Rigid Registration

Ziv Yaniv

1.1 Introduction

Rigid registration has been studied extensively over the past several decades, both for medical and non medical applications [28, 32]. Surprisingly, it is still an active area of research [26, 54, 64].

In the medical context, the use of a rigid transformation to align data is particularly attractive. The transformation itself is the simplest useful transformation, with only six degrees of freedom. It is invertible, has a simple composition rule, and, most importantly, there are rigid registration tasks that have analytic solutions with well understood error predictors.

This is why, currently, the majority of image guided navigation systems model *whole* anatomical structures as rigid bodies and employ rigid registration. Although this model is only correct for registration of intrapatient osseus structures, there are cases where it is *sufficiently accurate*[1] for procedures concerned with soft tissue, most notably neurosurgical procedures.

Registration methods are categorized according to many criteria that broadly fall into two general classes, input related, and algorithm related [58]. We describe rigid registration algorithms by first classifying them according to the data dimensionality, an input related criterion, and then classifying them according to the data type used by the registration process, an algorithm related criterion.

Currently the available input sources for registration include three dimensional images (e.g. CT), two dimensional projective images (e.g. X-ray fluoroscopy), two dimensional tomographic images (e.g. Ultrasound), and three dimensional digitized points obtained in several ways including, contact based digitization with a tracked probe, laser range scanning, and stereo imaging using standard video images.

Data types used for registration can be classified as either geometric objects, or image intensity data. The rationale for using geometric objects is that using sparse, accurately segmented, geometric data results in fast running times, and that intra-modal and inter-modal registration are not distinct cases. The rationale for using image intensities is that they do not require accurate segmentation and that the use of more information potentially

[1] By *sufficiently accurate* we mean that the registration error does not effect the procedure outcome. That is, the distance between correct and computed point locations is less than a pre-specified threshold as defined by a physician for a specific procedure (a subjective measure).

results in robustness to outlying intensity values.

Using the criteria described above, we first classify algorithms as either three dimensional to three dimensional (3D/3D), or two dimensional to three dimensional (2D/3D) and then classify them into geometry or intensity based algorithms. Algorithms that do not fall into these categories are also described at the end of each section.

1.2 3D/3D registration

Image guided navigation systems utilize 3D/3D rigid registration for two purposes, alignment of complementary image data sets for planning, and alignment of image data to the physical world for intraoperative navigation.

Definition Given two data sets defined over the domains $\Omega_l \subset \mathbb{R}^3, \Omega_r \subset \mathbb{R}^3$:

$$F_l : \mathbf{x_l} \in \Omega_l \mapsto F_l(\mathbf{x_l})$$
$$F_r : \mathbf{x_r} \in \Omega_r \mapsto F_r(\mathbf{x_r})$$

with, possibly, additional side information (e.g. pairing between corresponding features) find the rigid transformation:

$$\mathcal{T} : \mathbf{x_l} \mapsto \mathbf{x_r} \Leftrightarrow \mathcal{T}(\mathbf{x_l}) = \mathbf{x_r}$$

where $\mathbf{x_l} \in \Omega_l,\ \mathbf{x_r} \in \Omega_r$.

Specific registration instances are defined by the mappings F_l and F_r. Two common mapping combinations are: 1) F_l and F_r are the identity map, point set to point set registration; and 2) F_l and F_r map locations to intensity values, image to image registration.

1.2.1 Geometry based methods

Various geometric objects have been used for 3D/3D registration with point and surface data being the most popular, and thus the focus of this section.

To understand why these are the geometric objects of choice we evaluate them with regard to a set of desired characteristics:

1. Closed form solution - There is a registration algorithm based on this geometric object that has a closed form solution.
2. Number of detectable objects - The number of corresponding objects detectable across data sets is sufficient for registration.
3. Localization - Accurate object localization in all data sets is simple and has a low computational cost.
4. Correspondence - Establishing object correspondence across data sets is simple.

Point data arises from two sources, anatomical structures and fiducial markers. In both cases there are closed form solutions when correspondences between data sets are known

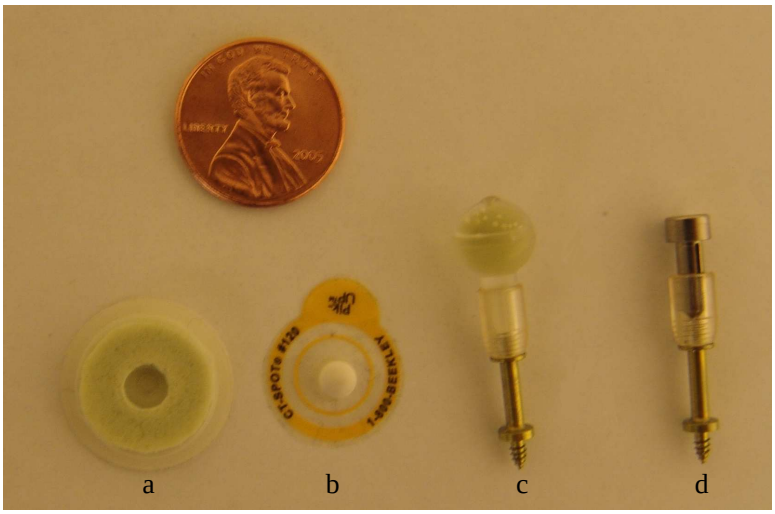


Figure 1.1: Registration fiducials: (a) CT/MR compatible adhesive skin marker from IZI Medical Products Corp., Baltimore MD, USA; (b) CT compatible adhesive skin marker from Beekley Corp., Bristol CT, USA; (c,d) Acustar II CT/MR compatible implantable markers with interchangeable caps from Z-kat, Hollywood Fla, USA, manufactured by Stryker, Kalamazoo MI, USA (c) with imaging cap and (d) with conical cap for pointer based digitization (implantable markers courtesy of J. M. Fitzpatrick).

[1, 18, 37, 38, 80, 87, 93]. This is the primary reason for the popularity of point based registration which is one of the few registration tasks that have a closed form solution. The number of distinct anatomical points that can be detected per organ across modalities is usually small and their accurate localization is a hard task which is still a subject of research [97]. Fiducial points, on the other hand, arise from markers that are designed to have the desired characteristics (see Figure 1.1 for examples). The number of points is set by the user and the fiducial design is such that detecting and accurately localizing it across the relevant modalities can be done efficiently [61, 94]. Finally, for both point types establishing correspondences can be done both automatically (e.g. [30]) or manually, with the manual approach most often used in image to physical space registration.

Surfaces arise from segmentation of anatomical structures in images, and from dense point digitization of exposed anatomical structures. Unlike point data, there is no closed form solution to surface based registration. In most cases a single surface in each data set is used for registration, providing that the detected surfaces have sufficient overlapping regions. The ease of iso-surface segmentation in medical images is the primary reason for the wide spread use of surface data for registration, most notably the marching-cubes algorithm [52]. Finally, establishing correspondences between surfaces is usually not an issue as in most cases only a single surface is involved.

Paired Point methods

Paired point registration arises when the mappings F_l and F_r in Definition 1.2 are the identity mapping and point correspondences are known. For a unique solution to exist the data must include three or more non collinear points.

In general, we are given two sets of corresponding 3D points obtained in two Cartesian coordinate systems using a measurement process having additive noise:

$$\begin{aligned} \mathbf{x_{li}} &= \widehat{\mathbf{x_{li}}} + \mathbf{e_{li}} \\ \mathbf{x_{ri}} &= \widehat{\mathbf{x_{ri}}} + \mathbf{e_{ri}} \end{aligned}$$

where $\widehat{\mathbf{x_{li}}}, \widehat{\mathbf{x_{ri}}}$ are the true point coordinates that satisfy the constraint $\mathcal{T}(\widehat{\mathbf{x_{li}}}) = \widehat{\mathbf{x_{ri}}}$, $\mathbf{x_{li}}, \mathbf{x_{ri}}$ are the observed coordinates, and $\mathbf{e_{li}}, \mathbf{e_{ri}}$ are the respective errors in the measurement process.

The most common solution to this problem is based on a least squares formulation:

$$\mathcal{T}: \sum_{i=1}^{n} \|\mathbf{x_{ri}} - \mathcal{T}(\mathbf{x_{li}})\|^2 \rightarrow \min$$

The transformation $\mathcal{T} = [\mathcal{R}, \mathbf{t}]$ is optimal when the errors, $\mathbf{e} = \mathcal{R}(\mathbf{e_l}) - \mathbf{e_r}$, are independent and have an identical zero mean Gaussian distribution. This assumes that the original errors are isotropic, Independent and Identically Distributed (IID) with a zero mean Gaussian distribution. In most cases these assumptions do not hold (e.g. optical tracking where the uncertainty is greater along the viewing direction). Several algorithms that relax these assumptions have been published. These include a generalized total least squares formulation [68], a hetroscedastic errors-in-variables regression formulation [60], a method for dealing with anisotropy arising from the use of reference frames [3], an extended Kalman filter formulation [70], and an unscented Kalman filter formulation [65].

Currently none of these methods is in wide spread use. The continued use of least squares solutions can be attributed to their mathematical and implementational simplicity and to the fact that for many procedures their accuracy is sufficient, although not optimal.

The primary difference between the various least squares solutions is in their representation of the transformation operator. These include the standard matrix representation [1, 38, 80, 87], a unit quaternion based rotation representation [18, 37], and a dual quaternion representation of the transformation [93]. While the mathematical derivation of the solution varies according to the transformation representation, the performance of all algorithms has been empirically found to be comparable [16].

