


10-1-2009

Statistical Hypothesis Testing for Postreconstructed and Postregistered Medical Images

Eugene Demidenko
Dartmouth College

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>

 Part of the [Bioimaging and Biomedical Optics Commons](#), [Diagnosis Commons](#), and the [Neoplasms Commons](#)

Recommended Citation

Demidenko, Eugene, "Statistical Hypothesis Testing for Postreconstructed and Postregistered Medical Images" (2009). *Open Dartmouth: Faculty Open Access Articles*. 1496.
<https://digitalcommons.dartmouth.edu/facoa/1496>

This Article is brought to you for free and open access by Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Faculty Open Access Articles by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.



Published in final edited form as:

SIAM J Imaging Sci. 2009 October 1; 2(4): 1049–1067. doi:10.1137/080722199.

Statistical Hypothesis Testing for Postreconstructed and Postregistered Medical Images*

Eugene Demidenko[†]

Eugene Demidenko: eugened@dartmouth.edu

[†] Section of Biostatistics and Epidemiology, Dartmouth Medical School and Departments of Mathematics and Computer Science, Dartmouth College, Hanover, NH 03755

Abstract

Postreconstructed and postregistered medical images are typically treated as the raw data, implicitly assuming that those operations are error free. We question this assumption and explore how the precision of reconstruction and affine registration can be assessed by the image covariance matrix and confidence interval, called the confidence eigenimage, using a statistical model-based approach. Various hypotheses may be tested after image reconstruction and registration using classical statistical hypothesis testing vehicles: Is there a statistically significant difference between images? Does the intensity at a specific location or area of interest belong to the “normal” range? Is there a tumor? Does the image require rigid registration? We illustrate statistical hypothesis testing with three examples: breast computed tomography, breast near infrared linear reconstruction, and brain magnetic resonance imaging.

Keywords

breast cancer detection; computed tomography; magnetic resonance imaging; near infrared; confidence eigenimage; image covariance matrix; image comparison; tumor detection

1. Introduction

The use of postreconstructed and postregistered medical images has become routine for clinical diagnostics, treatment, and evaluation. Typically, those images are treated as the raw data, implicitly assuming that those postprocessing techniques are error free and therefore do not affect further image evaluation [2]. The goal of the present work is to explore the impact of image reconstruction and registration on image comparisons and testing through computation of the image covariance matrix and confidence eigenimage. The main assumption of the statistical approach is that the image we look at is *random* and should be considered as one of many possible images. The image may be random because its acquisition is subject to noise (within-subject image variation), images of the same organ may vary between subjects, etc. Therefore, the image should be accompanied by its standard error or other characteristic which reflects how precisely the image at hand has been derived. The statistical approach is widely used for such image reconstruction as positron emission tomography (PET) imaging [37,3], but less for other modalities such as computed tomography (CT) and magnetic resonance imaging (MRI). Perhaps functional MRI is the most advanced in terms of statistical hypothesis testing, especially when it comes to identifying and comparing zones of brain activation [30]. An

*This work was supported by NIH grants RO1 CA130880 and RO1 EB007966.

AMS subject classifications. 62B5, 68U10, 94A08, 93E10, 46N30

advantage of a statistical approach is that it allows the assessment of the uncertainty introduced by reconstruction and registration through computation of the image covariance matrix [16]. Although a statistical model-based image registration has been used in several papers including [8,9,38], it has not been fully exploited in terms of statistical hypothesis testing, such as statistical image comparison after registration, consequence of image registration on abnormality detection, etc.

The goal of this paper is to fully exploit the statistical model for image reconstruction and registration by showing how relevant medical hypotheses may be tested statistically. The culmination of our approach is the computation of the p -value as an indication of statistical significance. This work is a continuation of our development of the statistical approach to image analysis [11], image comparison [12], and inverse problem for electrical impedance tomography [6,13].

As a motivating example, we offer Figure 1 with phantom images, where the image at left is the CT image reconstruction of the breast obtained before treatment with a diagnosed tumor located in the upper left quadrant. The image in the center is taken after chemotherapy treatment. Did treatment help, did the tumor disappear, should the patient undergo surgery? More precisely, is the difference in images statistically significant in the region of interest (ROI) outlined by a white circle? These are typical questions of clinical oncology. Without denying medical doctors' expertise, can image science contribute to those vital decisions? Statistical hypothesis testing may become a very useful tool for objectively comparing medical images in clinical settings. We describe in section 2.1 how to test the difference for statistical significance.

Another example is shown in Figure 6: in the top row, two brain MRI images of the same person are compared. A barely visible condensed mass in the top right corner of the brain seems suspicious. Is that a tumor? In statistical language, is a higher intensity at this location statistically significantly different, or does it belong to the normal range?

We will use these image examples to illustrate various medical hypothesis tests such as Are images before and after treatment the same? and Is there a tumor?

We have deliberately made simplifying assumptions, such as a Gaussian distribution, independence of the noise, and an affine registration. A generalization of the hypothesis testing under more realistic assumptions including a non-Gaussian distribution, Markov random field, and soft registration is needed. The goal of the present paper is mostly *methodological*, but we intend to continue this work with applications to medical images under more realistic assumptions.

It is routine in statistical applications to present an estimate with its standard error and the p -value. It is impossible to publish a paper on a new treatment in a medical journal without presenting the p -value. I hope that someday this will become a routine in imaging science.

The paper is organized as follows. In the next section, we underscore the advantage of a statistical model for image analysis, namely, the ability to assess the precision of the image reconstruction and registration. In particular, we demonstrate how this precision, expressed through a covariance matrix, can be employed to test various hypotheses and compute the p -value. In section 3, we introduce the concept of the confidence eigenimage and apply statistical testing to image reconstruction from projections. We illustrate statistical hypothesis testing with a breast cancer detection and localization example. A real-life example using linearized near infrared image reconstruction for studying optical properties of the breast is discussed in section 3.4. Hypothesis testing for postregistered images with an example of tumor detection using brain MRI is discussed in section 4.

2. Why a statistical model?

In this introductory section, we advocate a statistical modeling approach rather than a criterion or cost-function approach for medical image reconstruction. In particular, we illustrate how to exploit statistical hypothesis testing for postreconstructed images. This type of analysis is underused in applied inverse problems. Hypothesis testing should be an integral part of medical diagnostics.

To simplify, we deal with a linear system with n observations and m unknowns. In the case of a $P \times Q$ image reconstruction, the unknowns are the gray intensities at voxel locations $p = 1, \dots, P$ and $q = 1, \dots, Q$. Combining the $m = PQ$ intensities in an $m \times 1$ vector, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$, we assume that for the i th observation the system output is a linear function $\sum_{j=1}^m x_{ij}\theta_j$ with known coefficients $\{x_{ij}, i = 1, \dots, n, j = 1, \dots, m\}$ that define the hardware configuration. Usually, we use i to denote the observation and j to denote the component of the vector image $\boldsymbol{\theta}$. To make image reconstruction solvable (identifiable), the number of observations should be greater than the number of unknowns, $n > m$. For more detail, we refer the reader to a rich literature on image reconstruction including [24,19,15]. More details on CT reconstruction are given in section 3.

A few comments regarding the notation: We use the boldface font to denote vectors and matrices; we use lowercase for vectors and uppercase for matrices. The Greek letters are used to denote unknown parameters to be estimated from the data. Prime (') means vector or matrix transposition. If \mathbf{a} and \mathbf{b} are $m \times 1$ vector columns, $\mathbf{a}'\mathbf{b} = \sum_{i=1}^m a_i b_i$ is the dot (scalar) product and $\mathbf{a}\mathbf{b}'$ is an $m \times m$ matrix.

