An image registration technique for recovering rotation, scale and translation parameters

Morgan McGuire

Feb 19, 1998

M. McGuire is at the Massachusetts Institute of Technology, Cambridge MA. This work was performed at and funded by the NEC Research Institute, Princeton NJ. E-mail: morganm@mit.edu

Abstract

This paper presents a new method for obtaining a rotation and translation-invariant scale signature of an image, a scale and translation-invariant rotation signature, and a technique for recovering the rotation, translation and scale transformation parameters that relate two images of similar scenes using these signatures. The transformation parameters can be used to coregister the subbands of a composite image or to register time-series images for further analysis, common tasks in both satellite imagery and radiology research. In the finite discrete case where the invariances do not hold completely, compared to previous techniques the technique is extremely robust in the presence of occlusion due to alignment (image data moving in and out of the frame), resampling, linear and constant luminance changes, and noise. This robustness comes from the use of filters that minimize transformation artifacts in the Fourier-Mellin domain. The algorithm to implement the technique is $O((n_s N \log N)/4^k + Nk)$ where N is the area of the image in pixels, k is the order of the wavelet transform employed, and n_s is a small, image dependent integer factor that emerges during the scale phase of registration. A naive algorithm would take $O(N^3)$ operations to recover the same information. Previous algorithms commonly operate between these bounds and have lower peak correlation values than we report.

In selected experiments on actual natural photographs, medical images, aerial photographs, and remote sensor images, our algorithm successfully recovered the parameters of arbitrary Rotation-Scale-Translation transformations. In these trials, the rotation parameter varied by up to 31 degrees, the enlargement factor varied from 0.70 to 1.40, and translation parameters varied up to approximately 10% percent of the dimensions of the image. Actual registration problems tend to be less extreme than these trials. Across the experiments, the normalized correlation values for recovered transformation parameters were between 0.31 and 1.00 with a mean value of 0.76. The RMS error between corresponding pixels after registration was between 0.00 and 3.95 pixels with a mean value of 1.92.

Keywords

Image registration, Coregistration, Rotation invariance, Scale invariance, Translation invariance, Scale signature, Rotation signature, Correlation, Masked correlation, Fourier-Mellin,

Wavelet transform.

I. Introduction

Registration algorithms attempt to align a pattern image over a reference image so that pixels present in both images are in the same location. This process is useful in the alignment of an acquired image over a template, a time series of images of the same scene, or the separate bands of a composite image (coregistration). Two practical applications of this process are the alignment of radiology images in medical imaging and alignment of satellite images for environmental study.

For typical image registration problems, the sources of differences between two images fall into four categories:

1. **Differences of alignment** between images are caused by a spatial mapping from one image to the other. Typical mappings involve translation, rotation, warping, and scaling. For infinite continuous domain images, these differences are a result of a spatial mapping from one image to the other. Changing the orientation or parameters of the imaging sensor can cause differences of alignment.
2. **Differences from occlusion** occur when part of a finite image moves out of the image frame or new data enters the image frame of a finite image due to an alignment difference, when sensor errors produce identifiably invalid data in an image, or when an obstruction comes between the imaging sensor and the object being imaged. For example, in satellite images, clouds frequently occlude the earth.
3. **Differences from noise** occur from sampling error and background noise in the sensor, and from unidentifiably invalid data introduced by sensor error.
4. **Differences due to change** are actual differences between the objects or scenes

being imaged. In satellite images, lighting, erosion, construction, and deforestation are examples of differences due to change. It may be impossible to distinguish between change and noise.

Because images are registered in order to detect the changes in a scene, successful registration detects and undoes or accounts for differences due to alignment, occlusion, and noise while preserving difference due to change. Registration algorithms must assume that change is small with respect to the content of the image; that is, the images being registered are assumed to be "visibly similar" after accounting for differences due to alignment, occlusion, and noise. In addition, a sufficient amount of the object or scene must be visible in both images. We assume that at least 50% of the content of the reference image is also present in the pattern to be registered against it. In practice, medical and satellite sensors can usually be oriented with enough precision for images to share 90% or more of their content.

In this paper we present an algorithm for recovering transformation parameters from two images that differ by a Rotation-Scale-Translation (RST) transformation in the presence of noise and occlusion from alignment. Any RST transformation may be expressed as a combination of a single translation, single rotation, and single scale factor, all operating in the plane of the image. We express such a transformation as a pixel-mapping function, M, that maps a reference image into a pattern image. In practice these functions operate on finite images and can only account for data that does not leave or enter the frame of the image during transformation. If a two-dimensional infinite continuous reference image r and pattern image p are related by an RST transformation such that $p = M(r)$, then each point $r(x_r, y_r)$ in r maps to a corresponding point $p(x_p, y_p)$ according to the matrix

equation

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} s\cos\phi & -s\sin\phi & \Delta x \\ s\sin\phi & s\cos\phi & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix}. \tag{1}$$

Equivalently, for any pixel $p(x, y)$ it is true that:

$$r(x, y) = p(\Delta x + s * (x\cos\phi - y\sin\phi), \Delta y + s * (x\sin\phi + y\cos\phi)) \tag{2}$$

In this notation, ϕ, s, Δx, and Δy are the parameters of the transformation, where ϕ is the angle of rotation in a counter clockwise direction, s is the scale factor, and (Δx, Δy) is the translation. For finite discrete r and p, assume r and p are square with pixel area N (size $\sqrt{N} \times \sqrt{N}$). Note that an RST transformation of a finite image introduces differences due to occlusions as some data moves into or out of the image frame.