We now briefly describe the two most popular solutions, with the rotation operator represented as a matrix, and unit quaternion. In both cases the estimation of the rotation and translation is decoupled by replacing the original measurements with their demeaned versions: $\mathbf{x'_{li}} = \mathbf{x_{li}} - \mu_\mathbf{l}$, $\mathbf{x'_{ri}} = \mathbf{x_{ri}} - \mu_\mathbf{r}$. The least squares formulation is then rewritten as

$$[\mathcal{R}, \mathbf{t}]: \sum_{i=1}^{n} \|\mathbf{x'_{ri}} - \mathcal{R}(\mathbf{x'_{li}}) - \mathbf{t'}\|^2 \rightarrow \min, \qquad \mathbf{t'} = \mathbf{t} - \mu_\mathbf{r} + \mathcal{R}(\mu_\mathbf{l})$$

with the optimal translation given by

$$\mathbf{t} = \mu_\mathbf{r} - \mathcal{R}(\mu_\mathbf{l})$$

In both cases computation of the optimal rotation utilizes the following correlation matrix

$$C = \sum_{i=1}^{n} \mathbf{x'_{li}} \mathbf{x'_{ri}}^T = \begin{bmatrix} C_{xx} & C_{xy} & C_{xz} \\ C_{yx} & C_{yy} & C_{yz} \\ C_{zx} & C_{zy} & C_{zz} \end{bmatrix}$$

with the matrix based solution using the Singular Value Decomposition, $SVD(C) = U\Sigma V^T$, yielding the optimal rotation matrix as

$$\mathcal{R} = U \begin{bmatrix} 1 & & \\ & 1 & \\ & & det(UV^T) \end{bmatrix} V^T$$

and the unit quaternion based solution given by the eigenvector corresponding to the maximal eigenvalue of the matrix

$$N = \begin{bmatrix} C_{xx}+C_{yy}+C_{zz} & C_{yz}-C_{zy} & C_{zx}-C_{xz} & C_{xy}-C_{yx} \\ C_{yz}-C_{zy} & C_{xx}-C_{yy}-C_{zz} & C_{xy}+C_{yx} & C_{zx}+C_{xz} \\ C_{zx}-C_{xz} & C_{xy}+C_{yx} & C_{yy}-C_{xx}-C_{zz} & C_{yz}+C_{zy} \\ C_{xy}-C_{yx} & C_{zx}+C_{xz} & C_{yz}+C_{zy} & C_{zz}-C_{xx}-C_{yy} \end{bmatrix}$$

Surface based methods

Surface registration arises when one or both of the mappings F_l and F_r in Definition 1.2 are indicator functions with a value of one at surface points and zero otherwise. While this description uses an implicit surface representation, in most cases surfaces are represented explicitly as triangular meshes, although other implicit (e.g. distance maps) and explicit (e.g. NURBS) representations are also possible.

We limit our description of surface based registration methods to the case where one or both surfaces are represented as a set of points, as in the medical context this is the most ubiquitous instance of surface registration. This is primarily due to the fact that it is easy to reduce all surface registration tasks to this specific task via sampling. For descriptions of additional surface based algorithms including deformable registration see [2].

One of the first successful surface based registration algorithms is the head-and-hat algorithm of Pelizzari et al. [69]. The algorithm aligns two surface representations of the brain segmented from complementary imaging modalities. The surface extracted from the lower resolution data, the "hat", is represented as a set of points, while the surface extracted from the higher resolution data, the "head", is represented as a set of stacked 2D closed contours. The algorithm iteratively minimizes the sum of distances between the rays originating at "hat" points going through the "head" centroid, and the "head" surface using the Powell's standard optimization method [75]. Note that this algorithm is tailored for registration of images of the head, as the point matching strategy best suites spherical objects. In many ways this algorithm is a precursor of what is currently the common approach to point set to surface registration, the Iterative Closest Point (ICP) algorithm.

The ICP algorithm was independently introduced in [100], [11], and [5], with the later coining the name ICP. In practice the acronym ICP is now used to describe algorithms that follow an Iterative Corresponding Point framework, rather than the original closest point approach. In essence the ICP framework is an iterative two step approach with the first step establishing point correspondences and the second step computing a transformation based on these matches, yielding incremental transformations whose composition is the registration result. Table 1.1 summarizes this general framework.

The popularity of the ICP framework is primarily due to its presentation in [5]. The algorithm is simple and a proof of local convergence when correspondence is based on

Input: point set P_l, surface S_r,
$\tau > 0$ - improvement threshold, n - maximal number of iterations

0. Initialization:
   (a) Set cumulative transformation, and apply to points.
   (b) Find *corresponding* points and compute similarity (e.g. root mean square distance).
1. Iterate:
   (a) Compute incremental transformation using the current correspondences (e.g. analytic least squares solution), update cumulative transformation and apply to points.
   (b) Find *corresponding* points and compute similarity.
   (c) If improvement in similarity is less than τ, or number of iterations has reached n terminate.

Table 1.1: Iterative Corresponding Point (ICP) framework

the Euclidean distance is given. Additionally, the technique is shown to be versatile and successful, as the authors both describe closest point computations using several surface representations and show successful registration results on a variety of geometric objects.

In practice, the original framework suffers from several deficiencies: (1) convergence is local, requiring an initial transformation near the global optimum; (2) point pairing is a computationally expensive operation; and (3) the framework uses a least squares method for computing the incremental transformations and is thus sensitive to outliers and makes the implicit assumption that noise is isotropic IID with a zero mean Gaussian distribution.

A large body of papers addressing these issues exists.

Methods for improving the probability of convergence to the global optimum include, initialization approaches such as manual alignment of the data or initial registration using inaccurately localized paired points such as anatomical landmarks [55], multiple registration runs with different initial transformations [5], and use of heuristic global optimization methods such as simulated annealing [25, 53, 73].

Methods for accelerating the running time include, use of the kd-tree spatial data structure [5, 29, 81], closest point caching [81], extrapolation of the transformation parameters [5, 81], approximate nearest neighbor searches [29], parallelized nearest neighbor searches [45], hierarchical coarse to fine data sampling approaches [100, 40], and combinations of these methods.

Methods for dealing with the deficiencies of the least squares approach provide robustness by replacing the least squares formulation with methods based on M-estimators [42, 55], least median of squares [59, 85], least trimmed squares [12], and weighted least squares [62, 86, 100]. To remove the assumption on the noise characteristics the least squares approach can be replaced by the generalized total least squares [17], the unscented Kalman filter [65], or any of the other methods described in the previous section.

Finally, it should be noted that some researchers [19] advocate the use of classical optimization approaches instead of the two step ICP framework. The rationale being that the large body of research on robust estimation using classical optimization tools can be brought to bear on the current task. The proposed algorithm replaces point pairing with a precomputed distance map and liner interpolation to estimate point to surface distances. Optimization is then carried out using the Levenberg-Marquardt algorithm with M-estimators replacing the the sum of squares objective function. While this approach was not evaluated using medical data, a similar solution has been used successfully in the medical context [4], registering point data extracted from ultrasound to a bone surface extracted from CT.

1.2.2 Intensity based methods

Intensity based registration arises when the mappings F_l and F_r in Definition 1.2 map locations to intensity values, image to image registration. These are either the original images as acquired by the imaging apparatus or images derived from them.

While most intensity based registration algorithms utilize the original intensity values directly, some methods perform a "fuzzy segmentation", attempting to strike a balance between the geometric feature based and purely intensity based approaches. Images derived from the original ones are created by mapping the original intensities to values that are loosely related to geometric structures found in the images. The new images are then used as input for the registration algorithm.

Examples of this approach include, registration of CT/MR brain images [57], registration of US/MR liver images [71], and registration of US/CT bone images [72]. In [57] the original intensities are replaced with values related to the likelihood of the spatial location being an edge or ridge, based on gradient and Laplacian information. In [71] and [72] the original intensities are replaced with values representing the probability that the spatial location is part of a blood vessel, in the former case, and a bone soft tissue interface, in the later case. The mapping function is learnt from a segmented training set using image features such as intensity values, gradient magnitude, and characteristics of intensity profiles along the beam direction in the case of US. In all cases the derived images replace the original ones as input for intensity based registration.

Intensity based registration is cast as an optimization task where the objective function is directly dependent on the image intensity values and the transformation parameters. Given that optimization is carried out over a continuous domain and intensity values are only available on a discrete grid the registration process invariably requires interpolation of intensity values at non grid locations. A registration algorithm is thus uniquely defined by the following components:

1. Similarity measure (objective function).
2. Optimization algorithm.
3. Interpolation scheme.

The first step when casting a problem as an optimization task is to define the objective function. Most research has focused on devising task specific similarity measures, as this

is the only component in the registration algorithm that is directly related to the data at hand.

Ideally a similarity measure is continuous, strictly monotonic and has a unique global optimum for the correct transformation parameters. In practice the only requirement from a similarity measure is that it have a local optimum for the correct parameters. Given this condition it is not surprising that many similarity measures have been proposed. In general similarity measures can be divided into two classes, measures that assume the existence of a deterministic functional relationship between intensities and measures that assume a stochastic relationship. Table 1.2 presents various image intensity relationships and appropriate similarity measures that are described in the following discussion.

The simplest intensity relationships arise in the context of intra-modality registration. Intensity values are assumed to be identical across images or related via an affine transformation (often referred to as linear).