Usually, when an approximation is sought, the squared Euclidean norm is used as the discrepancy criterion. For example, to solve the overspecified linear image reconstruction system, one derives the solution from the least squares (LS),

$$\sum_{i=1}^n \left(y_i - \sum_{j=1}^m x_{ij}\theta_j \right)^2 \Rightarrow \min_{\boldsymbol{\theta}}, \quad (2.1)$$

where $\{y_i\}$ are the outcome measurements, $\{x_{ij}\}$ are the elements of the projection matrix, and $\{\theta_j\}$ are the image intensities. Assuming that the system of vectors $\mathbf{x}_i = (x_{i1}, \dots, x_{im})'$ has full rank, we obtain the LS solution for the reconstructed image,

$$\widehat{\boldsymbol{\theta}} = \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i y_i \right). \quad (2.2)$$

In many publications on numerical mathematics and engineering, criterion (2.1) is treated as just a convenient and familiar way of approximating the observations with a theoretical model. But why LS? Why the Euclidean distance? To answer these questions, we need to appeal to statistics. Following the statistical approach, we assume that our observations, y_i , are random variables with mean $\boldsymbol{\theta}'\mathbf{x}_i$. Also, we assume that $\{y_i, i = 1, \dots, n\}$ are independent and have the same variance, σ^2 . These assumptions lead to a statistical model,

$$y_i = \theta' \mathbf{x}_i + \varepsilon_i, \quad i=1, \dots, n, \quad (2.3)$$

where ε_i is the random error term. This model is a well-established classic linear regression model [31]. It is customary to assume that $\{\varepsilon_i\}$ are independent and identically distributed (iid) with the Gaussian distribution, $\mathcal{N}(0, \sigma^2)$. Under this assumption, the maximum likelihood is equivalent to LS (2.1). Moreover, it is a textbook result that the LS solution, $\hat{\theta}$ (*estimator*, in statistical language), has minimum variance among all unbiased estimators of the true θ [28]. Hence, the difference between the numerical criterion (2.1) and the statistical model (2.3) is that the former treats observations $\{y_i\}$ as fixed numbers, while the latter implicitly assumes that $\{y_i\}$ is just one set of possible realizations. Since observations may vary, it is important to determine how the solution (2.2) is sensitive to different observation realizations; see subsection 2.1 for details.

The statistical model-based approach to image reconstruction implies the following. First, use of criterion (2.1) implicitly assumes that observations are continuous, do not correlate, have the same accuracy of measurement (σ), and have a Gaussian distribution. If the distribution is not Gaussian, other criteria would be more efficient. For example, if the distribution is double exponential, one should minimize the sum of absolute residuals, not squared residuals. Second, under the iid assumption, criterion (2.1) is optimal. Specifically, as follows from the Gauss–Markov theorem, the LS solution under model (2.3) is unbiased and has minimal variance among all linear estimators. Moreover, if the distribution of ε_i is Gaussian, the LS estimator is efficient among all unbiased estimators [29]. Third, and perhaps most important, one can take full advantage of a statistical model, such as (2.3), by computing the sensitivity of our image reconstruction—more precisely, the variance-covariance matrix of $\hat{\theta}$ —and subsequent *significance and hypothesis testing*, including the p -value for the images.

2.1. Statistical model and hypothesis testing

Since, under the statistical model, the set of observations $\{y_i\}$ is just one of a myriad of possible realizations, the reconstructed image, $\hat{\theta}$, should be treated as random. Therefore, in addition to the LS estimate itself, one may ask how sensitive $\hat{\theta}$ is to the observation data. An adequate measure of the sensitivity is the derivative, $\partial \hat{\theta} / \partial y_i$. The cumulative sensitivity can be defined as an $m \times m$ matrix:

$$\sum_{i=1}^n \left(\frac{\partial \hat{\theta}}{\partial y_i} \right) \left(\frac{\partial \hat{\theta}}{\partial y_i} \right)'. \quad (2.4)$$

Using the LS formula (2.2), we obtain that the cumulative sensitivity (2.4) can be expressed as

$$\left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right)^{-1}.$$

In fact, this matrix is proportional to the covariance matrix of the LS solution (2.2), namely,

$$\text{cov}(\widehat{\theta}) = \sigma^2 \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right)^{-1}, \quad (2.5)$$

where σ^2 is the variance of the system noise. The j th diagonal element of matrix (2.5) is the variance of the j th component of the reconstructed image. Clearly a smaller sensitivity (variance) yields a more reliable (statistically significant) estimate. We have to admit that computation of the covariance matrix (2.5) and the respective image variance (sensitivity) is a rare encounter in the literature on image reconstruction. As a result, a powerful hypothesis testing technique is wasted.

Before illustrating how the statistical approach can be used for inverse problems, we briefly discuss how to test a general hypothesis in the framework of linear model (2.3), as described in classical statistical texts such as [7] and [28].

The null hypothesis is formulated as a system of k independent linear equations,

$$H_0: \mathbf{H}\theta = \mathbf{q}, \quad (2.6)$$

where \mathbf{H} is a fixed $k \times m$ matrix of full rank ($k < m$) and \mathbf{q} is a fixed $k \times 1$ vector. To test H_0 , we need to compute two sums of squares. The first sum of squares, $S_{LS} = \min_{\theta} \sum (y_i - \theta' \mathbf{x}_i)^2$, is simply the residual sum of squares from the LS solution (2.2). The second sum of squares, $S_0 = \min_{\mathbf{H}\theta = \mathbf{q}} \sum (y_i - \theta' \mathbf{x}_i)^2$, is the minimum sum of squares under the linear restriction $\mathbf{H}\theta = \mathbf{q}$. If the null hypothesis is true, the random variable

$$F = \frac{(S_0 - S_{LS})/k}{S_{LS}/(n-m)} \quad (2.7)$$

has an F -distribution with k and $n - m$ degrees of freedom [14,29]. This means that if α is the significance level (say 5%) and $F_{1-\alpha}$ is the $(1 - \alpha)$ th quantile of the F -distribution with k and $n - m$ degrees of freedom, we reject H_0 if $F > F_{1-\alpha}$. This test has a clear interpretation: If the null is not true, then the restricted sum of squares, S_0 , takes a large value and consequently (2.7) would be large so that we reject H_0 .

Now we describe how to use the F -test to answer different questions in the framework of the linear image-reconstruction problem described by statistical model (2.3). When two images are involved, we assume that they are coregistered, so that a pixel-by-pixel difference is valid.

1. *Is the reconstructed image nothing but background?* This question translates into hypothesis (2.6) with $\theta_j = \text{const}$ or, in matrix form (assuming $m = 5$),

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

Here $k = m - 1$, and S_0 is the minimum sum of squares of the background,

$$S_0 = \sum_{i=1}^n (y_i - \widehat{\theta}_0 \sum_{j=1}^m x_{ij})^2, \text{ where}$$

$$\widehat{\theta}_0 = \frac{\sum_{i=1}^n \sum_{j=1}^m y_i x_{ij}}{\sum_{i=1}^n (\sum_{j=1}^m x_{ij})^2}$$

is the LS estimate under the null hypothesis.

