Review

Rigid 2D–3D Registration in Medical Imaging: A Survey

Lin Li *, Chaochao Zhou, Benjamin Albert, Junlin Guo, Junchao Zhu

Posted Date: 7 May 2026

doi: 10.20944/preprints202605.0431.v1

Keywords: 2D–3D image registration; medical image registration; image-guided interventions; digitally reconstructed radiograph (DRR); similarity metrics; optimization; deep learning; X-ray imaging

 doi:10.20944/preprints202605.0431.v1

Review

Rigid 2D–3D Registration in Medical Imaging: A Survey

Lin Li [1,*], Chaochao Zhou [2], Benjamin Albert [3], Junlin Guo [4] and Junchao Zhu [5]

[1] Department of Mathematics, Tulane University
[2] Department of Computer Science, University of Illinois Urbana-Champaign
[3] EOS, EOS Imaging Inc.
[4] Department of Electrical and Computer Engineering, Vanderbilt University
[5] Department of Computer Science, Vanderbilt University
* Correspondence: lli15@tulane.edu

Abstract

Rigid 2D–3D image registration plays a critical role in modern image-guided interventions by enabling the alignment of intraoperative X-ray images with preoperative volumetric data such as computed tomography (CT). Accurate 2D–3D registration allows clinicians to localize anatomical structures in three-dimensional space while relying on fast and low-dose intraoperative imaging, which is essential for applications including orthopedic surgery, spine navigation, radiation therapy, and interventional radiology. Despite significant progress over the past two decades, achieving robust and accurate registration remains challenging due to factors such as limited imaging viewpoints, occlusions, imaging noise, and the large search space of rigid transformations. This paper provides a comprehensive survey of rigid 2D–3D registration methods with a particular focus on X-ray–to–CT alignment. We first introduce the mathematical formulation of the registration problem and present a taxonomy of existing approaches, including intensity-based optimization methods, feature-based registration techniques, and recent learning-based frameworks. We then review the key technical components underlying modern registration systems, including similarity metrics, digitally reconstructed radiograph (DRR) generation techniques, and optimization strategies for pose estimation in $SE(3)$. In addition, we summarize commonly used datasets and evaluation protocols, discussing widely adopted metrics such as target registration error (TRE), pose error, and reprojection error. The survey also highlights representative clinical applications and analyzes the practical challenges that remain in real-world deployment, including robustness to imaging artifacts, variations in imaging dose, and real-time computational constraints. Finally, we discuss emerging research directions, such as differentiable rendering, deep learning–based pose estimation, and multi-view registration frameworks, which are expected to further improve the accuracy, robustness, and clinical applicability of 2D–3D registration methods.

Keywords: 2D–3D image registration; medical image registration; image-guided interventions; digitally reconstructed radiograph (DRR); similarity metrics; optimization; deep learning; X-ray imaging

1. Introduction

Two-dimensional (2D) and three-dimensional (3D) image registration is a fundamental problem in medical image computing, aiming to determine the spatial relationship between images acquired from different modalities or imaging geometries [1,2]. In particular, rigid 2D–3D registration seeks to align a preoperative volumetric image, such as computed tomography (CT) or magnetic resonance imaging (MRI), with one or more intraoperative projection images acquired using X-ray fluoroscopy or radiography [3]. Accurate alignment between these modalities enables clinicians to leverage rich anatomical information from 3D imaging during procedures that rely on real-time 2D imaging.

 doi:10.20944/preprints202605.0431.v1

Rigid 2D–3D registration plays a crucial role in a wide range of clinical applications. In orthopedic and trauma surgery, preoperative CT scans provide detailed anatomical structures that guide surgical planning, while intraoperative fluoroscopy is used for real-time instrument guidance [4,5]. Similarly, in spinal surgery and joint replacement procedures, accurate registration allows the projection of preoperative anatomical models onto intraoperative X-ray images, facilitating precise implant placement and reducing the risk of surgical errors [6]. Beyond orthopedics, 2D–3D registration is also used in radiation therapy for patient positioning [7], in interventional radiology for navigation during minimally invasive procedures [8], and in image-guided robotic surgery [3]. In these scenarios, accurate and robust registration is essential for ensuring the safe and effective use of image-guided technologies.