Two related similarity measures that assume intensities are only subject to zero mean Gaussian noise are the Sum of Squared Differences (SSD, squared L_2 norm of the difference image) and Sum of Absolute Differences (SAD, L_1 norm of the difference image). The primary distinction between them being that the L_1 norm degrades more gracefully in the presence of outliers, although it is not robust in the breakdown point sense. Another, more robust, measure that makes the same assumption is the entropy of the difference image [9], $H(F_l(\mathbf{x}) - F_r(\mathbf{x}))$, although it is less popular due to its greater computational complexity.

In practice, even when intensities are assumed to be identical the noise does not necessarily follow a zero mean Gaussian distribution. This is why more often an affine intensity mapping is assumed and the Normalized Cross Correlation (NCC), *Pearson's r* [75], is used. This is probably the most popular similarity measure to assume a deterministic functional relationship between intensities. It should be noted that simple variations of the NCC such as tessellating the images and using the sum of *local* NCC values as the similarity measure can accommodate more general functional relationships such as those relating MR and CT [95]. A mathematical model motivating such an approach can be derived if it is assumed that the intensity relationship between the images is characterized by a low frequency spatially varying multiplicative and additive bias field:

$$F_r(\mathbf{x}) = f_1(\mathbf{x})F_l + f_2(\mathbf{x})$$

using the Taylor expansion and taking only the first terms we get:

$$\begin{aligned} F_r(\mathbf{x_0} + \delta) &= [f_1(\mathbf{x_0}) + \nabla f_1(\mathbf{x_0})\delta + O(\delta^2)]F_l + \\ &\quad [f_2(\mathbf{x_0}) + \nabla f_2(\mathbf{x_0})\delta + O(\delta^2)] \\ &\simeq f_1(\mathbf{x_0})F_l + f_2(\mathbf{x_0}) \end{aligned}$$

The intensity relationship is to first order locally affine making the sum of NCC values in a tesselated image a natural similarity measure.

The most generic similarity measure to assume intensities are related via a deterministic functional is the correlation ratio [77]. Unlike the previous similarity measures the correlation ratio does not require an explicit functional. This is an advantage, as explicit

functionals are usually based on *approximations* of the physical imaging processes, which are often hard to model. While the correlation ratio is not in wide spread use, it has been successfully applied to rigid registration of PET/T1, CT/T1, and T1/T2 data sets [77].

More complex intensity relationships arise in the context of inter-modality registration, where the intensity relationships are stochastic and not deterministic. Similarity measures based on the information theoretic concept of entropy were introduced in the 1990's. These measures assume that intensities are related via a statistical process, and that the joint and marginal image entropies can be used to align the images. In most cases the measures use Shannon's definition of entropy, $H = -\sum_i p_i \log p_i$, although other definitions exist (e.g. [41]).

The two most widely used similarity measures in this category are Mutual Information (MI) and Normalized Mutual Information (NMI).

MI was independently proposed as a similarity measure by Collignon et al. [13, 56] and Viola and Wells [91, 90]. It has been successfully applied to various inter and intra modality registrations, including CT/MR, PET/MR, MR/MR, and CT/CT. NMI was proposed in [82] as an overlap invariant version of MI. The normalization results in a measure that does not depend on the changes in the marginal entropy values due to changes in the region of overlap.

An exhaustive survey of these similarity measures and their application to various registration tasks is given in [74]. This survey is perhaps one of the few papers to explicitly address the fact that implementation decisions can significantly influence registration performance. In the case of MI, its definition is unique, but its implementation relies on the estimation of the joint and marginal intensity probability distributions. Thus the choice of estimation method and its parameters can influence the registration results. The two common estimation approaches are joint histogram computation and Parzen windowing [15]. Interestingly, the original MI papers also differ in their implementation choices with Collignon et al. using a histogram based estimate and Viola and Wells using a Parzen windowing estimate.

While MI is applicable in many registration tasks it does not address all registration challenges, most notably the correct transformation parameters will usually correspond to a local optimum. This requires either good initialization when using general purpose optimization techniques and caution when applying heuristic global search methods, such as simulated annealing. In addition, the computational cost of MI is higher than that of most deterministic functional similarity measures.

Currently, there is no universal similarity measure that is optimal for all registration tasks, although without problem specific knowledge MI is at least applicable. Incorporating problem specific knowledge when devising or choosing a similarity measure usually improves all aspects of registration performance.

In most cases the "best" similarity measure is chosen from several applicable ones based on its computational complexity, robustness, accuracy, and convergence range. The first two criteria can be analyzed in a context independent manner based on the mathematical definition of the similarity measure. The remaining criteria, accuracy and convergence range, do depend on the specific imaging modalities, and are most often evaluated in one of two ways: assessing registration performance as a function of the similarity measure, and by exploring the behavior of the similarity measure as a function of the transformation

relationship between intensity values	similarity measure	
identity	Sum of Squared Differences (SSD)	$\frac{1}{n}\Sigma_{\mathbf{x}\in\Omega_{lr}}\|F_l(\mathbf{x}) - F_r(\mathbf{x})\|^2$
identity	Sum of Absolute Differences (SAD)	$\frac{1}{n}\Sigma_{\mathbf{x}\in\Omega_{lr}}\|F_l(\mathbf{x}) - F_r(\mathbf{x})\|$
affine	Normalized Cross Correlation (NCC)	$\frac{\Sigma_{\mathbf{x}\in\Omega_{lr}}(F_l(\mathbf{x})-\overline{F_l})(F_r(\mathbf{x})-\overline{F_r})}{\sqrt{\Sigma_{\mathbf{x}\in\Omega_{lr}}(F_l(\mathbf{x})-\overline{F_l})^2\Sigma_{\mathbf{x}\in\Omega_{lr}}(F_r(\mathbf{x})-\overline{F_r})^2}}$
general functional	Correlation Ratio (CR)	$1 - \frac{Var[F_l(\mathbf{x})-E(F_l(\mathbf{x})\|F_r(\mathbf{x}))]}{Var[F_l(\mathbf{x})]}$
stochastic	Mutual Information (MI)	$H(F_l(\mathbf{x})) + H(F_r(\mathbf{x})) - H(F_l(\mathbf{x}), F_r(\mathbf{x}))$
stochastic	Normalized Mutual Information (NMI)	$\frac{H(F_l(\mathbf{x}))+H(F_r(\mathbf{x}))}{H(F_l(\mathbf{x}),F_r(\mathbf{x}))}$

Table 1.2: Examples of intensity relationships and appropriate similarity measures. All similarity measures are defined over the domain $\Omega_{lr} = \mathcal{T}(\Omega_l) \cap \Omega_r$, where $\mathcal{T}$ is a rigid transformation and Ω_l, Ω_r are the original image domains. The MI and NMI similarity measures are expressed using the Shannon entropy, $H = -\sum_i p_i \log p_i$.

parameters. In the former case the similarity measure is evaluated based on its impact on the accuracy and convergence range of the registration by comparing registration results to a known "gold standard" transformation. In the later case two dimensional graphs of the similarity measure as a function of each of the transformation parameters are plotted. These are two dimensional axis aligned orthogonal slices through the six dimensional parameter space, centered on the known parameter values obtained from a "gold standard" transformation.

Both approaches are lacking. The registration based approach does not evaluate the similarity measure, rather it evaluates the combination of similarity measure and a specific optimization technique. The parameter space exploration approach is based on visual inspection and on the ability of the viewer to mentally reconstruct a six dimensional terrain based on six axis aligned orthogonal slices, a task which is nearly impossible. This approach is thus qualitative and not quantitative.

A quantitative parameter space exploration approach has been recently presented in [92]. This approach starts by exhaustively sampling the parameter space along diameters of a six dimensional hypershpere centered on the "gold standard" parameter values. After evaluating the similarity measure at the sample points several figures of merit are computed. These describe the function behavior based on the extrema points along the diameters and the similarity measures rate of change near these extrema.

The second component of iterative intensity based registration is the optimization algorithm. The majority of algorithms utilize general purpose iterative optimization algorithms, unconstrained or constrained. Starting from an initial transformation $\mathcal{T}^{(0)}$ in each iteration a new transformation $\mathcal{T}^{(i)}$ is estimated based on the similarity measure and the previous estimate. This sequence converges to a local optimum $\mathcal{T}^{(*)}$, with the iterations terminated when a user specified convergence test is satisfied. For detailed descriptions of standard

optimization methods the reader is referred to [23].

While many general purpose optimization algorithms are available, the no-free-lunch theorem [36] proves that without making prior assumptions on the objective function there is no single algorithm that is superior to all others across all possible inputs. Thus, the choice of optimization algorithm cannot be separated from that of the similarity measure, as an algorithm will outperform all others only if it incorporates prior information on the objective function.

Finally, a plethora of heuristic strategies for improving the registration convergence range are in use. Among others, these include multi-resolution strategies, multi-start strategies and heuristic search algorithms such as simulated annealing and genetic algorithms. Again, we emphasize that while these strategies can improve the probability of convergence, they should be used with care, incorporating problem specific constraints, as the correct transformation parameters often coincide with a local and not global extremum.

The third component required for iterative intensity based registration is an interpolation scheme. The majority of algorithms use linear interpolation, usually viewed as a compromise between computational complexity and accuracy of the estimated intensity values. In some cases this has an adverse effect on the similarity measure, for example MI and NMI estimates based on linear interpolation exhibit additional local extrema when the images are translated to grid locations [39], although this effect can be mitigated by blurring the images prior to registration [34].