2. Assume two medical images are taken, before and after treatment. See Figure 1 for an illustration (details are given in section 3.2). *Is there any difference between the images?* This question translates into the null hypothesis $H_0: \boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)}$. To test this hypothesis, we denote $\{y_i^{(1)}\}$ and $\{y_i^{(2)}\}$ as the pre- and posttreatment observations, so that each image is $\widehat{\boldsymbol{\theta}}^{(s)} = (\sum \mathbf{x}_i \mathbf{x}_i')^{-1} (\sum \mathbf{x}_i y_i^{(s)})$, $s = 1, 2$. Here, two types of hypothesis testing are available, the Wald test [28,5] and the F -test defined by test statistic (2.7). The former is convenient when only a part of the parameter vector/image from the ROI is tested. To test whether images in the ROI are the same up to some random noise, namely, $\boldsymbol{\theta}_{ROI}^{(1)} = \boldsymbol{\theta}_{ROI}^{(2)}$, we compute the quadratic form

$$(\widehat{\boldsymbol{\theta}}_{ROI}^{(1)} - \widehat{\boldsymbol{\theta}}_{ROI}^{(2)})' (\mathbf{C}_{ROI}^{(1)} + \mathbf{C}_{ROI}^{(2)})^{-1} (\widehat{\boldsymbol{\theta}}_{ROI}^{(1)} - \widehat{\boldsymbol{\theta}}_{ROI}^{(2)}) \sim \chi^2(k),$$

where $\mathbf{C}_{ROI}^{(s)}$ is the covariance matrix of $\widehat{\boldsymbol{\theta}}_{ROI}^{(s)}$ as a submatrix of (2.5) and $\chi^2(k)$ is the chi-distribution with k degrees of freedom, the number of pixels in the ROI. If the statistic on the left is smaller than the critical value of the chi-distribution, we accept the hypothesis: The tumor was killed by the chemotherapy. To obtain statistical significance using the F -test, we find the total sum of squares as $S_{LS} = S_{LS}^{(1)} + S_{LS}^{(2)}$. To compute the residual sum of squares under the null, we combine the data with the resulting image,

$$\widehat{\boldsymbol{\theta}} = \frac{1}{2} \left(\sum \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left(\sum \mathbf{x}_i' (y_i^{(1)} + y_i^{(2)}) \right),$$

and $S_0 = \sum (y_i - \sum x_{ij} \widehat{\theta}_j)^2$. We can test that the images are the same up to illumination. Then the hypothesis involves $m - 1$ independent linear equations,

$$\theta_1^{(1)} - \theta_1^{(2)} = \theta_2^{(1)} - \theta_2^{(2)} = \dots = \theta_m^{(1)} - \theta_m^{(2)}.$$

3. *Is the intensity of the reconstructed image at a specific location normal?* This question translates into hypothesis $H_0: \theta_j = \theta_{j0}$, where θ_{j0} is known (the intensity level of a normal patient at voxel j). This hypothesis is called simple and can be reduced to a statistical significance test: If $\widehat{\theta}_j$ is the reconstructed value, we compute the ratio $t = (\widehat{\theta}_j - \theta_{j0})/s_j$, where s_j is the standard error (SE), the (j, j) th element of the covariance matrix (2.5) with σ^2 substituted by $\widehat{\sigma}^2 = S_0/(n - m)$. Then, under the null hypothesis, t has a Student distribution with $n - m$ degrees of freedom: If $|t| > t_{1-\alpha}$, we reject the null hypothesis with the error probability α , where $t_{1-\alpha}$ is the critical value. Note that if testing is done at several locations, some sort of Bonferroni adjustment should be applied as is routinely required in functional MRI (fMRI) applications [26].

We shall illustrate these hypotheses in the next section.

3. LS reconstruction from projections

In this section, we consider the LS reconstruction in more detail. In particular, we illustrate postreconstructed image inference by statistical hypothesis testing for breast cancer diagnostics.

Reconstruction from projections constitutes the basis of CT. We refer the reader to numerous books on CT including [18,24,33]. A recent book by [15] provides a comprehensive account of the mathematical aspects of image reconstruction with medical applications. In this section, we deal with discrete, sometimes called algebraic, reconstruction [19].

A generic CT device consists of several sources and detectors located on the perimeter of a square or circle. Beams of X-rays or light come out of the sources at a given angle, penetrate the body, and are received at detectors on the opposite side. Following Beer's law of optics, if I_0 is the initial intensity of the beam, which comes in at one end of a homogeneous bar of length x , and I_1 is the intensity at the exit, with a certain degree of approximation we have $I_1 = I_0 e^{-\theta x}$, where θ is called the attenuation coefficient [3]. If a nonhomogeneous bar is composed of m homogeneous bars of lengths x_j with the respective attenuation coefficients θ_j , the intensity at the end is $I_1 = I_0 e^{-\theta' \mathbf{x}}$ or, on the log scale, $y = \theta' \mathbf{x}$, where $y = \ln(I_0/I_1)$. This simple formula gives rise to CT image reconstruction. Imagine that the body is divided into m small boxes and within each box the attenuation coefficient θ_j is constant, $j = 1, \dots, m$. If the beam comes out of the source at a given angle, we can compute the length of the ray within each box, so that we come to a linear statistical model (2.3). Since the beam angles are known, $\{x_{ij}, i = 1, \dots, n, j = 1, \dots, m\}$ are fixed numbers and can be derived from the CT hardware specification. Having n measurements, $\{y_i\}$, we reconstruct (*estimate* in statistical terminology) m attenuation coefficients, $\{\theta_j\}$. Plotting $\{\theta_j\}$ at appropriate locations yields a CT image, so the set of attenuation coefficients, $\{\theta_j\}$, is the *image*. The larger θ_j , the denser the tissue at that location.

3.1. Confidence eigenimage

It is customary to show the error bars or confidence intervals when displaying the data. We argue that besides the reconstructed image itself one needs to show the *precision* of the reconstruction. The confidence eigenimage is a generalization of the confidence interval and serves the same purpose, namely, as a visualization tool for quick statistical significance assessment. However, the generalization of the confidence interval to images is not obvious and admits multiple solutions. Here, we further develop the method of displaying entire images on the confidence ellipsoid started in [35]. In particular, we demonstrate how to visually represent the accuracy of the reconstructed image based on the idea of the eigenimage [34], although we use eigenimage in a different setting.

We rewrite linear model (2.3) in a compact form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad (3.1)$$

where $\mathbf{y} = (y_1, \dots, y_n)'$ is the vector of observations, \mathbf{X} is the $n \times m$ matrix with the i th row \mathbf{x}_i' , and $\boldsymbol{\varepsilon}$ is the error vector. Then the LS solution is $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Recall that the $(1 - \alpha)$ th confidence region for $\boldsymbol{\theta}$ in a multivariate linear regression model (3.1) is an ellipsoid defined as

$$\left\{ \theta \in R^m : (\theta - \widehat{\theta})' (\mathbf{X}' \mathbf{X})^{-1} (\theta - \widehat{\theta}) \leq m \widehat{\sigma}^2 F_{1-\alpha}(m, n - m) \right\}, \quad (3.2)$$

where $\widehat{\sigma}^2$ is an estimate of the variance of the error term,

$$\widehat{\sigma}^2 = \frac{1}{n - m} \|\mathbf{y} - \mathbf{X} \widehat{\theta}\|^2,$$

and $F_{1-\alpha}(m, n - m)$ is the $(1 - \alpha)$ th quantile of the F -distribution with m and $n - m$ degrees of freedom [31]. The confidence ellipsoid covers the unknown parameter vector, θ , with the probability $1 - \alpha$. Typically, α is set to 5%, so we are talking about the 95% confidence ellipsoid. We refer the reader to Figure 2 for a geometrical illustration.

When θ is an image, each point in the confidence ellipsoid is an image as well. We want to show a few “bracket” images from (3.2). Below, we describe these images in terms of principal components. Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ be eigenvalues and $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m$ the corresponding eigenvectors of matrix $\mathbf{X}' \mathbf{X}$, so that λ_1 is the minimal and λ_m the maximal eigenvalue. As follows from eigenvalue decomposition, the m principal components of the confidence ellipsoid are defined as

$$\theta = \widehat{\theta} \pm \widehat{\sigma} \sqrt{\lambda_k^{-1} m F_{1-\alpha}(m, n - m)} \mathbf{p}_k. \quad (3.3)$$

There are $2m$ eigenimages on the boundary of the ellipsoid. The first upper and lower eigenimages correspond to $k = 1$, the second to $k = 2$, etc. One could show only the first few eigenimages, but theoretically other eigenimages may be informative as well.