In trials on actual images, an implementation of our algorithm was able to accurately recover the rotation, scale, and translation parameters for a variety of RST transformations. A summary of selected trials is presented in Table 1. The images were selected from remote sensor data, photography, and medical imaging. For these trials, peak normalized correlations for rotation and translation varied from 0.58 to 1.00, with a mean of 0.87. Peak correlations for scale varied from 0.31 to 0.83 with a mean of 0.53. The largest registration error observed along any dimension was a 4-pixel RMS error across all image pixels for a 128 $\times$ 128 pixel image. From this we conclude that the algorithm is sufficient for solving RST registration problems with high accuracy and high signal-to-noise ratio. However, the mean peak scale-correlation of 0.53 is still significantly lower than the optimal 1.00. Fine tuning of filters used in the algorithm to a specific class of images might improve correlations and accuracy at the expense of generality.

The next section of this paper reviews the Fourier-Mellin invariant underlying our algorithm. Section III demonstrates the error introduced by using Fourier-Mellin techniques on finite images. Section IV presents our coregistration algorithm, scale signature, rotation signature, and filters for reducing the error observed in Section III. Section V presents an experimental analysis of the algorithm. Section VI discusses open research problems and makes some concluding remarks about our registration algorithm.

II. The Fourier-Mellin Invariant

The Fourier transform has certain properties under RST transformations that make it useful for registration problems. Let two two-dimensional infinite continuous images r and p obey the relationship given in Eq. (2). By the Fourier shift, scale, and rotation theorems, the relationship between F_r and F_p, the Fourier transforms of r and p, is given by:

$$F_r(\omega_x, \omega_y) = e^{j2\pi(\omega_x \Delta x + \omega_y \Delta y)/s} s^2 F_p((\omega_x \cos\phi + \omega_y \sin\phi)/s, (-\omega_x \sin\phi + \omega_y \cos\phi)/s) \quad (3)$$

Note that the complex magnitude of a Fourier transform, F_p, is s^2 times the magnitude of F_r and it is independent of Δx and Δy. Also the magnitude of F_p is derived from the magnitude of F_r by rotating F_r by $-\phi$ and shrinking its extent by a factor of s. This enables us to recover the parameters of rotation and scale through separate operations on the magnitude of F_p.

Equation (3) shows that rotating an image in the pixel domain by angle ϕ is equivalent to rotating the magnitude of its Fourier transform by $-\phi$. Expanding an image in the pixel domain by a scale factor of s is equivalent to shrinking the extent of the magnitude of its Fourier transform by s and multiplying the height (amplitude) of the magnitude of the Fourier transform by s^2. Translation in the pixel domain has no effect on the magnitude of the Fourier transform. Because of this invariance, the magnitude of a Fourier transform

is sometimes referred to as the *Fourier-Mellin* invariant, and the Fourier-magnitude space is referred to as the *Fourier-Mellin domain*. The Fourier-Mellin transforms of r and p are simply $R = |F_r|$ and $P = |F_p|$.

Many registration techniques [1], [2], [3], [4], [5] operate on the translation-invariant Fourier-Mellin space, then convert to polar-logarithmic coordinates so that rotation and scale effects appear as translational shifts along orthogonal θ and $\log \rho$ axes. In polar-log space, the normalized correlation coefficient of R and P as a function of shift along these axes is maximized at coordinate $(-\phi, -s)$. The 1D normalized correlation coefficient at shift j is given by:

$$\mathbf{C}(R,P)_j = \frac{\sum_{i=0}^{N-1} R(i+j)P(i) - \left(\frac{1}{N}\right)\left(\sum_{i=0}^{N-1} R(i+j)\right)\left(\sum_{i=0}^{N-1} P(i)\right)}{\sqrt{\left(\sum_{i=0}^{N-1} R(i+j)^2 - \left(\frac{1}{N}\right)\left(\sum_{i=0}^{N-1} R(i+j)\right)^2\right)\left(\sum_{i=0}^{M-1} P(i)^2 - \left(\frac{1}{N}\right)\left(\sum_{i=0}^{N-1} P(i)\right)^2\right)}}. \quad (4)$$

This extends simply to two dimensional polar-log space.

Equation (3) holds for infinite images but not for finite images. If it were true for finite images, it would cost $O(N \log N)$ operations to obtain the Fourier-Mellin polar-log coefficients, and $O(N \log N)$ operations to calculate the normalized correlation coefficient (by the Convolution theorem) for all cyclic shifts of the coefficients. Rotation and scale could thus be detected in $O(N \log N)$ time. Using discrete images instead of continuous ones causes some sampling error between the two images and in the calculation of the polar-log representation. In practice, using high-resolution images and inter-pixel interpolation can minimize these errors, so the theory holds for discrete finite images. Unfortunately, the theory does not hold for finite images for two reasons.

1. **Occlusion error.** Rotating, scaling, or translating a finite image causes some of the pixel data to move out of the image frame or some new pixel data to enter the frame.

2. **Tiling error.** The FFT of a finite image is taken by tiling the image infinitely in the plane. Rotation and scale do not commute with tiling.

If an image depicts a feature against a uniform and sufficiently large background, as in Fig. 1(a), only uniform background pixels enter and leave the image frame during transformation, so no data are lost. This is the case for some medical imaging tasks such as MRI, where the images under examination depict cross-sections of anatomy with a uniform background outside the anatomy. For images with nonuniform backgrounds or insufficient padding such as Fig. 2(a), transformations introduce occlusion error and the correlation peak may shift to a different location or suffer significant degradation.

As noted by Stone, Tao, and McGuire [6] for rotation and scale and Alliney and Morandi [7] for translation, the tiling error is unavoidable when taking the FFT of a tiled image, except for rotations that are an integer multiple of 90 degrees and for small translations of padded images. The Fourier transform of a discrete finite image contains a high frequency border between tiles that manifests itself in the Fourier-Mellin space as a high intensity "+" shape. (See Fig. 2(b)) This artifact is more significant than the coefficients from the remainder of the image content. DeCastro and Morandi [3] recommend using a rotationally symmetric image frame to avoid seeing this artifact in the Fourier-Mellin space. In Section IV of this paper, we present a more effective approach and present the results of an experiment comparing both methods to the the naive approach of using an unprocessed square image.