Evaluation studies of various interpolation schemes using medical images [48, 63] have shown that barring nearest neighbor, linear interpolation is the fastest interpolation method although its accuracy is lacking. Both studies show that the best tradeoff between accuracy and computational complexity is obtained for B-spline interpolation. Given these conclusions it is surprising to find that most registration algorithms still use linear interpolation. This preference has to do with the effect of the interpolation scheme on the optimization process, primarily its running time and accuracy. The impact of the interpolation scheme on the total running time is considerable, as interpolation is performed many times during optimization. The impact it has on accuracy is usually not as considerable. Given that the necessary condition for accurate registration is that the similarity measure have a local extremum for the correct transformation parameters, inaccurate interpolation may not be too detrimental as long as this condition is satisfied. Finally, as we are dealing with iterative optimization, interpolation schemes that introduce additional local extrema in the similarity measure reduce the convergence range which can greatly impact automated registration algorithms as is the case of linear interpolation and MI [39].

1.3 2D/3D registration

Image guided navigation systems utilize 2D/3D rigid registration to align image data to the physical world for intraoperative navigation.

Definition Given two data sets over the spatial domains $\Omega_L \subset \mathbb{R}^3, \Omega_{Ri} \subset \mathbb{R}^2$ (i=1 ... n),

$$F_L : \mathbf{x_L} \in \Omega_L \mapsto F_L(\mathbf{x_L})$$
$$F_R : \mathbf{x_{Ri}} \in \Omega_{Ri} \mapsto F_R(\mathbf{x_{Ri}})$$

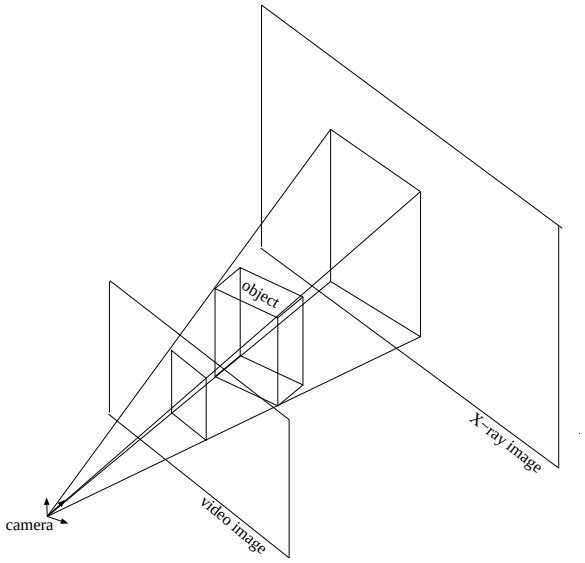


Figure 1.2: Pinhole camera model for X-ray and video images. Location of the image plane is behind the imaged object in the former case, and in front of the object in the later case.

and camera matrices, $K_i = P_i T_i$, where T_i is a rigid transformation and P_i a perspective projection matrix, find the rigid transformation:

$$\mathcal{T} : \mathbf{x_L} \mapsto \mathbf{x_{Ri}} \Leftrightarrow K_i(\mathcal{T}(\mathbf{x_L})) = \mathbf{x_{Ri}}$$

where $\mathbf{x_L} \in \Omega_L,\ \mathbf{x_{Ri}} \in \Omega_{Ri}$.

It should be noted that, in the medical context, the only modalities that are modeled as 2D perspective images are X-ray imaging and video images, such as those acquired by endoscopic systems. While the pin-hole camera model is applicable to both modalities they do differ in the location of the image plane relative to the imaged object. Figure 1.2 shows the relationships between camera, object and image plane in both cases.

Figure 1.3 describes the generic 2D/3D registration problem. One or more perspective images are acquired from known poses relative to a *world* (fixed) coordinate system W. Registration is then the task of estimating the transformation relating the volume coordinate system to the world coordinate system, $\mathcal{T} = T_V^W$. In general 2D/3D registration is formulated as an optimization task with the objective function being a sum of objective functions computed individually per each image.

Note that 2D/3D registration as defined above is closely related to the extensively studied computer vision problem of pose estimation. In pose estimation a single image is used as input and the world and camera coordinate systems coincide. Corresponding features, usually points or lines, are extracted from a geometric 3D model and the 2D image and used to compute the camera to model transformation. Feature extraction is performed only once, and the accuracy of the whole approach is highly dependent on accurate feature extraction and pairing, tasks which are difficult to perform using images of anatomical structures, as the number of distinct features is usually small.

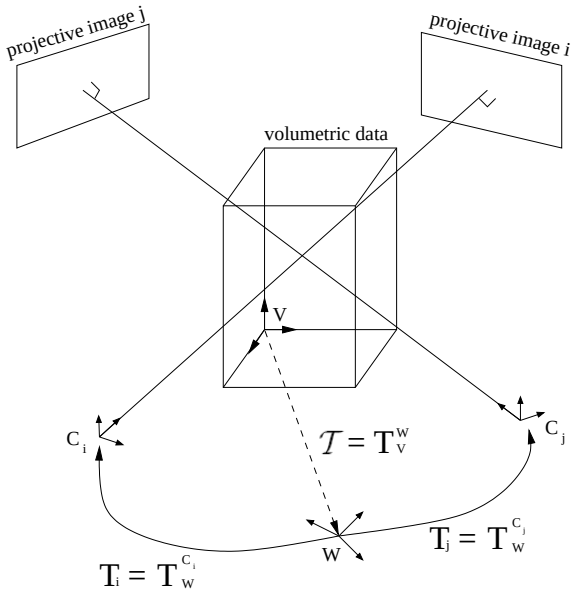


Figure 1.3: Coordinate systems used in 2D/3D registration. One or more images are acquired with known camera poses relative to a fixed, world, coordinate system. The estimated transformation relates the 3D volumetric coordinate system to the fixed coordinate system. Known transformations are marked by solid lines, estimated transformation by a dashed line. The figure uses a X-ray image plane to object relationship.

In the medical context, all 2D/3D rigid registration algorithms are iterative, requiring transformation initialization. Choice of initialization approach depends both on the available data, and on the expected convergence range of the registration algorithm. Several common initialization schemes exist:

1. manual initialization - the user interactively explores the transformation parameter space, visually comparing the medical images to computer generated images created using the current parameter values. The initial transformation parameters correspond to the most similar computer generated images, as visually assessed by the user.

2. clinical setup - knowledge of patient position (e.g. supine) and imaging apparatus poses define bounds on the transformation parameters. For example, in X-ray imaging each camera pose defines a bounding pyramid for the imaged object, with the image extent serving as the pyramid base and the camera location, its apex. For improved object localization it is possible to use the intersection of multiple pyramids as the bounding volume.

3. inaccurate paired point registration - using a paired point closed form solution with inaccurate point data. This approach can be used:

   - When the location of three or more anatomical landmarks is available. Anatomical landmark locations can be estimated in one of two ways: touching the landmark with a calibrated tracked probe, if the landmark is already exposed as part of the procedure, or by indicating the landmark location on multiple perspective

images and using triangulation to localize it in physical space. In both cases registration inaccuracy is primarily due to the difficulties in accurate identification of anatomical landmarks both on the patient anatomy [76], and across multiple images.

- When using skin adhesive fiducials. Fiducial locations are estimated by touching the fiducials with a calibrated tracked probe. In this case, registration inaccuracy is primarily due to fiducial motion between the imaging and digitization stages.

1.3.1 Geometry based methods

In many ways geometry based 2D/3D registration algorithms are very similar to their 3D/3D counterparts. The geometric objects are 3D surfaces, 2D image points corresponding to feature locations, and 3D rays emanating from the camera location and going through the spatial locations that correspond to the 2D features (back-projected rays). Figure 1.4 shows these geometric entities in the context of CT/X-ray registration.

In [6] known 2D/3D paired points are used to align CT and portal images for patient positioning in radiotherapy. The 3D points are projected onto the portal image and the distances between their projection and the known point locations are minimized in the image plane. Two closed form solutions are presented with the first assuming that the 3D points are approximately planar allowing for a weak perspective projection model, and the second assuming small rotations allowing for rotation linearization. This is in practice a simplified version of the classical computer vision task of pose estimation.

A registration algorithm for aligning X-ray and CT images of the head described in [67] can be viewed as a variant of the head-and-hat algorithm [69] with optimization carried out in 2D instead of 3D. In each iteration a Digitally Reconstructed Radiograph (DRR), simulated X-ray derived from CT, is generated. A set of 2D rays emanating from the center of the X-ray image ("head" centroid) and going towards the image boundary at regular angular intervals is defined. An edge detector is then used to determine the skull edge location along the ray direction in the DRR and X-ray images and this distance is minimized, along with the mean grey level value along the ray. Optimization is done with an algorithm that initially performs gradient descent and then switches to an approximation of the function using the first terms of the Taylor expansion.

A straightforward extension of the ICP algorithm to 2D/3D registration is described in [98]. In this case, instead of matching two sets of 3D points, back-projected rays are matched to the closest 3D model points. Once the pairs of ray-model points are established the incremental transformation is computed using the SVD based analytic solution for 3D/3D paired-point rigid registration. To improve robustness, the sum of squares objective function is replaced with its weighted version using the Huber M-estimator.

