MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **3**, Number **2**, April **2006**

http://www.mbejournal.org/

pp. **389–418**

MULTISCALE IMAGE REGISTRATION

Dana Paquin

Department of Mathematics, Stanford University
Stanford, CA 94305-2125

Doron Levy

Department of Mathematics, Stanford University
Stanford, CA 94305-2125

Eduard Schreibmann

Department of Radiation Oncology, Stanford University
Stanford, CA 94305-5947

Lei Xing

Department of Radiation Oncology, Stanford University
Stanford, CA 94305-5947

(Communicated by Yang Kuang)

Abstract. A multiscale image registration technique is presented for the registration of medical images that contain significant levels of noise. An overview of the medical image registration problem is presented, and various registration techniques are discussed. Experiments using mean squares, normalized correlation, and mutual information optimal linear registration are presented that determine the noise levels at which registration using these techniques fails. Further experiments in which classical denoising algorithms are applied prior to registration are presented, and it is shown that registration fails in this case for significantly high levels of noise, as well. The hierarchical multiscale image decomposition of E. Tadmor, S. Nezzar, and L. Vese [20] is presented, and accurate registration of noisy images is achieved by obtaining a hierarchical multiscale decomposition of the images and registering the resulting components. This approach enables successful registration of images that contain noise levels well beyond the level at which ordinary optimal linear registration fails. Image registration experiments demonstrate the accuracy and efficiency of the multiscale registration technique, and for all noise levels, the multiscale technique is as accurate as or more accurate than ordinary registration techniques.

1. **Introduction.** Often in image processing, images must be spatially aligned to allow practitioners to perform quantitative analyses of the images. The process of aligning images taken, for example, at different times, from different imaging devices, or from different perspectives, is called image registration. More precisely, image registration is the process of determining the optimal spatial transformation that maps one image to another. Typically, two images are taken as input, and the registration process is then the optimization problem which determines the

2000 *Mathematics Subject Classification.* Primary: 68U10; Secondary: 92C55, 62P10, 94A08.
Key words and phrases. image registration, multiscale analysis, noise, CT, MRI, mutual information.

geometric mapping that brings one image into spatial alignment with the other image. In practice, the particular type of transformation as well as the notion of optimal will depend on the specific application.

Examples of applications in which image registration is particularly important include astro- and geophysics, computer vision, remote sensing, and medicine. In this paper, we will focus on medical image registration. Image registration plays an important role in the analysis of medical images. For example, images taken from different sensors often contain complementary information. By bringing the two images into alignment so that anatomical features of one modality can be detected in the other modality, the information from the different modalities can be combined. In neurosurgery, for example, tumors are typically identified and diagnosed using magnetic resonance images (MRI), but stereotaxy technology (the use of surgical instruments to reach specified points) generally uses computed tomography (CT) images. Registration of these modalities allows the transfer of coordinates of tumors from the MRI images to the CT images. See [14] for a discussion of the applications of multimodality imaging to problems in neurosurgery. As another example, medical image data acquired prior to diagnosis can be compared with data acquired during or after treatment to determine the effectiveness of the treatment. To compare images taken at different times, however, the images must first be brought into spatial alignment so that actual differences in the data can be distinguished from differences that result from the image acquisition process.

In the context of medical imaging, the goal of the registration process is to remove artificial differences in the images introduced by patient movement, differences in imaging devices, etc., but at the same time, to retain real differences due to actual variations of the objects. Medical images, however, often contain significant levels of noise due to instrumentation imperfections, data acquisition techniques, image reconstruction methods, transmission and/or compression errors, and other factors. Although numerous successful image registration techniques have been published, we will see that ordinary image registration algorithms can fail to produce meaningful results when one or both of the images to be registered contains significant levels of noise.

Since noise is generally present in digital images, image denoising is a fundamental problem in image processing. Indeed, many approaches to image denoising have been presented. Thus a simple solution to the problem of image registration in the presence of noise would be to first apply a denoising algorithm to the noisy image(s), and then use existing image registration techniques to register the denoised images. However, common denoising algorithms, most notably spatial filtering algorithms, have the disadvantage that while they are successful in removing noise, they often remove edges as well. Additionally, most denoising procedures require a priori knowledge of the noise level, variance, and/or model, information not typically known in practice. For these and other reasons, we will demonstrate that ordinary image registration of noisy images fails to produce acceptable results even when classical denoising algorithms are applied to the noisy images prior to registration (for significantly high levels of noise). Moreover, even more advanced denoising techniques such as anisotropic diffusion (which was designed to remove noise while preserving edges) will be shown to fail to register noisy images. Thus, we seek a technique that enables successful image registration when one or both of the images to be registered is noisy.

Generally, we would like to consider an image f consisting of *coarse* and *fine* scales. The general shape and main features of an image are considered the coarse scales, and details and textures, such as noise, are the fine scales of the image. Separating the coarse and fine scales of an image, therefore, is an effective tool in denoising. Indeed, several denoising algorithms have been proposed using separation of the coarse and fine scales of an image, most notably [19], [18], [11], and [20]. The method presented in [20] presents a multiscale technique in which an image f is decomposed in a hierarchical expansion $f \sim \Sigma_j u_j$, where the u_j (called the components of f relative to the decomposition) resolve edges of f with increasing scales. More precisely, for small k, the sum $\Sigma_j^k u_j$ is a coarse representation of the image f, and as k increases, the sum captures more and more detail (and hence, noise) of the image.

In this paper, we present a multiscale image registration technique based on the multiscale decomposition of [20] that is particularly effective when one or both of the images to be registered contains significant levels of noise. Since the hierarchical expansion $f \sim \Sigma_j u_j$ decomposes the image f into components which contain increasingly fine scales, we expect a component-wise registration algorithm to produce accurate results for noisy images. That is, given a noisy image f, for small values of k, the component $\Sigma_j^k u_j$ retains the general shape of the image f while removing the details and noise of the image. Thus, if we wish to register the image f with another image, say g, we expect that registration of the components $\Sigma_j^k u_j$ with g will provide an accurate estimation of the actual transformation that brings the two images into spatial alignment with one another, for sufficiently small values of k. Similarly, if both f and g are noisy, we expect decomposing both images and performing component-wise registrations should accurately estimate the optimal transformation. We will demonstrate that multiscale image registration enables successful image registration for images that contain levels of noise that are significantly higher than the levels at which ordinary registration fails.

This paper is organized in the following way. In Section 2, we discuss the image registration problem and review standard image registration techniques. In Section 3, we present the problem of image registration in the presence of noise, and illustrate the failure of current techniques when one or both of the images to be registered contains high levels of noise. In Section 4, we briefly discuss classical and modern denoising techniques, and illustrate the failure of ordinary image registration of noisy images even when the images are denoised prior to registration. In Section 5, we review the multiscale image decomposition of [20], and illustrate the results of the hierarchical multiscale decomposition obtained upon applying the algorithm to noisy images. In Section 6, we present image registration techniques based upon the multiscale decomposition described in Section 5, and in Section 7, we present the results of our multiscale image registration experiments.

2. **The registration problem.**

Given a *fixed* and *moving* image, the registration problem is the process of finding an *optimal transformation* that brings the moving image into spatial alignment with the fixed image. While this problem is easy to state, it is difficult to solve. The main source of difficulty is that the problem is ill-posed, which means, for example, that the problem may not have a unique solution. Additionally, the notion of optimality may vary for each application: for example, some applications may require consideration only of rigid transformations, while other applications require non-rigid transformations. Finally, computation time

and data storage constraints place limitations on the complexity of models that can be used for describing the problem. This following discussion of image registration follows the presentation in [12].

2.1. **The mathematical setting.** A two-dimensional gray-scale image f is a mapping which assigns to every point $x \in \Omega \subset \mathbb{R}^2$ a gray value $f(x)$ (called the intensity value of the image at the point x). We will consider images as elements of the space $L^2(\mathbb{R}^2)$. Color images can be defined, for example, in terms of vector-valued functions $\mathbf{f} = (f_1, f_2, f_3)$ representing the RGB-color scales. For the medical imaging applications that we are interested in, images are in fact given in terms of discrete data, and the function f must be obtained by interpolation. We will not discuss this construction here, but we will assume that an interpolation method has been chosen.

Image registration is necessary, for example, for two images of the same object that are not spatially aligned. This occurs when the images are taken at different times, from different perspectives, or from different imaging devices. The basic input data to the registration process are two images: one is defined as the fixed image $f(x)$ and the other as the moving image $m(x)$. The goal is then to find a transformation ϕ such that the fixed image $f(x)$ and the deformed moving image $m_\phi(x) := m(\phi(x))$ are similar. To solve this problem in a mathematical way, the term *similar* needs to be defined appropriately. For example, if the images to be registered are taken from different devices, there may not be a correspondence between the intensities $f(x)$ and $m_\phi(x)$ for an optimal ϕ. Additionally, we may consider measures of similarity between the images which are not related to the intensities. Thus, the registration problem necessarily involves a discussion of the distance measures, or metrics, used to compare images. The general problem can then be stated as follows.

Given a distance measure $D : (L^2(\mathbb{R}^2))^2 \to \mathbb{R}$ and two images $f(x), m(x) \in L^2(\mathbb{R}^2)$, the solution ϕ of the registration problem is given by the following minimization problem:

$$\phi = \underset{\psi:\mathbb{R}^2\to\mathbb{R}^2}{\operatorname{argmin}} D(f, m_\psi). \tag{1}$$

In many applications, the set of allowable transformations to be considered in the minimization problem (1) is restricted to a strict subset of the set of all maps $\psi : \mathbb{R}^2 \to \mathbb{R}^2$. For example, we may require the transformation ϕ to be smooth, or we may impose specific parametric requirements, such as requiring the transformation to be rigid, affine, polynomial, etc.

