HHS Public Access

Author manuscript

IEEE Trans Med Imaging. Author manuscript; available in PMC 2018 October 01.

Published in final edited form as:

IEEE Trans Med Imaging. 2017 October ; 36(10): 1997–2009. doi:10.1109/TMI.2017.2725644.

Effects of Image Quality on the Fundamental Limits of Image Registration Accuracy

Michael D. Ketcha,
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205 USA

Tharindu De Silva,
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21205 USA

Runze Han,
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21205 USA

Ali Uneri,
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205 USA

Joseph Goerres,
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21205 USA

Matthew Jacobson,
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21205 USA

Sebastian Vogt,
Siemens Healthcare XP, Erlangen 91052, Germany

Gerhard Kleinszig, and
Siemens Healthcare XP, Erlangen 91052, Germany

Jeffrey H. Siewerdsen
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205 USA

Abstract

For image-guided procedures, the imaging task is often tied to the registration of intraoperative and preoperative images to a common coordinate system. While the accuracy of this registration is a vital factor in system performance, there is relatively little work that relates registration accuracy to image quality factors such as dose, noise, and spatial resolution. To create a theoretical model for such a relationship, we present a Fisher information approach to analyze registration performance in explicit dependence on the underlying image quality factors of image noise, spatial resolution, and signal power spectrum. The model yields analysis of the Cramer-Rao lower bound, CRLB, in registration accuracy as a function of factors governing image quality. Experiments were performed in simulation of CT low-contrast soft tissue images and high-contrast bone (head and neck) images to compare the measured accuracy [root mean squared error (RMSE) of the estimated transformations] with the theoretical lower bound. Analysis of the CRLB reveals that registration performance is closely related to the signal-to-noise ratio of the cross-correlation

Correspondence to: Jeffrey H. Siewerdsen.

space. While the lower bound is optimistic, it exhibits consistent trends with experimental findings and yields a method for comparing the performance of various registration methods and similarity metrics. Further analysis validated a method for determining optimal post-processing (image filtering) for registration. Two figures of merit (CRLB and RMSE) are presented that unify models of image quality with registration performance, providing an important guide to optimizing intraoperative imaging with respect to the task of registration.

Index Terms

Image-guided treatment; Image quality assessment; Registration; X-ray imaging and computed tomography

I. Introduction

In image-guided interventions, registration performance pertains to the accuracy with which the preoperative image (denoted I_1) and intraoperative image (denoted I_2) can be co-registered in a common coordinate system. In many scenarios, the ability to accurately register I_2 to I_1 (and planning data therein) may be even more important than the ability to visualize structures in I_2 directly. For example, Uneri *et al.* [1] developed a registration method to evaluate surgical screw placement relative to preoperative CT, enabling quantitative evaluation of screw malplacement in 3D rather than qualitative visualization interpretation of 2D projection radiographs. In this scenario, the primary task is accurate registration to CT (and overall perceptual image quality in the radiographs is of secondary importance). In related work, Uneri *et al.* [2] reported that accurate registration could be achieved even when the radiograph was acquired at a dose ~1/10th that of standard technique, indicating that the task of registration may be more robust against noise than the task of visualization; hence, imaging parameters that are optimal for visualization may not correspond to those that are optimal for registration. The development of imaging systems for interventional guidance therefore prompts consideration that the optimal I_2 imaging technique (i.e., factors governing image noise and spatial resolution) is that which provides a desired level of registration accuracy, rather than visual image quality. This consideration in turn motivates a quantitative framework to relate registration accuracy to image quality.

A considerable body of ongoing research seeks to establish reliable relationships between image quality and detection tasks [3], [4], helping to design new systems [5], [6], determine optimal imaging techniques [7], and substantiate claims of low-dose imaging performance [8], [9]. For example, in computed tomography (CT) and cone-beam CT (CBCT), the image quality characteristics are well characterized by information-theoretic descriptors such as noise-equivalent quanta (NEQ), which encompasses spatial-frequency-dependent descriptors of spatial resolution (modulation transfer function, MTF) and image noise (noise-power spectrum, NPS). Moreover, such analysis provides knowledgeable guidance of image acquisition and reconstruction techniques with respect to tasks of detection and visualization [10]–[12]. However, while CT and CBCT images are increasingly employed in image-guided interventions, there is comparatively little rigorous understanding of the relationship between these image quality factors and registration performance, leaving

largely unanswered fundamental questions in determining imaging techniques that achieve a desired level of registration accuracy.

Registration performance is commonly investigated by rigorous measurement and experimental evaluation of geometric accuracy in contexts appropriate to a particular application. Such investigation often involves registration repeated for either a large data set or simulated noise realizations, where the output transformation parameters are compared to the ground truth transformation. Results from such experiments provide an important basis for quantifying performance in support of the clinical application; however, they are still often performed using the general understanding that a "higher quality" image will give better registration performance– or that a level of image quality sufficient for visualization will in turn be sufficient for registration–without rigorous guidance of an analytical model for registration. As a result, there are untested opportunities for imaging methods that are best suited for the task of registration–e.g., methods that achieve a desired level of registration accuracy with reduced radiation dose.

We seek an analytical framework that will help to unify models of image quality (e.g., spatial resolution and noise) with models for registration performance, providing a rigorous basis and guide to selection of image acquisition protocols, reconstruction methods, and post-processing techniques sufficient (or optimal) for the task of registration. We approach the question by analyzing a simple model involving 2D translation-only registration to gain initial insight into the more complicated general registration problem. Future work will consider translation+rotation, affine, and potentially deformable registration. As detailed below, we build from well-established, image quality considerations for CT/CBCT image quality [13]–[15] and realize a framework that relates these factors to the task of image registration. While this framework is general to the broader field of image registration, it is presented here specifically in the context of medical imaging, relating registration performance to concepts of image noise, spatial resolution, and information-theoretic metrology (viz., image NPS, MTF, and NEQ) that are familiar and prevalent in medical image quality assessment – particularly in x-ray CT and CBCT.

II. Background Methods

Robinson and Milanfar [16] proposed a fundamental limit on registration accuracy by deriving the Cramer-Rao lower bound (CRLB) for 2D translation between two images. Their model considered registering images of the same content in the presence of additive white Gaussian noise (AWGN) and was further extended to rigid registration (including rotation) and 3D registration [17]. To examine the effect of spatial resolution on registration performance, Zhao *et al.* [18] similarly considered the AWGN assumption, finding that optimal performance could be achieved by blurring the higher-resolution image to match the MTF of the lower-resolution image. In the context of image-guided interventions, the framework for such lower bounds and spatial resolution analysis requires modification to account for both non-white noise (e.g., correlated noise that is intrinsic to CT and CBCT) as well as strong differences in image quality characteristics between the preoperative and intraoperative image (i.e., the I_2 CBCT typically exhibiting higher noise than the I_1 CT image). We therefore consider the following model for registration:

$$I_1[x,y]=g(x,y)+n_1(x,y) \quad (1)$$

$$I_2[x,y]=g(x-u,y-u)+n_2(x,y) \quad (2)$$

where g is the true underlying image function, and n_i are Gaussian additive noise terms that are independent of the signal and each other. The image I_2 is formed with a translation-only displacement of g, with the transformation $\boldsymbol{\theta} = [u, v]$ representing the unknown translation between the two images. The [·] bracket notation represents the process of discretely sampling the continuous underlying image functions. By convention, we use the discrete form using [·] when referring to the image domain and the continuous form (·) for the underlying spatial domain image functions or functions in the frequency domain.

III. Derivation of the Cramer-Rao lower bound (CRLB) For image registration

The CRLB is of fundamental interest to many parameter estimation problems, as it provides a theoretical statistical limit on the expected error of the estimator. In the case that the estimator is unbiased, the CRLB matrix (C_{LB}) is simply the inverse of the Fisher Information Matrix (FIM) which is derived from the log-likelihood function, log $L(\boldsymbol{I}|\boldsymbol{\theta})$, of the data conditioned on the parameter vector $\boldsymbol{\theta}$. The FIM examines the curvature of the likelihood function with respect to changes in $\boldsymbol{\theta}$, representing the intuitive concept that if the likelihood function is highly sensitive to perturbations of $\boldsymbol{\theta}$, then $\boldsymbol{\theta}$ can be estimated more accurately. Note that the FIM itself is independent of the estimator and bias. By definition [19], we have:

$$[FIM]_{ij}=E\left\{\frac{\partial \log(I|\theta)}{\partial \theta_i}\frac{\partial \log(I|\theta)}{\partial \theta_j}\right\} =-E\left\{\frac{\partial^2 \log L(\boldsymbol{I}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right\} \quad (3)$$

where the second equality holds under conditions satisfying interchangeable differentiation and integration of the log-likelihood function. Taking a 2D model for simplicity of notation, $\boldsymbol{I}$ refers to the image data, and $\boldsymbol{\theta}$ is the vector of transformation parameters given by $[u, v]^t$ for the translation-only case.