Despite its clinical importance, rigid 2D–3D registration remains a challenging computational problem. Unlike traditional 3D–3D registration, where both images share the same dimensionality, 2D–3D registration must reconcile the fundamental differences between projection images and volumetric data. A 2D X-ray image represents a line integral of X-ray attenuation along each projection ray, resulting in a loss of depth information and significant structural overlap [3,5]. Furthermore, fluoroscopic images often exhibit low contrast, substantial noise, and imaging artifacts such as scatter and metal occlusions from surgical instruments or implants. These factors make the registration problem highly ill-posed and prone to local minima in the optimization landscape.

Historically, most 2D–3D registration methods have relied on intensity-based optimization frameworks. In these approaches, a digitally reconstructed radiograph (DRR) is generated by simulating the X-ray projection of a volumetric image under a candidate rigid transformation [9]. The similarity between the synthesized DRR and the observed X-ray image is then evaluated using a similarity metric, such as normalized cross-correlation or mutual information [10,11]. The optimal transformation is obtained by solving a nonlinear optimization problem over the space of rigid transformations. These methods have demonstrated high accuracy in many applications and can achieve sub-millimeter registration performance under favorable imaging conditions [5,8].

However, 2D–3D registration is fundamentally challenging due to the loss of depth information inherent in projection imaging. As illustrated in Figure 1, distinct 3D structures can map to the same 2D detector location, making the problem intrinsically ambiguous, especially when only a limited number of views are available. Classical intensity-based approaches often suffer from significant computational cost due to repeated DRR generation and similarity evaluation. Moreover, the similarity landscape can be highly non-convex, making optimization sensitive to initialization and prone to convergence failures. To address these limitations, alternative approaches have been proposed that rely on geometric features extracted from both 2D and 3D images. Feature-based methods typically detect salient anatomical structures such as edges, contours, or landmarks, and estimate the transformation by matching corresponding features between modalities [6,10]. While these approaches can reduce computational complexity and improve robustness to intensity variations, their performance strongly depends on the reliability of feature extraction in noisy projection images.

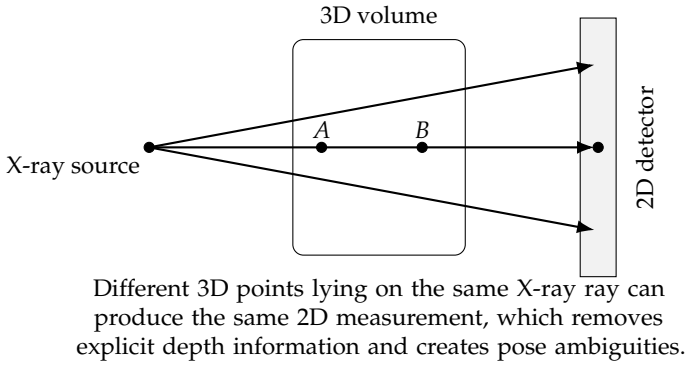


Figure 1. Illustration of the loss of depth information in 2D projection imaging. Distinct 3D points or structures can collapse onto the same detector location, which makes 2D–3D registration intrinsically ambiguous when only a limited number of views are available.

 doi:10.20944/preprints202605.0431.v1

More recently, the rapid development of deep learning has introduced new paradigms for 2D–3D registration. Learning-based approaches leverage convolutional neural networks to estimate transformation parameters, predict correspondences, or learn similarity metrics directly from data [12, 13]. In parallel, advances in differentiable rendering have enabled end-to-end optimization frameworks in which gradients can be propagated through the DRR generation process [14,15]. These methods combine the physical modeling of X-ray image formation with the optimization capabilities of modern machine learning frameworks, leading to improved robustness and computational efficiency.

Although substantial progress has been made in recent years, the literature on rigid 2D–3D registration remains fragmented across multiple methodological paradigms and application domains [3]. Existing studies differ significantly in terms of problem formulation, similarity metrics, DRR generation techniques, optimization strategies, and evaluation protocols. As a result, it can be difficult for researchers and practitioners to obtain a comprehensive understanding of the field and to compare different approaches in a unified framework.