Despite all of the sources of error, the infinite and finite cases are related closely enough for Fourier-Mellin techniques to work successfully on finite images. However techniques reported in the literature have low peak correlations and low signal-to-noise ratio in the correlation function [1], [8]. In Section IV we derive a method that achieves near unity

peaks and a high signal-to-noise ratio in the correlation function, which together greatly improve the accuracy of registration.

III. Fourier-Mellin Techniques on Finite Images

In this section we show that the cumulative error from all sources can be small for idealized images with specific characteristics when using naive Fourier-Mellin techniques. However the error is large for nonideal images.

The first experiment examines a member of the class of ideal images that behave much like an infinite-image Fourier-Mellin space. The high contrast, simple images with uniform background shown in Fig. 1(a) are representative of the class of ideal images with specific characteristics. The second experiment examines a member of the class of realistic satellite images in Fourier-Mellin space after the same transformations. The low-contrast, detailed images that fill the image frames shown in Fig. 2(a) are typical of satellite and some medical imaging and are representative of the class of realistic, nonideal images for which the error is nonnegligible.

Figure 1(a) shows four images arranged horizontally: a 64×64 idealized finite discrete reference image and rotated, scaled, and translated versions of that image. To minimize the error introduced by rotating and scaling discrete images, linear interpolation was used in calculating the pixel values for the rotated and scaled images. Figure 1(b) shows a contour plot of the nonzero Fourier-Mellin coefficients of the four images, where the effects of the transformations can be observed. Note that for this image, rotation by angle ϕ in the pixel domain is approximately equivalent to rotation by $-\phi$ in the Fourier-Mellin domain. Expanding by s in the pixel domain is approximately equivalent to shrinking by s and multiplying the amplitude by s^2 in the Fourier-Mellin domain. Translation has no

effect.

By inspection, it is clear that the Fourier-Mellin transforms in Fig. 1(b) exhibit properties that are close to those of transforms of infinite images to recover rotation and scale parameters. Figure 2 depicts the results of the same experiment performed on an actual satellite image. Figure 2(a) shows the satellite image in the same alignments as the simple image in Fig. 1(a). Figure 2(b) shows the corresponding Fourier-Mellin transforms. Note that the "+" shaped artifact obscures the movement of the coefficients that are rotated or scaled. The Fourier-Mellin transform of the translated image is not identical to the transform of the original image, so the Fourier-Mellin domain is *not* invariant to translation.

When the pixel-level transformation, M, is a combination of rotation, scaling, and translation, the differences from the ideal (infinite case) and the actual (finite case) coefficients in Fourier-Mellin space are cumulative, hence the poor performance of Fourier-Mellin based techniques in the literature.

IV. A New Registration Technique

Our registration technique operates in three phases, one for each transformation parameter: scale, rotation, and translation vector. The scale and rotation phases are independent and may be performed in parallel. The translation phase can proceed only after the rotation and scale are known. The following discussion describes how to use filters, transformations and correlations to recover the transformation parameters for infinite continuous images. The reasons why the theory does not hold for finite discrete images are presented together with methods that produce reasonable approximations. Section V presents experimental results that confirm show our technique has high accuracy and a

high signal-to-noise ratio.

A. Scale Signatures and Scale Phase of Registration

The goal of the scale phase is to recover the scale factor, s, in Eq. (2). To recover this parameter in the presence of a translation and rotation, we map the images into a translation and rotation invariant space. In this new space we compute a *scale signature* for the image. It changes with scale and is invariant to rotation or translation for finite images. For the finite discrete case, our experiments show that the scale signature is resistant to translational and rotational changes, and can be used to effectively recover the scale factor s.

As for previous approaches [1], [2], [3], [4], our mapping transform S relies fundamentally on the Fourier-Mellin translation-invariance property. In a log-polar mapping of the Fourier-Mellin domain, pixel scale and rotation operations act along orthogonal axes. Integrating along the θ axis and dividing by the product of the radius and the total image energy produces a one-dimensional function, $S_r(\log \rho)$, that is invariant to image rotation. This one-dimensional function is the scale signature. For infinite images, the scale signature has the property that expanding or scaling the reference image by a factor of s is equivalent to translating S_r by $-\log s$. Because we perform normalized correlations on the scale signature, the normalization factor is irrelevant, as it does not affect the location of correlation peaks. Ignoring the normalization, the transformation from $r(x, y)$ to $S_r(\log \rho)$ is given by:

$$R(\rho\cos\theta, \rho\sin\theta) = R(\omega_x, \omega_y) = \left| \int\int G(r(x,y)) e^{-j2\pi(\omega_x x + \omega_y y)} dxdy \right| \tag{5}$$

$$S(r) = S_r(\log\rho) = \frac{1}{\rho} \int_0^{\pi} R(\rho\cos\theta, \rho\sin\theta) d\theta \tag{6}$$

where G is a nonlinear filter used to correct for artifacts in the finite case. For the infinite case, G is the identity filter. Note that the Fourier transform of a real image has a 180^o rotational symmetry, so the integration bounds can be reduced by a factor of 2. Because any RST transformation in the pixel domain is equivalent to a translation (shift) in the S domain by $-\log s$, the normalized (noncircular) 1D correlation coefficient between the scale signatures of two images will be 1.0 at $-\log s$. This means that a simple 1D correlation can be used to recover the scale parameter.

For discrete images, the continuous function S_r is a discrete function whose coefficients represent a summation with respect to θ over interpolated points in log-polar Fourier-Mellin space. (Alliney [4] gives a direct Cartesian-image to polar-Fourier transform that can be used to minimize the error introduced by interpolation.) In the finite case, the S domain shift resulting from a scaling operation in the image domain may move coefficients outside of the original extent of S_r. The scale operation may also move some image pixels in or out of the original image frame, changing the overall amplitude of the Fourier transform as well as the individual components. The discontinuities produced by the implicit titling of the image during the finite Fourier transform add a high-amplitude scale-invariant "+" shaped feature to the Fourier-Mellin coefficients, which will be more significant than the scaling of the image-specific coefficients and make it difficult to observe any scale change.