2.2. **Landmark-based registration.** Landmark-based registration is an image registration technique based on a finite set of image features. The problem is to determine the transformation such that for a finite set of features, any feature of the moving image is mapped onto the corresponding features of the fixed image. More precisely, let $F(f, j)$ and $F(m, j)$, $j = 1, \ldots, m$ be given features of the fixed and moving images, respectively. The solution ϕ of the registration problem is then a map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$F(f, j) = \phi(F(m, j)), \ j = 1, \ldots m. \tag{2}$$

For a more general notion of landmark-based registration, we define the following distance measure:

$$D^{LM}(\phi) := \sum_{j=1}^{m} ||F(f,j) - \phi(F(m,j)||_l^2, \tag{3}$$

where $||\cdot||_l$ denotes a norm on the landmark, or feature, space. For example, if the features are locations of points, then $||\cdot||_l = ||\cdot||_{\mathbb{R}^2}$. We can then restate (2) as the minimization problem in which the solution $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ of the registration problem is given by:

$$\phi = \underset{\psi:\mathbb{R}^2\to\mathbb{R}^2}{\operatorname{argmin}} D^{LM}(\psi). \tag{4}$$

To solve this minimization problem, the transformation either is chosen to be an element of an n-dimensional space spanned, for example, by polynomials, splines, or wavelets, or it is required to be smooth in some sense. In the first case, the features to be mapped are the locations of a number of user-supplied landmarks. Let χ_k, $k = 1, \ldots n$ be the basis functions of the space. Then the minimization of

$$D^{LM}(\phi) := \sum_{j=1}^{m} ||F(f,j) - \phi(F(m,j)||_l^2$$

can be obtained upon expanding $\phi = (\phi_1, \phi_2)$ in terms of the basis functions χ_k and solving the resulting least squares problems.

In the case in which we require the transformation ϕ to be smooth, we introduce a functional which imposes smoothness restrictions on the transformation. That is, we look for a transformation ϕ which interpolates the features $F(f,j)$ and $F(m,j)$, and which is smooth in some sense. Such a transformation is called a minimal norm solution, and it turns out (see [8]) that the solution can be expressed in terms of radial basis functions.

Landmark-based registration is simple to implement, and the numerical solution requires only the solution of a linear system of equations. However, the main drawback of the landmark-based approach is that the registration process depends on the location of the landmarks. As the detection and mathematical characterization of landmarks (for example, anatomical landmarks in medical images) is not fully automated, the landmarks must be user-supplied, and this can be a time-consuming and difficult process, even for a medical expert; see, for example, [17]. Additionally, landmark-based registration does not always results in a physically meaningful registration. See [12, p. 44], for a simple example of a situation in which landmark-based registration fails to produce meaningful results.

2.3. **Principal-axes-based registration.** Principal-axes image registration is based on the idea of landmark-based registration, but it uses features that can be automatically detected. These features are constructed as follows. For an image $f : \mathbb{R}^2 \to \mathbb{R}$, and a function $g : \mathbb{R}^2 \to \mathbb{R}$, we define the expectation value of g with respect to f by

$$\mathbb{E}_f(g) := \frac{\int_{\mathbb{R}^2} g(x) f(x)\, dx}{\int_{\mathbb{R}^2} g(x)\, dx}. \tag{5}$$

If $u : \mathbb{R}^2 \to \mathbb{R}^{m\times n}$, we set $\mathbb{E}_f(u) := \mathbb{E}_f[u_{j,k}] \in \mathbb{R}^{m\times n}$. The center of an image f is defined by

$$C_f := \mathbb{E}_f[x] \in \mathbb{R}^2, \tag{6}$$

and the covariance by

$$\mathrm{Cov}_f := \mathbb{E}_f[(x - C_f)(x - C_f)^T] \in \mathbb{R}^{2\times 2}. \tag{7}$$

Given fixed and moving images, $f(x)$ and $m(x)$, the centers c_f and c_m and eigendecompositions of the covariance matrices Cov_f and Cov_m are used as the features F_i, and the registration problem is to compute $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that $F_i(m(\phi)) = F_i(f)$ for the features F_i.

This method is described in detail in [1]. The principal-axes method of image registration has the advantages that it is computationally fast and simple and requires few registration parameters, but has the disadvantages that it is not suitable for images of multiple modalities and that the solutions may be ambiguous. In particular, the principal-axes-based method cannot distinguish between images with the same center and covariance, even though images with very different structure and orientation may have the same center and/or covariance.

2.4. **Optimal parametric registration.** An alternative approach to registration is to use methods that are based on the minimization (or maximization) of some distance measure, or metric, D. The transformation ϕ is restricted to some parameterized space, and the registration can be obtained by minimizing (or maximizing) the distance D over the parameterized space. In particular, we will discuss metrics based on intensity, correlation, and mutual information. Given a metric D, a fixed image f, and a moving image m, optimal parametric registration is the problem of finding a transformation ϕ in some pre-specified parameterizable space such that $D(f, m(\phi))$ is minimized (or maximized in certain cases). Examples of commonly used parameterizable spaces in image registration are polynomial and spline spaces. We will primarily be interested in rigid and affine linear transformations. An affine linear map is a map of the form $\phi(x) = Ax + b$, $A \in \mathbb{R}^{2\times 2}$, $\det A > 0$, $b \in \mathbb{R}^2$. Such a map allows rotations, translations, scales, and shears of the coordinates. A translation (or rigid) transformation is a special case of an affine transformation which allows only rotations and translations of the coordinates, and in this case, the matrix A is required to be orthogonal with determinant 1. Optimal parametric registration is probably the most commonly used image registration technique.

To minimize $D(f, m(\phi))$, we must choose an optimization technique. That is, an optimal parametric registration technique is described by a metric to be minimized (or maximized) and an optimizer which controls the minimization (or maximization). The implementation of the registration algorithm works in the following way: at each iteration, the distance D between the two images is computed. An affine transformation is then applied to the moving image, and the distance between the images is recomputed. In theory, this process continues until the distance is minimized (or maximized), though in practice there is some stopping criterion.

At each stage, the optimizer determines the parameters of the transformation that will be applied to the moving image. Examples of commonly used optimizers include gradient descent and regular step gradient descent. Gradient descent optimization advances the parameters of the transformation in the direction of the gradient, where the step size is governed by a user-specified learning rate. Regular step gradient descent optimization advances the parameters of the transformation in the direction of the gradient where a bipartition scheme is used to compute the step size.

2.4.1. *The mean squares metric.* The mean squares metric computes the mean-squared pixel-wise difference in intensity between two images f and m:

$$MS(f,m) := \frac{1}{N}\sum_{i=1}^{N}(f_i - m_i)^2, \tag{8}$$

where N is the total number of pixels considered, f_i is the i^{th} pixel of image f, and m_i is the i^{th} pixel of image m. Note that the optimum value of the mean squares metric is 0, and poor matches between the images f and m result in large values of $MS(f,m)$. This metric has the advantage that it is computationally simple. It is based on the assumption that pixels in one image should have the same intensity as (spatially) corresponding pixels in the second image. Thus, the mean squares metric is restricted in practice to images of the same modality.

2.4.2. *The normalized correlation metric.* The normalized correlation metric computes pixel-wise cross-correlation and normalizes it by the square root of the auto-correlation function:

$$NC(f,m) := -\frac{\sum_{i=1}^{N}(f_i \cdot m_i)}{\sqrt{\sum_{i=1}^{N} f_i^2 \cdot \sum_{i=1}^{N} m_i^2}}, \tag{9}$$

where N, f_i, and m_i are as defined for the mean squares metric. The negative sign in (9) causes the optimum value of the metric to occur when the minimum is reached. Thus the optimal value of the normalized correlation metric is -1. As with the mean squares metric, the normalized correlation metric is restricted to images of the same modality.

2.4.3. *The mutual information metric.* Mutual information is an information-theoretic approach to image registration that was proposed independently by Viola and Wells [22] and Collignon et al. [4] in 1995. The idea is that mutual information computes the amount of information that one random variable (here, image intensity) gives about another random variable (here, intensity values of another image). More precisely, given a fixed image $f(x)$ and a moving image $m(x)$, we wish to compute the transformation ϕ which *maximizes* the mutual information; i.e.,

$$\phi = \arg\max_{\psi} I(f(x), m(\phi(x))). \tag{10}$$

Maximization of the mutual information criterion assumes that the statistical dependence between corresponding image intensity values is maximized when the images are geometrically aligned.

The mutual information $I(f(x), m(\phi(x))$ is defined in terms of entropy, where we consider x as a random variable over coordinate locations in the coordinate system of the fixed image. Let $h(\cdot)$ denote the entropy of a random variable: $h(x) := -\int p(x)\ln p(x)\,dx$, where $p(x)$ is the probability density function of the random variable x. Note that it is not clear how to construct $p(x)$; we will discuss methods for estimating the probability densities. The joint entropy of two random variables x and y is given by $h(x,y) = -\int p(x,y)\ln p(x,y)\,dx\,dx$, where $p(x,y)$ is the joint probability density function of the random variables x and y. Entropy can be considered as a measure of the uncertainty or complexity of a random variable.

If x and y are independent, then $p(x,y) = p(x)p(y)$, so $h(x,y) = h(x) + h(y)$. However, if there is any dependency (as would be the case if x and y are intensity values of images of the same object), then $h(x,y) < h(x) + h(y)$. The difference is defined to be *mutual information*:

$$I(f(x), m(\phi(x))) = h(f(x)) + h(m(\phi(x)) - h(f(x), m(\phi(x))). \tag{11}$$

The terms in (11) can be interpreted in the following way. The first term, $h(f(x))$, is the entropy of the fixed image and is independent of the transformation ϕ. The second term, $h(m(\phi(x)))$, is the entropy of $m(\phi(x))$, so maximization of mutual information encourages transformations ϕ for which $m(\phi(x)$ has a high level of complexity or uncertainty. The third term $-h(f(x), m(\phi(x)))$ is the negative joint entropy of $f(x)$ and $m(\phi(x))$, so maximization of mutual information is related to minimization of the joint entropy of $f(x)$ and $m(\phi(x))$. A detailed overview of mutual information based registration can be found in [16].

Mutual information has the following properties. Let $u(x)$ and $v(x)$ denote two images.

1. $I(u(x), v(x)) = I(v(x), u(x))$. Mutual information is symmetric.
2. $I(u(x), u(x)) = h(u(x))$. The information an image contains about itself is equal to the entropy of the image.
3. $I(u(x), v(x)) \leq h(u(x))$ and $I(u(x), v(x)) \leq h(v(x))$. The information that the images contain about each other can not be greater than the information contained in the individual images.
4. $I(u(x), v(x)) \geq 0$.
5. $I(u(x), v(x)) = 0$ if and only if $u(x)$ and $v(x)$ are independent. If the images $u(x)$ and $v(x)$ are independent, no information about one image is gained when the other image is known.

The entropies in equation (11) are defined in terms of integrals over the probability densities associated with the images $f(x)$ and $m(x)$. However, in a typical medical image registration problem, the probability densities are not directly accessible, and thus must be estimated from the image data. Parzen windowing, described in [5] and used in [22], is a common technique for density estimation. In this method, continuous density functions are constructed by a super-position of kernel functions $K(\cdot)$ centered at the elements of a sample of intensities taken from the image. The estimation of the probability density $p(z)$ is thus given by

$$p(x) \cong P^{\star}(z) = \frac{1}{N_S} \sum_{z_j \in S} K(z - z_j), \tag{12}$$

where N_S is the number of spatial samples in S and K is an appropriately chosen kernel function. The kernel function K must be smooth, symmetric, have zero mean, and unit mass. Examples of suitable candidates for K include the Gaussian density and the Cauchy density. In [22], Viola and Wells use a Gaussian density function with standard deviation σ to estimate the probability density functions. The optimal value of σ depends on the particular images to be registered.