The goal of this survey is to provide a comprehensive and structured overview of rigid 2D–3D registration methods in medical imaging. We first present a unified mathematical formulation of the registration problem and describe the underlying imaging geometry. We then organize existing approaches into several major methodological categories and analyze their key characteristics. In addition, we review commonly used similarity metrics, DRR generation techniques, and optimization strategies that form the core components of many registration algorithms. Finally, we summarize publicly available datasets and evaluation methodologies, discuss current challenges, and highlight promising directions for future research.

2. Rigid 2D–3D Registration Problem Formulation

Rigid 2D–3D registration aims to estimate the spatial transformation that aligns a three-dimensional (3D) volumetric image with one or more two-dimensional (2D) projection images acquired using an X-ray imaging system [3,5,16]. In this section, we present a unified mathematical formulation of the problem by describing the imaging geometry, the rigid transformation model, the physical principles underlying X-ray projection, and the resulting optimization formulation.

Figure 2 illustrates the typical pipeline of rigid 2D–3D registration [3,17]. Given a preoperative 3D volume, a candidate rigid transformation is applied to define the pose of the volume relative to the imaging system. A digitally reconstructed radiograph (DRR) is then generated by simulating the X-ray projection through the transformed volume. The generated DRR is compared with the observed 2D X-ray image using a similarity metric, and an optimization algorithm iteratively updates the transformation parameters until the similarity between the two images is maximized (or the dissimilarity is minimized) [3,5].

 doi:10.20944/preprints202605.0431.v1

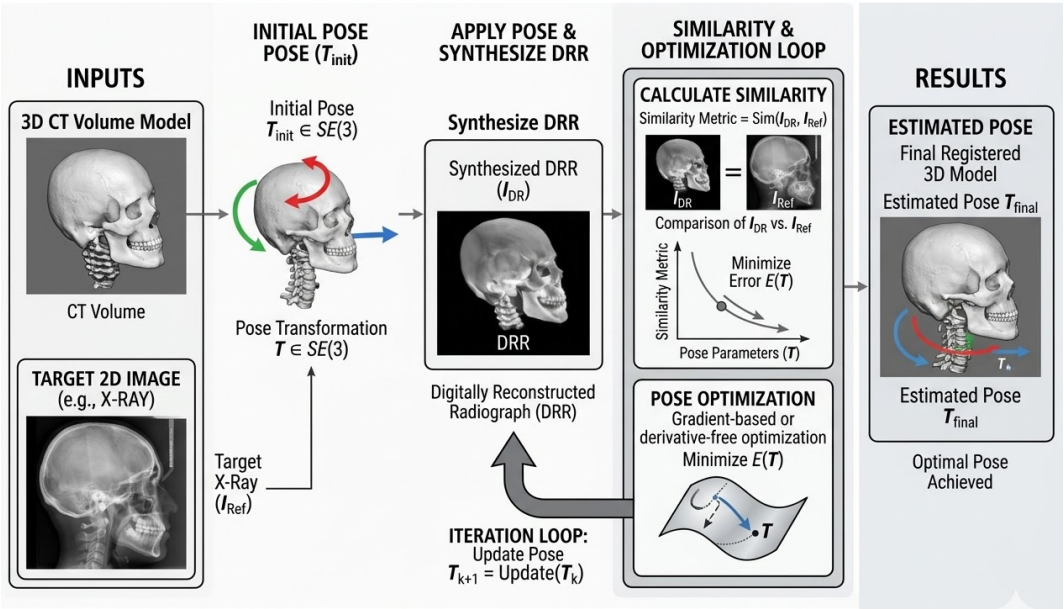


Figure 2. Typical pipeline of rigid 2D–3D registration. A 3D volume is transformed under a candidate pose, a digitally reconstructed radiograph (DRR) is generated, and a similarity metric between the DRR and the observed X-ray image is optimized to estimate the correct transformation.

2.1. X-Ray Projection Geometry

Most medical 2D–3D registration problems arise in imaging systems that follow a perspective projection model, such as C-arm fluoroscopy or cone-beam radiography [18,19]. In these systems, X-rays originate from a point source and propagate through the patient before being detected on a planar detector.