Although it is not always possible to recover the scale factor by examining the scale signatures of two finite discrete images, in practice an overall shift is often visible in the scale signatures if G is chosen to minimize the artifacts introduced by the finite transform. The scale phase of our registration technique, depicted in the left half of the system diagram given by Figure 5, uses this observation to attempt to recover the scale parameter.

The steps preceding the correlation implement the filtered S transform. W is a wavelet

decomposition used to reduce the computation bounds. The normalized correlation followed by the peak detector returns a list of S domain shifts for which the normalized correlation coefficient is a local maximum. By raising the inverse of the logarithm base used in the polar-log mapping to the power of each of these shifts, we obtain a vector of scale factors. Each of these scale factors is a likely value for the scale factor relating r and p. To determine the actual scale factor, recover the rotation parameter as described later in this section, and undo the rotation on the images, then choose the scale factor for which the translation correlation is maximized. In cases where an upper or lower bound is known on the scale factor, all candidates outside that range may be eliminated, reducing the computations required.

Assuming G is chosen such that it is $\mathrm{O}(N \log N)$, the Fourier-Mellin transform and polar-logarithmic mapping requires $\mathrm{O}(N \log N)$ computations. The projection onto the $\log \rho$ axis (integration by θ) requires $N - \sqrt{N}$ additions. Thus the total computational cost for obtaining the scale signatures S_r and S_p is $\mathrm{O}(N \log N)$. The 1D normalized correlation and peak detection can be performed in an additional $\mathrm{O}(\sqrt{N} \log N)$ operations. The entire computation can be increased in speed by a factor of 4^k by performing a wavelet decomposition of order k on r and p and using the low-low (thumbnail) subbands as the input to the Fourier-Mellin transform. The low-low subband of a k^{th} order wavelet decomposition is essentially the original image reduced in resolution by a factor of k along each dimension [9]. Because scaling transformations are independent of resolution, using a wavelet decomposition does not substantially affect the results of registration so long as k is not so large that a majority of the features in the images fall beneath the sampling resolution. By incorporating the use of wavelets, the total cost of the scale phase of registration is $\mathrm{O}((N \log N)/4^k)$.

B. Rotation Signatures and Rotation Phase of Registration

The goal of the rotation phase is to recover the rotation parameter, ϕ, which relates r and p in Eq. (2). Just as scale signatures can be used to recover scale information in a rotation and translation-invariant domain, a similar approach can be employed for rotation. Let Φ be the transform function such that $\Phi_r = \Phi(r)$ is the translation and scale invariant rotation signature of infinite continuous r. For finite discrete images, our experiments show that Φ is resistant to translation and scaling of the original image, although overall changes in magnitude may occur.

Integrating polar Fourier-Mellin coefficients along the ρ axis produces a one-dimensional function of angle that is translation and scale invariant. Although Φ is not normalized for total image energy, this will not affect the normalized correlation. $\Phi(r)$ is given by:

$$\Phi(r) = \Phi_r(\theta) = \int J(R(\rho\cos\theta, \rho\sin\theta))d\rho \tag{7}$$

where J is a pointwise weighting function to be applied in the finite case when it is known that there is no scale factor ($s = 1$). Assume J is 1 otherwise.

For the infinite continuous case, a rotation by angle ϕ in the image domain corresponds to a shift by $-\phi$ in the Φ domain. This means the normalized circular correlation of Φ_r and Φ_p will be unity at $-\phi$. Note that rotation signatures, like Fourier coefficients, have 180 degree rotational symmetry ($\Phi_r(\theta) = \Phi_r(\theta + \alpha\pi)$ where α is an integer). Thus, from the rotation signatures it is possible to determine the rotation parameter modulo 180^o. If it is not known whether the rotation is larger than $\pm 90^o$, it is necessary to perform a single pixel-level inverse rotation and image correlation to eliminate one of the choices. For rotationally symmetric images, the rotation signature is uniform and it is impossible to recover rotation.

The discussion of scale signatures for discrete and finite images holds for rotation signatures as well. The largest discrepancy observed between the infinite continuous and finite discrete cases is usually the "+" shaped artifact in the Fourier-Mellin coefficients. An appropriate choice of G reduces this artifact. In addition to the pixel-domain filter with G, we use J to weight Fourier-Mellin coefficients from edges in the image texture more strongly. Unlike the scale signatures, the correlations of rotation signatures frequently achieve a narrow peak with value near unity near the actual rotation parameter, and exhibit high signal-to-noise ratio. Because of this, it is often possible to determine the rotation parameter accurately modulo 180^o using only a single peak of the normalized correlation of the rotation signatures.

The Fourier-Mellin transform for the rotation phase of registration is identical to the one used for the scale phase, so the $\mathrm{O}(N \log N)$ computation need only be performed once. For rotation, the J weighting factor requires N additional multiplications if used. Projection onto the ρ axis requires $N - \sqrt{N}$ additions. Thus the rotation signatures can be obtained in $\mathrm{O}(N \log N)$ operations. The 1D circular correlation costs an additional $\mathrm{O}(\sqrt{N} \log N)$ operations. As with scale, rotation can be detected at any reasonable resolution of the image, so that the k^{th}-order wavelet transform of the input images allows recovery of the rotation parameter in $\mathrm{O}((N \log N)/4^k)$ operations. If the same filtered Fourier-Mellin-Wavelet transforms are used for the scale and rotation phases, the rotation phase requires only $\mathrm{O}(N/4^k)$ independent operations after performing the scale phase.

C. Translation Phase of Registration

The goal of the translation phase is to determine the translation parameters Δx and Δy, given the rotation and scale parameters. If there are data that are known to be invalid

in the images, these occluded areas can be ignored during the translation phase at no additional cost by using masks m and h for r and p respectively. These masks contain value 1 where the corresponding image pixel is valid and 0 where the corresponding pixel is invalid. We will allow these masks to take on fractional values produced by interpolation and wavelet decomposition. We derive the case where the scale factor is known and then show how it is extended to the case where there are a number of possible scale factors. The latter case matches the actual output of the scale phase of registration.