Upon estimating the probability densities using the Parzen windowing technique, the entropy integral $h(z) = -\int p(z) \ln(p(z))\, dz$ must be evaluated, for example, by using a sample mean:

$$h(z) \cong -\frac{1}{N_R} \sum_{z_j \in R} \ln(P^\star(z_j)), \tag{13}$$

where R is a second sample of intensities taken from the image. That is, two separate intensity samples S and R are taken from the image. The first is used to estimate the probability density, and the second is used to approximate the entropy.

The main advantage of the mutual information measure is that was shown to be generally applicable for multi-modality registration, whereas intensity-based measures are typically not applicable for multimodality registration. Mutual information registration has been successfully used for a number of complex applications. Most notably, mutual information has been shown to be highly accurate for MRI-CT registration; see, for example, [9], [15], and [21].

2.5. **Non-parametric image registration.** All of the image registration techniques that we have discussed so far have been based on certain parameters. For example, either the transformation ϕ can be expanded in terms of basis functions that span a specified finite-dimensional space, or the registration is controlled by a specified set of external features. Non-parametric techniques do not restrict the transformation to a parameterizable set. Given two images, a fixed image $f(x)$ and a moving image $m(x)$, we seek a transformation ϕ such that $m(\phi(x))$ is similar to $f(x)$ in a certain sense. Upon defining a suitable distance measure D, the registration problem is then to minimize the distance between $m(\phi(x))$ and $f(x)$. However, a direct minimization is often not possible in the non-parametric case. The problem is ill-posed: small changes in the input data may lead to large changes in the output. Additionally, the solution is not unique. Given these constraints, a stable numerical implementation is often impossible. To circumvent these problems, a regularizing, or smoothing, term S is introduced, and the registration problem becomes the minimization of the distance between $m(\phi(x))$ and $f(x)$ plus a smoothing term $S(\phi)$. That is, the registration is based on a regularized minimization of the distance between the images.

In the discussion of non-parametric image registration, the transformation $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is split into the trivial identity part and the deformation or displacement part u; i.e.,

$$\phi(x) = x - u(x). \tag{14}$$

Upon decomposing ϕ in this way, we have $m(\phi(x)) = m(x - u(x)) := m_u(x)$. Given a distance D and a smoother S, the elastic registration problem is then the minimization of $D(f(x), m_u(x)) + \alpha S(u)$, where $\alpha \in \mathbb{R}$ is a positive regularizing parameter.

The choice of smoother S typically depends on the particular application. Examples of non-parametric image registration techniques include elastic registration [3], fluid registration [2], and diffusion registration [6]. Elastic registration uses linear elasticity theory to model the deformation of an elastic body. In this case, the regularizing term $S(u)$ is the linearized elastic potential of the displacement u. In fluid registration, the regularization is based on the linearized elastic potential of the time derivative of u. Finally, diffusion registration uses a regularization that is based on spatial derivatives of the displacement.

Remark. *In this section, we presented a brief overview of the major image registration techniques currently used in image registration. In practice, the best*

registration method for a given set of images will depend on the particular features of the images themselves. However, numerous studies comparing the accuracy and performance of different image registration techniques for various applications have been presented. The most extensive of these is [24]*, which originally consisted of a comparison of 16 methods but has since been substantially expanded.*

3. **Registration in the presence of noise.** In this section, we study the effect of noise on image registration, and we determine the approximate noise level at which registration fails. This study is conducted on the brain proton density slice images shown in Figure 1 below. The image on the right is the result of translating the image on the left by 13 mm to the right in the X-direction and 17 mm downward in the Y-direction. Let I denote the original image, and let T denote the translated image.

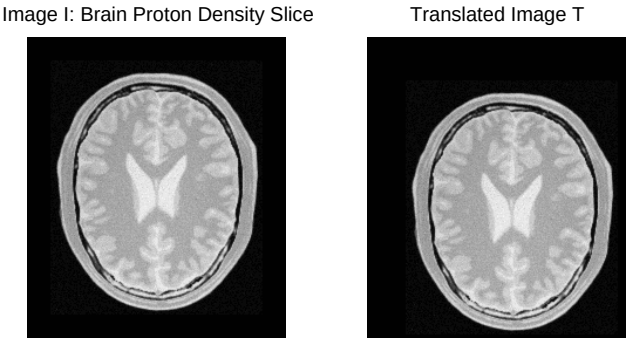

FIGURE 1. Original image I and translated image T.

Initially, we will consider the registration problem in which one of the images (here, the fixed image) is noisy. We will add increasing levels of noise to the image I and register the non-noisy translated image T with the noisy images. Our goal is to determine the approximate noise levels at which various image registration techniques fail, and to develop an algorithm that will enable registration beyond these levels. Since we know the exact transformation that brings T into spatial alignment with I, we can effectively evaluate and compare the accuracy of various registration techniques. We will demonstrate that our multiscale registration technique enables accurate registration of the translated image T with images that contain significant levels of noise. Eventually, we will also apply our techniques to the case in which both the fixed and the moving images contain high levels of noise. Before we present these results, we discuss the notion of noise in some detail.

REMARK. *In this paper we present the results only for registration experiments using the images I and T shown in Figure 1. We have performed numerous experiments using other images, and we obtained results similar to those presented in this paper. For the sake of brevity, we limit the results presented in this paper to the experiments using the images in Figure 1.*

3.1. **Noise.** Digital images are often degraded by random noise. In imaging, the term noise refers to random fluctuations in intensity values that occur during image capture, transmission, or processing, and that may distort the information given by

the image. Image noise is not part of the ideal signal and may be caused by a wide range of sources, such as detector sensitivity, environmental radiation, transmission errors, discretization effects, etc. Noise is generally classified as either independent noise or noise which is dependent on the image data.

Independent noise can often be described by an additive noise model, in which the observed image $f(x)$ is the sum of the true image $s(x)$ and the noise $n(x)$:

$$f(x) = s(x) + n(x). \tag{15}$$

Within this framework of additive noise, the noise $n(x)$ is commonly modeled by Gaussian noise of mean m and variance v. A multiplicative noise model describes noise that is dependent on the image data. This is often referred to as speckle noise:

$$f(x) = s(x) + s(x)n(x) = s(x)(1 + n(x)). \tag{16}$$

In this case, $n(x)$ is uniformly distributed random noise with mean m and variance v. Impulse noise, or salt-and-pepper noise, is noise that resembles salt and pepper granules randomly distributed over the image. Impulse noise is typically defined by the following model. We let $s(x)$ denote the actual image, and $f(x)$ denote the observed image. Then

$$f(x) = \begin{cases} s(x), & \text{with probability } 1 - \delta, \\ \eta(x), & \text{with probability } \delta, \end{cases} \tag{17}$$

where $\eta(x)$ is an identically distributed, independent random process. With this model, an arbitrary pixel $x \in \Omega \subset \mathbb{R}^2$ is affected by noise with probability δ, and not affected with probability $1 - \delta$. We will refer to δ as the impulse noise density, as adding impulse noise of density δ to an image $f(x)$ affects approximately $\delta \cdot \text{size}(f)$ pixels. The random process $\eta(x)$ is typically such that the corrupted pixels are either set to the maximum value, have single bits flipped over, or are set alternatively to zero or to the maximum value. This last case results in a salt-and-pepper appearance. Note that unaffected pixels always remain unchanged.

In Figure 2, we add additive Gaussian noise of mean 0 and variance 0.2, multiplicative speckle noise of mean 0 and variance 0.2, and impulse noise of density 0.2 to the brain proton density slice image I.

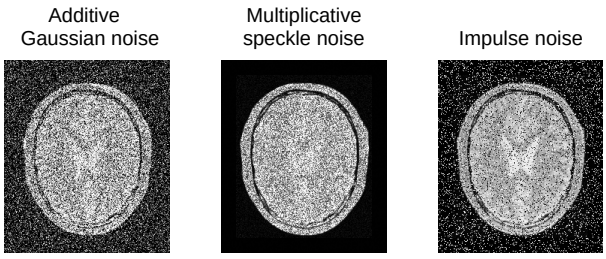

FIGURE 2. An illustration of the addition of various types of noise to the image I.

In this paper, we will study the problem of image registration in the presence of high levels of impulse noise. We will determine the impulse noise density level at which ordinary registration methods fail, and we will present a multiscale registration algorithm that enables accurate registration for noise levels higher than those at which ordinary methods fail. To study the effect of varying noise densities on the registration process, we add impulse noise of increasing densities δ to the brain proton density slice image I, and register the (non-noisy) translated image T with the noisy images. Let I_δ denote the image I with added impulse noise of density δ. In Figure 3, we illustrate the noisy images I_δ for increasing values of δ.

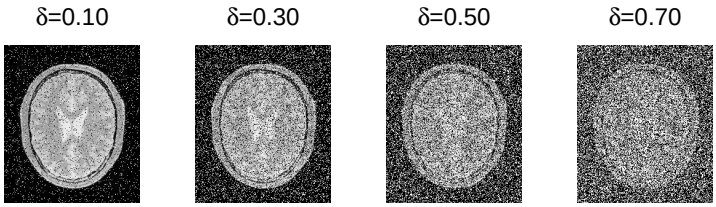

FIGURE 3. An illustration of adding impulse noise of increasing densities δ to the image I.

REMARK. Although in this paper we present the results of image registration experiments using only impulse noise, we have also conducted numerous experiments using other types of noise, including additive Gaussian noise and speckle noise. The results obtained with all other types of noise are similar.

3.2. **Registration results.** For increasing noise densities δ, we register T with I_δ using various registration methods. Recall that the image T is the result of translating the original image I 13 units in X and 17 units in Y, and that I_δ is the result of adding uniform impulse noise of density δ to the image I. Since T is a rigid transformation of I, we will restrict the registration process to linear transformations; i.e., we will consider optimal linear registrations. The optimal transformation ϕ produced by the optimal linear registration process will consist of two parameters, namely X- and Y-translation values. We will let ϕ_X and ϕ_Y denote the X- and Y-translation parameters, respectively, of the optimal transformation ϕ. For comparison purposes, we will perform the optimal linear registration using the mean squares, normalized correlation, and mutual information metrics.