Let $X = (x, y, z)^\top \in \mathbb{R}^3$ denote a point in the 3D volume coordinate system. Under a perspective projection model, the corresponding point (u, v) on the detector plane can be expressed as

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X \\ 1 \end{bmatrix}, \tag{1}$$

where s is a scaling factor and $\mathbf{P} \in \mathbb{R}^{3\times4}$ is the projection matrix [18]. The projection matrix encodes both the intrinsic parameters[20] of the imaging system (e.g., focal length and detector spacing) and the extrinsic parameters describing the relative pose between the X-ray source and the volume coordinate system.

In practice, the projection matrix can be decomposed as

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}], \tag{2}$$

where $\mathbf{K}$ denotes the intrinsic calibration matrix of the detector, and $(\mathbf{R}, \mathbf{t})$ represent the rotation and translation that define the pose of the imaging system relative to the 3D volume [18,19].

Accurate knowledge of the projection geometry is essential for 2D–3D registration, as it determines how a 3D point in the volume maps onto the detector plane [3,16].

2.2. Rigid Transformation Model

The goal of rigid 2D–3D registration is to estimate the rigid transformation that aligns the coordinate system of the 3D volume with the coordinate system of the projection image.

Rigid transformations belong to the special Euclidean group

 doi:10.20944/preprints202605.0431.v1

$$SE(3) = \{(R, t) \mid R \in SO(3),\, t \in \mathbb{R}^3\}, \tag{3}$$

where $SO(3)$ denotes the group of 3D rotation matrices [21,22]. A rigid transformation can be represented in homogeneous coordinates as

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}. \tag{4}$$

Applying the transformation T to a point X yields the transformed point

$$X' = RX + t. \tag{5}$$

In practice, the transformation can be parameterized using six degrees of freedom (DoF):

- three translations (t_x, t_y, t_z)
- three rotations $(\theta_x, \theta_y, \theta_z)$

Various rotation parameterizations have been used in the literature, including Euler angles, axis-angle representations, and Lie algebra representations based on $\mathfrak{se}(3)$ [21,22]. The choice of parameterization can influence the numerical stability and efficiency of the optimization process [3,23].

2.3. Digitally Reconstructed Radiographs

A central component of many 2D–3D registration methods is the generation of digitally reconstructed radiographs (DRRs), which simulate X-ray projections from a 3D volumetric image [3,24].

Let $\mu(x)$ denote the linear attenuation coefficient at position x in the volume. According to the Beer–Lambert law, the intensity of an X-ray beam after passing through the object can be expressed as

$$I = I_0 \exp\left(-\int_{\mathcal{R}} \mu(x)\, dx\right), \tag{6}$$

where I_0 is the incident X-ray intensity and $\mathcal{R}$ denotes the path of the ray through the volume [19,25,26].

In practice, DRR generation typically computes the line integral of the attenuation coefficients along each projection ray:

$$I_{\text{DRR}}(u, v; T) = \int_{\mathcal{R}(u,v;T)} \mu(x)\, dx, \tag{7}$$

where $\mathcal{R}(u, v; T)$ denotes the ray corresponding to detector pixel (u, v) under transformation T.

Several algorithms have been proposed for efficient DRR generation, including ray-casting methods, the Siddon algorithm for computing ray-voxel intersections [27], and GPU-accelerated rendering techniques [24,28]. More recently, differentiable rendering methods have been introduced to enable gradient-based optimization through the DRR generation process [29,30].

2.4. Registration Objective

Given an observed X-ray image I_{obs} and a volumetric image V, the objective of rigid 2D–3D registration is to estimate the transformation T that best aligns the projection of the volume with the observed image [3,5].

This problem is commonly formulated as an optimization problem:

$$\hat{T} = \arg\min_{T \in SE(3)} \mathcal{S}(I_{obs}, I_{\text{DRR}}(T)), \tag{8}$$

where $\mathcal{S}(\cdot)$ denotes a similarity metric that measures the discrepancy between the observed X-ray image and the synthesized DRR [31,32].

 doi:10.20944/preprints202605.0431.v1

Depending on the chosen similarity metric, the objective function may be highly non-convex and contain multiple local minima. Consequently, the success of the registration algorithm often depends on both the choice of similarity metric and the optimization strategy used to search the transformation space [3,16].

In cases where multiple projection images are available, the objective can be extended to incorporate information from several views:

$$\hat{T} = \arg\min_{T \in SE(3)} \sum_{i=1}^{N} \mathcal{S}\left(I_i, I_{\mathrm{DRR}}^{(i)}(T)\right), \tag{9}$$

where N denotes the number of projection images. Multi-view formulations generally improve observability and reduce pose ambiguity relative to single-view registration [3,33].