A. CRLB for Signal-Known-Exactly (SKE) With White Noise

As shown in [20], the FIM result presented in [16] applies to the simplified case of a noiseless I_1 [simplified from (1–2)] and therefore the signal-known-exactly (SKE) scenario (i.e., exactly known $g[x, y] = I_1[x, y]$) described by:

$$I_1[x,y]=g(x,y) \quad (4)$$

$$I_2[x,y]=g(x-u,y-u)+n_2(x,y) \quad (5)$$

in which only the shifted image is considered to be contaminated by AWGN with variance σ^2. As shown in [20], the subtraction of the images at the true shift leaves only the AWGN noise term; therefore, $L(\boldsymbol{I}|\boldsymbol{\theta})$ is simply the product of Gaussian probability density functions, giving:

$$FIM_{SKE,AWGN}=C_{LB}^{-1}=\frac{1}{\sigma^2}\begin{bmatrix}\sum_{x,y}g_x^2 & \sum_{x,y}g_xg_y\\ \sum_{x,y}g_xg_y & \sum_{x,y}g_y^2\end{bmatrix} \quad (6)$$

where $g_x[x, y]$, $g_y[x, y]$ are the partial derivative images with respect to x, y. Examining the FIM, we see that registration performance depends generally on two primary components: (i) the image noise (i.e., variance, which is governed largely by image acquisition technique factors such as the level of radiation dose); and (ii) the sum-of-squared image gradients (which are governed by the contrast and frequency content of the subject). While this formulation provides useful basic insight, it is limited in that it does not account for the presence of noise in both the I_1 fixed and I_2 moving images; nor does it account for correlated noise. In the following section, we address this limitation by deriving the lower bound for the scenario in which the noise terms have different magnitude and frequency content.

B. CRLB for Image Registration

A typical scenario in image-guided procedures involves the registration of a high-quality preoperative image to a lower-quality (noisy and/or blurry) intraoperative image, requiring an alternate formulation of the FIM to explicitly address the presence of noise in both images and allowing the noise in each image to carry disparate (heteroscedastic) magnitude and correlation. We assume linear systems having stationary jointly Gaussian signal and noise for the setting of (1–2), where n_i is zero-mean and independent of g and $n_{j\neq i}$ - a common starting point in formulation of image statistics that can be extended to 'local' approximation for nonlinear, nonstationary systems (discussed below and in [11] and [21]). Appendix A demonstrates a proof analogous to that derived for 1D time-delay estimation (TDE) in [22]–[24] to achieve the following key result, which for the purposes of analytical tractability is derived in the frequency domain:

$$FIM = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$
$$\text{where } \gamma_{ij} = \int\int_{-f_{Nyq}}^{f_{Nyq}} f_i f_j \frac{G^2}{GN_1 + GN_2 + N_1 N_2} df_x df_y \quad (7)$$

where $G(f_x, f_y)$ is the power spectrum of $g(x, y)$, the $N_i(f_x, f_y)$ are the NPS of images I_i, A is the image area, and f_{Nyq} refers to the Nyquist frequency. For notational convenience, we leave out the explicit frequency (f_x, f_y) dependence in G and N_i. Equation (7) provides the necessary framework to analyze registration performance bounds when the two images have separate noise forms that are not necessarily white. We rewrite (7) as follows to more explicitly show the dependence of Fisher information on the image signal and noise:

$$FIM = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$
$$\text{where } \gamma_{ij} = \int\int_{-f_{Nyq}}^{f_{Nyq}} f_i f_j SNR df_x df_y \quad (8)$$

$$\text{and } SNR\ (f_x, f_y) = \frac{G^2}{GN_1 + GN_2 + N_1 N_2} \quad (9)$$

Equation (9) carries intuitive dependencies of registration performance on SNR, namely: the numerator scales with signal feature strength (i.e., contrast and gradient magnitude), and the denominator scales with image noise, including cross terms corresponding to noise in cross-correlation (examined further in section IV).

One can see that the general form in (7) reduces to that of (6) in the simplified case of noiseless I_1 ($N_1 = 0$) and AWGN (N_2 is "white" with magnitude σ^2). From (7) this suggests:

$$FIM_{SKE,AWGN} = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$
$$\text{where } \gamma_{ij} = \int\int_{-f_{Nyq}}^{f_{Nyq}} f_i f_j \frac{G}{N_2} df_x df_y$$
$$= \frac{1}{\sigma^2} \int\int_{-f_{Nyq}}^{f_{Nyq}} f_i f_j G df_x df_y \quad (10)$$

which is equivalent to (6) via Parseval's Theorem and the Fourier derivative theorem.

C. Strong Signal Approximation

A simplifying approximation of the FIM is obtained when the denominator of the SNR is not a function of G, as in the simple case of (10). This form is achieved when $N_1 N_2 \ll GN_1 + GN_2$ allowing approximation of the denominator in (9) as $GN_1 + GN_2$ and giving:

$$FIM_{\hat{N}\ll G}=(2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$
$$\text{where}\,\gamma_{ij}=\int\int_{-f_{Nyq}}^{f_{Nyq}} f_i f_j SNR_{\hat{N}\ll G} df_x df_y$$
$$\text{with}\,SNR_{\hat{N}\ll G}\ (f_x,f_y)=\frac{G}{N_1+N_2} \quad (11)$$

The approximation can also be written as $N_1 N_2/(N_1 + N_2) = \hat{N} \ll G$, showing that this approximation holds when the signal is strong compared to the noise. An interesting observation is that this approximation is equivalent to treating the noise as being contained in only one image:

$$I_1[x,y]=g(x,y) \quad (12)$$

$$I_2[x,y]=g(x-u,y-u)+n_2(x,y)+n_1(x,y) \quad (13)$$

indicating that (for low noise levels) it is not important how the noise is distributed between the two images, and only the sum $(N_1 + N_2)$ affects registration performance. For the simple AWGN case, (11) amounts to the scenario in (4–5) and the FIM becomes:

$$FIM_{\hat{N}\ll G,AWGN}=\frac{1}{\sigma_1^2+\sigma_2^2}\begin{bmatrix} \sum_{x,y} g_x^2 & \sum_{x,y} g_x g_y \\ \sum_{x,y} g_x g_y & \sum_{x,y} g_y^2 \end{bmatrix} \quad (14)$$

where we now have the sum of the variances from the two images entering the denominator of the FIM, simplifying the explicit relationship between registration performance and total image variance. In the case of equal variance (homoscedastic), this amounts to an increase in the CRLB by a factor of 2 compared to (6).

In this reduced form, the $FIM_{\hat{N}\ll G}$ captures intuitive dependencies between noise and registration performance (i.e., the CRLB scales directly with total variance). However, we will see that it further yields simple relationships for NEQ and DQE (shown in Section V). The $N_1 N_2/(N_1 + N_2) = \hat{N} \ll G$ condition corresponds to at least three relevant scenarios: (i) the signal is strong compared to the noise sum (i.e, $G \gg (N_1 + N_2) > N_1 N_2/(N_1 + N_2)$, indicating that $SNR_{\hat{N}\ll G}$ is large; (ii) I_1 is noiseless ($N_1 \approx 0$); or (iii) I_2 is much noisier than I_1 (i.e., $N_1 \ll N_2$) *and* the signal power is strong compared to the noise in I_1 (i.e., $N_1 \ll G$). These are suitable approximations, for example, in registration of high-contrast bone anatomy (high signal power in g) from a high quality preoperative CT (low N_1) to a lower quality (high N_2) intraoperative CBCT.

IV. cross-Correlation and Optimal Filtering

An important question that an analytical registration model must also address is the impact of (optional) post-processing blur. It can be seen in (7) that blur by a simple linear filter (described by the MTF^2 implicit in both the G and N_i terms) will exactly cancel out for invertible filters (with non-zero MTF), exemplifying the information-theoretic data processing inequality: application of a linear blur filter does not reduce the CRLB in registration performance. However, in practice the benefits of post-processing blur are well known (and shown in section VII-C) for reducing the impact of high-frequency noise on registration performance. Therefore, to more fully examine the question of spatial resolution and optimal filtering, the registration method itself must be examined. In this section we focus on the commonly used registration method of maximizing cross-correlation:

$$\begin{aligned} r(\tau,\varphi) &= (I_1 \otimes I_2)(\tau,\varphi) \\ &= \sum_{x,y} I_1[x,y]\, I_2[x+\tau, y+\varphi] \end{aligned} \tag{15}$$

where an interpolation over I_2 must take place to achieve the continuous cross-correlation function. By examining a local region near the peak of the cross-correlation function, we observe that an equivalent estimator of $\boldsymbol{\theta} = [u, v]$ is one that solves for $\hat{\boldsymbol{\theta}} = [\hat{u}, \hat{v}]$ such that $r_\tau(\hat{u},\hat{v}) = \frac{\partial r(\hat{u},\hat{v})}{\partial \tau} = 0$ and $r_\varphi(\hat{u},\hat{v}) = \frac{\partial r(\hat{u},\hat{v})}{\partial \varphi} = 0$. If we assume errors in the estimation to be contained within the linear region near the true solution (similar to [25] in 1D TDE), a first-order Taylor series approximation near $\boldsymbol{\theta}$ yields the error estimate:

$$\begin{bmatrix} (\hat{u}-u) \\ (\hat{v}-v) \end{bmatrix}^t \approx \begin{bmatrix} r_\tau(u,v) \\ r_\varphi(u,v) \end{bmatrix}^t \begin{bmatrix} r_{\tau\tau}(u,v) & r_{\tau\varphi}(u,v) \\ r_{\tau\varphi}(u,v) & r_{\varphi\varphi}(u,v) \end{bmatrix}^{-1} \tag{16}$$

where $r_{\tau\tau}(u,v) = \frac{\partial^2 r(u,v)}{\partial \tau^2}$, and similarly for the other 2nd derivative terms. The root mean square error (RMSE) of the estimate is obtained by computing the magnitude of the expectation of this error. We show in Appendix B that under the same assumptions of stationarity and independence used for the FIM derivation and by assuming $r_{\tau\varphi}(u, v)$ to be small in comparison to the diagonal terms (appropriate when the image gradients are evenly distributed in orientation), the RMSE for the maximizing cross-correlation estimate is approximately:

$$\begin{aligned} RMSE &\approx \sqrt{1/\rho_x + 1/\rho_y}, \text{ where} \\ \rho_i &= \frac{(2\pi)^2 A \left[\int\int_{-f_{Nyq}}^{f_{Nyq}} f_i^2 G\, df_x df_y\right]^2}{\int\int_{-f_{Nyq}}^{f_{Nyq}} f_i^2 (GN_1 + GN_2 + N_1N_2)\, df_x df_y} \end{aligned} \tag{17}$$