We use the following parameters for the registration algorithms. For the mean squares and normalized correlation registration algorithms, we use the regular step gradient descent optimizer. Due to the stochastic nature of the metric computation in the mutual information algorithm, the regular step gradient descent optimizer does not work well in the case of mutual information. Instead, we use the gradient descent optimizer with a user-specified learning rate of 20.0. See [7] for a detailed

TABLE 1. The results obtained upon registering the translated image T with the noisy image I_δ, where δ is the impulse noise density; ϕ_X and ϕ_Y denote the X- and Y-translation values of the optimal transformation ϕ produced by the registration algorithm, and n is the number of iterations until convergence. The actual translation values are 13 units in X and 17 units in Y.

	Mean Squares			Normalized Correlation			Mutual Information		
δ	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n
0.00	12.99	17.00	18	13.01	17.00	18	12.75	17.03	200
0.10	12.99	17.01	28	12.99	17.01	20	12.83	16.88	200
0.20	13.03	16.98	17	13.04	16.98	19	12.98	16.64	200
0.30	12.97	17.03	28	13.02	17.02	11	13.02	17.02	200
0.40	18.89	7.16	15	8.05	1.30	13	11.08	9.72	200
0.50	2.16	7.06	19	9.09	2.18	8	9.72	7.12	200
0.60	29.81	3.19	40	4.08	0.24	7	4.57	5.17	200
0.70	2.08	1.14	13	3.11	2.13	12	3.08	2.86	200

discussion of these parameters. Finally, we set the maximum number of iterations for each algorithm to 200. As we shall see, mean squares and normalized correlation registrations typically converge very quickly to the optimal value. Mutual information, on the other hand, often does not actually reach the true optimal solution but instead oscillates within one or two pixels of the optimal solution (generally after 100-150 iterations). By reducing the learning rate, we can increase the likelihood of convergence, but this increases the computation time significantly without improving the accuracy of the solution.

For each of these three registration algorithms, and for each δ we record the X- and Y-translation parameters, denoted by ϕ_X and ϕ_Y, respectively, of the optimal transformation ϕ produced by the registration process. We also record the number of iterations n until convergence. The results are shown in Table 1. Recall that the actual translation values are 13 units in X and 17 units in Y. We also record the number of iterations until convergence, which we denote by n.

The results presented in Table 1 indicate that optimal linear registration in the presence of impulse noise fails when the impulse noise density in the fixed image reaches approximately 0.40, regardless of the metric used.

4. **Denoising.**

4.1. **Denoising techniques.** In this section, we discuss various denoising techniques. Image denoising is a fundamental problem in image processing, and there has been much research and progress on the subject. As our primary interest is not denoising but the problem of image registration of noisy images, we do not focus on the general problem of image denoising. Instead, we present a few of the most common and computationally simple denoising techniques. We will then apply these techniques to one of our noisy images and study the effect of denoising on the image registration techniques. In particular, in Section 3, we saw that ordinary optimal linear registration of noisy images failed when the impulse noise density was greater than 0.40. Also in this section, we shall determine whether or not denoising prior

to registration enables successful registration of noisy images for which registration failed previously.

Spatial filtering is the traditional approach to removing noise from images. Spatial filters use the assumption that noise occupies the higher regions of the frequency spectrum, and thus they attenuate high spatial frequencies. Local spatial filtering is a process in which the value of a given pixel in the filtered image is computed by applying some algorithm to the pixel values in a neighborhood of the given pixel. Typical implementations of spatial filters include mean filtering, median filtering, and Gaussian smoothing. Mean filtering computes the value of each output pixel by computing the statistical mean of the neighborhood of the corresponding input pixel. Thus, applying a mean filter to a noisy image reduces the amount of variation in gray-level intensity between pixels. Although this filter is computationally easy to implement, it is sensitive to the presence of outliers. Median filtering, which computes the value of each output pixel by computing the statistical median of the neighborhood of the corresponding input pixel, is more robust to the presence of outliers, and is thus commonly used for removing impulse noise from images. Convolution with a Gaussian kernel is another commonly used spatial filtering technique. See [23] for an overview of classical spatial filtering techniques.

In Figure 4, we illustrate the effect of applying a mean, median, and Gaussian convolution filter to the noisy image $I_{0.70}$, the brain proton density slice image with impulse noise of density 0.70. As is indicated in Figure 4, spatial filters smooth the data to remove noise but also blur edges.

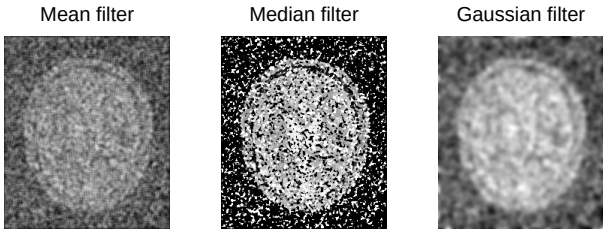

FIGURE 4. The results of applying mean, median, and Gaussian filters to the brain proton density slice image with impulse noise of density 0.70.

More advanced denoising techniques that remove noise more effectively while preserving edges include wavelet-based methods [10], total variation methods [19], and PDE-based anisotropic diffusion methods [13], to name a few. Total variation denoising reduces the total variation of the image, and thus removes noise, textures, and fine-scale details while preserving edges. In Figure 5, we illustrate the effect of applying these denoising techniques to the noisy image $I_{0.70}$, the brain proton density slice image with impulse noise of density 0.70.

4.2. **Registration results after denoising.** In this section, we register the translated image T with the denoised images illustrated in Figures 4 and 5. As in Section

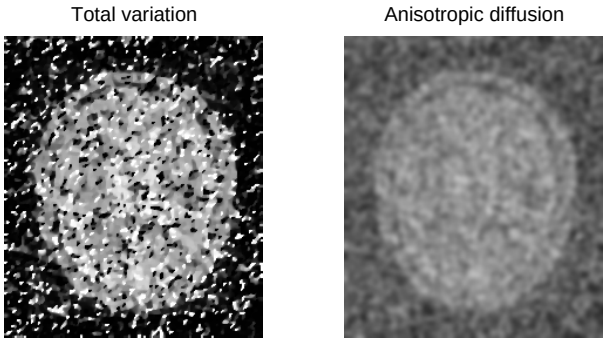

FIGURE 5. The results of applying the Osher-Rudin total variation and the Perona-Malik anisotropic diffusion denoising algorithms to the brain proton density slice image with impulse noise of density 0.70.

TABLE 2. The results obtained upon registering the translated image T with the denoised images obtained upon applying median, mean, and Gaussian convolution filters to the noisy image $I_{0.70}$. ϕ_X and ϕ_Y are the X- and Y-translation values of the optimal transformation ϕ produced by the registration algorithm, n is the number of iterations until convergence. The actual translation values are 13 units in X and 17 units in Y.

Denoising Technique	Mean Squares			Normalized Correlation			Mutual Information		
	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n
Mean Filtering	31.83	1.15	46	16.88	1.11	29	5.39	5.30	200
Median Filtering	18.87	1.26	31	2.38	6.90	34	4.39	4.06	200
Gaussian Filtering	18.86	-0.76	31	2.19	0.25	11	7.38	7.37	200
Total Variation	6.11	4.26	19	5.29	9.15	15	6.30	6.23	15
Anisotropic Diffusion	2.10	1.13	11	4.09	6.22	10	10.62	14.77	200

3, we use mean squares, normalized correlation, and mutual information optimal linear registration. For each registration method, we let ϕ denote the optimal transformation produced by the registration algorithm, and we let ϕ_X and ϕ_Y the X- and Y-translation parameters of the optimal transformation ϕ. We denote by n the number of iterations of each registration algorithm until convergence. We record the results in Table 2. The moving image in each case is the translated image T; recall that the actual translation values are 13 in X and 17 in Y.

The results presented in Table 2 indicate that the application of some of the classical as well as modern denoising techniques prior to registration does not enable successful registration of the noisy image $I_{0.70}$ with the translated image T. Although the more advanced denoising techniques such as total variation and anisotropic diffusion result in translation values that are closer to the actual values, particularly when mutual information registration is used, we conclude from Table 2 that denoising prior to registration does not produce accurate registration results for images that contain high levels of noise.

5. **Multiscale decomposition.** In this section, we present the multiscale image representation using hierarchical (BV, L^2) decompositions of [20]. The multiscale decomposition will provide a hierarchical expansion of an image that separates the essential features of the image (such as large shapes and edges) from the fine scales of the image (such as details and noise). The decomposition is hierarchical in the sense that it will produce a series of expansions of the image that resolve increasingly finer scales, and hence will include increasing levels of detail. We will eventually apply the multiscale decomposition algorithm to the problem of image registration in the presence of noise, and will demonstrate the accuracy of the multiscale registration technique for noisy images such as those that were considered in Sections 3 and 4.

We will use the following mathematical spaces in the decomposition algorithm. The space of functions of bounded variation, BV, is defined by

$$BV = \left\{ f \quad \middle| \quad ||f||_{BV} := \sup_{h \neq 0} |h|^{-1} ||f(\cdot + h) - f(\cdot)||_{L^1} < \infty \right\}.$$

We will also use the Sobolev space $W^{-1,\infty}$ with norm given by:

$$||f||_{W^{-1,\infty}} := \sup_{g} \left[\int \frac{f(x)g(x)}{||g||_{W^{1,1}}} \, dx \right],$$

where $||g||_{W^{1,1}} := ||\nabla g||_{L^1}$.

5.1. **The hierarchical decomposition.** Define the J-functional $J(f, \lambda)$ as follows:

$$J(f, \lambda) := \inf_{u+v=f} \lambda ||v||_{L^2}^2 + ||u||_{BV}, \tag{18}$$

where $\lambda > 0$ is a scaling parameter that separates the L^2 and BV terms. This functional $J(f, \lambda)$ was introduced in the context of image processing by Rudin, Osher, and Fatemi [19]. They suggested the following. Let $[u_\lambda, v_\lambda]$ denote the minimizer of $J(f, \lambda)$. The BV component, u_λ, captures the coarse features of the image f, while the L^2 component, v_λ, captures the finer features of f such as noise. This model is effective in denoising images while preserving edges, though it requires prior knowledge of the noise scaling λ.