This formulation provides a general framework that encompasses most existing rigid 2D–3D registration methods. Different algorithms can be understood as variations of this framework, distinguished primarily by their choice of similarity metric, DRR generation technique, and optimization strategy [3,16].

3. Taxonomy of Rigid 2D–3D Registration Methods

Over the past two decades, a wide variety of algorithms have been proposed for rigid 2D–3D registration in medical imaging [1–3]. Despite their apparent diversity, most existing approaches can be understood within a common conceptual framework derived from the optimization formulation introduced in the previous section. In particular, different methods primarily differ in how they represent image information, how correspondences between modalities are established, and how the transformation parameters are estimated [5,8,9].

To provide a structured overview of the literature, we categorize rigid 2D–3D registration methods into four major methodological families: (1) intensity-based DRR optimization methods, (2) feature-based methods, (3) learning-based methods, and (4) hybrid pipelines that combine multiple paradigms. Although the boundaries between these categories are not always strict, this taxonomy captures the dominant design principles underlying most published approaches [3,14]. Figure 3 provides an overview of this classification.

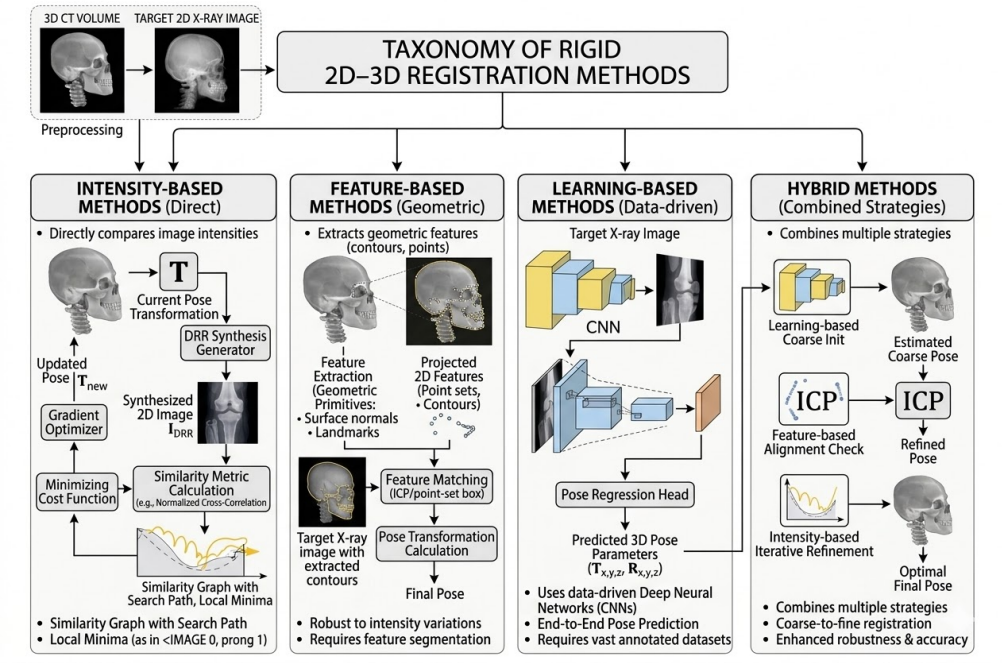


Figure 3. Taxonomy of rigid 2D–3D registration methods. Existing approaches can generally be categorized into intensity-based, feature-based, learning-based, and hybrid methods.

 doi:10.20944/preprints202605.0431.v1

3.1. Intensity-Based DRR Optimization Methods

Intensity-based methods represent the classical and historically dominant paradigm for rigid 2D–3D registration [3,5,9,26]. These approaches directly compare the intensity patterns of the observed projection image with a simulated projection generated from the 3D volume. Typically, a digitally reconstructed radiograph (DRR) is generated from the volume under a candidate transformation, and a similarity metric is evaluated between the synthesized DRR and the observed X-ray image [6,8].

