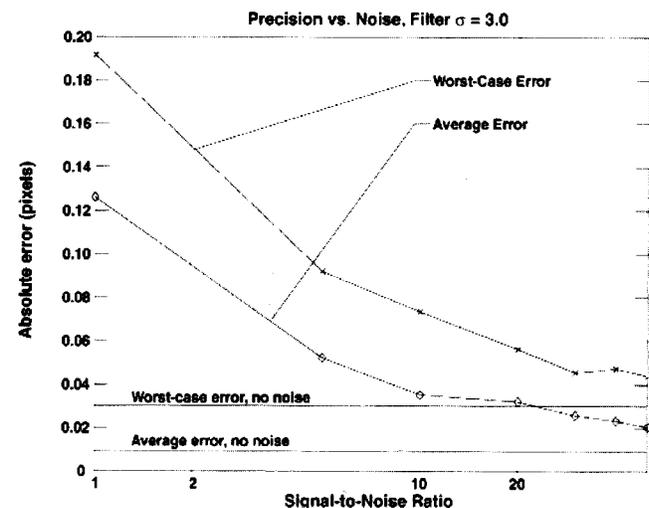

Fig. 6. Plot of displacement as a function of SNR for additive white Gaussian noise.

tained essentially the same results. Hence, the algorithm is insensitive to such changes of intensity, as the theory suggests it should be.

To deal with the question of how to set the α threshold, Fig. 5 plots the displacement estimate as a function of α threshold for the registration of a specific image for one offset with σs of 2 and 3. The figure verifies that there is a region of the threshold parameter space for which the displacement is almost independent of threshold. In this case, we thresholded magnitudes by eliminating frequency components whose magnitudes fell below $\alpha \cdot p_{\text{RMS}}$, where p_{RMS} is the RMS magnitude of this spectrum at frequencies that lie in a 5×5 region around the central peak (any threshold function that orders the magnitudes and accepts the K largest can be substituted here).

The algorithm also performs well in the presence of noise. Fig. 6 plots the precision of the algorithm for an image filtered with $\sigma = 3$ in the presence of additive white Gaussian noise for various SNRs. The horizontal lines in the plot indicate the precision in the absence of noise. The performance is excellent

TABLE I
COMPLEXITY OF REGISTRATION ALGORITHMS

Algorithm	Iterative/Direct	Complexity	Major Cost
Abdou. polynomial fit	Direct	$O(N \log N)$	Correlations via FFTs.
Abdou, Gaussian fit	Direct	$O(N \log N)$	Correlations via FFTs.
Abdou, Frequency interpolation	Iterative	$O(N \log N)$ once, $O(N)$ per iteration	Five FFTs once, five cross-spectral products per iteration.
Shekarforoush *et al.*	Direct	$O(N \log N)$	Two FFTs
Kim and Su	Iterative	$O(N \log N)$ once, $O(N)$ per iteration	Two FFTs once, several image interpolations per iteration.
Thévenaz *et al.*	Iterative	$O(N)$ per iteration	One or more image interpolations per iteration
Frequency Masking	Direct	$O(N \log N)$	2 FFTs and 1 sort.

for both the average and worst-case errors. This performance is comparable to the performance reported for the iterative algorithm in [21].

We also applied the algorithm to examples of multispectral images to measure its ability to register satellite images taken in different spectra. URLs of the sample images are [44]–[46]. To test the algorithm we registered the three color planes to each other in pairs. The expected result is that $\delta_{1,2} + \delta_{2,3} = \delta_{1,3}$ where $\delta_{i,j}$ is the measured displacement between color planes i and j. We found that this equation was satisfied to within a few hundredths of a pixel when the common features of all image planes were plentiful and accounted for most features present, and is not satisfied when the registration fails because of the absence of common features. The algorithm is sensitive to edges, and is not very sensitive to actual intensity values. However, mutlispectral images do not always share features across spectra. For images in which there were significant differences in features among the color planes, the algorithm failed to yield consistent pairwise registrations. We do not recommend the use of the algorithm in such cases.

C. Computational Complexity Comparisons

Table I contains a summary of the computational complexity of each of the algorithms studied as a function of N, the number of pixels in one image. The table also shows whether the algorithm is iterative or direct. For direct algorithms the complexity shown is total complexity. For iterative algorithms, the complexity given is the complexity per iteration.