The result bears many similarities to the FIM in (7), particularly with respect to the noise term $G N_1 + G N_2 + N_1 N_2$. From the associative property of cross correlation, we see that I_1

$\otimes I_2 = g \otimes g + g \otimes n_2 + n_1 \otimes g + n_1 \otimes n_2$, which comprises two primary terms: (i) the true cross correlation of the signal $g \otimes g$; and (ii) the remaining terms associated with additive noise in the cross-correlation. The power associated with these noise terms is represented in the Fourier domain as $G N_1 + G N_2 + N_1 N_2$, which appears in both (7) and (17). Equation (17) is particularly interesting in that the signal and noise terms (numerator and denominator of ρ_i, respectively) are integrated separately, analogous to a non-prewhitening (NPW) model, for which an "optimal blur" can be derived - distinct from the simple data processing concept noted above. Noting that blurring the images leads to a blur of the cross-correlation [i.e., $(h_1 * I_1) \otimes (h_2 * I_2) = (h_1 \otimes h_2) * (I_1 \otimes I_2) = h * r$], we may consider a blur kernel (h) with Fourier transform H, giving:

$$RMSE \approx \sqrt{1/\rho_x + 1/\rho_y} \text{ where,}$$
$$\rho_i = \frac{(2\pi)^2 A \left[\int\int_{-f_{Nyq}}^{f_{Nyq}} f_i^2 H \, G df_x df_y \right]^2}{\int\int_{-f_{Nyq}}^{f_{Nyq}} f_i^2 H^2 (G N_1 + G N_2 + N_1 N_2) df_x df_y} \quad (18)$$

This result then can be minimized as a function of H. In the 1D TDE case, Knapp and Carter [25] showed that the maximum likelihood estimate is achieved by filtering according to the data-dependent Hannan and Thompson method, giving:

$$H_{HT} = \frac{G/(N_1 N_2)}{1 + G/N_1 + G/N_2} \quad (19)$$

In practice, however, blurring of I_1 or I_2 is typically achieved using a simple kernel (e.g., symmetric Gaussian blur); thus, we consider a blurring function H_b with kernel width σ_b:

$$H_b(f_x, f_y) = e^{-4\pi^2 (f_x^2 + f_y^2)\sigma_b^2} \quad (20)$$

where H_b represents blurring both images by a Gaussian kernel of width σ_b. Therefore, we can solve for the optimal blur by minimizing (18) with respect to σ_b. Depending on the registration method, additional blur can be included explicitly (such as the blur incurred with interpolation, described by H_{interp}), giving a combined $H = H_{interp} H_b$ for the total blur appearing in (18).

V. Connection to Models of Image Quality

As derived above, the CRLB for image registration depends explicitly on the noise and resolution characteristics of the imaging system(s). Recent decades have seen the development of accurate models for the image quality characteristics of CT and CBCT imaging systems [13]–[15], including the *MTF*, *NPS*, and their dependence on each factor in the imaging chain, such as dose, system geometry, acquisition technique, and reconstruction technique. Such models consider the propagation of signal and noise through

the imaging chain to describe the *MTF* and *NPS*. Simplifying from the 3D case in [13] to the 2D case considered here, a simple form for the *ideal* axial CT image NPS (i.e., a deterministic system featuring quantum noise without aliasing or electronic noise) can be written:

$$NPS_{ideal}\ (f_x, f_y) = \frac{\pi \sqrt{f_x^2 + f_y^2}}{m\bar{q}M^2\Gamma} MTF^2\ (f_x, f_y) \quad (21)$$

(with the full form of the NPS detailed in [13]) where the dose is related to the number of projections (m) and incident x-ray fluence ($\bar{q}$), M refers to system magnification, and is the system gain. Considering the SKE white noise registration model of (6) and the image quality model of (21), we reach an immediate finding: since noise-power [and thus σ^2, given by the integral of (21)] is inversely proportional to dose (via the $m\bar{q}$ term in the denominator), and the CRLB is proportional to σ^2, then the lower bound on registration error scales in inverse proportion to dose. Incorporating noise terms in both images shows a more complex relationship that depends not only on the total dose, but the relative dose in each image.

We can consider such relationships further in terms of metrics of fidelity that incorporate both the *NPS* and *MTF*. The performance of an imaging system is commonly described in terms of the *NEQ*, representing the effective number of incident photons contributing to each spatial frequency [13]:

$$NEQ\ (f_x, f_y) = \pi \sqrt{f_x^2 + f_y^2} \frac{MTF^2\ (f_x, f_y)}{NPS\ (f_x, f_y)} \quad (22)$$

Arranging terms from (22) and combining with the $FIM_{\hat{N} \ll G}$ formulation of (11) yields a relationship between image quality and registration performance. For example, in a scenario where the two images are produced by the same imaging system (equivalent *MTF*), the $FIM_{\hat{N} \ll G}$ in (11) becomes:

$$FIM_{\hat{N} \ll G} = 4\pi\ A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$

where

$$\gamma_{ij} = \int\int_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i\, f_j}{\sqrt{f_x^2 + f_y^2}} \frac{NEQ_1\, NEQ_2}{NEQ_1 + NEQ_1} G_{obj}\, df_x\, df_y \quad (23)$$

where G_{obj} refers to the power spectrum of the object rather than the image of the object (equal to G divided by MTF^2). In this form, we see that registration performance is

dependent on high-frequency weighting [carried by the $f_i\, f_j/\mathrm{sqrt}(f_x^2+f_y^2)$ term] of the object, in turn modified by the effective *NEQ* term.

Alternatively, the *DQE* describes the performance in terms of the dose, *MTF*, and *NPS* as:

$$DQE\ (f_x, f_y) = \frac{\pi\sqrt{f_x^2+f_y^2}}{m\bar{q}} \frac{MTF^2\ (f_x, f_y)}{NPS\ (f_x, f_y)} \quad (24)$$

Similarly, taking the example of two images produced by the same system (equivalent *MTF*), rearranging (24) allows the $FIM_{\hat{N}\ll G}$ (11) to be written as a function of *DQE*:

$$FIM_{\hat{N}\ll G} = 4\pi\, A \frac{(m\bar{q})_1\,(m\bar{q})_2}{(m\bar{q})_1+(m\bar{q})_2} \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$
$$\text{where } \gamma_{ij} = \int\int_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i\, f_j}{\sqrt{f_x^2+f_y^2}} DQE \cdot G_{obj}\, df_x df_y \quad (25)$$

Examination of (18) similarly elucidates the dependence of registration accuracy on resolution characteristics, particularly for the case of maximizing cross correlation. Reduced system *MTF* (via system blur and/or coarser voxel size) carries the benefit of reduced noise but also reduces the strength of image gradients via *H*; therefore, the lower bound on registration accuracy follows a non-monotonic dependence on spatial resolution, suggesting an 'optimal' resolution (alternatively, an optimal post-processing filter) that balances the tradeoffs between noise and gradient strength.

VI. Experimental Methods

A. Formation of Test Images

Experiments were conducted based on two digitally simulated axial CT images: (i) a soft-tissue model and (ii) an anthropomorphic head phantom. The soft-tissue model was based on a power-law noise distribution with frequency content following the distribution $F\{g\}(f) \propto f^{-\beta}$, as common in statistical modeling of anatomical "clutter" [26]–[28]. A value of $\beta = 3$ has been shown to model a stochastic arrangement of self-similar, soft-tissue anatomy [27]. A realization of such power-law distribution in 3D was generated and taken as ground truth soft-tissue anatomical structure for CT simulation. For the head image, ground truth was measured from a high-quality CT scan of an anthropomorphic head phantom (The Phantom Laboratory, Greenwich, NY) with soft-tissue manually segmented [10] and set to a constant value of 40 HU.

Simulated CT images of these two models were computed over a broad range in dose by digitally forward-projecting the ground truth images, scaling the fluence in proportion to dose, and adding Poisson noise in proportion to $1/\sqrt{(1+SPR)\times Dose}$, where nominal values for scatter-to-primary ratio (SPR) were chosen: SPR = 2 for the soft-tissue image and SPR = 9 for the higher attenuation head image [20], [21]. In each case, dose is specified in

terms of the x-ray tube current - time product (mAs), which is proportional to absorbed dose via the fluence per unit exposure (q/X), exposure per mAs (X/mAs), and exposure-to-dose conversion (f-factor, cGy/X), all of which are constant for a fixed beam energy (in these studies, a 100 kV spectrum computed using the SPEKTR x-ray simulation toolkit [29])

Each image was simulated from $m = 720$ forward projections over 360°. The fluence was scaled according to total x-ray tube output (mAs) at a beam energy of 100 kV. Images were reconstructed by filtered backprojection, and central 2D axial slices (241 × 241 pixels for the soft-tissue model and 485 × 390 for the head) at 0.5 mm × 0.5 mm voxel size were extracted. Example soft-tissue and head images at various dose levels are shown in Fig. 1. A total of 22 independent image realizations were generated for each phantom and dose level, each taken as input to the registration process, described below.