Within the general formulation introduced in Section 2, intensity-based methods solve

$$\hat{T} = \arg\min_{T \in SE(3)} \mathcal{S}(I_{obs}, I_{\mathrm{DRR}}(T)), \tag{10}$$

where $\mathcal{S}$ denotes a similarity metric that quantifies the agreement between the observed image and the simulated projection.

A wide range of similarity metrics have been proposed for this purpose, including intensity-based measures such as sum of squared differences (SSD) and normalized cross-correlation (NCC), information-theoretic measures such as mutual information (MI) [11,34], as well as gradient-based metrics and structural similarity measures [5,8]. Several studies have also investigated robust similarity metrics designed specifically for X-ray projection images [3,5].

The optimization of the transformation parameters is typically performed using nonlinear optimization techniques in the six-dimensional transformation space [9]. Early approaches frequently relied on gradient-free optimization methods, such as Powell's method or evolutionary strategies, due to the difficulty of computing analytic gradients of the similarity function [5]. More recent methods leverage differentiable rendering techniques that enable gradient-based optimization through the DRR generation process [14,15].

Intensity-based approaches have demonstrated high registration accuracy in many clinical applications, particularly when the imaging geometry is well calibrated and the initial transformation estimate is reasonably close to the true solution [3,5,8]. However, these methods often suffer from high computational cost due to repeated DRR generation and similarity evaluation. Moreover, the resulting similarity landscape is frequently highly non-convex, making the optimization sensitive to initialization and prone to local minima [3].

While DRR-based methods dominate the literature, alternative formulations have been proposed that avoid explicit forward projection by exploiting tomographic consistency conditions. These approaches are discussed next.

3.2. Radon and Consistency-Based Methods

In addition to intensity- and feature-based approaches, a distinct class of 2D–3D registration methods is based on tomographic consistency conditions and Radon space representations. These methods exploit the fact that both projection data and volumetric images can be linked through integral geometry, avoiding explicit comparison in the image domain.

A key theoretical foundation in this category is *Grangeat's relation*, which connects cone-beam projections to the derivative of the three-dimensional Radon transform of the object [35]. Specifically, each X-ray projection encodes information about plane integrals of the underlying attenuation field, enabling a mapping from projection space to a common intermediate domain defined over planes in 3D space.

Building on this principle, recent approaches reformulate 2D–3D registration by transforming both the 2D projection and the 3D volume into comparable representations in Radon space [36]. The projection is mapped via a differentiated 2D Radon transform, while the volume is represented through the derivative of its 3D Radon transform. Under ideal conditions, these intermediate representations are theoretically equivalent, and registration can be achieved by minimizing their discrepancy.

Compared to classical DRR-based methods, this formulation avoids repeated forward projection during optimization. Instead, the computationally expensive transformation of the 3D volume can be precomputed once, and subsequent evaluation only requires resampling of the intermediate represen-

 doi:10.20944/preprints202605.0431.v1

tation according to the candidate transformation. This significantly reduces computational cost and enables faster registration while maintaining competitive accuracy.

However, consistency-based methods rely on assumptions such as complete projection data and accurate imaging geometry. In practice, factors such as truncation, noise, or deviations from ideal acquisition conditions can violate these assumptions and degrade performance, requiring additional robustness strategies [36].

Overall, Radon and consistency-based approaches provide an alternative perspective on 2D–3D registration by shifting the problem from image-domain similarity to geometric consistency in transform space, offering advantages in computational efficiency and theoretical grounding.

3.3. Feature-Based Methods

Feature-based registration methods attempt to overcome some of the limitations of intensity-based approaches by focusing on geometric structures that are shared between the 3D volume and the 2D projection image [3,6,10,37]. Instead of comparing raw image intensities, these methods extract salient features from both modalities and estimate the transformation by establishing correspondences between them.

Commonly used features include edges, contours, anatomical landmarks, and skeletal structures [6,8]. In projection images, edges and contours often correspond to boundaries of anatomical structures, while in volumetric images these structures can be extracted using gradient operators or segmentation techniques.

Feature-based registration methods generally follow a two-stage pipeline consisting of feature extraction and transformation estimation. In this framework, the transformation is typically estimated by minimizing a geometric distance between projected 3D features and detected 2D features [6,10].