A similar ICP like approach is described in [31]. Incremental transformations are computed with a weighted version of the sum of squares objective function using the Tukey M-estimator, and an analytic solution of the 3D/3D paired point based on the Cayley rotation representation is derived. In addition only a small subset of model points are used in the closest point computation stage. The authors observe that at the correct position all back-projected rays are tangent to the model surface defining a 3D space curve on the surface of the model. As the rays are defined by the object's 2D boundary, its apparent

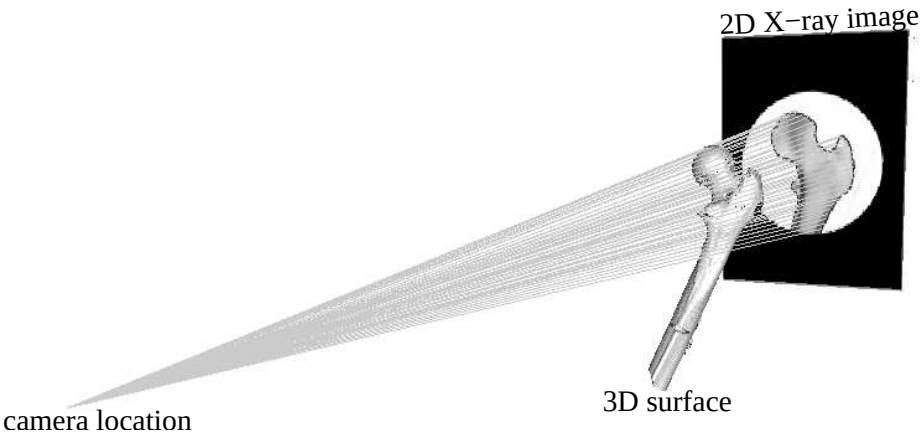


Figure 1.4: Geometric entities used in 2D/3D (CT/X-ray) geometry based registration of a femur bone. These include the 3D bone surface extracted from CT, the spatial locations of the 2D bone contour points on the X-ray image and the corresponding back-projected rays. Bone model is displayed in its correct position with back-projected rays tangent to the model surface.

contour, they should only match points on the 3D space curve that separates front-facing from back-facing regions, the contour generator. At each iteration the contour generator is computed based on the object's current pose. The closest points are then constrained to lie on this 3D curve.

A feature based two stage algorithm for registering endoscopic video images to CT is described in [8]. In the first stage the endoscopic camera motion and a scaled 3D model are estimated. The estimate is based on the identification of three or more corresponding points across multiple video images. In the second stage the scaled 3D model is registered to a surface derived from the CT. Initially the scale of the reconstructed model is estimated using Principle Component Analysis of the model and a roughly corresponding local CT surface patch. Once the model's scale is known the problem is reduced to 3D/3D point to surface registration and the transformation is estimated using the ICP algorithm.

In [47] X-ray images are registered to CT data using back-projected rays corresponding to object contours in the X-ray image, and an implicit surface representation. Observing that for the correct registration parameters the back-projected rays are tangent to the object surface, the squared distance between the back-projected rays and surface is used as an objective function. Minimization is then performed with the Levenberg-Marquardt method. The novel aspect of this approach is in the use of an adaptive, octree-based, distance map to represent the surface. Figure 1.5 visually illustrates this spatial data structure. It should be noted that the algorithm assumes that *all* image contour points arise from projections of the object's surface points, an assumption which is rarely valid due to the presence of multiple objects in an image. In [33] this restriction is mitigated by using a cooperative registration-segmentation approach. Instead of using all contour points to generate the back-projected rays, the surface model is projected onto the image using the initial registration estimate and only contour points that are within a specified distance from the projected surface are utilized. This in turn implies that the initial registration estimate is close to the correct one. Interestingly, the basic algorithm can be viewed as the 2D/3D equivalent of the 3D/3D

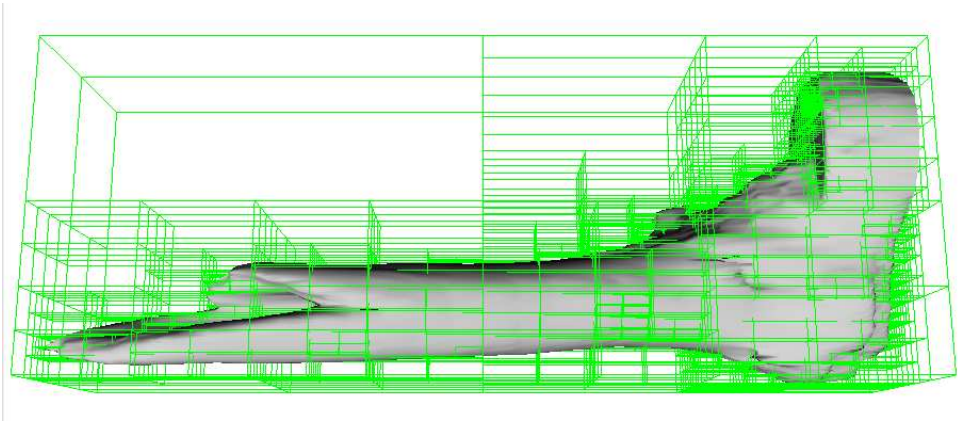


Figure 1.5: Surface of femur embedded inside an octree spatial data structure, providing an adaptive distance map. The distance to the object surface is computed for all vertices of the octree's terminal nodes (leaves). Given a query point its distance is computed using linear interpolation of the vertex distances for the terminal node that contains it.

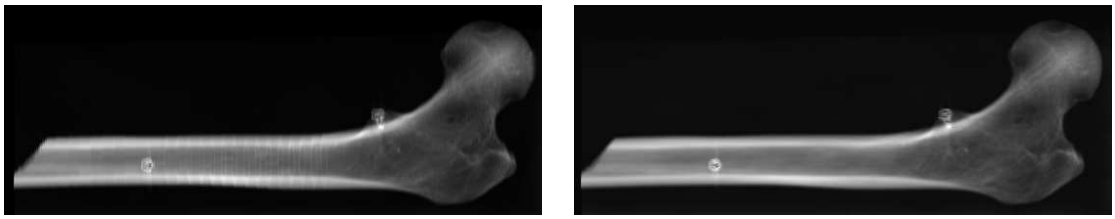
registration approach proposed by Fitzgibbon in [19] and described above. The primary difference between these methods is in the implementation of the distance map. In the 3D/3D case the distance map is unsigned and spatially uniform while in the 2D/3D case it represents a signed distance and is spatially non-uniform.

1.3.2 Intensity based methods

Intensity based 2D/3D registration methods use the same iterative optimization framework as their 3D/3D counterparts with a slight modification due to the dimensionality difference between data sets. As direct comparison of the 2D and 3D images is not possible, an intermediate step is introduced in which 2D simulated images are generated using the 3D image and the current transformation parameters. These simulated images are then compared with the actual ones using a similarity measure as was done in the 3D/3D case. It should be noted that image interpolation is performed during the 2D image generation stage, and no interpolation is required as part of the similarity measure computation.

While this approach is the rule there is an exception [83]. Instead of simulating 2D X-ray images from the 3D data, the 2D images are used to reconstruct a low quality 3D volume. This reduces the task of 2D/3D registration to 3D/3D registration, which is then solved using a multi-feature mutual information based similarity measure.

The task of generating a 2D perspective image from a 3D image, direct or indirect volume rendering, has been studied extensively in the computer graphics domain [24]. In computer graphics the end result is the image, and the goal is to achieve visual realism. That is, the result of the simulation should be as similar as possible to an actual 2D image as obtained with an imaging apparatus. Luckily, for the purpose of 2D/3D registration a perfect simulation is not necessary, and interestingly, it is not sufficient. In most cases even a perfect simulation will not provide an exact match with the 2D images, as the 3D data does not represent the same physical reality as the 2D images. These two physical situations most often differ by the presence of tools and shift of soft tissue structures. Thus,



(a) (b)

Figure 1.6: Digitally Reconstructed Radiograph of dry femur from CT using a ray casting approach with (a) nearest neighbor, and (b) linear interpolation of the intensity values. The interval between consecutive samples is one millimeter. Note the visible discontinuities corresponding to CT slice locations when using the nearest neighbor interpolator.

the quality of the simulation primarily impacts the choice of the similarity measure and does not directly determine the quality of the registration.

Most of the research to date has focused on registration of various types of X-ray images to volumetric, primarily CT, data. In this context a Digitally Reconstructed Radiograph (DRR) is most often generated by casting rays from the known camera position toward the image plane and summing the volume intensity values found along each ray's path. In practice the ray segment contained within the volume is sampled at regular intervals and the intensity values at these spatial locations are estimated via interpolation. Figure 1.6 illustrates this approach and the effect of the interpolation on the quality of the simulated image.

Other direct volume rendering methods are also applicable but are less common. One such method is the splatting technique which was used for DRR generation in [7]. This is an object order direct volume rendering method where the voxels, or more often a kernel function that is placed at the voxel locations, are projected onto the image plane. A DRR is created by summing the intensity values of the projected voxels at each pixel location.

One of the first papers describing intensity based registration of X-ray and CT was the work of Lemieux et. al [49]. They generate DRR's using the ray casting rendering method with registration performed using the NCC similarity measure and Powell's standard optimization technique. This work identified the main limitation of intensity based methods, the high computational complexity associated with DRR rendering. This is perhaps the primary factor that has delayed the introduction of intensity based 2D/3D registration methods into clinical use.

In [35] digital subtraction angiography images (X-ray angiography) are registered to a magnetic resonance angiography image (MR angiography). To enable the use of DRR image generation via ray casting the MR angiography intensity values are mapped to new values, resulting in a CT-like angiographic image. Intensity mapping is performed using various hard, zero one, and fuzzy, probabilistic, segmentation methods. This is similar to the "fuzzy segmentation" approach described for 3D/3D registration. Interestingly, a decade after the initial work on intensity based registration the high computational complexity of DRR generation via ray casting was still identified as the a limiting factor of this registration approach.