Let r and p be infinite continuous images. Let p' be the image p after undoing the effects of rotation and scale. Let h' be the mask h after undoing the effects of rotation and scale. If the ϕ and s are accurately recovered then p' differs from r by only a translational shift. For the finite case, p' and r may also differ by occlusions introduced when data moves into or out of the image frame. These occlusions are automatically taken into account by h'.

Stone [10] shows how to compute the normalized correlation coefficient in the Fourier domain using binary validity masks for the case where the extent of the pattern mask is smaller than the extent of the image mask. The normalized correlation coefficient as a function of translation of the pattern will have a near unity value at the shift that aligns the images properly. McGuire and Stone [11] extend this technique to use fractional validity masks for reduced resolution representations of images. The normalized circular correlation coefficient for all translational shifts of images r and p', with validity masks m and h' is:

$$\mathbf{C}\left(r,p'\right)_{x,y} = \frac{\left(\tilde{\mathbf{r}}\odot\tilde{\mathbf{p}'}\right)_{x,y} - \left(\frac{1}{\left(\mathbf{m}\odot\mathbf{h}'\right)_{x,y}}\right)\left(\tilde{\mathbf{r}}\odot\mathbf{h}'\right)_{x,y}\left(\mathbf{m}\odot\tilde{\mathbf{p}'}\right)_{x,y}}{\sqrt{\left(\left(\widetilde{\mathbf{r}^{(2)}}\odot\mathbf{h}'\right)_{x,y} - \left(\frac{1}{\left(\mathbf{m}\odot\mathbf{h}'\right)_{x,y}}\right)\left(\tilde{\mathbf{r}}\odot\mathbf{h}'\right)^2_{x,y}\right)\left(\left(\mathbf{m}\odot\widetilde{\mathbf{p}'^{(2)}}\right)_{x,y} - \left(\frac{1}{\left(\mathbf{m}\odot\mathbf{h}\right)_{x,y}}\right)\left(\mathbf{m}\odot\tilde{\mathbf{p}'}\right)^2_{x,y}\right)}} \tag{8}$$

where $\odot$ denotes circular correlation, $\widetilde{x} = r. * m$, and $\widetilde{r^{(2)}} = r. * r. * m$, (where $.*$ is point-by-point array multiplication). By the Convolution Theorem, this process can be efficiently computed in $O(N \log N)$ time for finite images.

Note that for finite images, Eq. (8) assumes that the pattern-mask nonzero extent is less than one half of the image mask size along any one dimension. If this assumption is violated, cyclic shifts are no longer equivalent to translations and the computation cannot be performed in the Fourier domain. Computing the correlations in the pixel domain costs $O(N^2)$ operations. To avoid incurring this cost when the images are of similar size, the correlations can be computed for versions of the images and masks that are padded to double their size in each dimension. Even with this padding, Fourier techniques are usually faster than pixel techniques on unpadded images.

The translation phase of our registration technique uses the recovered parameters ϕ and s to undo the rotation and scale on the pattern, producing p' and h'. The normalized correlation coefficient for all translations is then computed by Eq. (8). The location of this peak is directly related to the translation difference between r and p. If the peak occurs at a shift of (Dx, Dy), then the translation parameters for Eq. (2) are:

$$\begin{bmatrix} \Delta x \\ \Delta y \\ 1 \end{bmatrix} \begin{bmatrix} s\cos\phi & -s\sin\phi & 0 \\ s\sin\phi & s\cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} Dx \\ Dy \\ 1 \end{bmatrix} \tag{9}$$

Some care must be taken in handling the translations for which only a small portion of p' overlaps r, as these translations have very few pixels participating in the correlation, and can produce misleading results.

The translation phase requires $O(N)$ operations to undo the scale and rotation and $O(N \log N)$ operations to perform the normalized correlation. We recommend performing

registration using a k^{th} order wavelet to determine a coarse estimate of translation. A k^{th} order wavelet decomposition resolves the registration point to within $\pm 2^{k-1}$ pixels along any dimension. Given this coarse estimate, a refined search for the exact translation can be performed at the full resolution over the 4^k possible fine translations that map to the same coarse translation. Under the assumption that the correlation function is smooth near the peak value, this can be reduced to examining only $2k$ translations with a binary search algorithm. This property is image dependent, and generally holds for images that display continuity. Using this procedure, the total cost for the translation phase is $\mathrm{O}((N \log N)/4^k + Nk)$.

The above discussion considered the case where the scale factor is known. The scale phase of the algorithm does not produce a single scale factor, however. It produces a list of likely scale factors. To determine which of the scale factors is correct, the translation phase is repeated for each of the scale factors, and the scale factor that produces the highest translation correlation peak is selected. The fine resolution search need only be applied to the scale factor that produces the highest peak in the coarse resolution search. If n_s is the number of scale factors returned by the scale phase of registration, then the true cost of the translation phase is $\mathrm{O}((n_s N \log N)/4^k + Nk)$. The total cost of recovering rotation, scale and translation is $\mathrm{O}((n_s N \log N)/4^k + Nk)$.

D. Choosing Filters G, H and J

Filter G is used to minimize the effects of the implicit tiling of finite images when the Fourier transform is computed for the rotation and scale phases of registration. Figure 3(a) shows an unfiltered image. Figure 4(a) shows the nonzero coefficients of the Fourier-Mellin transform of this image. Note the "+" shaped artifact superimposed over the central

coefficients which actually represent the texture of the image.