Tadmor, et al. propose in [20] an alternative point of view in which the minimization of $J(f, \lambda)$ is interpreted as a decomposition $f = u_\lambda + v_\lambda$, where u_λ extracts the edges of f and v_λ extracts the textures of f. This interpretation depends on the scale λ, since texture at scale λ consists of edges when viewed under a refined scale (2^λ, for example). We refer to $v_\lambda = f - u_\lambda$ as the residual of the decomposition. Upon decomposing $f = u_\lambda + v_\lambda$, we proceed to decompose v_λ as follows:

$$v_\lambda = u_{2\lambda} + v_{2\lambda},$$

where

$$[u_{2\lambda}, v_{2\lambda}] = \underset{u+v=v_\lambda}{\operatorname{arginf}} J(v_\lambda, 2\lambda).$$

Thus, we obtain a two-scale representation of f given by $f \cong u_\lambda + u_{2\lambda}$, where now $v_{2\lambda} = f - (u_\lambda + u_{2\lambda})$ is the residual. Next we decompose $v_{2\lambda}$ and continue this process, which results in the following hierarchical multiscale decomposition of f. Starting with an initial scale $\lambda = \lambda_0$, we obtain an initial decomposition of the image f:

$$f = u_0 + v_0, \quad [u_0, v_0] = \underset{u+v=f}{\operatorname{arginf}} J(f, \lambda_0).$$

We then refine this decomposition to obtain

$$v_j = u_{j+1} + v_{j+1}, \quad [u_{j+1}, v_{j+1}] = \underset{u+v=v_j}{\operatorname{arginf}} J(v_j, \lambda_0 2^{j+1}), \quad j = 0, 1, \ldots$$

After k steps of this process, we have:

$$f = u_0 + v_0 = u_0 + u_1 + v_1 = u_0 + u_1 + u_2 + v_2 = \ldots = u_0 + u_1 + \ldots + u_k + v_k,$$

which is a multiscale image decomposition $f \sim u_0 + u_1 + \ldots + u_k$, with a residual v_k. As k increases, the u_k components resolve edges with increasing scales $\lambda_k = \lambda_0 2^k$.

5.2. **Implementation.**

5.2.1. *Initialization.* As described in [20], the initial scale λ_0 should capture the smallest oscillatory scale in f, given by

$$\frac{1}{2\lambda_0} \leq ||f||_{W^{-1,\infty}} \leq \frac{1}{\lambda_0}. \tag{19}$$

However, in practice, we may not be able to determine the size of $||f||_{W^{-1,\infty}}$, so we determine the initial choice of λ_0 experimentally. Following [20], for the applications presented in this paper, we will use $\lambda_0 = 0.01$ and $\lambda_j = \lambda_0 2^j$.

5.2.2. *Numerical discretization.* We follow the numerical algorithm of [20] for the construction of our hierarchical decomposition. In each step, we use finite-difference discretization of the Euler-Lagrange equations associated with the $J(v_j, \lambda_{j+1})$ to obtain the next term, u_{j+1}, in the decomposition of the image f. The Euler-Lagrange equation associated with the minimization of the functional $J(f, \lambda)$ given in equation (18) is

$$u_\lambda - \frac{1}{2\lambda} \operatorname{div}\left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\right) = f,$$

with the Neumann boundary conditions:

$$\left.\frac{\partial u_\lambda}{\partial n}\right|_{\partial\Omega} = 0, \tag{20}$$

where $\partial\Omega$ is the boundary of the domain Ω and n is the unit outward normal.

We thus obtain an expansion $f \sim \sum_{j=0}^{k} u_j$, where the u_j are constructed as approximate solutions of the recursive relation given by the following elliptic PDE:

$$u_{j+1} - \frac{1}{2\lambda_{j+1}}\text{div}\left(\frac{\nabla u_{j+1}}{|\nabla u_{j+1}|}\right) = -\frac{1}{2\lambda_j}\text{div}\left(\frac{\nabla u_j}{\nabla |u_j|}\right). \tag{21}$$

Note that $J(f,\lambda)$ contains a singularity when $|\nabla u_\lambda| = 0$. To remove this singularity, we replace $J(f,\lambda)$ by the regularized functional

$$J^\epsilon(f,\lambda) := \inf_{u+v=f}\left\{\lambda||v||_{L^2}^2 + \int_\Omega \sqrt{\epsilon^2 + |\nabla u|^2}\, dx\, dy\right\}, \tag{22}$$

and at each step, we find the minimizer u_λ of J^ϵ. The Euler-Lagrange equation for the regularized J^ϵ functional is

$$u_\lambda - \frac{1}{2\lambda}\text{div}\left(\frac{\nabla u_\lambda}{\sqrt{\epsilon^2 + |\nabla u_\lambda|^2}}\right) = f \in \Omega,$$

with Neumann boundary conditions.

To numerically implement the method, we cover the domain Ω with a grid ($x_i := ih, y_j := jh$), and discretize the elliptic PDE of equation (21) as follows:.

$$\begin{aligned}
u_{i,j} &= f_{i,j} + \frac{1}{2\lambda}D_{-x}\left[\frac{1}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j})^2 + (D_{0y}u_{i,j})^2}}D_{+x}u_{i,j}\right] \\
&+ \frac{1}{2\lambda}D_{-y}\left[\frac{1}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j})^2 + (D_{+y}u_{i,j})^2}}D_{+y}u_{i,j}\right] \\
&= f_{i,j} + \frac{1}{2h^2}\left[\frac{u_{i+1,j} - u_{i,j}}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j})^2 + (D_{0y}u_{i,j})^2}}\right. \\
&\left. - \frac{u_{i,j} - u_{i-1,j}}{\sqrt{\epsilon^2 + (D_{-x}u_{i,j})^2 + (D_{0y}u_{i-1,j})^2}}\right] \\
&+ \frac{1}{2h^2}\left[\frac{u_{i,j+1} - u_{i,j}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j})^2 + (D_{+y}u_{i,j})^2}}\right. \\
&\left. - \frac{u_{i,j} - u_{i,j-1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j-1})^2 + (D_{-y}u_{i,j})^2}}\right],
\end{aligned} \tag{23}$$

where D_+, D_-, and D_0 denote the forward, backward, and centered divided differences, respectively. To solve the discrete regularized Euler-Lagrange equations (24), we use the Gauss-Siedel iterative method to obtain:

$$u_{i,j}^{n+1} = f_{i,j} + \frac{1}{2h^2}\left[\frac{u_{i+1,j}^n - u_{i,j}^{n+1}}{\sqrt{\epsilon^2 + (D_{+x}u_{i,j}^n)^2 + (D_{0y}u_{i,j}^n)^2}} - \frac{u_{i,j}^{n+1} - u_{i-1,j}^n}{\sqrt{\epsilon^2 + (D_{-x}u_{i,j}^n)^2 + (D_{0y}u_{i-1,j}^n)^2}}\right] + \frac{1}{2h^2}\left[\frac{u_{i,j+1}^n - u_{i,j}^{n+1}}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j}^n)^2 + (D_{+y}u_{i,j}^n)^2}} - \frac{u_{i,j}^{n+1} - u_{i,j-1}^n}{\sqrt{\epsilon^2 + (D_{0x}u_{i,j-1}^n)^2 + (D_{-y}u_{i,j}^n)^2}}\right]. \tag{24}$$

To satisfy the Neumann boundary conditions (20), we first reflect f outside Ω by adding grid lines on all sides of Ω. As the initial condition, we set $u_{i,j}^0 = f_{i,j}$. We iterate this numerical scheme for $n = 0, 1, \ldots N$ until $||u^{n_\infty} - u^{n_\infty - 1}||$ is less than some preassigned value so that $u_{i,j}^{n_\infty}$ is an accurate approximation of the fixed point steady solution u_λ.

Finally, we denote the final solution $u_\lambda := \{u_{i,j}^{n_\infty}\}_{i,j}$. To obtain the hierarchical multiscale decomposition, we reiterate this process, each time updating f and λ in the following way:

$$\begin{aligned} f_{new} &\leftarrow f_{current} - u_\lambda, \\ \lambda_{new} &\leftarrow 2\lambda_{current}. \end{aligned} \tag{25}$$

That is, at each step, we apply the $J(f_{current} - u_\lambda, 2\lambda)$ minimization to the residual $f_{current} - u_\lambda$ of the previous step. Taking $\lambda_j = \lambda_0 2^j$, we obtain after k steps a hierarchical multiscale decomposition $f = u_{\lambda_0} + u_{\lambda_1} + \ldots + u_{\lambda_k} + v_{\lambda_k}$, where we write $u_{\lambda_j} = u_j$. We call the u_j, $j = 1, 2, \ldots, k$ the components of f and the v_k the residuals.

EXAMPLE 1. ***Decomposition of a noisy image.*** *We apply the hierarchical multiscale decomposition of* [20] *as described in Section 5 to the noisy image* $I_{0.70}$ *in Figure 3, using the following parameters:* $m = 12$ *hierarchical steps,* $\lambda_0 = 0.01$, $\lambda_j = \lambda_0 2^j$, $\epsilon = 0.001$, $n = 10$, *and* $h = 1$. *In Figures 6 and 7, we illustrate the components* u_{λ_j} *and the residuals* v_{λ_j} *for this decomposition. Note that in each hierarchical step, an additional amount of texture is seen in the components. Further, the noise is not seen in the first few components, while most of the texture is kept, and the noise only reappears as the refined scales reach the same scales as the noise itself. Our goal is to use this multiscale decomposition to register the noisy image* $I_{0.70}$ *with the translated image* T.

6. **Multiscale registration.** Consider again the noisy images I_δ shown in Figure 3 with impulse noise of increasing densities δ. Recall that in Section 3, we demonstrated that registration of the the translated image T with the noisy image I_δ

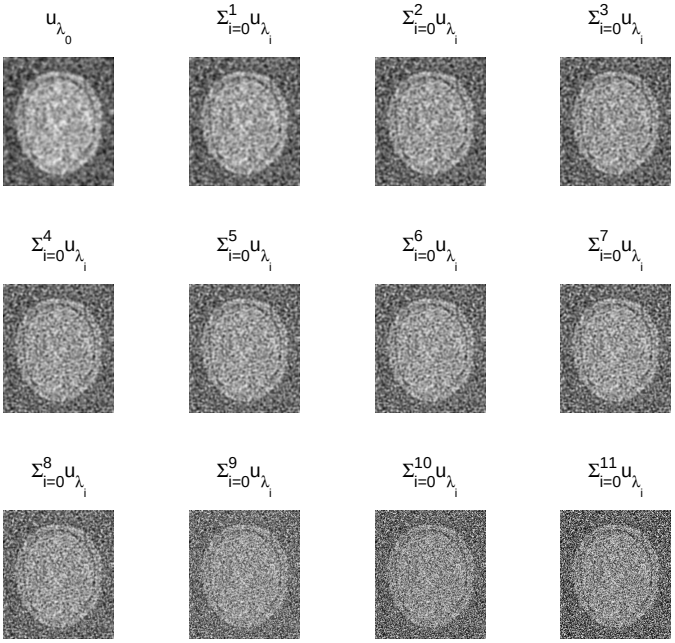

FIGURE 6. Multiscale decomposition of the noisy image $I_{0.70}$ shown in Figure 3.

failed when $\delta \geq 0.40$, regardless of the metric used in the optimal linear registration process. Moreover, registration using these classical methods failed even after denoising the noisy image using various standard denoising techniques, as demonstrated in Section 4. In this section, we present new methods for image registration that allow for a successful registration of the translated image T with the noisy images I_δ for values of the noise density δ significantly greater than the levels at which classical registration and registration after denoising fail. These registration techniques will be based on the hierarchical multiscale decomposition described in Section 5.