The high computational cost of rendering volumetric images has long been a subject of

research in the computer graphics community. A more recent development on the border of computer graphics and computer vision is the subject of image based rendering. This subject covers all methods that use a set of images to render novel views from camera poses not included in the original set of images. A prime example of this approach is the data structure known as the Lumigraph [27] or light field [50]. This is a 4D parameterized representation of the flow of light through space in a static scene, as computed from a set of images of the scene.

This approach was first adopted for fast DRR generation for X-ray CT registration in [46] under the name Transgraph, with a reported hundred fold improvement in the rendering time. In this context the images used to construct the light field are DRR's rendered prior to the registration process, transferring the bulk of the work to the data structure construction phase. This approach was used later on in [44] and more recently in [79].

While the light field does improve the rendering speed it has two limitations. The size of the data structure is large, due to the amount of information required for rendering novel views, and the novel views are limited to camera poses that are close to those used to generate the images in the data structure construction phase. This second limitation is the more constrictive as it requires that the pose of the 2D X-ray imaging apparatus be roughly known in advance. In clinical practice this is an assumption that can often be accommodated as there are standard viewing poses (e.g. anterior-posterior). An approach that mitigates the camera pose limitation at the cost of a smaller improvement in DRR rendering speed was presented in [78]. Instead of constructing the light field in advance it is constructed progressively during the registration process when the pose is roughly known.

Another approach to reducing the rendering speed is based on the use of the graphics processor unit (GPU). DRR's are generated by loading the CT volume into texture memory, rasterizing parallel polygonal slices that are perpendicular to the viewing direction and blending them using a transparency value [10]. In [46] this approach enabled the generation of 200×200 DRR images from CT volumes of up to $256 \times 256 \times 256$ voxels in approximately $25ms$. In [43] a similar approach enabled the generation of 256×256 DRR images from CT volumes of up to $256 \times 256 \times 216$ voxels in approximately $60ms$.

While the majority of algorithms deal with 2D X-ray images, intensity based registration of 2D video images from bronchoscopy to CT data has also been studied in [66]. The registration framework remains the same, with the only difference being the rendering method. DRR generation is replaced by a standard surface rendering of a polygonal model extracted from the CT data.

1.3.3 Gradient based methods

Gradient based methods cast registration as an optimization task where the objective function is directly dependent on image or surface gradients. This approach has been applied to registration of X-ray images to CT [51, 84], X-ray images to MR [84], and bronchoscopic video images to CT [14].

In [51] and [84] it is shown that there is a direct relationship between the 3D gradient vectors computed from CT intensity values and the 2D gradient vectors computed from an X-ray image of the same physical object. The relationship is derived using the physical model of X-ray propagation through material and assumes a logarithmic sensor response, and is as

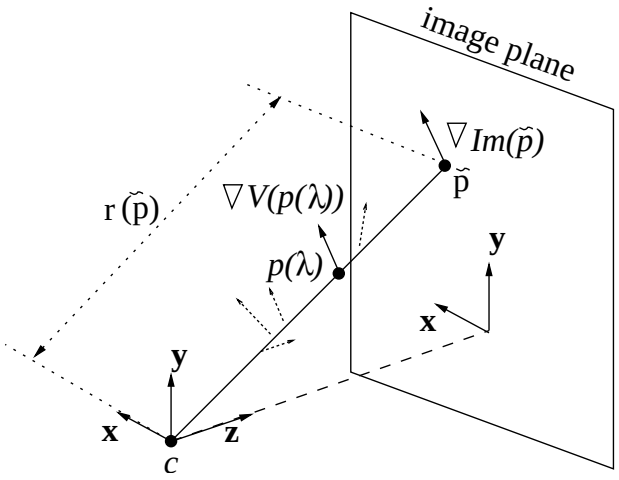


Figure 1.7: Relationship between 3D gradient vectors, $\nabla V(\mathbf{p}(\lambda))$, along the ray connecting camera location c and image point $\widetilde{\mathbf{p}}$, and the corresponding 2D gradient vector $\nabla Im(\widetilde{\mathbf{p}})$.

follows. The 2D image gradient is proportional to the projection onto the image plane of the weighted sum of 3D gradient vectors along the ray:

$$\nabla Im(\widetilde{\mathbf{p}}) \propto \begin{pmatrix} \mathbf{x}^T \\ \mathbf{y}^T \end{pmatrix} r(\widetilde{\mathbf{p}}) \int_{\lambda} \lambda \nabla V(\mathbf{p}(\lambda)) \mathrm{d}\lambda$$

where $\mathbf{x}$ and $\mathbf{y}$ are the camera coordinate system's axis with the viewing direction $\mathbf{z}$, $\widetilde{\mathbf{p}}$ is a point on the image plane, $\mathbf{p}(\lambda)$ is a point on the ray connecting the camera location and $\widetilde{\mathbf{p}}$, and $r(\widetilde{\mathbf{p}})$ is the distance between the camera location and $\widetilde{\mathbf{p}}$. Figure 1.7 illustrates these relationships.

In [51] this observation is used to register X-ray images to CT as follows. For each X-ray image a set of back-projected rays is defined via edge detection. The gradient of the CT volume is computed and the objective function is defined as a function of the magnitude of the volumetric gradient vectors projected onto the image plane along the back-projected rays. The optimal transformation is the one that maximizes the magnitude of these projections, which means that the rays are close to the 3D surface in the volume.

In [84] the outlook is reversed. Surface points and their normals are obtained via edge detection from CT or MR. These points define a set of rays that emanate from the camera location. The gradient of each of the X-ray images is computed and the objective function is defined as a function of the magnitudes of the 3D point normals and back-projections of the 2D gradients at the locations where the rays intersect the image plane, and the angular deviation between these two vectors. The optimal transformation is the one that maximizes the magnitude of the 2D back-projected gradients and the 3D normals while at the same time minimizing their angular deviation.

Interestingly this later approach is utilizing the rate of change in depth along the $\mathbf{x}$ and $\mathbf{y}$ camera axes at a specific surface point, which is known in the shape from shading literature as the surface gradient,$(p, q) = (\partial z/\partial x, \partial z/\partial y)$ [99]. The use of the surface gradient for registration of bronchoscopic video images to CT is described in [14]. In this case a 3D model

is extracted from the CT and its surface gradient from the current estimated viewpoint is computed. An estimate of the surface gradient at each video pixel location is obtained using a shape from shading technique. The objective function is then defined as a function of the angular deviation between the two estimates of the surface gradient. The optimal transformation is the one that minimizes this deviation.

1.4 Registration Evaluation

To evaluate registration we start by defining the characteristics of an ideal registration algorithm and then describe the criteria for evaluating registration as they relate to these characteristics. An ideal registration algorithm is:

- Fast: the result is obtained in real time.
- Accurate: after successful registration the distance between corresponding points in the region of interest is less than $0.1mm$. This is the well known Target Registration Error (TRE) introduced in [61].
- Robust: has a breakdown point of $\frac{N}{2}$, more than half of the data elements must be outliers in order to throw the registration outside of reasonable bounds.
- Automatic: no user interaction is required.
- Reliable: given the expected clinical input the registration always succeeds.

Given these characteristics registration is evaluated using the following criteria:1) execution time, theoretical analysis and performance on in-vivo or in-vitro data sets; 2) accuracy in the region of interest, a theoretical analysis, and performance on in-vivo or in-vitro data sets with a known "gold standard"; 3) breakdown point, a theoretical analysis; 4) automation, a binary classification of algorithms as either semi-automatic or automatic; 5) reliability of the algorithm, if iterative provide proof of global convergence, and an estimate of the convergence range using in-vitro or in-vivo data sets with a known "gold standard" transformation. These are the maximal parameter deviations from the "gold standard" that result in a 95% success rate as judged by the accuracy in the region of interest and the limits imposed by the specific clinical application.

Given these evaluation criteria, it is clear that evaluation necessitates data sets with a known "gold standard". Preferably these are in-vivo, reflecting a clinical situation. In practice, most evaluation studies use in-vitro data, as the common approach to obtaining a "gold standard" is based on implanting fiducials which is not possible in most in-vivo situations. More importantly, these data sets are usually not publicly available, precluding comparative algorithm evaluations.

The Retrospective Image Registration Evaluation project [20] addresses this issue for 3D/3D registration, by providing CT/MR and PET/MR in-vivo data sets with known "gold standard" transformations. This is a well established database that has been used to evaluate a wide variety of algorithms from more than twenty research groups [96, 20].

The Standardized Evaluation Methodology for 2-D3-D Registration database [88] is a more recent effort, providing in-vitro data sets with known "gold standards" for evaluation of

2D/3D registration. The database includes pairs of 2D X-ray images and MR/CT/3DRX (C-arm CT) data sets. As this is a rather recent effort it has only been used to evaluate a small number of algorithms [89].

Returning to our evaluation criteria, we divide them into two categories, criteria that are directly observable and those that are not. The former include the algorithm's execution time, its breakdown point, and its degree of automation. The later include algorithm accuracy and reliability.

An algorithm's execution time is directly observable. In some cases theoretical analysis of the algorithm's computational complexity is also possible, although most evaluation studies only report empirical results. For most clinical procedures real time performance is ideal, but not necessary, rather, a maximal execution time is defined and all algorithms whose execution time is below this threshold can be considered equivalent. It should be emphasized that the execution time criterion includes the total running time. It is often the case that iterative algorithms do not report the running time of their initialization phase. This is somewhat misleading as it is the total running time required for obtaining the registration result that ascertains its applicability.