DeCastro and Morandi [3] recommend using a rotationally symmetric image frame to avoid seeing this artifact. A simple experiment confirms that this is effective. Figure 3(b) shows an image after a rotationally symmetric image frame has been applied. In this case G is a mask shaped like a disk, which zeros out pixels outside a certain radius. Figure 4(b) shows that the Fourier-Mellin transform of this filtered image does not bear the "+" shaped artifact. However, remaining coefficients do not exactly match the ones from the original image after removing the "+" from the transform. Comparing Figs. 3(a) and 3(b), note that Fig. 3(b) has a large circular edge and approximately 25% fewer pixels. Correspondingly, there is some radially distributed noise and a general change of shape in Fig. 4(b), compared to Fig. 4(a). This reduces the correlation peak and may affect accuracy, but it is superior to the unfiltered version because the "+" artifact is removed.

We recommend using a filter that blurs the borders of an image against the opposite borders. Very few pixels need be altered by this filter. Figure 3(c) shows an image treated by such a filter. Note that within a few pixels of each border there is no effect, so that a majority of pixels are unchanged. Figure 4(c) shows the nonzero Fourier-mellin coefficients of this image. Like the round mask, this filter successfully removes the "+" artifact. Note that Figure 4c is almost identical to Figure 4a without the "+". From this we conclude that a filter G that blurs edges cyclically around an image is excellent for removing the "+" artifact without significantly altering the Fourier-Mellin coefficients.

The H filter is used during the scale phase to increase the signal to noise ratio of the scale signatures before correlation. The content of a finite image may change significantly under a scale transformation, so H is necessary to stabilize the scale signature. A filter that removes the first-order trend from the scale signature by subtracting a moving average

achieves this goal.

Filter J is used to selectively weight the Fourier-Mellin coefficients that correspond to edges in the image. It is best to apply J only if there is no scale factor because expansion in the pixel domain shrinks the Fourier-Mellin coefficients, and they will be weighted improperly. In a polar representation of the Fourier-Mellin transform, the coefficients that correspond to image texture edges lie near the center of the ρ axis, parallel to the θ axis. The coefficients where ρ is small encode the DC offset and low frequencies in the image, and the coefficients where ρ is large represent very high frequency information and noise. A gaussian weighting that emphasizes the edge coefficients while minimizing the DC and noise components will suppress the remains of the "+" artifact, intensity shifts in the image, and noise while emphasising the edges of features. J may be tuned to a specific class of images to enhance performance, or eliminated for the general case of arbitrary images with arbitrary scale transformations.

V. Experimental results

Table 1 shows the results of six selected experiments. The first column lists the name of the file used. The "suburban" image is an aerial photograph of a suburban area at approximately 4 meter resolution. The "fish" image is a photograph of a school of fish. The "mammogram" image is a mammogram. The "essai" image is a Landsat thematic mapper image of Washington state. The ϕ, s, and $(\Delta x, \Delta y)$ columns display the actual transformation parameter, the parameter predicted by the algorithm ("Recovered"), and the value of the peak correlation corresponding to that parameter. The choice of the image and values for the actual parameters are the input to a trial, the recovered parameters are the output and the peak correlation is a measure of how well the parameters account for

the differences in the reference and pattern images. Table 2 reports the root mean square (RMS) error along any one dimension for each of the trials listed in Table 1. The 1D RMS error is an estimate of the distance between a pixel transformed using the actual and recovered parameters. A good result has recovered parameters that accurately predict the actual parameters, peak correlation values near unity, and RMS error near zero.

For each trial, a reference and pattern image were created from a common source image. The center 256 $\times$ 256 pixel subimage from the 768 $\times$ 768 pixel source image was cropped then subsampled by a factor of 2 to form the reference image. The pattern image was created by transforming the source image by the actual parameters, then cropping the center 256 $\times$ 256 image and subsampling by a factor of two. For moderate transformation parameters, this process creates test images that accurately represent the registration problem because they have the following properties:

1. Data that move out of the image frame after transformation are lost (occluded). Other data that move into the image frame correspond to parts of the image that were previously occluded by the frame, and are **not** uniform.
2. An enlarged image has approximately the same frequency range as the original image.

These properties can be observed in images used in real-world registration. If these properties are not maintained, the registration problem is significantly easier, so it is important to model them in the test images.

Trials 1-3 demonstrate the performance of the algorithm on rotation-only, scale-only, and translation-only transformations. Table 1 shows that the algorithm was able to recover the transformation parameters within reasonable tolerance. For trials 1 and 2, the recovered parameters were exactly equal to the actual parameters. In trial 2, the scale-only case, there was a slight error in the estimated parameters. This error corresponds to a 2.164-

pixel RMS displacement as shown in Table 2. Throughout all testing, estimation of the scale factor was the most difficult and error prone phase of the algorithm. Unfortunately, the algorithm uses the recovered scale factor to recover the translation parameters, so an error in the scale factor usually resulted in an error in the translation as well. For trials 1-3, the rotation and translation parameters were recovered with high correlation values. The scale correlation values were significantly lower than desired.

Trials 4-6 are selected RST transformations where all parameters vary, and are applied to images from different disciplines. These are relatively extreme with respect to the differences that are actually observed in practice. The algorithm recovered the transformation parameters with very small error in all cases. As with trials 1-3, the rotation and translation parameters were recovered with high correlations, and the scale correlation was lower than desired.

VI. Summary and Conclusions

The new technique performs well on the difficult case of realistic image pairs and extreme transformations. Separating the detection of rotation, scale, and translation reduces the size of the search space. The use of progressive wavelet search and the calculation of correlations in the Fourier domain make all of the computations very efficient. During testing, the time to produce test image pairs was longer than the time the algorithm used to register them. By introducing pixel and Fourier-Mellin filters, we reduce the artifacts of finite discrete Fourier transforms and achieve accurate registration.

We suggest that superior results can be obtained by tuning the G, H, and J filters to images from a specific discipline. As shown in trials 1-3 from Tables 1 and 2, the algorithm performs exceptionally if scale differences can be removed from images prior to

registration.

In addition to RST transformations, the algorithm should perform well for constant or linear luminance differences between the reference and pattern images. These luminance changes are constant linear functions in the Fourier-Mellin domain, so they are ignored by the normalized correlation function. It is well known that the value of the normalized pixel correlations used to recover translation is invariant to linear and constant luminance changes.