Consider two images A and B, and suppose that we want to register image B with image A. Suppose that one or both of the images contains a significant amount of noise. If only one of the images is noisy, we assume that it is image A. We propose the following multiscale registration method. First, we apply the multiscale hierarchical decomposition to both images. Let m denote the number of hierarchical steps used for the multiscale decompositions. For ease of notation, given an image f, we let

$$C_k(f) := \sum_{i=0}^{k} u_{\lambda_k} \tag{26}$$

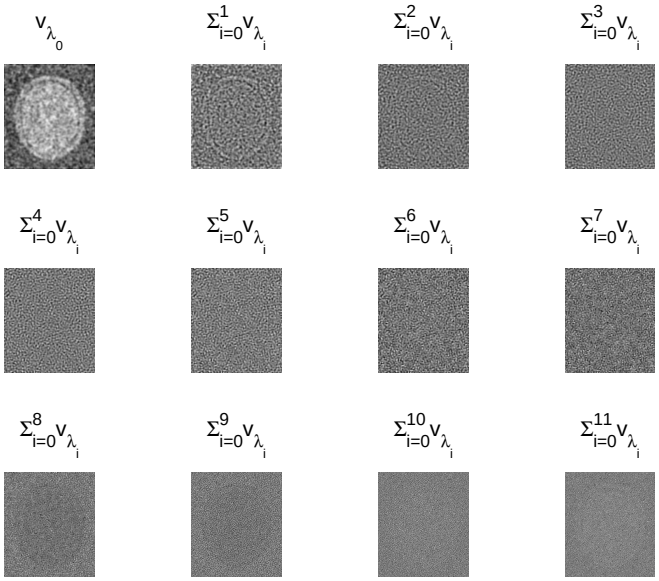

FIGURE 7. The residuals of the multiscale decomposition of the noisy image $I_{0.70}$ shown in Figure 3.

denote the k^{th} component of the image f, $k = 0, 1, \ldots, m-1$, obtained as in Section 5. Thus $C_k(A)$ will denote the k^{th} component of the image A, and $C_k(B)$ will denote the k^{th} component of image B.

We will present two algorithms; in the first, we register image B with the components of image A, and in the second, we register the components of image B with the components of image A.

6.1. **Algorithm I: One-node multiscale registration.** In our first multiscale registration algorithm, we register image B with the k^{th} component of A, for $k = 0, 1, \ldots, m-1$. This is illustrated by the schematic in Figure 8.

We refer to this algorithm as a *one-node multiscale registration algorithm* because in each of the m registrations prescribed by the algorithm, the moving image is always the image B. We only use the multiscale components of the fixed image A for the one-node algorithm.

Let ϕ_k denote the optimal transformation produced by the registration algorithm upon registering B with $C_k(A)$, $k = 0, 1, \ldots, m-1$. Recall that $C_0(A)$ contains only the coarsest scales of the image A, and as k increases, $C_k(A)$ contains increasing levels of detail (and hence, noise) of the image A. Thus, we expect that registration of image B with $C_k(A)$ should give an improvement compared to ordinary registration for the first few values of k. As k increases, however, we expect that eventually the component $C_k(A)$ will become too noisy to give successful registration.

Upon determining the transformations ϕ_k with a suitable registration algorithm (e.g., an optimal linear registration), we have several options for defining the optimal transformation Φ that should bring the image B into spatial alignment with the

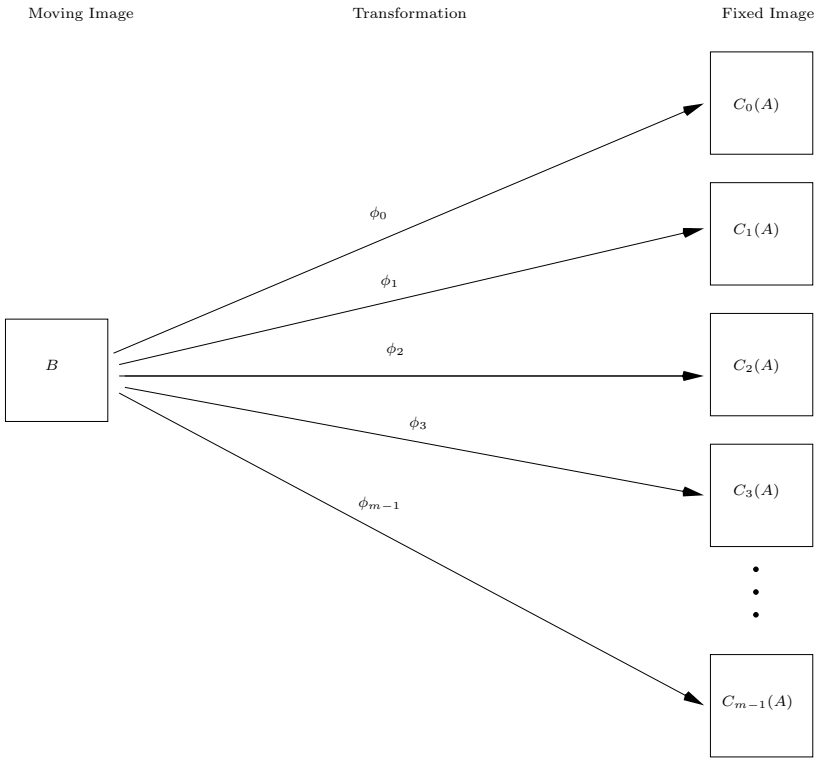

FIGURE 8. This schematic represents a one-node multiscale image registration algorithm in which we register the moving image B with the k^{th} component of the fixed image A, for $k = 0, 1, \ldots, m-1$, where m is the number of hierarchical steps used for the multiscale decompositions.

image A. The first option would be to take into account the registration parameters corresponding to the coarse scales only, i.e., the first few values of k, for which we expect a more accurate registration. Upon determining the number of registrations that we wish to take into account, we could then estimate Φ by averaging the registration parameters corresponding to those coarse scale registrations. A second option would be to define Φ as a weighted average of the ϕ_k; i.e.,

$$\Phi := \frac{1}{m} \sum_{k=0}^{m-1} a_k \phi_k, \tag{27}$$

where the weights a_k are appropriately chosen non-negative real numbers such that $\Sigma a_k = m$. For example, we could perform a statistical analysis on the registration parameters corresponding to the ϕ_k, and use the mean and standard deviation (or the mean and standard deviation of the first several values) to determine the weights a_k.

6.2. **Algorithm II: Multi-node multiscale registration.** In our second multiscale registration algorithm, we register the k^{th} component of image B with the

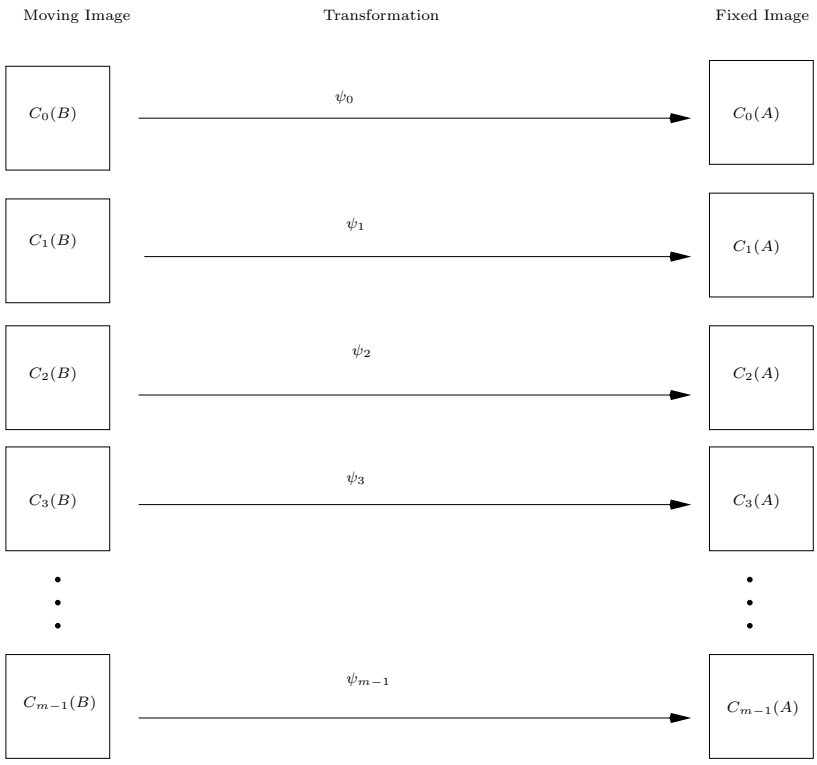

FIGURE 9. This schematic represents a multi-node multiscale image registration algorithm in which we register the k^{th} component of the moving image B with the k^{th} component of the fixed image A, for $k = 0, 1, \ldots, m-1$, where m is the number of hierarchical steps used for the multiscale decompositions.

k^{th} component of image A, for $k = 0, 1, 2, \ldots m-1$, as illustrated by the schematic in Figure 9.

We refer to this algorithm as a *multi-node multiscale registration algorithm* because in each of the m registrations prescribed by the algorithm, we consider both the components of the fixed image A and the components of the moving image B.

Let ψ_k denote the optimal transformation produced by the registration algorithm upon registering $C_k(B)$ with $C_k(A)$, $k = 0, 1, \ldots, m-1$. As before, we expect that registration of $C_k(B)$ with $C_k(A)$ should give an improvement compared to ordinary registration for the first few values of k. As k increases, however, we expect that eventually the components $C_k(A)$ and $C_k(B)$ will become too noisy to register successfully. Since this algorithm considers components of both images, we expect that it will be particularly successful in the case in which both images are noisy.