Typically, Monte Carlo simulations are carried out to get the range of reconstructed images. The difference with our approach is that Monte Carlo simulations give a *sample* of images but the eigenimages give the *bracket* images with the specified confidence probability.

3.2. Breast cancer detection and eigenimage

In this section, we use a breast cancer imaging example to illustrate statistical hypothesis testing and eigenimages. A large body of literature exists on breast computer-aided diagnostics (CAD) using image processing methods for both mammographic and ultrasound images [27]. These techniques rely on either the analysis of the grayscale distribution of the entire breast or a region of interest (ROI) and are based on the observation that tumors and other abnormalities are more dense and look lighter on a mammogram; see [20,17,23] to name a few. In addition, since tumors typically have less regular (rounded) contours, shape analysis can be employed to discriminate benign from cancerous abnormalities. Several CAD breast cancer detection software packages are available on the market, such as *Second Look* (Montreal, Canada) [25, 10]. In our example, we deliberately make simplifying assumptions to illustrate how statistical hypothesis testing can be used for significance testing and p -value computation.

In Figure 1, we show three reconstructed breast images. The breast density is generated on the square (shown within the circle) $x = 0, \dots, 50$, and $y = 0, \dots, 50$, proportional to the Gaussian circular density, namely,

$$\theta(x, y) = 10 + 2e^{-(x-25)^2/(2 \cdot 18.3^2) - (y-25)^2/(2 \cdot 18.3^2)}, \quad (3.4)$$

where the standard deviation is 18.3, which defines the width of the dark spot around the nipple. The tumor is also generated as a Gaussian circular density with the standard deviation 1.4 at 10:30 o'clock with the center approximately one half-radius from the center. We assume that $2 \times 51^2 = 5202$ projections are made and that the observations are generated from model (2.3), where θ is composed of 51^2 true image intensities with variance noise $\text{var}(\varepsilon_i) = (0.4)^2$. Since the difference between maximum and minimum image intensity is approximately 2, the signal-to-noise ratio is $2/0.4 = 5$. The breast image reconstructed by LS (2.2) (at left, *Before treatment*) reveals a tumor. After the first session of chemotherapy treatment (*After treatment*), we want to know whether the treatment has helped—vital information in the decision to continue the treatment or undergo surgery. Visual comparison of the images at left and in the center is inconclusive. We want to test whether the two images are the same using test 2 from section 2.1. In our case $n = 5202$ and $m = 2601$, $k = 2600$. We compute three sums of squares; the first two are the residual sums of squares from the first two images, $S_{LS}^{(1)} = 412.5$ and $S_{LS}^{(2)} = 423.9$. The third, under the alternative, is the total sum of squares $S_{LS} = 412.3 + 423.9 = 836.2$. Under the null, we need to combine observations to obtain the *Combined* image with the sum $S_0 = 1765.5$. These values give $F = 1.11$ with the p -value 0.0009. This means that, with error 9/10,000, the images before and after treatment are different: Treatment was successful.

Now we discuss the confidence eigenimage for the *Before treatment* image; see Figure 3. As the reader can see, the blob is seen on all four images—therefore it is statistically significant. In summary, eigenimages play the same role as error bars: If an object of interest is seen in all eigenimages, its presence is statistically significant.

3.3. Breast cancer localization

In this section, we illustrate a statistical hypothesis test for comparing left and right breast images using a nonlinear tumor model. We refer the reader to Figure 4. The first two images are coregistered CT images of the left and right breasts. (This could equally well be CT images of the same breast taken at different times, say a year apart.) There is a suspicious mass in the lower right quadrant, and the pixel-by-pixel difference reveals it. Is this spot statistically significant? To answer this question, we fit the difference image with a function similar to (3.4):

$$D(x, y) = f(x, y; \theta) + \varepsilon(x, y), \quad (3.5)$$

where $D(x, y)$ is the difference image and

$$f(x, y; \theta) = \theta_1 e^{-(x-\theta_2)^2/\theta_4^2 - (y-\theta_3)^2/\theta_4^2} \quad (3.6)$$

is the Gaussian-type function with $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$, the vector parameter to estimate by nonlinear LS. Noise $\varepsilon(x, y)$ is assumed spatially uncorrelated and homogeneous with Gaussian distribution $N(0, \sigma^2)$. Several papers, including [21,40], consider a robust model-based approach for abnormality detection; we simplify the problem by focusing on the image postprocessing hypothesis testing.

As follows from nonlinear regression analysis [32,4], the covariance matrix of the vector is estimated by

$$\text{cov}(\widehat{\theta}) = \widehat{\sigma}^2 \left[\sum_{x,y=1}^n \left(\frac{\partial f(x,y;\widehat{\theta})}{\partial \theta} \right) \left(\frac{\partial f(x,y;\widehat{\theta})}{\partial \theta} \right)' \right]^{-1}. \quad (3.7)$$

Recall that the SEs for the parameters are the square roots of this matrix.

After five iterations of the Gauss–Newton algorithm, we obtain parameter estimates; see Table 1. There is no tumor in D if and only if $\theta_1 = 0$; that is our null hypothesis. This hypothesis may be tested by computing the Z -statistic and respective p -value. Since the p -value is less than 0.05, we assert that the presence of a tumor is statistically proven.

One can even estimate the volume of the tumor based on the Gaussian model (3.6) as

$$\begin{aligned} V &= \theta_1 \int e^{-(x^2+y^2+z^3)/\theta_4^2} dx dy dz = \pi^{3/2} \theta_1 \theta_4^3 \\ &= \pi^{3/2} \times 0.631 \times 3.817^3 = 195.4. \end{aligned}$$

Assuming that the breast has a 10 cm diameter, the tumor volume is $195.4/(50/10)^3 = 1.563$ cm³.

3.4. Near infrared breast imaging

Near infrared (NIR) breast imaging is an emerging alternative to established imaging technologies such as CT and MRI [36]. The beam of light penetrates the tissue and, by measuring the attenuation of the light intensity and change in the phase at the interior of the breast, the spatial optical properties of the breast are reconstructed. The complete forward model involves the solution of the elliptic partial differential equation which governs the light propagation and scattering. Several linearizations have been proposed; perhaps the simplest is based on Beer's law, previously mentioned in section 3. This implies that after taking the logarithm of the intensity value, recorded at the detectors, one can apply a linear reconstruction. In Figure 5 we show an example of the linearized NIR breast reconstruction on the 8×8 square. The normal breast is squeezed between two plates; each plate contains 8 equidistant sources and detectors at locations shown in the left plot. For example, when source 3 emits the NIR light, detector 14 receives the attenuated light intensity (arrow). The beam of light is sent at each of 16 locations and received at 15 other detectors (the location of sources and detectors are the same; overall there are $n = 15 \times 16 = 240$ data points). Following Beer's law, we assume that the log drop of the intensity is equal to the sum of elements of the 8×8 image matrix that are intersected by the arrow from source to detector. This assumption leads us to the linear model (3.1), where θ is the 64×1 vector of attenuated coefficients and X is the 240×64 design matrix with each entry either 0 or 1.