B. Registration Methods and Similarity Metrics

Experiments were performed using 3 categories of registration (translation-only): (i) The first category involved intensity interpolation, which optimizes a similarity metric over $\hat{\boldsymbol{\theta}}$ and resamples the image at each iteration under the specified interpolation model (here, cubic B-spline [30]). This was accomplished with SimpleITK [31] using 3 similarity metrics: (a) mean-square-difference (MSD), (b) Matte's mutual information (MMI) [32] (50 bins), and (c) joint histogram mutual information (JMI) [33] (50 bins, 1.5 σ). (ii) The second category involved metric interpolation, which was computed by evaluating the maximum of a parabolic fit to the normalized cross correlation (NCC-fit) metric at the pixel-shift peak location and its surrounding 8 pixel-shift neighbors. (iii) Finally, a Fourier-based method of phase correlation (PC) [34] was evaluated. For a pair of images (generated as in Section VI-A.), registration was performed using each of these methods after introducing a known shift $\boldsymbol{\theta}$. Prior to inducing the shift, a constant intensity (equal to the mean over the image edges) was subtracted from both images to reduce the effect of zero-padding used for the image transformations.

C. Performance Evaluation

Registration performance was evaluated in terms of the root-mean-square error (RMSE) of the translation estimate $\hat{\boldsymbol{\theta}}$. For each of the categories described above, the RMSE was compared to the two forms of lower bound derived in Section III: the CRLB (Eq. 7) and the $\mathrm{CRLB}_{\hat{N} \ll G}$ (Eq. 11). The CRLB for an unbiased estimator can be written in terms of RMSE as:

$$RMSE \geq \sqrt{\mathrm{trace}\,(C_{LB})} = \sqrt{\mathrm{trace}(FIM^{-1})} \quad (26)$$

To estimate the power spectra required to calculate the CRLB, the NPS at each mAs (i.e., N_i) was computed by averaging periodograms from a set of 20 instances of simulated noisy images with the mean image subtracted. In computing G, since we only have access to noisy images in the context of (1–2), we do not truly know g. In this work g was formed by computing the mean over 20 images simulated at 500 mAs. We then computed G using the 2D Welch periodogram method [35] using 16 windows (4 increments in each dimension)

with an overlap ratio 0.5 and a Hamming window to reduce spectral leakage. While this method is suitable to image simulation, it may not be practical to acquire many instances of an image to compute g. Other methods for approximating G (not investigated in this work) include (i) computing the power spectrum from a low-noise (e.g., pre-operative) image assuming minimal noise, (ii) computing the image power spectrum and subtracting a model estimation of the *NPS*, or (iii) using a model estimation of the signal power spectrum (e.g., power-law model as fairly common in describing tissue parenchyma [26]–[28]). RMSE was analyzed as a function dose (proportional to mAs) and total image variance $\sigma_1^2+\sigma_2^2$ (computed by integrating $N_1 + N_2$).

We further evaluated each registration method in terms of the statistical registration efficiency (denoted SRE), defined as:

$$SRE=\frac{\text{trace}\,(C_{LB})}{(RMSE)^2} \quad (27)$$

Written this way, the SRE is bounded (SRE < 1) and describes the ratio of the CRLB to the measured mean squared error performance. The SRE was evaluated as a function of dose for each category of registration and similarity metric mentioned above.

D. Registration Cases

1) Registration of Images at Equivalent Dose (Homoscedastic)—Analysis was first performed for image registration in which the noise characteristics of both images were equivalent. Each registration followed the method in section VI-B, with a known shift of $\boldsymbol{\theta}$ = [1.2 pix, 1.2 pix] introduced to the moving image. 231 (i.e., 22 choose 2) registrations were performed between the 22 image realizations formed at the same dose level. 13 dose levels were considered with mAs ranging over 3 orders of magnitude (0.5 – 500 mAs).

2) Registration of a High-Dose Image to a Low-Dose Image (Heteroscedastic)—A common scenario in image-guided interventions was simulated in which a high-dose (i.e., high quality) preoperative image is registered to a low-dose (i.e., lower quality) intraoperative image. The experiment of section VI-D-1 was repeated using the MSD similarity metric, considering a fixed dose for the fixed image and varying the dose for the higher-dose moving image. Performance was evaluated in terms of RMSE as a function of the total noise magnitude in the registered images.

3) Effect of Image Blur on Registration Performance—We further examined the effect of image blur on image registration performance. Factors affecting blur [described by the image quality model leading to (21)] include the imaging system configuration (e.g., x-ray focal spot size, detector pixel size, and system geometry), reconstruction method (filter kernel), and optional post-processing and/or interpolation filters. The derivation in Section IV, leading to (18), exposed the non-trivial relationship between system blur and registration performance, suggesting an optimum tradeoff between high-frequency noise magnitude and image signal (i.e., gradient) power. To investigate the effect, the experiment of Section V-

D-1 was repeated with an additional post-processing Gaussian blur kernel of width σ_b ranging from 0.5 to 7 pixels applied to both images. Results were compared to theoretical predictions for optimal Gaussian blurring using (18).

VII. Results

A. Registration Accuracy: Homoscedastic Images

Figure 2 shows the performance for the various categories of registration: (2A) metrics MSD, MMI, and JMI; and (2B) methods NCC-Fit and PC – each in comparison to the theoretical lower bound predicted by CRLB (Eq. 7) and $\text{CRLB}_{\hat{N}\ll G}$ (Eq. 11). In Fig 2A, registration performance is seen to improve (i.e., RMSE decreases) with dose for each of the interpolation-based similarity metrics. Each metric performs equivalently at high dose, and each exhibits a low-dose threshold below which registration fails, with MSD demonstrating the strongest robustness to noise and JMI performing the worst. The threshold reflects the noise level at which point estimation errors lie outside of the main lobe of the optimization search space (causing a “failed registration”) and leading to arbitrarily large registration errors.

The theoretical lower bound predictions appear to be optimistic - i.e., none of the methods achieve the lower bound. However, the overall trend with dose is similar, and the estimators (MSD and MMI in particular) adhere to the trend with similar slope across the broad range of dose levels.

Figure 2B shows the same for the NCC-fit and PC registration methods. (The CRLB and $\text{CRLB}_{\hat{N}\ll G}$ curves are the same as in Fig. 2A.) Interestingly, these methods do not exhibit a low-dose threshold for registration failure, instead following the general trend of the CRLB. The NCC-Fit method appears more robust against noise than PC and follows the CRLB down to the lowest dose levels investigated. This behavior is attributed in part to the brute-force sampling of NCC-Fit (thus avoiding local minima) and in part to the highly constrained search of NCC-Fit (evaluated only at pixel shifts within a fixed distance from solution), which avoids the large RMSE registration failure threshold effect. The PC registration method exhibits poorer robustness to noise (higher negative slope) and steady degradation to registration failure with reduced dose.

Figures 2C and 2D illustrate the extent to which various methods achieve the CRLB by evaluating the SRE versus dose. Figure 2C shows the SRE for the intensity-interpolation metrics (MSD, MMI, and JMI), showing that each approaches SRE ~0.04 at high dose, but efficiency falls by more than an order of magnitude at the low-dose threshold identified in Fig. 2A. Figure 2D shows that the NCC-Fit method maintains SRE over the entire range of dose investigated (again, likely attributed to the highly constrained search), whereas the PC method shows a steep degradation in SRE with reduced dose.

At high dose, each metric and method achieved SRE of only ~0.04 for the soft-tissue phantom (and ~0.11 for the anthropomorphic head, not shown for brevity). When considering this fairly low level of efficiency, it should first be noted that the CRLB is generally not guaranteed to be obtainable. Moreover, even when it is obtainable, only a

selection of estimators may be able to achieve the bound, and often only asymptotically -i.e., in a manner that requires larger and larger data size to achieve the bound. Since the images used in this study were relatively small and optimal estimators were not examined [e.g., optimal filtering to minimize (18)], we do not expect the result to achieve an SRE of 1; however, we note the increased upper limit in SRE for the head phantom, as expected for a larger image.

Examining (25), we see that in the homoscedastic (equal-dose) case, registration error in the strong signal approximation is proportional to 1/sqrt(dose), which is evident in the slope of the $\mathrm{CRLB}_{\hat{N} \ll G}$ curve (Fig. 2A and 2B). For the soft-tissue image case (dominated by low-frequency signal power), the strong signal approximation appears to hold well only at high dose (i.e., low noise). For the head image (which exhibits a greater proportion of mid- and high-frequency signal power), the approximation holds within 15% of CRLB over a broader dose range – down to ~2.5 mAs. In the lower dose range, disagreement between CRLB with $\mathrm{CRLB}_{\hat{N} \ll G}$ arises due to increased influence of the N_1N_2cross-correlation noise term.

B. Registration Accuracy: Heteroscedastic Images

Figure 3 shows MSD registration performance as a function of (total) noise magnitude for the heteroscedastic case in which a low-noise (i.e., higher-dose) image is registered to a noisier (i.e., lower-dose) image. The RMSE is plotted versus total variance ($\sigma_1^2+\sigma_2^2$) for the (A) soft-tissue and (B) head phantom images of Fig. 1. The CRLB and $\mathrm{CRLB}_{\hat{N} \ll G}$ theoretical lower limits are shown as dashed lines. The color scale on the plot symbols and curve fits refers to the mAs of the lower-dose image: for example, the lower-left of each plot shows the (red) case for which both images were formed at the maximum dose (500 mAs), whereas the upper-right of each plot shows the (black) case for which the lower-dose image was formed at just 0.5 mAs.