As in the case of the one-node multiscale registration algorithm, we can define the optimal transformation Ψ that should bring image B into spatial alignment with image A either by taking into account only the first few registration results corresponding to registration of the coarse scales, and averaging the registration parameters corresponding to these first few registrations, or by computing a weighted average:

TABLE 3. The registration results upon registering T with $I_{0.70}$ using Algorithm I. Here, we use $m = 12$ hierarchical steps to decompose the noisy image, so we perform $m = 12$ registration simulations. The transformation parameters ϕ_X and ϕ_Y are the X- and Y-translation parameters of the optimal transformation ϕ produced by the registration algorithm. The actual translation values are 13 in X and 17 in Y. The moving image in all simulations is the translated image T.

	Mean Squares			Normalized Correlation			Mutual Information		
Fixed Image	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n	ϕ_X	ϕ_Y	n
$I_{0.70}$	4.57	5.17	200	2.08	1.14	7	4.08	0.24	7
$C_0(I_{0.70})$	12.65	16.36	200	3.08	1.11	12	3.11	0.17	9
$C_1(I_{0.70})$	12.69	16.78	200	2.08	3.08	14	2.13	2.12	12
$C_2(I_{0.70})$	12.56	16.79	200	2.11	3.08	14	2.14	3.11	15
$C_3(I_{0.70})$	12.53	16.76	200	3.08	2.11	14	3.11	2.14	7
$C_4(I_{0.70})$	12.48	16.76	200	24.88	1.16	36	18.86	1.18	30
$C_5(I_{0.70})$	12.46	16.78	200	40.80	1.07	52	0.21	1.18	11
$C_6(I_{0.70})$	12.43	16.80	200	28.86	0.15	46	27.84	2.19	42
$C_7(I_{0.70})$	12.43	16.79	200	-2.87	4.11	15	0.18	3.14	12
$C_8(I_{0.70})$	12.43	16.74	200	25.89	3.12	40	-1.84	4.12	14
$C_9(I_{0.70})$	9.33	9.41	200	6.05	4.12	12	7.99	2.08	16
$C_{10}(I_{0.70})$	8.44	8.32	200	-3.92	8.12	21	4.09	3.15	16
$C_{11}(I_{0.70})$	6.96	6.46	200	8.97	6.13	13	3.65	1.17	27

$$\Psi := \frac{1}{m} \sum_{k=0}^{m-1} b_k \psi_k, \tag{28}$$

where the weights b_k are appropriately chosen non-negative real numbers such that $\Sigma b_k = m$.

7. **Examples of multiscale registration.**

7.1. **A noisy fixed image.** In this section, we use the multiscale registration technique described in Section 6 to register the translated (non-noisy) image T with the noisy image $I_{0.70}$. Recall that $I_{0.70}$ is the image obtained upon adding impulse noise of density 0.70 to the brain proton density slice image I. As before, let $C_k(I_{0.70})$ denote the k^{th} component in the multiscale decomposition of $I_{0.70}$, for $k = 0, 1, \ldots m$, obtained as in Section 6. We perform the multiscale decomposition using $m = 12$ hierarchical steps, $\lambda_0 = 0.01$, and $\lambda_j = \lambda_0 2^j$. In Table 3, we present the results of $m = 12$ registration simulations, obtained upon registering T with $C_k(I_{0.70})$, $k = 0, 1, \ldots, 11$, using Algorithm I of Section 6.1. For each registration, we let ϕ denote the optimal transformation produced by the registration algorithm, and we let ϕ_X and ϕ_Y the X- and Y-translation parameters of the optimal transformation ϕ. The moving image in each registration is the translated image T. For reference, we also include in the first line of Table 3 the parameters obtained using ordinary registration.

It is clear from the results presented in Table 3 that the results obtained using mean squares and normalized correlation methods are completely inaccurate. Thus, the one-node multiscale algorithm did not produce meaningful results for these metrics. For mutual information, however, the X and Y translation parameters are clustered around 12.5 units in X and 16.8 units in Y for $k = 0, 1, \dots 8$, but then are significantly different for the remaining values of k. We expected that the multiscale registration results would be an accurate approximation of the actual transformation Φ for small values of k, but then would deviate as k became sufficiently large, because as k becomes large, increasing scales of detail (and hence, noise) appear in the component C_k. Thus, even without knowing the actual values of the X- and Y-translations, it makes sense to take into account only the parameters corresponding to the first 9 registrations ($k = 0, 1, \dots 8$). Averaging the translation parameters for the first 9 registrations, we obtain $\Phi_X = 12.52$ and $\Phi_Y = 16.73$. Since the actual values are 13 in X and 17 in Y, we see that multiscale mutual information registration produced very accurate results in this case, and indeed is a significant improvement compared to ordinary registration as well as to classical and modern denoising followed by registration.

Next, we provide the results obtained with Algorithm II by registering the multiscale components of the translated image T with the multiscale components of the noisy image $I_{0.70}$. Let $C_k(T)$ and $C_k(I_{0.70})$ denote the multiscale components of T and $I_{0.70}$, respectively, obtained through the multiscale decomposition presented in Section 5. As before, we use $m = 12$ hierarchical steps, $\lambda_0 = 0.01$, and $\lambda_j = \lambda_0 2^j$ to perform the decomposition. In Table 4, we present the results of $m = 12$ registration simulations, obtained upon registering $C_k(T)$ with $C_k(I_{0.70})$, $k = 0, 1, \dots, 11$. For each registration, we let ψ denote the optimal transformation produced by the registration algorithm, and let ψ_X and ψ_Y denote the X- and Y-translation parameters of the optimal transformation ψ. For reference, we also include in the first line of Table 4 the parameters obtained using ordinary registration.

To estimate the transform parameters Ψ_X and Ψ_Y, we note that for mutual information, the translation parameters ψ_x and ψ_y are clustered together for the first 9 registrations, and for mean squares and normalized correlation, the values are clustered together for the first 2 registrations. Thus for mutual information we determine Ψ by averaging the parameters corresponding to the first 9 registrations, and for mean squares and normalized correlation, we average the first 2 values. In Table 5, we present the X- and Y-translation values corresponding to these averages.

REMARK. *Since the actual translation values are 13 in X and 17 in Y, we see that the multinode multiscale registration of the translated image T with the noisy image $I_{0.70}$ produces very accurate results for each of the three optimal linear registration metrics considered here (mean squares, normalized correlation, and mutual information). The main difference between the results obtained with Algorithm I and Algorithm II is the accurate registration of the coarse scales obtained with Algorithm II.*

7.2. **Noisy fixed and moving images.** In this section, we consider the registration problem in which both the fixed and moving images are noisy. Consider the noisy images $I_{0.40}$ and $T_{0.40}$, where T, as before, is the result of translating I 13 units in X and 17 units in Y, and A_δ denotes the image obtained by adding impulse noise of density δ to the image A. The noisy images are shown in Figure 10.

TABLE 4. The registration results obtained with Algorithm II. Here we register the k^{th} multiscale component $C_k(T)$ of the translated image T with the k^{th} multiscale component $C_k(I_{0.70})$ of the noisy image $I_{0.70}$ obtained via the multiscale decomposition discussed in Section 5. Here, we use $m = 12$ hierarchical steps to decompose the noisy image, so we perform $m = 12$ registration simulations. The transformation parameters ψ_X and ψ_Y are the X- and Y-translation parameters of the optimal transformation ψ produced by the registration algorithm. The actual translation values are 13 in X and 17 in Y.

Fixed and Moving Images	Mean Squares		Normalized Correlation		Mutual Information	
	ϕ_X	ϕ_Y	ϕ_X	ϕ_Y	ϕ_X	ϕ_Y
$I_{0.70}$ and T	4.57	5.18	2.08	1.14	4.08	0.24
$C_0(I_{0.70})$ and $C_0(T)$	12.69	16.66	12.29	17.72	12.96	17.08
$C_1(I_{0.70})$ and $C_1(T)$	12.67	16.87	13.70	17.75	12.99	17.67
$C_2(I_{0.70})$ and $C_2(T)$	12.59	16.86	20.77	5.20	16.84	4.31
$C_3(I_{0.70})$ and $C_3(T)$	12.55	16.82	3.19	0.31	4.20	4.23
$C_4(I_{0.70})$ and $C_4(T)$	12.52	16.83	2.20	2.24	26.74	5.18
$C_5(I_{0.70})$ and $C_5(T)$	12.51	16.84	31.65	2.23	14.90	6.27
$C_6(I_{0.70})$ and $C_6(T)$	12.49	16.87	30.69	6.16	19.87	4.29
$C_7(I_{0.70})$ and $C_7(T)$	12.48	16.85	33.64	3.16	29.64	3.32
$C_8(I_{0.70})$ and $C_8(T)$	12.53	16.71	28.81	3.22	1.26	1.29
$C_9(I_{0.70})$ and $C_9(T)$	9.26	9.36	2.13	3.13	17.93	3.21
$C_{10}(I_{0.70})$ and $C_{10}(T)$	8.80	8.61	2.12	3.12	32.63	3.14
$C_{11}(I_{0.70})$ and $C_{11}(T)$	6.95	6.34	34.74	2.10	4.13	5.08

TABLE 5. The translation parameters Ψ_X and Ψ_Y obtained by averaging the parameters corresponding to the coarse scale registrations. The actual translation values are 13 in X and 17 in Y.

	Mean Squares	Normalized Correlation	Mutual Information
Ψ_X	12.56	12.99	12.98
Ψ_Y	16.82	17.74	17.37

Before applying our multiscale registration algorithm, we attempt to register $T_{0.40}$ with $I_{0.40}$ using the three registration methods mean squares, normalized correlation, and mutual information. The results shown in Table 6 indicate that registration of the noisy images fails, regardless of the metric used in the optimal linear registration algorithm.

Since ordinary registration of the noisy images fails, we register the images using Algorithm II, the multi-node multiscale registration technique. First, we perform the multiscale decomposition discussed in Section 5 to both noisy images, again using $m = 12$ hierarchical steps, initial scale $\lambda_0 = 0.01$, and $\lambda_j = 2^j\lambda_0$. Let $C_k(I_{0.40})$ and $C_k(T_{0.40})$ denote the k^{th} component in the multiscale decomposition of $I_{0.40}$

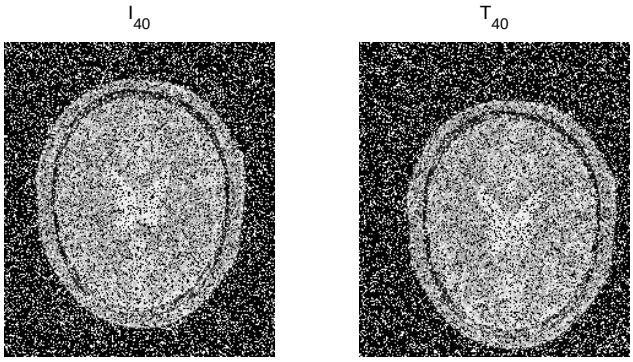

FIGURE 10. The original image I and translated image T with impulse noise of density $\delta = 0.40$

TABLE 6. The results of registering the noisy translated image $T_{0.40}$ with the noisy image $I_{0.40}$, using three different metrics. The actual translation values are 13 in X and 17 in Y.