Feature-based approaches often exhibit improved robustness to intensity variations and imaging artifacts because geometric structures are typically less sensitive to noise and contrast differences than raw intensity values [3]. However, their performance strongly depends on the reliability of feature extraction in both modalities, which can be challenging in fluoroscopic images due to low contrast and overlapping anatomical structures [8].

3.4. Learning-Based Methods

The rapid development of deep learning has led to a growing body of research on learning-based approaches to 2D–3D registration [12–14,38–40]. These methods leverage neural networks to learn representations or transformations directly from data, reducing the reliance on hand-crafted similarity metrics or explicit feature extraction.

Learning-based methods can be broadly divided into several subcategories depending on how the transformation is estimated. One common strategy is direct pose regression, in which a neural network predicts the rigid transformation parameters from the input images [12].

Another class of approaches focuses on learning similarity metrics or feature embeddings that facilitate registration [13]. In these methods, the network learns a feature representation in which corresponding structures between the DRR and the observed image become easier to compare.

More recently, differentiable rendering frameworks have enabled end-to-end learning and optimization pipelines in which gradients can be propagated through the DRR generation process [14,15]. These methods combine physically-based projection models with deep learning, allowing the transformation parameters to be optimized using gradient-based methods within modern deep learning frameworks.

3.5. Hybrid Registration Pipelines

In practice, many modern 2D–3D registration systems combine multiple methodological paradigms in order to leverage the strengths of each approach [3,14]. These hybrid pipelines typically integrate learning-based components with classical optimization techniques.

 doi:10.20944/preprints202605.0431.v1

A common strategy is to use a neural network to provide an initial estimate of the transformation parameters, which is then refined using an intensity-based optimization framework [13]. Another class of hybrid methods combines feature-based matching with intensity-based refinement [6].

Hybrid approaches have become increasingly popular because they provide a practical balance between computational efficiency, robustness, and registration accuracy. By combining data-driven models with physics-based image formation models, these methods can exploit complementary sources of information and achieve improved performance in challenging clinical scenarios [14].

Table 1. Representative quantitative comparison across the four major categories of rigid 2D–3D registration methods. The reported accuracy and runtime ranges are intended as high-level summaries of representative studies rather than strictly controlled head-to-head benchmarks, since datasets, anatomies, initialization ranges, and hardware differ substantially across publications.

Category	Representative algorithms	Typical reported accuracy	Typical speed	Comments
Intensity-based	Powell or gradient-based DRR optimization with NCC, MI, or gradient-difference metrics [3,5,8]	Often sub-millimeter to low-millimeter TRE in well-calibrated single-structure settings	Typically seconds per case; historically longer without GPU DRR [3,24]	High final accuracy is possible, but optimization is sensitive to initialization and repeated DRR generation remains computationally expensive.
Feature-based	Contour-, landmark-, or vertebral edge-based registration pipelines [6,10]	Commonly low-millimeter accuracy when salient anatomy is reliably extracted	Often faster than dense intensity matching because optimization is performed on sparse geometric cues	Performance depends strongly on robust feature detection, which can degrade under occlusion, overlap, or poor fluoroscopic contrast.
Learning-based	CNN pose regression and policy-learning frameworks [12,13,39]	Usually competitive low-millimeter or low-error pose estimation within the trained operating regime	Near real-time inference, often tens of milliseconds to a few hundred milliseconds on GPUs [12,13]	Fast at test time, but accuracy and robustness depend on training data coverage, simulation realism, and domain generalization.
Hybrid / differentiable	Learned initialization followed by differentiable DRR refinement in $SE(3)$ [14,15,41]	Sub-millimeter accuracy has been reported in recent surgical datasets [41]	Faster than classical exhaustive optimization while retaining test-time refinement; typically near intraoperative speed on modern GPUs	Combines the capture range of learned priors with the precision of physics-based optimization, but introduces additional implementation and deployment complexity.

 doi:10.20944/preprints202605.0431.v1

4. Similarity Metrics for 2D–3D Registration

Similarity metrics play a central role in most rigid 2D–3D registration algorithms, particularly in intensity-based optimization frameworks. Given an observed projection image I_{obs} and a simulated projection $I_{\mathrm{DRR}}(T)$ generated from the 3D volume under transformation T, the similarity metric $\mathcal{S}$ measures how well the two images correspond. The choice of similarity metric significantly influences both the robustness and accuracy of the registration process, as it determines the shape of the optimization landscape.