Whereas Figs. 2A–B demonstrated that RMSE is proportional to 1/sqrt(dose) and thus $\mathrm{sqrt}(\sigma_1^2+\sigma_2^2)$, this simple relationship is lost in the heteroscedastic case shown in Fig. 3. In Fig. 3A, we observe a highly non-linear dependence on the total noise; however, this non-linearity is predicted very well by the CRLB. The $\mathrm{CRLB}_{\hat{N} \ll G}$ approximation, however, only describes the effect of relative dose through the $(m\bar{q})_1$ $(m\bar{q})_2/[(m\bar{q})_1 + (m\bar{q})_2]$ term, appropriate only at low-noise (high-dose) conditions for the soft-tissue model.

Figure 3B shows similar trends in registration performance for the head phantom image. In this case, however, the image exhibits sufficiently large high-frequency signal power, such that $N_1N_2 \ll GN_1 + GN_2$. As a result, the CRLB is roughly proportional to $\mathrm{sqrt}(\sigma_1^2+\sigma_2^2)$ and agrees with the $\mathrm{CRLB}_{\hat{N} \ll G}$ approximation over a much broader range of dose.

C. Registration Accuracy: Effect of Image Blur

Figure 4A shows the registration performance (RMSE) for the heteroscedastic soft-tissue case (as in Fig. 3A), comparing the RMSE achieved by the MSD method (which can be shown to be nearly equivalent to maximizing cross-correlation) with that predicted by (17). We immediately see that while (17) somewhat underestimates the magnitude of RMSE, it trends well with the measured dependence of registration performance on dose, yielding a

correlation coefficient of R = 0.988 between the predicted and measured RMSE in the higher dose region [where the Taylor approximation in (17) is appropriate].

Figure 4B summarizes the findings of optimal post-processing blur for registration of the soft-tissue image at various dose levels. Each curve represents the RMSE (at a given dose level) as a function of blur width (σ_b). The blue star marks the measured minimum in RMSE (i.e., optimal blur), and the magenta circle marks the theoretical minimum as predicted by minimizing (18) with respect to H_b. As expected, post-processing blur is most beneficial under high-noise (low-dose) conditions (black curves). On the other hand, for low-noise (high-dose) conditions (red curves), excessive blur is seen to degrade registration performance. The measured and predicted values for optimal blur agree with this trend and match fairly well across a broad range of dose, further validating the model of (18) as a figure of merit for registration.

Figure 5 further investigates the predicted benefit of post-processing blur on registration performance for several similarity metrics. The data correspond to the soft-tissue image model, homoscedastic image registration, and theoretically optimal Gaussian blur (OGB) derived by minimization of (18) with respect to H_b. The SRE is plotted versus dose, and we observe that application of an optimal Gaussian blur maintains optimality (again at a level of SRE ~0.04) across the range of dose levels investigated for MMI and MSD, bearing in mind that (18) applies directly only to cross-correlation based metrics, e.g., MSD. Close inspection of Fig. 5B suggests a slight increase in SRE at the lower dose levels - a somewhat surprising result that is in agreement with the theoretical prediction. The increase in efficiency is because Gaussian blur is not a truly optimal filter as described by H_{HT} of (19) in minimizing (18); however, with respect to Gaussian filters, the OGB may more closely approximate H_{HT} under low-dose/high-noise conditions, leading to the increase in efficiency.

VIII. Conclusion

To paraphrase Barrett *et al.* [3], an image is always acquired for a particular task, and evaluation of imaging system performance should be with respect to that task. For many scenarios in image-guided interventions, the task may relate to registration of image information more so than to visualization. The framework described above provides a means by which to evaluate an imaging system with respect to registration performance, providing a basis for assessing the performance of various registration methods and selecting optimal image acquisition and reconstruction techniques.

As a first step in establishing this framework, we presented theoretical lower bounds for registration performance and investigated several sub-pixel estimators as a function of dose and noise magnitude. Following this analysis, we examined the registration method of maximizing cross-correlation to analyze the effect of spatial resolution on registration performance, thereby deriving an RMSE figure of merit, (18). The model was shown to agree well with measurements of registration accuracy for various choices of post-processing linear filters (blur), providing understanding beyond the basic information-

theoretic data processing effect (in which linear filters have no effect on SNR) and guiding selection of optimal filters that extend registration performance to lower dose levels.

In the derivation of the *FIM*, we assumed stationary Gaussian characteristics of the signal and noise. As discussed in [36], the Gaussian assumption is not a particularly strong requirement, since (by the Central Limit Theorem) even when the signal is not Gaussian the Fourier coefficients of the signal will tend toward a Gaussian distribution as the number of image samples increases. Nonetheless, further investigation is warranted to more fully account for nonstationary *MTF* and *NPS* characteristics in CT images [37] as well as non-Gaussian characteristics. However, in line with similar approximations for analysis of detectability index [3], we see in Section VII-B that despite the approximation the *FIM* proves a useful predictor of performance trends, even in the case of a high-contrast (head) image that exhibits highly non-Gaussian, non-stationary characteristics.

The current work examined the simple case of 2D translation-only registration. Extension of the formulation to 3D translation is straightforward, as shown in Appendix A. In future work, we will extend the analysis to include translation and rotation, providing a more general framework for analysis of rigid registration. Further, because image quality models (such as [21] and [11]) permit analysis of the spatially varying (i.e., nonstationary) *local MTF* and *NPS*, the analysis shown above can be similarly extended to description of local registration accuracy in regions differing in image quality. Using these local approximations to extend the analysis to deformable image registration is an exciting possibility, though beyond the current scope of investigation.

Throughout this work is the assumption that the registration method is unbiased (intrinsic to $C_{LB} = FIM^{-1}$). This assumption breaks down at least in part for the NCC-fit registration method [16] due to the parabolic fit and has yet to be rigorously investigated for the other optimization methods. While it is likely that there are small biases in the other methods as well (owing to choice of interpolation, optimizer, etc.), the assumption of unbiased estimators appears to be reasonable, as the observed RMSE was dominated by the variance term (rather than the bias), and the experiments demonstrated similar trends as the CRLB (whereas a plateau effect would likely be observed in the low-noise region for a system dominated by bias).

The *FIM* provides a framework that is independent of the particular registration method - whether biased or unbiased. We extended previous work in CRLB estimation in such problems by generalizing to 2D and 3D, allowing for disparate noise in the I_1 and I_2images, and including image blur as well as noise correlation, demonstrating results beyond a simple approach of AWGN (which is a poor approximation to noise in CT / CBCT). The resulting analytical framework leverages well-established models describing image quality in CT/CBCT [13]–[15] as in (23) and is consistent with the theme of task-based imaging performance. With respect to image-guided interventions, the analysis provides a new framework for understanding the performance of imaging systems with respect to the task of image registration.

Acknowledgments

The research was supported by NIH grant R01-EB-017226 and research collaboration with Siemens Healthineers (XP and AT groups@comma Forcheim Germany).

Thanks to Dr. Jennifer Xu for discussion of image quality models and Michael Mow and Hao Dang (Department of Biomedical Engineering, Johns Hopkins University) for assistance with CT noise simulation. Thanks also to Dr. Jerry Prince (Department of Biomedical Engineering, Johns Hopkins University) for valuable discussion of image registration methods.

References

1. Uneri A, et al. Intraoperative evaluation of device placement in spine surgery using known-component 3D-2D image registration. Phys Med Biol. 2017
2. Uneri A, et al. Evaluation of low-dose limits in 3D-2D rigid registration for surgical guidance. Phys Med Biol. 2014; 59(18):5329–5345. [PubMed: 25146673]
3. Barrett HH, Denny JL, Wagner RF, Myers KJ. Objective assessment of image quality II Fisher information, Fourier crosstalk, and figures of merit for task performance. J Opt Soc Am A. 1995; 12(5):834.
4. ICRU. ICRU Report 54 - Medical Imaging - The Assessment of Image Quality. 1995.
5. Prakash P, et al. Task-based modeling and optimization of a cone-beam CT scanner for musculoskeletal imaging. Med Phys. 2011; 38(10):5612–29. [PubMed: 21992379]
6. Xu J, et al. Modeling and design of a cone-beam CT head scanner using task-based imaging performance optimization. Phys Med Biol. 2016; 61(8):3180–3207. [PubMed: 27025783]
7. Gang GJ, Siewerdsen J, Stayman J. Task-driven optimization of CT tube current modulation and regularization in model-based iterative reconstruction. Phys Med Biol. 2017
8. Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. Med Phys. 2012; 39(7):4115–22. [PubMed: 22830744]
9. Yu L, Leng S, Chen L, Kofler JM, Carter RE, McCollough CH. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms. Med Phys. 2013; 40(4):41908.
10. Dang JHH, Stayman JW, Xu J, Zbijewski W, Sisniega A, Mow M, Wang X, Foos DH, Aygun N, Koliatsos VE, Siewerdsen. Task-based regularization design for high-quality cone-beam CT of the head. Phys Med Bio. Submitted, 2017.
11. Gang GJ, Stayman JW, Zbijewski W, Siewerdsen JH. Task-based detectability in CT image reconstruction by filtered backprojection and penalized likelihood estimation. Med Phys. 2014; 41(8):81902.
12. Gang GJ, et al. Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance. Med Phys. 2011; 38(4):1754–1768. [PubMed: 21626910]
13. Tward DJ, Siewerdsen JH. Cascaded systems analysis of the 3D noise transfer characteristics of flat-panel cone-beam CT. Med Phys. 2008; 35(12):5510. [PubMed: 19175110]
14. Siewerdsen JH, et al. Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology. Med Phys. 1997; 24(1):71–89. [PubMed: 9029542]
15. Kijewski MF, Judy PF. The noise power spectrum of CT images. Phys Med Biol. 1987; 32(5):565–575. [PubMed: 3588670]
16. Robinson D, Member S, Milanfar P, Member S. Fundamental Perfromance Limits in Image Registration. IEEE Trans Image Process. 2004; 13(9):1185–1199. [PubMed: 15449581]
17. Yetik IS, Nehorai A. Performance bounds on image registration. IEEE Trans Signal Process. 2006; 54(5):1737–1749.
18. Zhao C, Carass A, Jog A, Prince JL. Effects of spatial resolution on image registration. Proc SPIE. 2016; 9784:97840Y–97840Y–9.