Registration Method	ϕ_X	ϕ_Y
Mean Squares	11.02	7.04
Normalized Correlation	3.05	0.99
Mutual Information	5.03	2.54

and $T_{0.40}$, respectively. Since both images are noisy, we register the k^{th} component $C_k(T_{0.40})$ with the k^{th} component $C_k(I_{0.40})$. For each registration simulation, denote by ψ the optimal transformation produced by the registration algorithm, and denote by ψ_X and ψ_Y the corresponding X- and Y-translation parameters of the optimal transformation ψ. We present the results of this multiscale registration in Table 7.

To estimate the transformation Ψ, for mutual information we average the parameters corresponding to registration of the first 7 scales. For mean squares, we average the results of the first 2 registrations, and for normalized correlation, we average the registration results from the first 4 registrations. In Table 8, we present the X- and Y-translation values Ψ_X and Ψ_Y.

Note that since the actual translation values are 13 in X and 17 in Y, our multiscale registration technique provides accurate results in the case in which both the fixed and moving images contain significant levels of noise.

REMARKS.

1. *For the sake of brevity, we presented only the multiscale registration results for registration of images that contain levels of noise greater than the level at*

TABLE 7. The results of registering $T_{0.40}$ with $I_{0.40}$ using Algorithm II. Here, we use $m = 12$ hierarchical steps to decompose the noisy image, so we perform $m = 12$ registration simulations. The actual translation values are 13 in X and 17 in Y.

Fixed Image	Moving Image	Mean Squares		Normalized Correlation		Mutual Information	
		ϕ_X	ϕ_Y	ϕ_X	ϕ_Y	ϕ_X	ϕ_Y
$I_{0.40}$	$T_{0.40}$	5.03	2.54	11.02	7.04	3.05	0.99
$C_0(I_{0.40})$	$C_0(T_{0.40})$	13.06	16.92	13.05	16.92	13.05	16.92
$C_1(I_{0.40})$	$C_1(T_{0.40})$	13.05	16.93	13.02	16.22	13.06	16.92
$C_2(I_{0.40})$	$C_2(T_{0.40})$	13.03	16.93	8.11	5.29	13.02	16.27
$C_3(I_{0.40})$	$C_3(T_{0.40})$	13.02	16.94	5.40	12.19	13.02	16.25
$C_4(I_{0.40})$	$C_4(T_{0.40})$	13.02	16.94	2.20	8.00	2.23	5.09
$C_5(I_{0.40})$	$C_5(T_{0.40})$	13.01	16.93	26.76	1.21	1.17	7.00
$C_6(I_{0.40})$	$C_6(T_{0.40})$	12.99	16.81	23.83	4.11	1.22	2.17
$C_7(I_{0.40})$	$C_7(T_{0.40})$	7.05	6.08	0.20	3.15	0.20	4.15
$C_8(I_{0.40})$	$C_8(T_{0.40})$	6.78	5.05	6.04	2.09	6.04	6.05
$C_9(I_{0.40})$	$C_9(T_{0.40})$	3.05	1.02	9.98	1.10	5.06	10.01
$C_{10}(I_{0.40})$	$C_{10}(T_{0.40})$	12.20	14.01	-1.97	0.99	-3.93	3.04
$C_{11}(I_{0.40})$	$C_{11}(T_{0.40})$	4.80	3.19	1.01	5.98	3.91	0.72

TABLE 8. The translation parameters Ψ_X and Ψ_Y for registration of $T_{0.40}$ with $I_{0.40}$ obtained by averaging the translation parameters of the coarse scale registrations. The actual translation values are 13 in X and 17 in Y.

	Mean Squares	Normalized Correlation	Mutual Information
Ψ_X	13.03	13.03	13.04
Ψ_Y	16.92	16.57	16.59

which ordinary registration methods fail. However, we also performed multiscale registration simulations for noise densities lower than those presented here, and in all cases, the multiscale technique was either as accurate as or more accurate than ordinary registration techniques.

2. *The method of estimating the translation parameters in X and Y by averaging the parameters corresponding to the coarse scale registrations is based on determining the scales that should be taken into consideration. For the results presented in this paper, as well as for all of the other simulations that we studied, we found a drastic jump in the translation parameters such as that between the eighth and ninth scales in the mutual information column of Table 4. In such cases, the natural choice is to average the parameters corresponding to the coarse scales before the jump, and to exclude the remaining values; indeed, in all cases considered for this study, we have found that estimating the parameters in this way yields extremely accurate results. More generally, we expect that for most problems of this type, there should be a noticeable jump in the multiscale registration parameters, thus enabling a determination of the*

coarse scales that should be averaged. This jump occurs because once a certain level of detail, and hence noise, appears in the scales, the registration process fails. More specifically, the optimization of the registration metric does not produce meaningful results if enough noise is present, and the registration parameters that result differ significantly from those that result from registration of images in which less noise is present.

8. **Summary.** While there are many existing medical image registration techniques, common approaches are shown to fail to give accurate results when one or more of the images to be registered contains high levels of noise. Further, if the noise level is significantly high, image registration can fail even when a denoising algorithm is applied to the noisy images before registration. We have presented an image registration technique based on the hierarchical multiscale decomposition of [20] of the images to be registered. The multiscale decomposition of an image results in a hierarchical representation that separates the coarse and fine scales of the image. Upon obtaining the decomposition of one or both of the images to be registered, we register the components of the moving image with the components of the fixed image. Since the coarse scale components of an image contain the essential features and shapes of the image, registration of the coarse scale components of the moving image with the coarse scale components of the fixed image provides an accurate estimate of the actual transformation that brings the moving image into spatial alignment with the fixed image. Using images in which the precise transformation that maps one to the other is known, we have shown that the multiscale approach is indeed accurate for levels of noise much higher than the noise levels at which ordinary optimal linear registration and denoising prior to ordinary registration methods fail; moreover, for all levels of noise, the multiscale technique either matches or outperforms ordinary registration techniques. Finally, we hope to extend these techniques to other image registration problems in which ordinary registration techniques are not successful due to degradation or other factors present in the images to be registered.

Acknowledgments. The work of D. Levy was supported in part by the National Science Foundation under Career Grant No. DMS-0133511. The work of L. Xing was supported in part by the Department of Defense under Grant No. PC040282 and the National Cancer Institute under Grant No. 5R01 CA98523-01.

REFERENCES

[1] N.M. Alpert, J.F. Bradshaw, D. Kennedy, and J.A. Correia, *The principal axes transformation–A method for image registration*, Journal of Nuclear Medicine, vol. 31, no.10, pp. 1717-1722, 1990.

[2] M. Bro-Nielsen and C. Gramkow, *Fast fluid registration of medical images*, Lecture Notes in Computer Science, 1131, pp. 267-276, Springer, Berlin and Heidelberg, 1996.

[3] C. Broit, *Optimal Registration of Deformed Images*, Ph.D. thesis, Computer and Information Science, University of Pennsylvania, 1981.

[4] A. Collignon, D. Vadermeulen, P. Suetens, and G. Marchal, *3d multi-modality medical image registration based on information theory*, Computational Imaging and Vision, vol. 3, pp. 263-274, 1995.

[5] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

[6] B. Fischer and J. Modersitzki, *Fast inversion of matrices arising in image processing*, Numerical Algorithms, vol. 22, pp. 1-11, 2001.

[7] *Insight Segmentation and Registration Toolkit (ITK) Software Guide*, http://www.itk.org.

[8] W.A. Light, *Variational methods for interpolation, particularly by radial basis functions*, Numerical Analysis. pp. 94-106, 1995.
[9] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, *Multimodality image registration by maximization of mutual information*, IEEE Transactions on Medical Imaging, vol. 16, no. 2, pp. 187-198, 1997.
[10] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, Paris, 1998.
[11] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, University Lecture Series 22, AMS, Providence, RI, 2002.
[12] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford, 2004.
[13] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pp. 629-639, 1990.
[14] T. M. Peters, B. L. K. Davey, P. Munger, R. M. Comeau, A. Evans, and A. Oliver, *Three-dimensional multi-modal image-guidance for neurosurgery*, IEEE Transactions on Medical Imaging, vol. 15, no. 2, pp. 121-128, 1996.
[15] J. P. W. Pluim, J .B. A. Maintz, and M. A. Viergever, *Image registration by maximization of combined mutual information and gradient information*, IEEE Transactions on Medical Imaging, vol 19, no. 8, pp. 809-814, 2000.
[16] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, *Mutual information based registration of medical images: A survey*, IEEE Transactions on Medical Imaging, vol. 22, no. 8 , pp. 986-1004, 2003.
[17] K. Rohr, *Landmark-based Image Analysis*, Computational Imaging and Vision, Kluwer Academic, Dordrecht, 2001.
[18] L. Rudin and V. Caselles, *Image recovery via multiscale total variation*, Proceedings of the Second European Conference on Image Processing, Palma, Spain, 1995.
[19] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, vol. 60, pp. 259-268, 1992.
[20] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical* (BV, L^2) *decompositions*, Multiscale Modeling and Simulations, vol. 2, no. 4, pp. 554-579, 2004.
[21] P. Thévenaz and M. Unser, *Optimization of mutual information for multiresolution image registration*, IEEE Transactions on Image Processing, vol. 9, no. 12, pp. 2083-2099, 2000.
[22] P. Viola, W. Wells, H. Atsumi, S. Nakajima, and R. Kikinis, *Multi-modal volume registration by maximization of mutual information,* Medical Image Analysis, vol. 1, no. 1, pp. 35-51, 1995.
[23] A. R. Weeks, *Fundamentals of Electronic Image Processing*, SPIE Optical Engineering Press and IEEE Press, 1996.
[24] J. West, et al., *Comparison and evaluation of retrospective intermodality brain image registration techniques*, Journal of Computer Assisted Tomography, vol. 21, no. 4, pp. 554-566, 1997.

Received on December 29, 2005. Accepted on January 26, 2006.

E-mail address: dpaquin@stanford.edu
E-mail address: dlevy@math.stanford.edu
E-mail address: eduards@stanford.edu
E-mail address: lei@reyes.stanford.edu