FT complexity of $O(N \log N)$ dominates the cost for Abdou's [2] correlation-based algorithms, Shekarforoush's algorithm and our Frequency Masking algorithm. It also is present in the initialization of Abdou's [2] iterative algorithm and of Kim and Su's algorithm. Only the iterative algorithm of Thévenaz *et al.* [21] does not incur this directly.

Measures of relative time depend on the number of iterations and on constant factors, both of which are image and implementation dependent. An upper bound on the number of iterations for the algorithm of Thévenaz *et al.* [21] is $O(\log N)$ because the sum of N independently distributed random residuals has a variance that grows with N, and thus the number of bits in the sum that have to be eliminated tends to grow as $\log N$. The Marquardt–Levenberg strategy used in [21] has slow convergence at first, and then becomes quadratic near the end. In the region of linear convergence, the number of iterations required to reduce $\log N$ bits may grow as $\log N$. As convergence changes from linear to quadratic, the number of required iterations remaining falls very quickly, and is essentially a constant in practice. The complexity of the full registration algorithm thus may be as high as $O(N \log N)$ or as low as $O(N)$ depending on whether the number of iterations grows as $\log N$ or is constant. We do not have data on which to base fair estimates of the constant factors and relative running times for efficient implementations of the various algorithms.

V. SUMMARY AND CONCLUSIONS

The subpixel registration algorithm presented here enjoys very high precision in the presence of aliasing. It is conceptually simple to implement and is very efficient because its complexity is essentially that of two FTs. Also the algorithm uses only the frequencies in the central quadrant, and it is possible to speed the computation by producing only those frequencies when computing the FT. The new algorithm is competitive with the algorithm of [21] in registration precision. Relative computation times are difficult to compare because our algorithm is direct and the other algorithm is iterative.

Our algorithm can be adapted to deal with rotations and scale changes by using Fourier–Mellin invariants as described in [33], but to do so requires interpolations either in the Fourier or in the image domain. The algorithm becomes iterative in that setting rather than direct.

The new algorithm is useful in applications in which illumination and translational differences between images have to be discovered accurately and efficiently. It also may have applications to multispectral registration when image pairs from different spectra have many edge features in common.

REFERENCES

[1] J. R. G. Townshend, C. O. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 1054–1060, Sept. 1992.

[2] I. E. Abdou, "Practical approach to the registration of multiple video images," *Proc. SPIE*, vol. 3563, pp. 371–382, Jan. 1999.

[3] R. L. Allen, F. A. Kamangar, and E. M. Stokely, "Laplacian and orthogonal wavelet pyramid decompositions in coarse-to-fine registration," *IEEE Trans. Signal Processing*, vol. 41, pp. 3536–3541, Dec. 1993.

[4] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia, "Reconstruction of high resolution 3D visual information," in *Comput. Vis. Pattern Recognit.'94*, June 1994, pp. 654–657.

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

[6] M. Corvi and G. Nicchiotti, "Multiresolution image registration," in *Proc. 1995 IEEE Int. Conf. Image Processing*, 1995, pp. 224–227.

[7] I. J. Cox, M. L. Miller, S. W. Omohundro, and P. N. Yianilos, "Target testing and the PicHunter Bayesian multimedia retrieval system," in *Proc. Advances in Digital Libraries*, Washington, DC, May 1996, pp. 66–75.

[8] A. P. Cracknell and K. Paithoonwattanakij, "Pixel and sub-pixel accuracy in geometrical correction of AVHRR imagery," *Int. J. Remote Sensing*, vol. 10, no. 4–5, pp. 661–667, 1989.

[9] J. P. Djamdji, A. Bijaoui, and R. Maniere, "Geometrical registration of images: The multiresolution approach," *Photogramm. Eng. Remote Sensing J.*, vol. 59, no. 5, pp. 645–653, 1993.

[10] D. M. Etter and S. D. Stearns, "Adaptive estimation of time delays in sampled data systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 582–587, June 1981.

[11] C. M. Fan, N. M. Namazi, and P. B. Penafiel, "A new image motion estimation algorithm based on the EM technique," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 348–352, Mar. 1996.

[12] P. L. Feintuch, N. J. Bershad, and F. A. Reed, "Time delay estimation using the LMS adaptive filter—Dynamic behavior," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 571–576, June 1981.