19. Kay, SM. Fundamentals of Statistical Signal Processing: Estimation Theory. Upper Saddle River, NJ, USA: Prentice-Hall, Inc; 1993.
20. Ketcha MD, et al. Fundamental limits of image registration performance: effects of image noise and resolution in CT-guided interventions. Proc SPIE. 2017; 10135:1013508.
21. Pineda AR, Tward DJ, Gonzalez A, Siewerdsen JH. Beyond noise power in 3D computed tomography: The local NPS and off-diagonal elements of the Fourier domain covariance matrix. Med Phys. 2012; 39(6):3240. [PubMed: 22755707]
22. Bangs, WJI. Array Processing with Generalized Beamformers. Yale University; 1971.
23. Weiss A, Weinstein E. Fundamental limitations in passive time delay estimation–Part I: Narrow-band systems. IEEE Trans Acoust. 1983; 31(2):472–486.
24. MacDonald VH, Schultheiss Peter M. Optimum Passive Bearing Estimation in a Spatially Incoherent Noise Environment. J Acoust Soc Am. 1969; 46(1A):37–43.
25. Knapp CH, Carter GC. The Generalized Correlation Method for Estimation of Time Delay. IEEE Trans Acoust. 1976; 24(4):320–327.
26. Burgess AE. Mammographic structure: data preparation and spatial statistics analysis. Proc SPIE. 1999; 3661:642–653.
27. Gang GJ, Tward DJ, Lee J, Siewerdsen JH. Anatomical background and generalized detectability in tomosynthesis and cone-beam CT. Med Phys. 2010; 37(5):1948–1965. [PubMed: 20527529]
28. Bochud FO, Valley JF, Verdun FR, Hessler C, Schnyder P. Estimation of the noisy component of anatomical backgrounds. Med Phys. 1999; 26(7):1365–1370. [PubMed: 10435539]
29. Punnoose J, Xu J, Sisniega A, Zbijewski W, Siewerdsen JH. Technical Note: spektr 3.0—A computational tool for x-ray spectrum modeling and analysis. Med Phys Med Phys Med Phys Med Phys Med Phys. 2016; 43(3057):4711–41906. [PubMed: 27487888]
30. Unser M, Aldroubi A, Eden M. B-spline signal processing. I. Theory. IEEE Trans Signal Process. 1993; 41(2):821–833.
31. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The Design of SimpleITK. Front Neuroinform. 2013; 7(December):45. [PubMed: 24416015]
32. Mattes D, Haynor DR, Vesselle H, Lewellyn TK, Eubank W. Nonrigid multimodality image registration. Proc SPIE. 2001; 4322:1609–1620.
33. Thévenaz P, Unser M. Optimization of mutual information for multiresolution image registration. IEEE Trans Image Process. 2000; 9(12):2083–2099. [PubMed: 18262946]
34. Foroosh H, Zerubia JB, Berthod M. Extension of phase correlation to subpixel registration. IEEE Trans Image Process. 2002; 11(3):188–199. [PubMed: 18244623]
35. Welch PD. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. IEEE Trans Audio and electroacoustic. 1967; 15:70–73.
36. Carter GC. Time delay estimation. Nav Underw Syst Cent. 1976
37. Baek J, Pelc NJ. The noise power spectrum in CT with direct fan beam reconstruction. Med Phys. 2010; 37(5):2074–2081. [PubMed: 20527540]
38. Papoulis, A., Pillai, SU. Probability, random variables, and stochastic processes. McGraw-Hill; 2002.

Appendix A

In this appendix, we derive the CRLB for image registration. We present a derivation that extends work from 1D TDE to the scenario of 2D and 3D image registration, and thus is similar in concept to derivations presented for TDE in [22]–[24]. Here we allow the two images to exhibit disparate noise levels as well as noise correlation. Taking the 2D case, we seek to determine $\boldsymbol{\theta} = [u, v]$ for the following scenario:

$$I_1[x,y]=g(x,y)+n_1(x,y) \quad \text{(A1)}$$

$$I_2[x,y]=g(x-u,y-v)+n_2(x,y) \quad \text{(A2)}$$

where g is the true image function and the n_i are additive noise terms. Due to the presence of correlated noise terms in both images, we lose the simple form for the likelihood function presented in [20], causing the spatial domain analysis to become less tractable. A Fourier representation, however, facilitates the analysis. We take $Z_1[f_x, f_y]$, $Z_2[f_x, f_y]$ as the 2D Fourier transforms ($\mathcal{F}\{\cdot\}$) of I_1, I_2, where for purposes of the proof we break from the convention of the $(\cdot)$ notation for continuous Fourier domain functions to directly represent the discrete Fourier transform of the image. We first note that for a signal bandlimited below f_{Nyq} (i.e., no aliasing), Z_1 comprises the sampled frequency representation of the signal and noise so that:

$$Z_1[f_x,f_y]=\mathcal{F}\{g(x,y)\}(f_x,f_y)+\mathcal{F}\{n_1(x,y)\}(f_x,f_y) \quad \text{(A3)}$$

and by the shift property we have:

$$Z_2[f_x,f_y]=e^{-j2\pi(f_x u+f_y v)}\mathcal{F}\{g(x,y)\}(f_x,f_y)+\mathcal{F}\{n_2(x,y)\}(f_x,f_y) \quad \text{(A4)}$$

$$K_{mn}=E\{X_{mn}X_{mn}^{*}\boldsymbol{\theta}\} \text{ where } X_{mn}=\begin{bmatrix} Z_1\left[f_x^{(m)},f_y^{(n)}\right] \\ Z_2\left[f_x^{(m)},f_y^{(n)}\right] \end{bmatrix}$$
$$=\begin{bmatrix} G\left[f_x^{(m)},f_y^{(n)}\right]+N_1\left[f_x^{(m)},f_y^{(n)}\right] & G\left[f_x^{(m)},f_y^{(n)}\right]e^{-j2\pi(f_x^{(m)}u+f_y^{(n)}v)} \\ G\left[f_x^{(m)},f_y^{(n)}\right]e^{j2\pi(f_x^{(m)}u+f_y^{(n)}v)} & G\left[f_x^{(m)},f_y^{(n)}\right]+N_2\left[f_x^{(m)},f_y^{(n)}\right] \end{bmatrix} \quad \text{(A5)}$$

If we examine a particular frequency $f_x^{(m)}$, $f_y^{(n)}$ (where m, n refer to the indexed frequency samples) we see that the covariance of $Z_1\left[f_x^{(m)},f_y^{(n)}\right]$, $Z_2\left[f_x^{(m)},f_y^{(n)}\right]$ is (A5), as shown at the bottom of this page, where G and N_i are the power spectra of g and n_i, respectively, and * denotes the complex conjugate transpose. With this in mind we wish to represent the entirety of the data in a single vector. Therefore we concatenate Z_1 and Z_2 into the vector $\mathbf{X}$ in a manner that the corresponding frequencies in Z_1 and Z_2 are adjacent to each other:

$$\mathbf{X}=[Z_1\left[f_x^{(1)},f_y^{(1)}\right],Z_2\left[f_x^{(1)},f_y^{(1)}\right],\ldots Z_1\left[f_x^{(M)},f_y^{(1)}\right],Z_2\left[f_x^{(M)},f_y^{(1)}\right],Z_1\left[f_x^{(1)},f_y^{(2)}\right],Z_2\left[f_x^{(1)},f_y^{(2)}\right],\ldots Z_1\left[f_x^{(M)},f_y^{(N)}\right],Z_2$$

(A6)

where $f_x^{(i)}, f_y^{(i)}$ are the indexed frequency samples $\in [-f_{Nyq}, f_{Nyq}]$. Under the assumption of stationary signal and noise, the frequency components of Z_i are statistically independent [38], and the covariance matrix of **X** has a block diagonal form:

$$K=E\{\boldsymbol{X}\boldsymbol{X}^*\boldsymbol{\theta}\}=\begin{bmatrix} K_{11} & & & & & \\ & \ddots & & & \mathbf{0} & \\ & & K_{M1} & & & \\ & & & K_{12} & & \\ & \mathbf{0} & & & \ddots & \\ & & & & & K_{MN} \end{bmatrix} \quad \text{(A7)}$$

where the block-diagonal components are described in (A5). By assuming both the image function and noise to be zero-mean Gaussian processes, the frequency components of the Fourier representation are also jointly Gaussian, and the likelihood function for **X** can be written:

$$L\left(\mathbf{X}|\boldsymbol{\theta}\right)=\frac{1}{\det\left(\pi K\right)}\exp(-\mathbf{X}^*K^{-1}\mathbf{X}) \quad \text{(A8)}$$