We ask a basic question: Is the spatial distribution of the light attenuation uniform across the breast or is there an artifact? This question translates into whether the reconstructed image is nothing but background or, in statistical terms, $H_0: \theta_j = \theta_0, j = 1, 2, \dots, 64$ ($k = 63$), as in section 2.1. As follows from the F -test (2.7), we need to compute the residual sum of squares of the unrestricted and restricted models:

$$\mathbf{y} = \theta_0(\mathbf{X}\mathbf{1}) + \boldsymbol{\varepsilon},$$

where $\mathbf{1}$ is the 64×1 vector filled with 1. We have $S_{LS} = 29.19$ and $S_0 = 311.76$, which yields the F -statistic:

$$F = \frac{(311.76 - 29.19)/63}{29.19/(240 - 64)} = 27.04.$$

This value is larger than the 95% quantile of the F -distribution with degrees of freedom 63 and 176, meaning that we reject the null hypothesis that the optical properties of the breast are homogeneous. We notice that the breast tissue has lesser attenuation coefficients in the center than at the sides. We attribute this to the fact that the breast was squeezed; since breast is a soft tissue, the maximum density is at the sides. This phenomenon should be taken into account when NIR is used to detect breast abnormalities [36]. Other statistical testing may be of interest: Are the attenuation coefficients the same for 1 and 8; 1, 2, 3; 6, 7, 8; etc.?

4. Postregistered image analysis

In the previous discussion, we assumed that images are perfectly registered, so that the pixel-by-pixel comparison and image difference are valid. Usually, medical images should be registered before comparison. When comparing the post-registered images, the fact that the images were registered is usually ignored. We, however, argue that image registration brings in errors and those errors should be taken into account when images are compared afterwards. Following the line of our approach, we make a statistical model for image registration that allows the computation of the respective errors and rigorously carry out hypothesis testing. The aim of this section is to develop a sound statistical methodology. To simplify, two-dimensional images of the same size with affine/rigid registration are considered.

Let $M[p, q]$ and $N[p, q]$ be two $P \times Q$ images (or more precisely, matrices of image intensities). A comment regarding the notation: We use the brackets to indicate that the image is a function of indices treated as continuous arguments (for example, after linear interpolation). Assuming affine image transformation, we want to find six parameters, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_6)$, from minimization of the sum of squares:

$$\sum_{p=1}^P \sum_{q=1}^Q (M[p, q] - N[\theta_1 + \theta_2 p + \theta_3 q, \theta_4 + \theta_5 p + \theta_6 q])^2. \quad (4.1)$$

This criterion can be derived from nonlinear regression model (3.5), which in the case of the image registration problem is written as

$$M[p, q] = N[\theta_1 + \theta_2 p + \theta_3 q, \theta_4 + \theta_5 p + \theta_6 q] + \varepsilon[p, q], \quad (4.2)$$

where $n = PQ$. In (4.1) it is implicitly assumed that images differ by random noise with zero mean and constant variance. One can use the simpler assumption that the original image M and the registered image N differ by random noise up to illumination and contrast:

$$M[p, q] = \alpha + \beta N[\theta_1 + \theta_2 p + \theta_3 q, \theta_4 + \theta_5 p + \theta_6 q] + \varepsilon[p, q]. \quad (4.3)$$

Here α and β are called the illumination and contrast parameters. It is elementary to prove that after parameters α and β are eliminated with linear LS formulas, the criterion (4.1) reduces to maximization of the correlation coefficient between $M[p, q]$ and $N[\theta_1 + \theta_2 p + \theta_3 q, \theta_4 + \theta_5 p + \theta_6 q]$. The latter criterion is called the *similarity criterion* in the literature on image registration [22]. In other words, the similarity criterion is equivalent to the LS criterion allowing images to have different illumination and contrast.

To estimate the unknown registration parameter vector, $\hat{\theta}$, we apply a minimization algorithm for (4.1) or (4.3), such as Gauss–Newton or Levenberg–Marquardt. As a word of caution, these algorithms require a continuous and differentiable regression function which is not available in this case because N is defined on discrete values. Although continuity may be achieved by a linear interpolation and differentiability may be achieved by a nonlinear interpolation, the computational burden increases. A derivative-free algorithm for image registration was developed in [11]. We, however, pretend that matrix N is a continuous and differentiable function and therefore standard results on nonlinear regression apply. Similarly to the CT reconstruction problem of section 3, the statistical approach based on model (4.2) implies that the LS estimate, $\hat{\theta}$, is a random variable with a covariance matrix similar to (3.7):

$$\text{cov}(\hat{\theta}) = \hat{\sigma}^2 \left[\sum_{p=1}^P \sum_{q=1}^Q \left(\frac{\partial N[p, q]}{\partial \theta} \right) \left(\frac{\partial N[p, q]}{\partial \theta} \right)' \right]^{-1}, \quad (4.4)$$

where $\hat{\sigma}^2$ is the estimate of the variance noise computed as the residual sum of squares divided by the degrees of freedom, $PQ - 6$. With no interpolation, one needs to approximate the derivative by finite difference using a generic formula $(N(x+h) - N(x))/h$, where h is the step.

The statistical model for image registration, (4.2) or (4.3), may be used for various hypothesis tests as in section 2.1. Two types of general tests may be used to test (2.6): the Wald test, which is based on the covariance matrix (4.4), or the likelihood ratio test, which does not involve derivative computation [28,7]. Letting $\mathbf{C} = \text{cov}(\hat{\theta})$ in the Wald test, we compute

$$(\mathbf{q} - \mathbf{H}\hat{\theta})' (\mathbf{HCH}')^{-1} (\mathbf{q} - \mathbf{H}\hat{\theta}) \sim \chi^2(k), \quad (4.5)$$

meaning that under the null hypothesis the left-hand side is distributed approximately as the chi-squared distribution with k degrees of freedom. For a special case when $k = 1$, $H = 1$, and $q = \theta_0$, the distribution $Z = (\hat{\theta}_j - \theta_0) / \sqrt{C_{jj}}$ is approximately normal with zero mean and unit variance. In the likelihood ratio test, we do not compute the derivatives but do an additional registration under hypothesis (2.6). Then, if S is the residual sum of squares (4.1) and S_0 is the residual sum of squares under the restriction $\mathbf{H}\theta = \mathbf{q}$, we have

$$n \ln \frac{S_0}{S} \sim \chi^2(k). \quad (4.6)$$

For the linear statistical model, the F -test (2.7) and the χ^2 -test (4.6) are asymptotically equivalent. Since we deal with a nondifferentiable function, the likelihood ratio may be more attractive. Also, this test is convenient when testing a nonlinear hypothesis.

The following hypothesis tests may be of interest in the framework of image registration.

1. *Do we need a registration at all?* Let some registration software give a set of theta-parameters with fairly small values. Maybe they are, in fact, all zeros. So is no registration required? This question translates into the statistical hypothesis $H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = \theta_6 = 0$.
2. *Is only translation involved?* The null hypothesis can be expressed as $H_0: \theta_2 = \theta_3 = \theta_5 = \theta_6 = 0$. Two methods for testing this hypothesis may be suggested. By the Wald test, we evaluate the covariance matrix (4.4) at $\theta = \hat{\theta}$ and compute $\eta = \tau' C_{\tau}^{-1} \tau$, where τ is the 4×1 vector with components $\theta_2, \theta_3, \theta_5, \theta_6$ and C_{τ} is a 4×4 symmetric submatrix of matrix (4.4) with rows and columns 2, 3, 5, and 6. Under the null, η has a chi-squared distribution with four degrees of freedom. Thus, a large value for η would indicate that more than just translation is involved. This hypothesis can be tested via the likelihood ratio test. The same approach can be applied to model (4.3) to allow for different image illumination and contrast.
3. *Does the image registration involve only translation, rotation, and resizing?* This question translates into $H_0: \theta_3 = -\theta_5, \theta_2 = \theta_6$.
4. *Is registration rigid?* To answer this question we introduce squared terms and test whether the respective coefficients are zero.