An algorithm's breakdown point, its robustness to outlying data, is analyzed in a theoretical manner. Interestingly, even a low breakdown point does not imply that an algorithm is not useful, only that a prior outlier rejection phase is required if the algorithm is used. This is clearly the case with the analytic paired-point registration algorithms whose breakdown point is one, yet are in widespread use.

We classify registration algorithms as either automatic or semi-automatic, requiring some form of user interaction. Analytic algorithms are automatic. Iterative algorithms, on the other hand, may require user interaction as part of their initialization. In most cases this fact is downplayed with emphasis placed on the automatic nature of the main body of the algorithm. To gage an algorithm's applicability for a specific procedure the whole framework is classified, and not just the algorithm's main part.

Algorithm accuracy is the single most important evaluation criterion as it directly addresses the goal of registration, alignment of corresponding points. Two aspects of this criterion should be noted, accuracy is spatially variant, and it cannot be directly inferred from the optimal value of the objective function. Very few analytic estimates of accuracy have been developed. Currently these are limited to paired-point registration algorithms. In the seminal work of Fitzpatrick et al. [22] a first order approximation of the root mean square value of the TRE was developed. Later on, Fitzpatrick and West [21] introduced a first order approximation of the distribution of the TRE. More recently, Moghari and Abolmaesumi [64] introduced a second order approximation of the distribution of the TRE based on the Unscented Kalman Filter. For all other registration algorithms accuracy is evaluated empirically. That is, evaluation studies are carried out in one of two ways, with or without a known "gold standard" transformation. When a "gold standard" is available accuracy is readily quantified across the region of interest. This approach provides insight into an algorithm's behavior but it cannot be used to estimate the accuracy that is obtained on data sets not included in the study. When a "gold standard" is not available accuracy is often assessed qualitatively by visual inspection, and quantitatively by manually identifying corresponding points. While this approach does provide accuracy information at specific point locations it depends on the accuracy of manual localization and the information is

limited to the specific spatial locations.

Finally the reliability of iterative algorithms is most often evaluated empirically using data sets that have a "gold standard" transformation. By initializing the registration with parameter values sampled in the region of the "gold standard" values the convergence range can be estimated. As this is an empirical approach its conclusions are only valid if the input data sets represent the distribution of the general data set population. Ensuring this requires the use of a large number of data sets.

1.5 Conclusions

Rigid registration has been studied extensively, with current research interest shifting to non-rigid registration. While most problems in this field have been solved there is still work to be done, primarily in evaluation and validation.

Currently, the only registration algorithms that are truly complete are the paired-point algorithms. They have analytic solutions, their computational complexity is known, and most importantly the expected TRE can be computed. That is, we are guaranteed to find the optimum, we know how long it will take, and once we have the transformation we can estimate the error at our target. This is not the case with all other registration algorithms. The probability to find the optimum subject to certain constraints is high, the computational complexity is known, but we do not have methods for estimating the TRE.

Interestingly, for 2D/3D registration of bronchsopic/endoscopic video images to CT it is still not clear how well the existing algorithms deal with clinical data, as most of the reported evaluation studies have been done on phantoms. It is still also not clear if rigid registration is sufficiently accurate for these types of procedures as it is an approximation of the true physical reality.

Finally, we still need to establish standard *in-vivo* data sets with known "gold standard" transformations for evaluating registration of US data to 3D images, 2D/3D registration of X-ray images to 3D images, and bronchoscopic/endoscopic images to 3D images.

Bibliography

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(5):698–700, 1987.

[2] M. A. Audette, F. P. Ferrie, and T. M. Peters. An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis*, 4(3):201–217, 2000.

[3] R. Balachandran, J. M. Fitzpatrick, and R. F. Labadie. Fiducial registration for tracking systems that employ coordinate reference frames. In *SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Display*, volume 5744, pages 134–145, 2005.

[4] D. C. Barratt, G. P. Penney, C. S. K. Chan, M. Slomczykowski, T. J. Carter, P. J. Edwards, and D. J. Hawkes. Self-calibrating 3D-ultrasound-based bone registration for minimally invasive orthopedic surgery. *IEEE Trans. Med. Imag.*, 25(3):312–323, 2006.

[5] P. J. Besl and N. D. McKay. A method for registration of 3D shapes. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(2):239–255, 1992.

[6] J. Bijhold. Three-dimensional verification of patient placement during radiotherapy using portal images. *Med. Phys.*, 20(2):347–356, 1993.

[7] W. Birkfellner et al. Wobbled splatting - a fast perspective volume rendering method for simulation of X-ray images from CT. *Phys. Med. Biol.*, 50(9):N73–N84, 2005.

[8] D. Burschka, M. Li, M. Ishii, R. H. Taylor, and G. D. Hager. Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. *Medical Image Analysis*, 9(5):413–426, 2005.

[9] T. M. Buzug, J. Weese, C. Fassnacht, and C. Lorenz. Image registration: convex weighting functions for histogram-based similarity measures. In *CVRMed-MRCAS*, pages 203–212, 1997.

[10] B. Cabral, N. Cam, and J. Foran. Accelerated volume rendering and tomographic reconstruction using texture mapping hardware. In *VVS '94: Proceedings of the 1994 symposium on Volume visualization*, pages 91–98, 1994.

[11] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[12] D. Chetverikov, D. Stepanov, and P. Krsek. Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing*, 23(3):299–309, 2005.

[13] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3D multi-modality medical image registration using feature space clustering. In *Computer Vision, Virtual Reality and Robotics in Medicine*, pages 195–204, 1995.

[14] F. Deligianni, A. Chung, and G.-Z. Yang. Patient-specific bronchoscope simulation with pq-space-based 2D/3D registration. *Computer Aided Surgery*, 9(5/6):1–12, 2004.

[15] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2'nd edition, 2001.

[16] D. W. Eggert, A. Lorusso, and R. B. Fisher. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Mach. Vis. Appl.*, 9(5/6):272–290, 1997.

[17] R. S. J. Estépar, A. Brun, and C.-F. Westin. Robust generalized total least squares iterative closest point registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2004.

[18] O. D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-D objects. *Int. J. Rob. Res.*, 5(3):27–52, 1986.

[19] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. *Image and Vision Computing*, 21(13–14):1145–1153, 2003.

[20] J. M. Fitzpatrick. Retrospective image registration evaluation project. http://www.insight-journal.org/rire/index.html, accessed February 13, 2007.

[21] J. M. Fitzpatrick and J. B. West. The distribution of target registration error in rigid-body, point-based registration. *IEEE Trans. Med. Imag.*, 20(9):917–927, 2001.

[22] J. M. Fitzpatrick, J. B. West, and C. R. Maurer, Jr. Predicting error in rigid-body, point-based registration. *IEEE Trans. Med. Imag.*, 17(5):694–702, 1998.

[23] R. Fletcher. *Practical methods of optimization; (2nd ed.)*. Wiley-Interscience, 1987.

[24] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer graphics (2nd ed. in C): principles and practice*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1996.

[25] J. Gong, R. Bächler, M. Sati, and L. P. Nolte. Restricted surface matching, a new approach to registration in computer assisted surgery. In *Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-assisted Surgery*, pages 597–605, 1997.

[26] R. H. Gong, J. Stewart, and P. Abolmaesumi. A new method for CT to fluoroscopes registration based on unscented kalman filter. In *Medical Image Computing and Computer-Assisted Intervention*, pages 891–898, 2006.

[27] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *SIGGRAPH: Computer graphics and interactive techniques*, pages 43–54, 1996.

[28] A. A. Goshtasby. *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications*. Wiley, 2005.

[29] M. Greenspan and M. Yurick. Approximate K-D tree search for efficient ICP. In *International Conference on 3-D Digital Imaging and Modeling*, pages 442–448, 2003.

[30] W. E. L. Grimson and T. Lozano-Pérez. Localizing overlapping parts by searching the interpretation tree. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(4):469–482, 1987.

[31] A. Guéziec, P. Kazanzides, B. Williamson, and R. H. Taylor. Anatomy based registration of CT-scan and intraoperative X-ray images for guiding a surgical robot. *IEEE Trans. Med. Imag.*, 17(5):715–728, 1998.

[32] J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes, editors. *Medical Image Registration*. CRC Press, 2001.

[33] A. Hamadeh, S. Lavallée, and P. Cinquin. Automated 3-dimensional computed tomographic and fluoroscopic image registration. *Computer Aided Surgery*, 3(1):11–19, 1998.

[34] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46(3):R1–R45, 2001.

[35] J. H. Hipwell, G. P. Penney, R. A. McLaughlin, K. Rohde, P. Summers, T. C. Cox, J. V. Byrne, A. Noble, and D. J. Hawkes. Intensity-based 2D-3D registration of cerebral angiograms. *IEEE Trans. Med. Imag.*, 22(11):1417–1426, 2003.

[36] Y. C. Ho and D. L. Pepyne. Simple explanation of the no-free-lunch theorem and its implications. *J. Optimization Theory and Applications*, 115(3):549–570, 2002.

[37] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4(4):629–642, April 1987.

[38] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A*, 5(7):1127–1135, 1988.

[39] J. B. A. M. Josien P. W. Pluim and M. A. Viergever. Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding*, 77(2):211–223, 2000.

[40] T. Jost and H. Hügli. A multi-resolution ICP with heuristic closest point search for fast and robust 3D registration of range images. In *International Conference on 3-D Digital Imaging and Modeling*, pages 427–433, 2003.

[41] G. Jumarie. A new information theoretic approach to the entropy of non-random discrete maps relation to fractal dimension and temperature of curves. *Chaos, Solitons and Fractals*, 8(6):953–970, 1997.