As shown in [22], the *FIM* for this complex Gaussian likelihood function can be reduced to:

$$[FIM]_{ij}=\mathrm{Tr}\left[\left(K^{-1}\frac{\partial K}{\partial\theta_i}\right)\left(K^{-1}\frac{\partial K}{\partial\theta_j}\right)\right] \quad \text{(A9)}$$

$$[FIM]_{ij}=\frac{1}{2}\sum_{m,n}\mathrm{Tr}\left[\left(K_{mn}^{-1}\frac{\partial K_{mn}}{\partial\theta_i}\right)\left(K_{mn}^{-1}\frac{\partial K_{mn}}{\partial\theta_j}\right)\right]$$
$$=\frac{1}{2}\sum_{m,n}\frac{2\left(2\pi\right)^2 f_i^{(m)}f_j^{(n)}G^2\left[f_x^{(m)},f_y^{(n)}\right]}{G\left[f_x^{(m)},f_y^{(n)}\right]N_1\left[f_x^{(m)},f_y^{(n)}\right]+G\left[f_x^{(m)},f_y^{(n)}\right]N_2\left[f_x^{(m)},f_y^{(n)}\right]+N_2\left[f_x^{(m)},f_y^{(n)}\right]N_1\left[f_x^{(m)},f_y^{(n)}\right]}$$

(A10)

$$\gamma_{ij}=\int\int_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j G^2\,(f_x,f_y)\,df_x df_y}{G\,(f_x,f_y)\,N_1\,(f_x,f_y)+G\,(f_x,f_y)\,N_2\,(f_x,f_y)+N_1\,(f_x,f_y)\,N_2\,(f_x,f_y)} \quad \text{(A11)}$$

Combining (A5), (A7), and (A9) gives the 2 × 2 *FIM*:

where we have denoted $f_1 = f_x$ and $f_2 = f_y$ in reference to $f_i^{(m)}, f_j^{(n)}$ when computing $[FIM]_{ij}$, and the 1/2 term prior to the sum is included to compensate for symmetry in the Fourier domain (thus over-representing the information by a factor of 2). Given sufficiently high sampling density, we may approximate the summation as an integral to write (A10), shown at the top of this page as:

$$FIM=(2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$

where (A11) is shown at the top of this page, where A is the image area. To simplify notation, we remove explicit (f_x, f_y) dependence:

$$FIM=(2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix},$$

where

$$\gamma_{ij}=\int\int_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j G^2}{GN_1+GN_2+N_1N_2} df_x df_y \quad \text{(A12)}$$

Extension to a pair of 3D images is straightforward, giving:

$$FIM=(2\pi)^2 V \begin{bmatrix} \gamma_{xx} & \gamma_{xy} & \gamma_{xz} \\ \gamma_{xy} & \gamma_{yy} & \gamma_{yz} \\ \gamma_{xz} & \gamma_{yz} & \gamma_{zz} \end{bmatrix},$$

where

$$\gamma_{ij}=\int\int\int_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j G^2}{GN_1+GN_2+N_1N_2} df_x df_y df_z \quad \text{(A13)}$$

where V is the image volume.

Appendix B

This appendix expands the derivation of (16) and (17) for the RMSE estimate of the cross-correlator, expanding a result for 1D TDE shown in [25]. We begin by simplifying (16) by assuming $r_{\tau\varphi}(u, v)$ to be small in comparison to the diagonal terms, giving:

$$(\hat{u}-u) \approx r_{\tau}(u,v)/r_{\tau\tau}(u,v) \quad \text{(B1)}$$

$$(\hat{v}-v) \approx r_{\varphi}(u,v)/r_{\varphi\varphi}(u,v) \quad \text{(B2)}$$

By the associative property of cross correlation:

$$\begin{aligned} r(u,v) &= (I_1 \otimes I_2)(u,v) \\ &= (g \otimes g)(0,0) + (g \otimes n_2)(u,v) + (n_1 \otimes g)(0,0) + (n_1 \otimes n_2)(u,v) \end{aligned} \quad \text{(B3)}$$

and without loss of generality:

$$r(u,v) = (g \otimes g)(0,0) + (g \otimes n_2)(0,0) + + (n_1 \otimes g)(0,0) + (n_1 \otimes n_2)(0,0)$$

where the second equality carries the discretization of g and n_i in forming I_i as well as the implicit $[u, v]$ shift of g in I_2 so that $g \otimes g$ and $n_1 \otimes g$ are evaluated at (0, 0). Further, this equality neglects interpolation and assumes an un-aliased signal. The third equality, which does not affect the result (as the expectation is unaffected), is introduced for notational convenience. From examination of (B3), we note that the numerators of (B1) and (B2) have expected value equal to zero. Further, by definition $\partial(g \otimes g)/\partial\theta_i(0, 0) = 0$. Therefore, near solution the numerator will be dominated by the remaining terms in the derivative of (B3). On the other hand, the expectation of the denominator is non-zero and depends only on the signal term, indicating that near solution the denominator is dominated by the signal term alone. Therefore, we have (B1) to first-order approximation:

$$\begin{aligned} (\hat{u}-u) &\approx \frac{\frac{\partial}{\partial\tau}(g \otimes n_2 + n_1 \otimes g + n_1 \otimes n_2)(0,0)}{\frac{\partial^2(g \otimes g)}{\partial\tau^2}(0,0)} \\ &= \frac{\sum_{m.n}\left(j2\pi f_x^{(m)}\right)(F_g\overline{F}_{n_2} + F_{n_1}\overline{F}_g + F_{n_1}\overline{F}_{n_2})}{\sum_{m,n}\left(j2\pi f_x^{(m)}\right)^2 F_g\overline{F}_g} \end{aligned} \quad \text{(B4)}$$

with the second equality following from the Fourier derivative property and Parseval's Theorem, where $F_g\left[f_x^{(m)}, f_y^{(n)}\right]$ denotes the Fourier coefficients of the discretized g (similarly for F_{ni}) where $f_x^{(m)}, f_y^{(n)}$ are the indexed frequency samples $\in [-f_{Nyq}, f_{Nyq}]$, and

the hat denotes complex conjugation. Explicit notation of the frequency dependence on the F_i terms is excluded for notational convenience in (B4) and the equations below. As the expectation of the error is zero, we need only to examine the variance:

$$\mathrm{Var}\ (\hat{u}) \approx \frac{\mathrm{Var}\left(\sum_{m.n}\left(jf_x^{(m)}\right)\ (F_g\overline{F}_{n_2}+F_{n_1}\overline{F}_g+F_{n_1}\overline{F}_{n_2})\right)}{(2\pi)^2\left[\sum_{m,n}\left(f_x^{(m)}\right)^2 F_g\overline{F}_g\right]^2} \tag{B5}$$

From the assumption of stationarity, the frequency components of the Fourier terms in the numerator are independent [38], leaving only the sum over the variance terms.

$$\mathrm{Var}\ (\hat{u}) \approx \frac{\sum_{m,n}\left(f_x^{(m)}\right)^2 E\left\{(F_g\overline{F}_{n_2}+F_{n_1}\overline{F}_g+F_{n_1}\overline{F}_{n_2})\,(\overline{F}_gF_{n_2}+\overline{F}_{n_1}F_g+\overline{F}_{n_1}F_{n_2})\right\}}{(2\pi)^2\left[\sum_{m,n}\left(f_x^{(m)}\right)^2 F_g\overline{F}_g\right]^2} \tag{B6}$$

In expectation, only the associated power spectra terms remain non-zero giving:

$$\mathrm{Var}\ (\hat{u}) \approx \frac{\sum_{m,n}\left(f_x^{(m)}\right)^2 (GN_1+GN_2+N_1N_2)}{(2\pi)^2\left[\sum_{m,n}\left(f_x^{(m)}\right)^2 G\right]^2} \tag{B7}$$

where $G\left[f_x^{(m)}, f_y^{(n)}\right]$ and $N_i\left[f_x^{(m)}, f_y^{(n)}\right]$ are the signal and noise-power spectra, respectively. By analogous derivation for Var ($\hat{\upsilon}$), and approximating the sum as an integral, the RMSE is then:

$$RMSE \approx \sqrt{1/\rho_x+1/\rho_y},\ \text{where}$$
$$\rho_i=\frac{(2\pi)^2 A\left[\int\int_{-f_{Nyq}}^{f_{Nyq}} f_i^2 G df_x df_y\right]^2}{\int\int_{-f_{Nyq}}^{f_{Nyq}} f_i^2(GN_1+GN_2+N_1N_2)df_x df_y} \tag{B8}$$

and A is the image area.

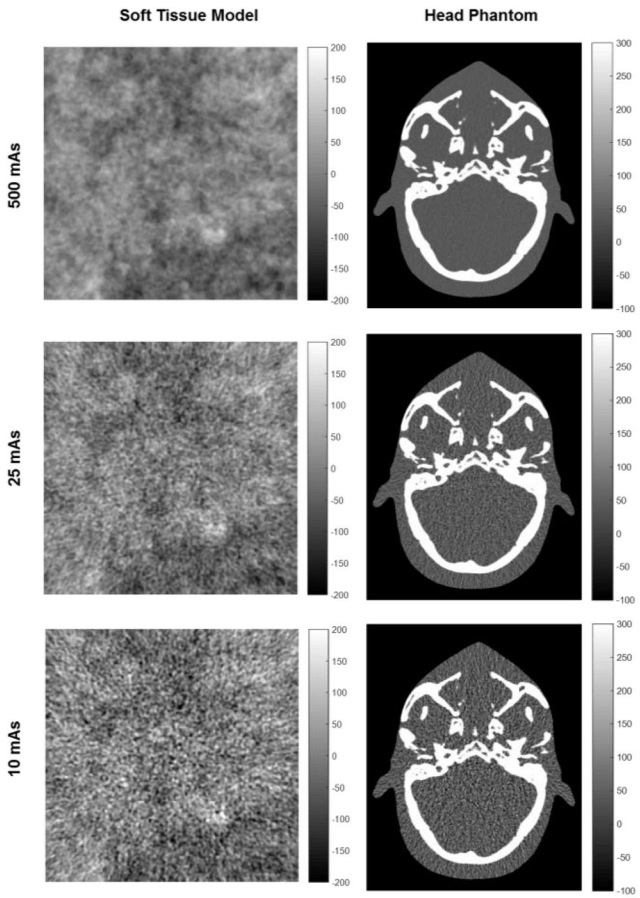

Fig. 1.