We illustrate some of these tests with brain MRI images in the following example.

4.1. Example: Brain MRI imaging registration

We illustrate statistical hypothesis testing for postregistered medical images with clinical brain MRI images; see Figure 6. Two brain images of the same person are taken approximately three years apart. The pixel-by-pixel image difference reveals that the images should be aligned before comparison. After affine registration a dense mass is seen. Is this a brain tumor?

In Table 2, we show image registration statistics. Also, we show the step we used to compute the derivative. The SE for each coefficient is computed as the square root of the diagonal matrix (4.4), where the residual variance is $\hat{\sigma}^2 = 138.3$. Column q ($=\theta_0$) shows the appropriate hypothesis testing. For example, the first hypothesis tests whether there is no horizontal translation or, in statistical terms, whether $\hat{\theta}_1$ is statistically significant. Since the Z-statistic value is larger than 3, we conclude that the horizontal translation is statistically significant. Indeed, all registration coefficients are statistically significant.

Now we answer question 3 from the previous section. For this hypothesis $k = 2$,

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad \text{and } \mathbf{q} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since the left-hand side of (4.5) is greater than 5.99, the critical value for the chi-squared distribution with two degrees of freedom, we infer that the registration involves more than just translation, rotation, and resizing.

Now we come to the question posed at the beginning of this section, Is there a tumor? We test this hypothesis by taking the intensity of the difference after registration in the middle of the suspicious region: 39. Since the difference between the images in the top row after registration is approximately a white noise with mean 0 and variance 138.3, we compute $Z = 39/11.8 = 3.3$. Under the null hypothesis that there is no tumor, we have $Z \sim \mathcal{N}(0, 11.8^2)$ and therefore we reject the hypothesis that the suspicious spot looks normal with error

$$\frac{2}{\sqrt{2\pi}} \int_{3.3}^{\infty} e^{-t^2/2} dt = 0.001,$$

the false positive rate of tumor detection.

We should mention that the one-point image comparison should have a stringent p -value when searching for a localized difference. There is plenty of literature on this topic in the framework of multiple comparison going back to [1]; a recent reference in the context of brain imaging is [39]. Typically, the threshold is about 4–5, much higher than the commonly used 1.96, which corresponds to a 5% type I error. To test whether the image differs from the background with a higher Z -value, we could take a number of points within a circle and test whether the mean is statistically different from 0.

5. Summary points and discussion

We advocate a statistical model-based approach to image reconstruction and registration. The main advantage of a statistical model is that the accuracy of the solution, the postprocessed image, can be assessed via a covariance matrix and confidence eigenimages. Although many authors in the field of image reconstruction use a model-based approach, they do not exploit the full potential this approach brings, namely, assessing how the solution is sensitive to the noise.

In this paper, the statistical theory of linear and nonlinear regression analysis was used to compute the image covariance matrix, which is pivotal to image comparison. Various hypotheses may be tested using the classic vehicle of statistical hypothesis testing: Are images the same? Does a suspicious spot belong to the “normal” range of the intensity values? This paper touches on the basics of statistical image analysis. More work should be done to make image comparison a valuable computer-assisted tool in the clinical setting when a longitudinal ensemble of images is available, to account for patient age, or when images are compared across different modalities. Our examples are purely illustrative; more serious image hypothesis testing is a topic for the future.

Acknowledgments

We are grateful to Dr. Subhadra Srinivasan of the Thayer School of Engineering at Dartmouth for providing the data on NIR breast imaging.

References

1. Adler, RJ. *The Geometry of Random Fields*. Wiley; New York: 1981.
2. Andrews, HC.; Hunt, BR. *Digital Image Restoration*. Prentice–Hall; Englewood Cliffs, NJ: 1997.
3. Barrett, HH.; Myers, KJ. *Foundations of Image Science*. Wiley; New York: 2004.
4. Bates, DM.; Watts, DG. *Nonlinear Regression and Its Applications*. Wiley; New York: 1988.
5. Bickel, PJ.; Doksum, KA. *Mathematical Statistics*. 2. Vol. 1. Prentice–Hall; Upper Saddle River, NJ: 2001.
6. Borcea, L. *Inverse Problems*. Vol. 18. 2002. Electrical impedance tomography; p. R99-R136.
7. Casella, G.; Berger, RL. *Statistical Inference*. Duxbury Press; New York: 2001.
8. Christensen GE, Johnson HJ. Consistent image registration. *IEEE Trans Med Imaging* 2001;20:568–582. [PubMed: 11465464]
9. Christensen GE, Sarang CJ, Miller MI. Volumetric transformation of brain anatomy. *IEEE Trans Med Imaging* 1997;16:864–877. [PubMed: 9533586]

10. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: Prospective study of 104 cancers. *Am J Roentgenol* 2006;187:20–28. [PubMed: 16794150]
11. Demidenko, E. *Mixed Models: Theory and Applications*. Wiley; New York: 2004.
12. Demidenko, E. *Computational Science and Its Applications, Lecture Notes in Comput Sci. Vol. 3046*. Springer; Berlin, Heidelberg: 2004. Kolmogorov-Smirnov test for image comparison; p. 933-939.
13. Demidenko E, Hartov A, Paulsen K. Statistical estimation of resistance/conductance by electrical impedance tomography measurements. *IEEE Trans Med Imaging* 2004;23:829–838. [PubMed: 15250635]
14. Draper, NR.; Smith, H. *Applied Regression Analysis*. Wiley; New York: 1998.
15. Epstein, CL. *Introduction to the Mathematics of Medical Imaging*. Pearson Education; Upper Saddle River, NJ: 2003.
16. Evans SE, Stark PB. Inverse problems in statistics. *Inverse Problems* 2002;18:R55–R97.
17. Fischer U, Hermann KP, Baum F. Digital mammography: Current state and future aspects. *Eur Radiol* 2006;16:38–44. [PubMed: 16132935]
18. Hall, EL. *Computer Image Processing and Recognition*. Academic Press; New York: 1979.
19. Herman, GT. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press; New York: 1980.
20. Houssami N, Irwig L, Ung O. Review of complex breast cysts: Implications for cancer detection and clinical practice. *ANZ J Surg* 2005;75:1080–1085. [PubMed: 16398815]
21. Mahadevan V, Narasimha-Iyer H, Roysam B, Tanenbaum HL. Robust model-based vasculature detection in noisy biomedical images. *IEEE Trans Inform Technol Biomed* 2004;8:360–376.
22. Maintz JBA, Viergever MA. A survey of medical image registration. *Med Image Anal* 1998;2:1–36. [PubMed: 10638851]
23. Mendez AJ, Tahoces PG, Lado MJ, Souto M, Vidal JJ. Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms. *Med Phys* 1998;25:957–964. [PubMed: 9650186]
24. Parker, JA. *Image Reconstruction in Radiology*. CRC Press; Boca Raton, FL: 1990.
25. Pisano ED, Yaffe MJ. Digital mammography. *Radiology* 2005;234:353–362. [PubMed: 15670993]
26. Poldrack RA, Fletcher PC, Henson RN, Worsley KJ, Brett M, Nichol TE. Guidelines for reporting an fMRI study. *NeuroImage* 2008;40:409–414. [PubMed: 18191585]
27. Rangayyan, R. *Biomedical Image Analysis*. CRC Press; Boca Raton, FL: 2005.
28. Rao, CR. *Linear Statistical Inference and Its Application*. Wiley; New York: 1973.
29. Rao, CR.; Toutenberg, H. *Linear Models. Least Squares and Alternatives*. Springer; New York: 1995.
30. *Statistical Parametric Mapping (SPM)*. 2009. <http://www.fil.ion.ucl.ac.uk/spm>
31. Searle, SR. *Linear Models*. Wiley; New York: 1973.
32. Seber, GAF.; Wild, CJ. *Nonlinear Regression*. Wiley; New York: 1989.
33. Seeram, E. *Computed Tomography: Physical Principles, Clinical Applications and Quality Control*. W. B. Saunders; Philadelphia: 1974.
34. Soltenian-Zade H, Saigal R, Windham JP, Yagle AE, Hearshen DO. Optimization of MRI protocols and pulse sequence parameters for eigenimage filtering. *IEEE Trans Med Imaging* 1994;13:161–175. [PubMed: 18218494]
35. Tosteson TD, Pogue BW, Demidenko E, Troy OM, Paulsen K. Confidence maps and confidence intervals for near infrared images in breast cancer. *IEEE Trans Med Imaging* 1999;18:1188–1193. [PubMed: 10695531]
36. Tromberg BJ, Pogue BW, Paulsen KD, Yodh AG, Boas DA, Cerussi AE. Assessing the future of diffuse optical imaging technologies for breast cancer management. *Med Phys* 2008;35:2443–2451. [PubMed: 18649477]
37. Vogel, CR. *Computational Methods for Inverse Problems*. SIAM; Philadelphia: 2002.
38. Wang HS, Feng F, Yeh E, Huang SC. Objective assessment of image registration results using statistical confidence intervals. *IEEE Trans Nuclear Sci* 2001;48:106–110.

39. Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. Unified univariate and multivariate random field theory. *NeuroImage* 2004;23:189–195.
40. Zhang, X-Q.; Froment, J. *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Lecture Notes in Comput Sci. Vol. 3757. Springer; Berlin, Heidelberg: 2005. Constrained total variation minimization and application in computerized tomography; p. 456-472.

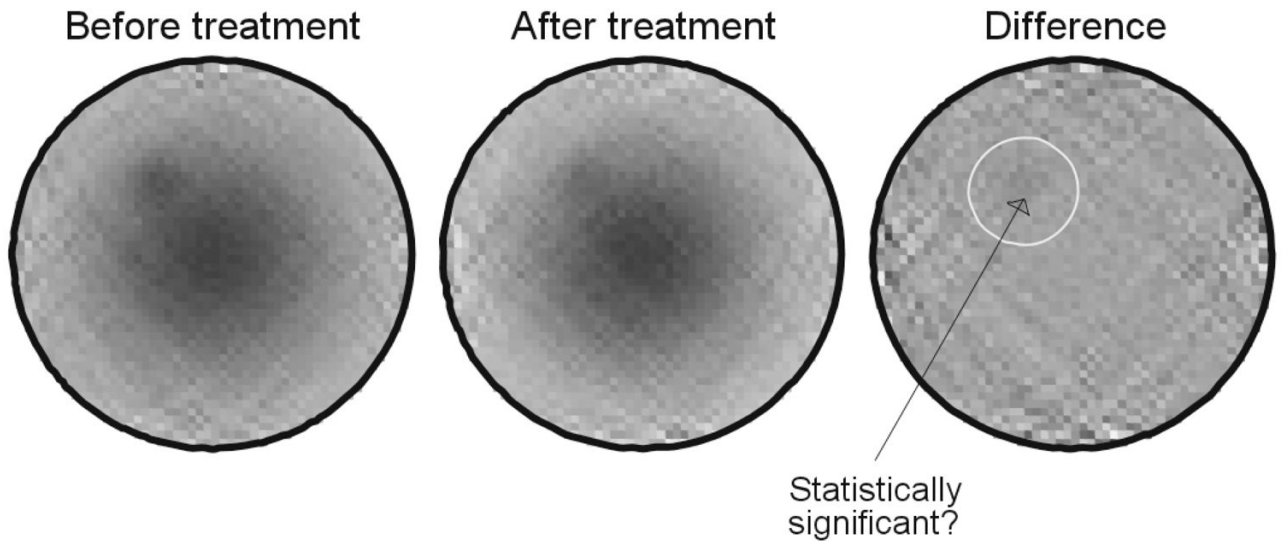


Figure 1.

Three CT breast phantom images reconstructed from projections by least squares. In the *Before treatment* image, a dark spot indicating the presence of a tumor is visible in the upper left quadrant. In the *After treatment* image, the same breast is reconstructed after the chemotherapy treatment. A dark spot appears to be seen in the *Difference* image. Is the difference statistically significant in the region of interest (white circle)? To answer this question one needs to know the precision with which the images are obtained.

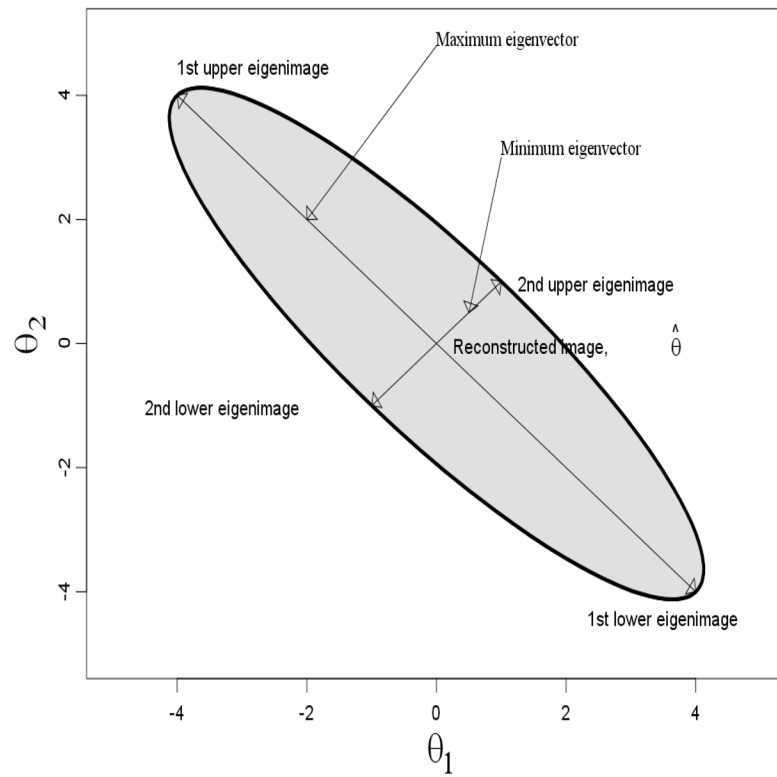


Figure 2. Four confidence eigenimages correspond to the boundary points of the confidence ellipsoid.