[13] C. R. Guarino, "A novel method for two-dimensional phase estimation," in *Geoscience and Remote Sensing Symp., 1994*, vol. 4, Aug. 1994, pp. 2279–2281.

[14] W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inform. Technol.*, vol. IT-19, pp. 608–614, 1973.

[15] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Proc. Computer Vision and Pattern Recognition'88*, June 1988, pp. 742–746.

[16] J. Le Moigne, W. J. Campbell, and R. F. Cromp, "An automated parallel image registration technique of multiple source remote sensing data," *IEEE Trans. Geosci. Remote Sensing*, to be published.

[17] N. M. Namazi and J. A. Stuller, "A new approach to signal representation with the emphasis on variable time delay estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1649–1660, Dec. 1987.

[18] N. M. Namazi and B. Biswal, "Variable time-delay estimation in colored and correlated noise environment," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 597–604, Apr. 1992.

[19] F. A. Reed, P. L. Feintuch, and N. J. Bershad, "Time delay estimation using the LMS adaptive filter—Static behavior," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 561–571, June 1981.

[20] J. A. Stuller and N. Hubing, "New perspectives for maximum-likelihood time-delay estimation," *IEEE Trans. Signal Processing*, vol. 45, pp. 513–525, Mar. 1997.

[21] P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Processing*, vol. 7, pp. 27–41, Jan. 1998.

[22] Q. Zheng and R. Chellappa, "A computational vision approach to image registration," *IEEE Trans. Image Processing*, vol. 2, pp. 311–326, Mar. 1993.

[23] S. Alliney, "Digital analysis of rotated images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 499–504, May 1993.

[24] S. Alliney, G. Cortelazzo, and G. A. Mian, "On the registrations of an object translating on a static background," *Pattern Recognit.*, vol. 29, no. 1, pp. 131–141, Jan. 1996.

[25] S. Alliney and C. Morandi, "Digital image registration using projections," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 222–233, Mar. 1986.

[26] P. E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast-Fourier transform techniques," *IEEE Trans. Geosci. Electron.*, vol. GE-8, pp. 353–368, Oct. 1970.

[27] Q.-S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier–Mellin transforms for image registration and recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 1156–1168, Dec. 1994.

[28] B. Dasgupta and B. N. Chatterji, "Fourier Mellin transform based image matching algorithm," *J. Inst. Electron. Telecommun. Eng.*, vol. 42, no. 1, pp. 3–9, 1996.

[29] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 700–703, May 1987.

[30] S. P. Kim and W. Y. Su, "Subpixel accuracy image registration by spectrum cancellation," in *Proc. ICASSP '93*, vol. V, April 1993, pp. 153–156.

[31] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *IEEE 1975 Conf. Cybernetics and Society*, Sept. 1975, pp. 163–165.

[32] D. V. Papadimitriou and T. J. Dennis, "Stereo disparity using phase correlation," *Electron. Lett.*, vol. 30, no. 18, pp. 1475–1477, Sept. 1994.

[33] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Processing*, vol. 5, pp. 1266–1270, Aug. 1996.

[34] H. Shekarforoush, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of cross power spectrum," in *Comput. Vis. Pattern Recognit.'96*, June 1994, pp. 532–537.

[35] H. S. Stone, "Progressive wavelet correlation using Fourier methods," *IEEE Trans. Signal Processing*, vol. 47, pp. 97–107, Jan. 1999.

[36] J. W. Wong and E. L. Hall, "Scene matching with invariant moments," *Comput. Graph. Image Processing*, vol. 8, pp. 16–24, 1978.

[37] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[38] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[39] F. Artigas, R. Holowczak, S. A. Chun, J. Cho, and H. Stone, "An experimental study on content-based image classification for satellite image databases," *IEEE Trans. Geosci. Remote Sensing*, to be published.

[40] G. L. Gordon, "On the tracking of featureless objects with occlusion," in *Proc. Visual Motion Workshop*, Mar. 1989, pp. 13–20.

[41] E. D. Dickmanns, "Dynamic scene analysis and applications," in *1989 Multidimensional Signal Processing Workshop*, Sept. 1989, pp. 95–96.

[42] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 382–387, Mar. 1994.