The use of realistic image pairs for testing is very important. If images are not modeled as cropped subsampled portions of infinite frequency, infinite size signals, the registration problem is dramatically simpler. For unrealistic test cases, rotation and translation can be recovered by observing the movement of the edge between the image texture and the background padding. Scale can be recovered by observing the radial shift of frequencies in the Fourier-Mellin domain.

The correlation of scale signatures generally fails to produce a single optimal scale factor. In practice, we find it is necessary to check 1 to 5 peaks to identify the scale factor. It is an open challenge to identify a robust technique which can achieve high correlations and accurately identify a single, optimal scale factor.

Fourier-Mellin approaches appear limited to the case of RST transformations on mostly unoccluded images. Because the value of a single pixel in the pixel domain affects all of the Fourier-Mellin coefficients, it is improbable that a Fourier-Mellin technique like the one we have presented will emerge that allows selective consideration of separate pixels. The Fourier-Mellin invariances do not extend to arbitrary deformations, so it is also unlikely that these techniques will be extended to handle non-RST transformations. New techniques must be found that can achieve comparable results under arbitrary deformations.

Acknowledgments

The author is indebted to Harold Stone of the NEC Institute for his continual advice and editing of this paper. In addition the author would like to thank Bo Tao of Princeton University for sharing his insights into the properties of the Fourier-Mellin domain, which led to this research. The South Africa satellite images in this paper were supplied by Jacqueline Le Moigne from CESDIS.

References

[1] B. S. Reddy and B. N. Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. on Image Processing*, vol. 3, no. 8, pp. 1266–1270, August 1996.

[2] D. J. Lee, S. Mitra, and T. F. Krile, "Analysis of sequential complex images, using feature extraction and two-dimensional cepstrum techniques," *J. Opt. Soc. Am. A*, vol. 6, no. 6, pp. 863–871, June 1989.

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

[4] S. Alliney, "Spatial registration of multispectral and multitemporal digital imagery using fast-fourier transform techniques," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 5, pp. 499–504, May 1993.

[5] Q.-S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1156–1168, December 1994.

[6] H. S. Stone, B. Tao, and M. McGuire, "A note on translation, rotation, and scale-invariant image registration," submitted to *IEEE Trans. on Image Processing*, 1997.

[7] S. Alliney and C. Morandi, "Digital image registration using projections," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 2, pp. 222–233, March 1986.

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

[9] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Box 812060, Wellesley, MA USA, 1996.

[10] H. S. Stone, "Progressive wavelet correlation using fourier methods," submitted to *IEEE Transactions on Signal Processing*, 1997.

[11] M. McGuire and H. S. Stone, "Techniques for multi-resolution image registration in the presence of occlusions," in *Proceedings of the 1997 Image Registration Workshop*, J. Le Moigne, Ed., November 1997, pp. 101–122.

Table 1
Selected Results from Registration Problems

(Trial) Image		ϕ (degrees)	s	$(\Delta x, \Delta y)$ (pixels)
(1) suburban	Actual	31.00	1.00	(0.00, 0.00)
	Recovered	31.00	1.00	(0.00, 0.00)
	Peak Corr.	0.90	0.48	0.97
(2) suburban	Actual	0.00	1.40	(0.00, 0.00)
	Recovered	0.25	1.37	(0.00, 0.00)
	Peak Corr.	0.86	0.31	0.84
(3) suburban	Actual	0.00	1.00	(10.00, -3.00)
	Recovered	0.00	1.00	(10.00, -3.00)
	Peak Corr.	0.91	0.83	1.00
(4) fish	Actual	24.00	0.70	(7.00, 12.00)
	Recovered	25.00	0.72	(7.06, 11.99)
	Peak Corr.	0.78	0.45	0.95
(5) mammogram	Actual	-10.00	1.20	(10.00, 9.00)
	Recovered	-10.25	1.15	(9.76, 8.75)
	Peak Corr.	0.94	0.73	0.58
(6) essai	Actual	-15.00	1.50	(-10.00, -10.00)
	Recovered	-15.75	1.56	(-11.33, -11.40)
	Peak Corr.	0.94	0.40	0.78

Table 2 RMS Error from Registration Problems

(Trial) Image	RMS Error in Pixels
(1) suburban	0.000
(2) suburban	2.164
(3) suburban	0.000
(4) fish	1.501
(5) mammogram	3.947
(6) essai	3.923

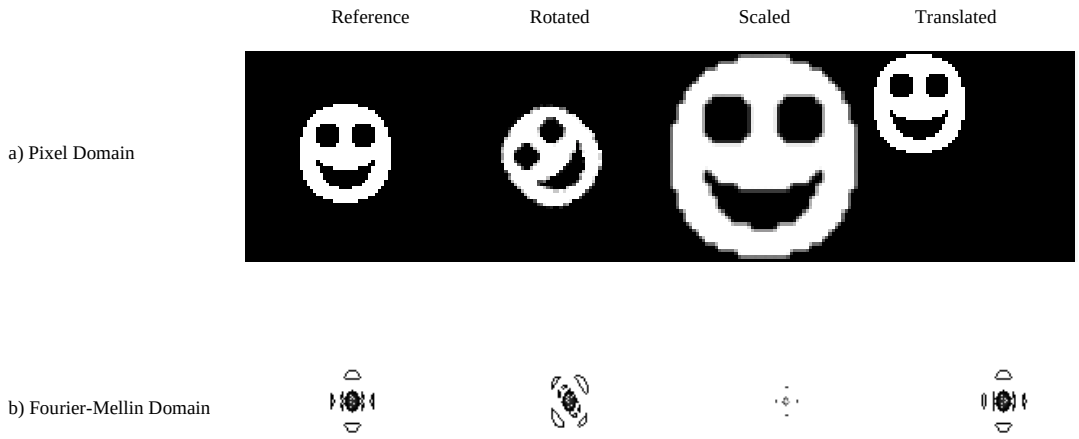

Figure 1.
a) 64x64 Idealized Reference image, Reference rotated by 45º, Reference scaled by 2.00, Reference translated by (-16, -16) **b)** Contour plot of magnitude of FFT's of corresponding images Linear interpolation was used in generation of the Rotated and Scaled images.