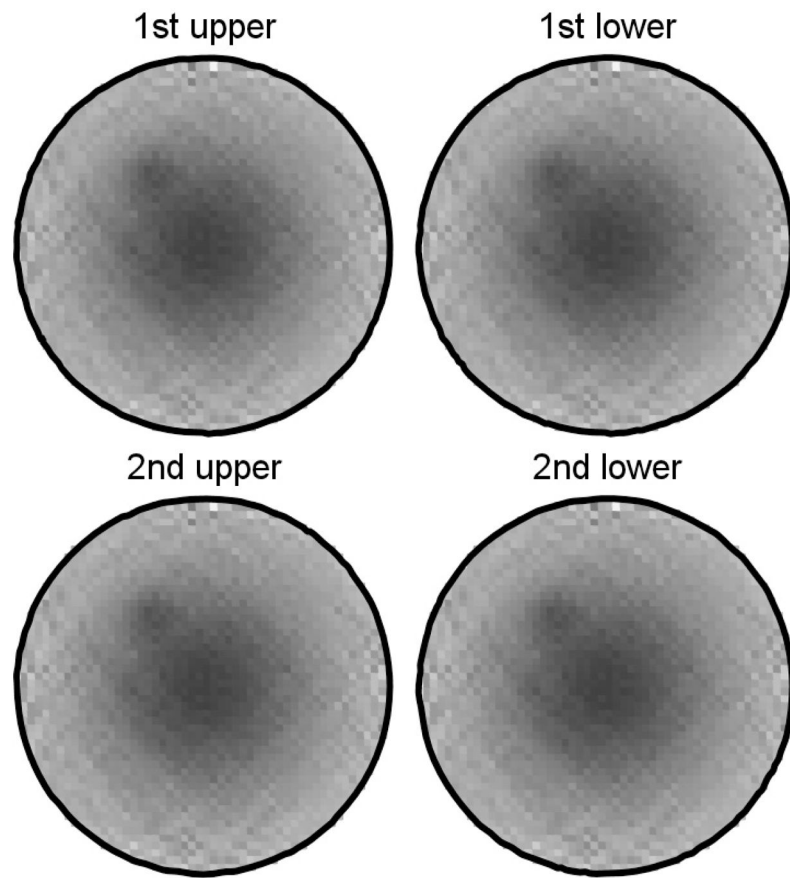


Figure 3. Four confidence eigenimages for the *Before treatment* breast images. The blob is seen on all images; therefore it is statistically significant.

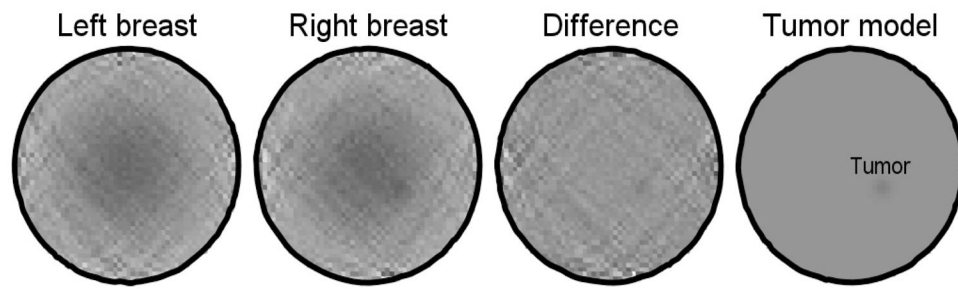


Figure 4.
Breast tumor localization using a nonlinear regression model.

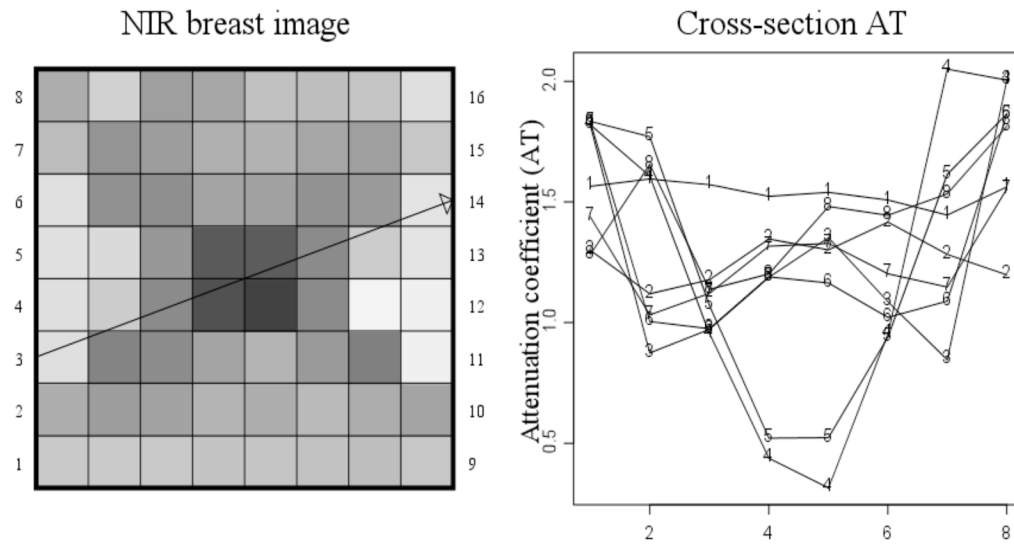


Figure 5. Linearized NIR 8×8 breast image reconstruction with two arrays of sources and detectors. The darker the image cell, the lower the attenuation coefficient. These values are depicted in the right plot as the curve for horizontal cross-section attenuation coefficient; the number indicates the source index in the left image.

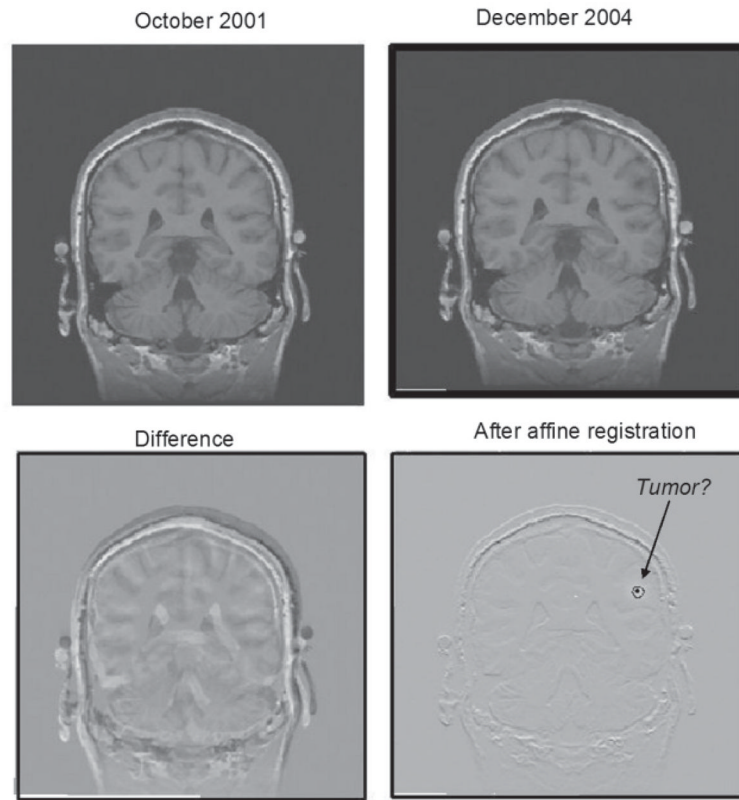


Figure 6. Brain tumor diagnostic comparing the MRI images of the same person taken three years apart. The difference does not reveal any suspicious regions. After affine registration, a suspicious blob is seen in the upper right brain. Is this a tumor?

Table 1

Estimation of nonlinear tumor regression (3.5).

Parameter	θ_1	θ_2	θ_3	θ_4	σ^2
Estimate	0.631	20.890	32.790	3.817	0.222
SE	0.272	0.596	0.596	2.329	
Z-statistic	2.327	35.040	55.010	1.638	
p-value	0.021	0.000	0.000	0.101	

Table 2

Affine brain MRI image registration results.

Estimate	Step, h	SE	q	Z
$\hat{\theta}_1 = 3.45$	2	0.014	0	239
$\hat{\theta}_2 = 1.015$	0.01	6.1×10^{-5}	1	242
$\hat{\theta}_3 = -0.008$	0.01	7.2×10^{-5}	0	-110
$\hat{\theta}_4 = -6.78$	2	0.014	0	-486
$\hat{\theta}_5 = 0.980$	0.01	7.8×10^{-5}	1	-257
$\hat{\theta}_6 = 0.010$	0.01	8.0×10^{-5}	0	129