[42] S. Kaneko, T. Kondo, and A. Miyamoto. Robust matching of 3D contours using iterative closest point algorithm improved by M-estimation. *Pattern Recognition*, 36(9):2041–2047, 2003.

[43] A. Khamene, P. Bloch, W. Wein, M. Svatos, and F. Saur. Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy. *Medical Image Analysis*, 10(1):96–112, 2006.

[44] D. Knaan and L. Joskowicz. Effective intensity-based 2D/3D rigid registration between fluoroscopic X-ray and CT. In *Medical Image Computing and Computer-Assisted Intervention*, pages 351–358, 2003.

[45] C. Langis, M. Greenspan, and G. Godin. The parallel iterative closest point algorithm. In *International Conference on 3-D Digital Imaging and Modeling*, pages 195–204, 2001.

[46] D. A. LaRose. *Iterative X-ray/CT Registration Using Accelerated Volume Rendering.* PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2001.

[47] S. Lavallée and R. Szeliski. Recovering the position and orientation of free-form objects from image contours using 3D distance maps. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(4):378–390, 1995.

[48] T. M. Lehmann, C. Gönner, and K. Spitzer. Survey: Interpolation methods in medical image processing. *IEEE Trans. Med. Imag.*, 18(11):1049–1075, 1999.

[49] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and G. T. Thomas. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Med. Phys.*, 21(11):1749–1760, 1994.

[50] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH: Computer graphics and interactive techniques*, pages 31–42, 1996.

[51] H. Livyatan, Z. Yaniv, and L. Joskowicz. Gradient-based 2D/3D rigid registration of fluoroscopic X-ray to CT. *IEEE Trans. Med. Imag.*, 22(11):1395–1406, 2003.

[52] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.

[53] J. P. Luck, C. Q. Little, and W. Hoff. Registration of range data using a hybrid simulated annealing and iterative closest point algorithm. In *International Conference on Robotics and Automation*, pages 3739–3744, 2000.

[54] B. Ma and R. Ellis. Analytic expressions for fiducial and surface target registration error. In *Medical Image Computing and Computer-Assisted Intervention*, pages 637–644, 2006.

[55] B. Ma and R. E. Ellis. Robust registration for computer-integrated orthopedic surgery: Laboratory validation and clinical experience. *Medical Image Analysis*, 7(3):237–250, 2003.

[56] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16(2):187–198, 1997.

[57] J. B. A. Maintz, P. A. van den Elsen, and M. A. Viergever. Comparison of edge-based and ridge-based registration of CT and MR brain images. *Medical Image Analysis*, 1(2):151–161, 1996.

[58] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–37, 1998.

[59] T. Masuda and N. Yokoya. A robust method for registration and segmentation of multiple range images. *Computer Vision and Image Understanding*, 61(3):295–307, 1995.

[60] B. Matei and P. Meer. Optimal rigid motion estimation and performance evaluation with bootstrap. In *CVPR*, pages 339–347, 1999.

[61] C. R. Maurer, Jr., J. M. Fitzpatrick, M. Y. Wang, R. L. Galloway, Jr., R. J. Maciunas, and G. S. Allen. Registration of head volume images using implantable fiducial markers. *IEEE Trans. Med. Imag.*, 16(4):447–462, 1997.

[62] C. R. Maurer, Jr., R. J. Maciunas, and J. M. Fitzpatrick. Registration of head CT images to physical space using a weighted combination of points and surfaces. *IEEE Trans. Med. Imag.*, 17(5):753–761, 1998.

[63] E. H. Meijering, W. J. Niessen, and M. A. Viergever. Quantitative evaluation of convolution-based methods for medical image interpolation. *Medical Image Analysis*, 5(2):111–126, 2001.

[64] M. Moghari and P. Abolmaesumi. A high-order solution for the distribution of target registration error in rigid-body point-based registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 603–611, 2006.

[65] M. H. Moghari and P. Abolmaesumi. A novel incremental technique for ultrasound to CT bone surface registration using unscented kalman filtering. In *Medical Image Computing and Computer-Assisted Intervention*, pages 197–204, 2005.

[66] K. Mori et al. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Medical Image Analysis*, 6(3):321–336, 2002.

[67] M. J. Murphy. An automatic six-degree-of-freedom registration algorithm for image-guided frameless stereotaxic radiosurgery. *Med. Phys.*, 24(6):857–866, 1997.

[68] N. Ohta and K. Kanatani. Optimal estimation of three-dimensional rotation and reliability evaluation. In *Computer Vision ECCV'98*, volume 1406/1998 of *LNCS*, pages 175–187, 1998.

[69] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chen. Accurate three-dimensional registration of CT, PET and MRI images of the brain. *Journal of Computer Assisted Tomography*, 13(1):20–26, 1989.

[70] X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3-D registration methods based on points and frames. *International Journal of Computer Vision*, 25(3):pp. 203–229, 1997.

[71] G. Penney, J. Blackall, M. Hamady, T. Sabharwal, A. Adam, and D. Hawkes. Registration of freehand 3D ultrasound and magnetic resonance liver images. *Medical Image Analysis*, 8(1):81–91, 2004.

[72] G. P. Penney, D. C. Barratt, C. S. K. Chan, M. Slomczykowski, T. J. Carter, P. J. Edwards, and D. J. Hawkes. Cadaver validation of intensity-based ultrasound to CT registration. *Medical Image Analysis*, 10(3):385–395, 2006.

[73] G. P. Penney, P. J. Edwards, A. P. King, J. M. Blackall, P. G. Batchelor, and D. J. Hawkes. A stochastic iterative closest point algorithm (stochastICP). In *Medical Image Computing and Computer-Assisted Intervention*, pages 762–769, 2001.

[74] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: A survey. *IEEE Trans. Med. Imag.*, 22(8):986–1004, 2003.

[75] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.

[76] M. Robinson, D. G. Eckhoff, K. D. Reinig, M. M. Bagur, and J. M. Bach. Variability of landmark identification in total knee arthroplasty. *Clinical Orthopaedics and Related Research*, 442:57–62, 2006.

[77] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 1115–1124, 1998.

[78] T. Rohlfing, D. B. Russakoff, J. Denzler, K. Mori, and C. R. Maurer, Jr. Progressive attenuation fields: Fast 2D-3D registration without precomputation. *Med. Phys.*, 32(9):2870–2880, 2005.

[79] D. B. Russakoff, T. Rohlfing, K. Mori, D. Rueckert, J. R. Adler, Jr., and C. R. Maurer, Jr. Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2D-3D image registration. *IEEE Trans. Med. Imag.*, 24(11):1441–1454, 2005.

[80] P. H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31:1–10, 1966.

[81] D. Simon. *Fast and Accurate Shape-Based Registration*. PhD thesis, Carnegie Mellon University, 1996.

[82] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[83] D. Tomaževič, B. Likar, and F. Pernuš. 3-D/2-D registration by integrating 2-D information in 3-D. *IEEE Trans. Med. Imag.*, 25(1):17–27, 2006.

[84] D. Tomaževič, B. Likar, T. Slivnik, and F. Pernuš. 3-D/2-D registration of CT and MR to X-ray images. *IEEE Trans. Med. Imag.*, 22(11):1407–1416, 2003.

[85] E. Trucco, A. Fusiello, and V. Roberto. Robust motion and correspondence of noisy 3-D point sets with missing data. *Pattern Recognition Letters*, 20(9):889–898, 1999.

[86] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *SIGGRAPH: Computer graphics and interactive techniques*, pages 311–318, 1994.

[87] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(4):376–380, 1991.

[88] E. van de Kraats. Standardized evaluation methodology for 2D-3D registration. http://www.isi.uu.nl/Research/Databases/GS/, accessed February 13, 2007.

[89] E. B. van de Kraats, G. P. Penney, D. Tomaževič, T. van Walsum, and W. J. Niessen. Standardized evaluation methodology for 2-D3-D registration. *IEEE Trans. Med. Imag.*, 24(9):1177–1189, 2005.

[90] P. Viola and W. M. W. III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[91] P. A. Viola and W. M. W. III. Alignment by maximization of mutual information. In *IEEE International Conference on Computer Vision*, pages 16–23, 1995.

[92] D. Škerl, B. Likar, and F. Pernuš. A protocol for evaluation of similarity measures for rigid registration. *IEEE Trans. Med. Imag.*, 25(6):779–791, 2006.

[93] M. W. Walker, L. Shao, and R. A. Volz. Estimating 3-D location parameters using dual number quaternions. *CVGIP: Image Understanding*, 54(3):358–367, 1991.

[94] M. Y. Wang, C. R. Maurer, Jr., J. M. Fitzpatrick, and R. J. Maciunas. An automatic technique for finding and localizing externally attached markers in CT and MR volume images of the head. *IEEE Trans. Bio-Med. Eng.*, 43(6):627–637, 1996.

[95] J. Weese, P. Rösch, T. Netsch, T. Blaffert, and M. Quist. Gray-value based registration of CT and MR images by maximization of local correlation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 656–663, 1999.

[96] J. West et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 4(4):554–568, 1997.

[97] S. Wörz and K. Rohr. Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models. *Medical Image Analysis*, 10(1):41–58, 2006.

[98] P. Wunsch and G. Hirzinger. Registration of CAD-models to images by iterative inverse perspective matching. In *International Conference on Pattern Recognition*, pages 78–83, 1996.

[99] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Trans. Pattern Anal. Machine Intell.*, 21(8):690–706, 1999.

[100] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994.