[43] X. Dai and S. Khorram, "Development of a feature-based approach to automated image registration for multitemporal and multisensor remotely sensed imagery," in *Proc. Int. Geoscience and Remote Sensing Symp.'97*, Singapore, Aug. 1997, pp. 243–245.

[44] [Online]. Available: http://www.spaceimaging.com/carterra/images/losalamos_msi_fc.jpg.

[45] [Online]. Available: http://www.spaceimaging.com/carterra/images/hongkong_msi.jpg.

[46] [Online]. Available: http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/LAND_BIO/LANDSAT_FTP/Me_MtKatahdin_821208_234.jpg.

Harold S. Stone (S'61–M'63–SM'81–F'87) received the B.S.E.E. degree from Princeton University, Princeton, NJ, in 1960, and the Ph.D. degree in electrical engineering from the University of California, Berkeley (UCB), in 1963.

He is a Fellow at NEC Research Institute, Princeton, NJ, where he is engaged in research in image search and retrieval. He is the author and co-author of eight textbooks. He has been a Faculty Member at Stanford University, Stanford, CA, and the University of Massachusettes, Amherst, and has taught at UCB, Cornell University, Ithaca, New York University, and Princeton University. His work on the perfect shuffle interconnections is well known in computer architecture. He also collaborated on the IEEE floating-point standard now embedded in commercial microprocessors. He served as Associate Editor of the *Journal of the Association of Computing Machinery*, Technical Editor of *Computers*, and Associate Editor of the TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS.

Dr. Stone has been active in the IEEE Computer Society, IEEE Communications Society, and the ACM. He has been a Computer Society Governing Board Member, Vice President for Publications, and Technical Committee Chair. He is a Fellow of the ACM, and received the Charles Babbage Award in 1991, the IEEE Emanuel Piore Award for contributions to parallel computation and education, and the Taylor Booth Award for contributions to education.

Michael T. Orchard (F'00) was born in Shanghai, China, and grew up in New York, NY. He received the B.S. and M.S. degrees in electrical engineering from San Diego State University, San Diego, CA, in 1980 and 1986, respectively, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1988 and 1990.

He was with the Government Products Division of Scientific Atlanta, Atlanta, GA, from 1982 to 1986, developing passive sonar DSP applications, and he has consulted with the Visual Communications Department of AT&T Bell Laboratories, Florham Park, NJ, since 1988. From 1990 to 1995, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Illinois, Urbana, where he served as Associate Director of the Image Laboratory of the Beckman Institute. Since 1995, he has been an Associate Professor with the Department of Electrical Engineering, Princeton University. During the spring of 2000, he served as Texas Instruments Visiting Professor at Rice University, Houston, TX.

Dr. Orchard received the National Science Foundation Young Investigator Award in 1993, the Army Research Office Young Investigator Award in 1996, and was elected IEEE Fellow in 2000 for "contributions to the theory and development of image and video compression algorithms."

Stephen A. Martucci (S'86–M'93–SM'00) was born in New York state in 1960. He received the B.E.E. degree (highest honor), the M.S.E.E. degree, and the Ph.D. degree in electrical engineering, all from the Georgia Institute of Technology, Atlanta, GA, in 1982, 1987, and 1993, respectively.

He was with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung of the Technische Universitaet Hannover, Hannover, Germany, on German Academic Exchange Service (DAAD) scholarship from October 1988 to December 1989. He is currently a Research Engineer with NEC Research Institute, Princeton, NJ. Prior to that position, he worked at Front Porch Digital, Cherry Hill, NJ, NEC Electronics, Scitex Digital Video, and the Sarnoff Corporation. His research interests include digital video, discrete transforms, image and video compression, and compressed-domain processing. Some of his work can be found in the JPEG-LS and MPEG-4 international standards.

Dr. Martucci is a member of Sigma Xi, Phi Kappa Phi, and Eta Kappa Nu. He received the SPIE Young Investigator Award for best paper at VCIP'94.

Ee-Chien Chang received the B.Sc. and M.Sc. degrees from National University of Singapore, Singapore, in 1991 and 1993, respectively, and the Ph.D. degree from New York University in 1998.

In 1999, he was a Postdoctoral Fellow with the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), Rutgers University, New Brunswick, NJ, and NEC Research Institute, Princeton, NJ. Currently, he is an Assistant Professor in the Department of Computational Science, National University of Singapore.

His research interests include wavelet, digital watermarking, and image authentication.