Example images of the soft-tissue model (left) and anthropomorphic head phantom (right) at various levels of dose (mAs).

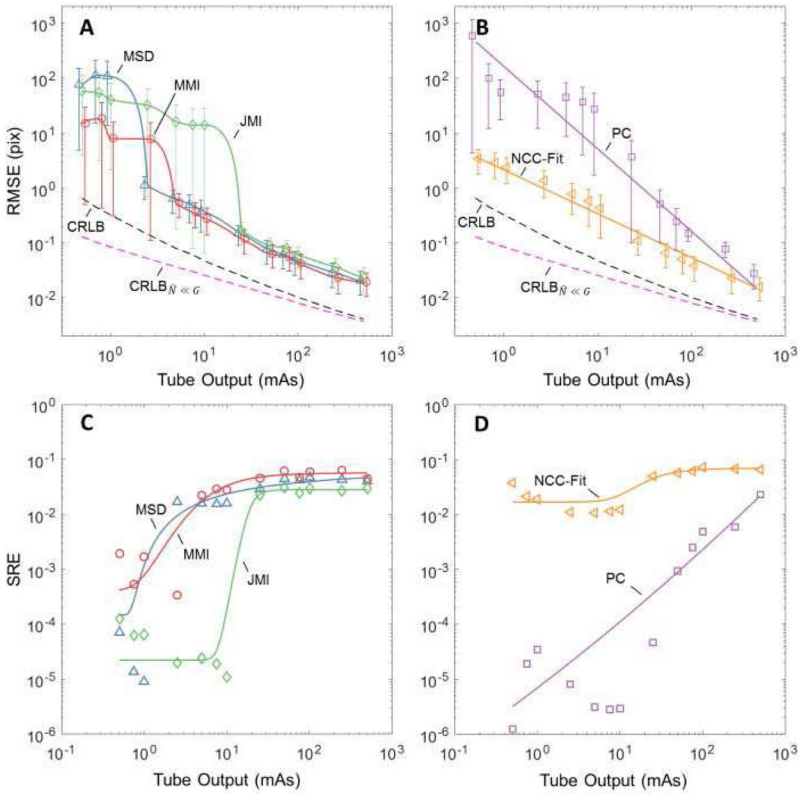

Fig. 2.

Effect of dose on registration performance for the “equal-dose” case (i.e., images with equivalent noise characteristics). Each case is for the soft-tissue images in Fig. 1. The dashed curves in (A) and (B) mark the lower-bound in registration accuracy predicted by the CRLB and $CRLB_{\hat{N} \ll G}$. (A) RMSE for intensity-interpolation registration using the MSD, MMI, and JMI similarity metrics. (B) RMSE for the NCC-fit and PC registration methods. (C) SRE versus dose for the MSD, MMI, and JMI metrics. (D) SRE for the NCC-fit and PC methods.

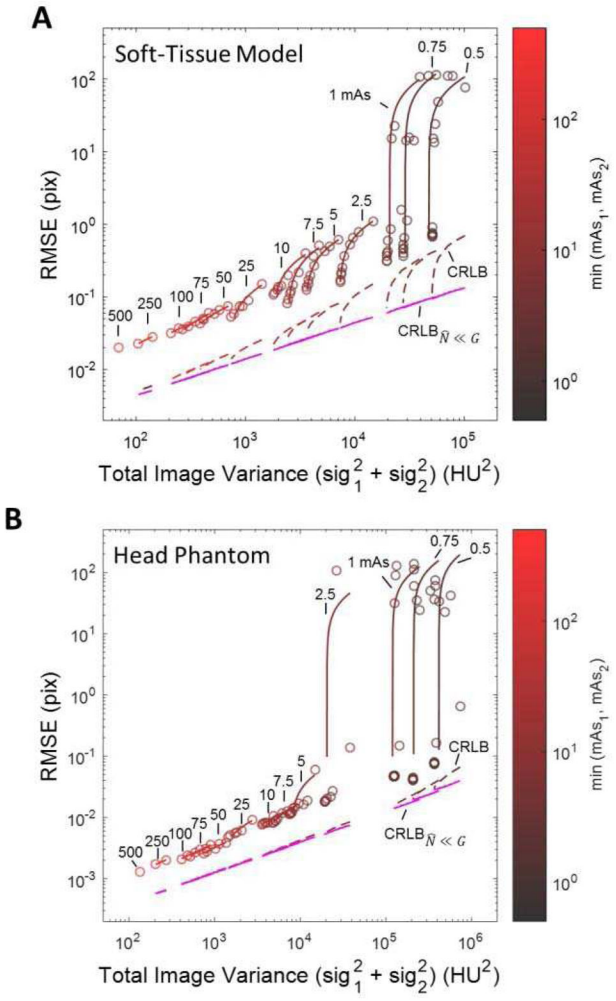

Fig. 3.

Registration performance (using MSD) versus total image noise for the heteroscedastic case: (A) soft tissue image and (B) head phantom image. Each circle represents the RMSE for a specific I_1, I_2 dose level combination, with connected circles of the same color indicating the same mAs for the low-dose image. The colorscale and labels denote the mAs for the lower-dose image. The CRLB (dashed) and CRLB$_{\hat{N} \ll G}$ (magenta) formulations are also plotted.

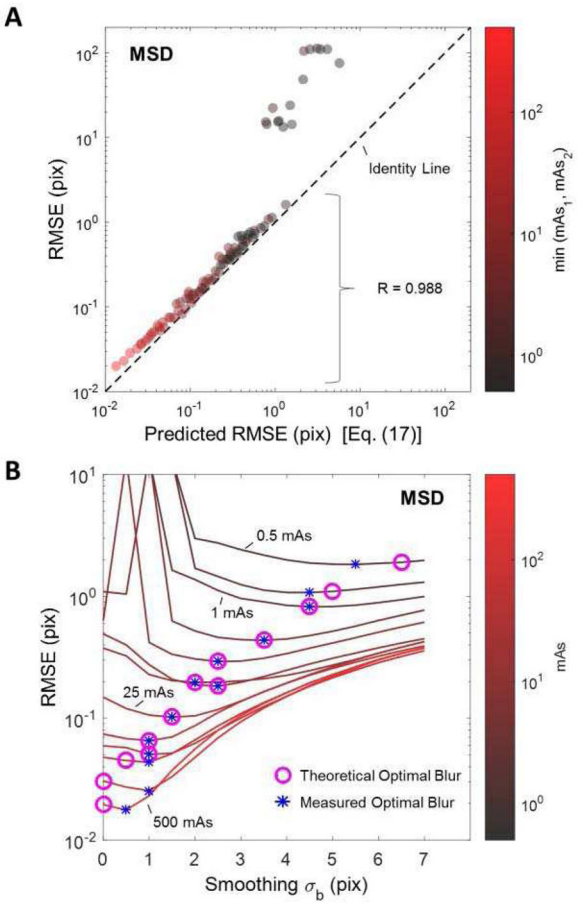

Fig. 4.

(A) Error in soft-tissue image registration compared to the performance predicted by Eq. (17). (B) Registration performance as a function of post-processing blur at various dose levels. The results pertain to the MSD registration method, and dose reflected in the mAs colorscale. For each curve, the magenta circle represents the predicted optimal blur level, and the blue star represents the measured optimal blur.

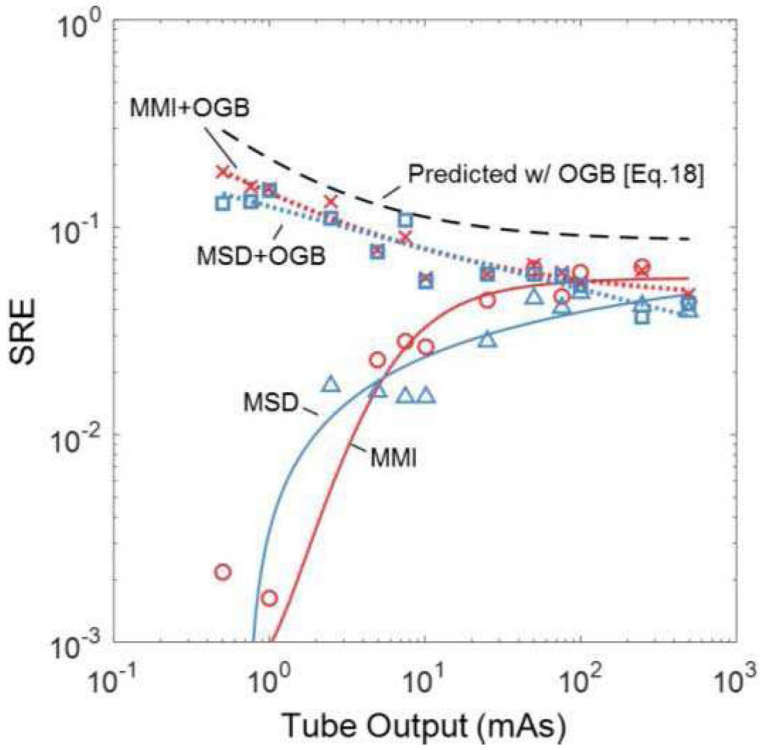

Fig. 5.

SRE evaluated as function of dose for MSD (blue) and MMI (red) with and without optimal Gaussian blur (OGB). The predicted SRE (with OGB) is shown as the black dashed line, demonstrating a similar dose dependence as the measurements with optimal blur.