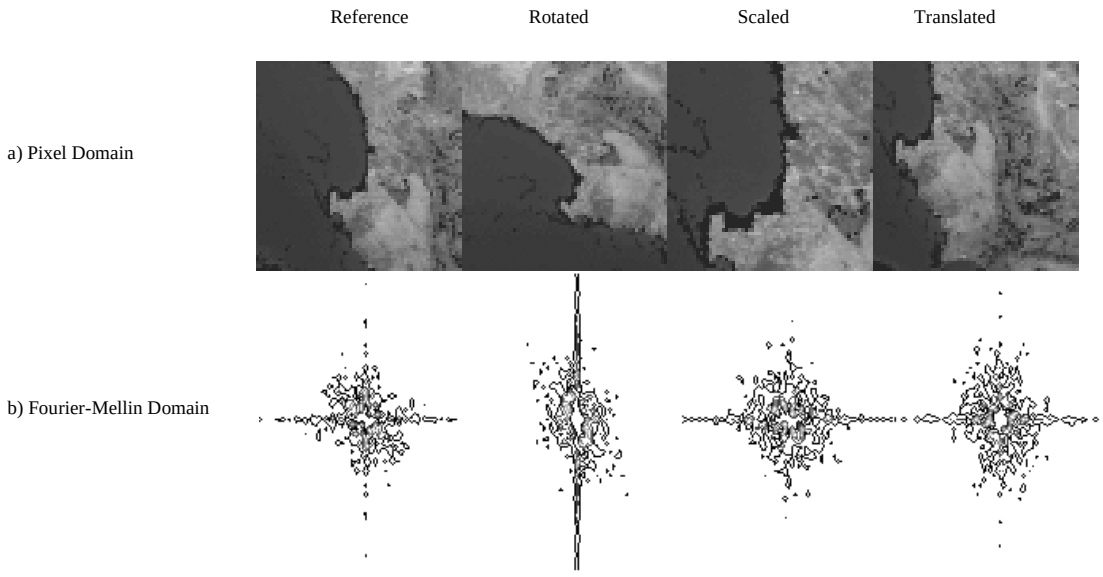

Figure 2.
a) 64x64 Reference image (thumbnail of actual satellite image of the coast of South Africa), Reference rotated by 45º, Reference scaled by 2.00, Reference translated by (-16, -16) **b)** Contour plot of magnitude of FFT's of corresponding images. For Rotated and Scaled versions of image, transformations were applied at 2x resolution and thumbnails were constructed.

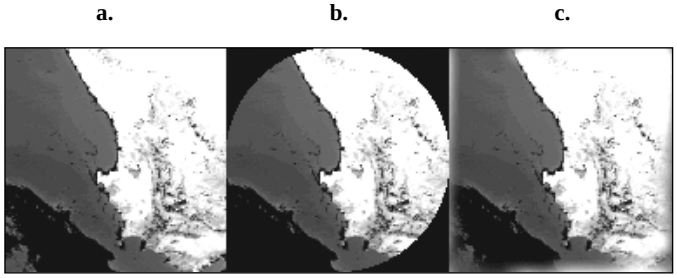

Figure 3.
a) No filter **b)** Round mask **c)** Cyclically blurred borders. Black borders are added for viewing purposes only-- they are not pixel data.

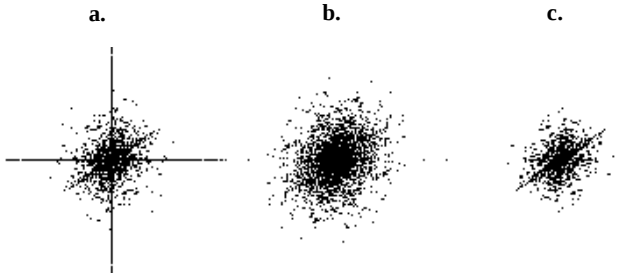

Figure 4.
Non-zero Fourier-Mellin coefficients of the images from Figure 3.

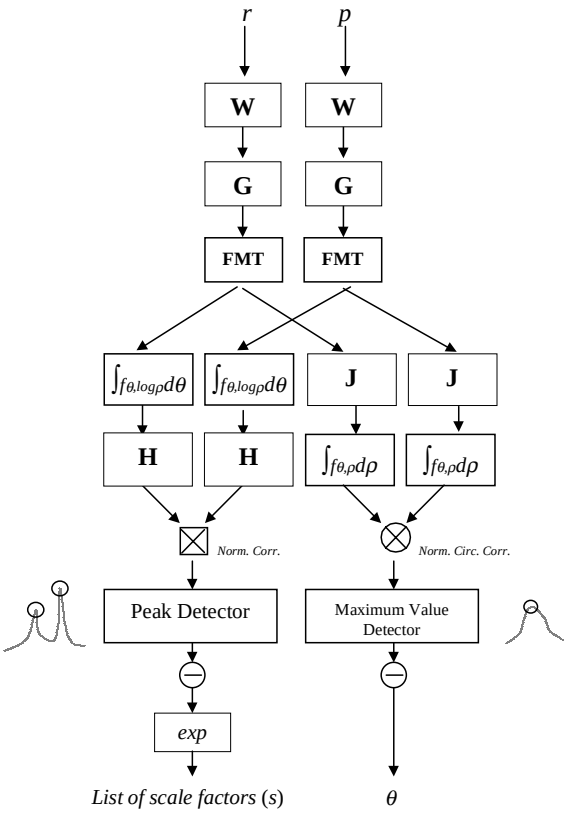

Figure 5.
System diagram of the first two phases of the coregistration algorithm. *W* is the wavelet decomposition filter, *G*, *H*, and *J* are pre- and post-processing filters to remove finite artifacts and improve accuracy.