In practice, similarity metrics must address several challenges that arise in projection imaging. First, X-ray images and DRRs often exhibit different intensity distributions due to variations in imaging physics, scatter, and detector response. Second, fluoroscopic images typically contain substantial noise and artifacts. Third, the projection process introduces overlapping anatomical structures that can complicate direct intensity comparisons. Consequently, a large variety of similarity measures have been proposed to improve robustness under these conditions.

In this section, we review several classes of similarity metrics that have been widely used in rigid 2D–3D registration, including intensity-based metrics, information-theoretic measures, gradient-based metrics, phase-based representations, and more recent learned metrics. Table 2 summarizes several commonly used similarity metrics and highlights their typical properties in terms of modality compatibility, robustness, and computational efficiency.

Table 2. Common similarity metrics used in 2D–3D registration.

Metric	Modality	Robustness	Speed
Sum of Squared Differences (SSD)	Mono-modal	Low	High
Normalized Cross Correlation (NCC)	Mono-modal	Medium	High
Mutual Information (MI)	Multi-modal	High	Medium
Gradient Correlation (GC)	X-ray/CT	High	Medium
Normalized Gradient Information (NGI)	X-ray/CT	High	Medium

4.1. Intensity-Based Metrics

Intensity-based similarity measures directly compare pixel intensity values between the observed projection image and the simulated DRR. These metrics are simple to compute and often serve as baseline methods in registration frameworks.

4.1.1. Sum of Squared Differences (SSD)

The sum of squared differences is one of the simplest similarity metrics and is defined as

$$\mathrm{SSD}(T) = \sum_{u,v} (I_{obs}(u,v) - I_{\mathrm{DRR}}(u,v;T))^2. \tag{11}$$

Minimizing SSD corresponds to finding the transformation that minimizes the squared intensity differences between the two images. SSD works well when the images have similar intensity distributions and noise levels. However, in 2D–3D registration the intensity relationship between DRRs and real X-ray images is often inconsistent due to differences in image formation models and detector responses. Consequently, SSD is rarely used in isolation for multi-modal registration problems.

4.1.2. Normalized Cross-Correlation (NCC)

Normalized cross-correlation improves robustness to global intensity scaling and offset by normalizing the intensity values:

$$\mathrm{NCC}(T) = \frac{\sum_{u,v}(I_{obs} - \bar{I}_{obs})(I_{\mathrm{DRR}} - \bar{I}_{\mathrm{DRR}})}{\sqrt{\sum_{u,v}(I_{obs} - \bar{I}_{obs})^2}\sqrt{\sum_{u,v}(I_{\mathrm{DRR}} - \bar{I}_{\mathrm{DRR}})^2}}. \tag{12}$$

 doi:10.20944/preprints202605.0431.v1

Here $\bar{I}_{obs}$ and $\bar{I}_{\mathrm{DRR}}$ denote the mean intensity values of the respective images. NCC is widely used in mono-modal registration and has also been applied in early 2D–3D registration studies [5]. Nevertheless, its performance may degrade when the intensity relationship between images becomes nonlinear.

4.2. Information-Theoretic Metrics

Information-theoretic measures have become a standard choice for multi-modal image registration. These metrics quantify the statistical dependence between the intensity distributions of two images rather than relying on direct intensity differences.

4.2.1. Mutual Information (MI)

Mutual information measures the amount of shared information between two images and is defined as

$$\mathrm{MI}(I_{obs}, I_{\mathrm{DRR}}) = H(I_{obs}) + H(I_{\mathrm{DRR}}) - H(I_{obs}, I_{\mathrm{DRR}}), \tag{13}$$

where $H(\cdot)$ denotes Shannon entropy and $H(\cdot,\cdot)$ denotes joint entropy.

MI has been widely used in multi-modal registration problems because it does not assume a linear relationship between image intensities. Several studies have successfully applied mutual information to 2D–3D registration tasks [11,34]. However, MI can be sensitive to noise and interpolation artifacts introduced during DRR generation.

Variants such as normalized mutual information (NMI) and entropy correlation coefficient (ECC) have been proposed to improve robustness and reduce sensitivity to overlap variations.

4.3. Gradient-Based Metrics

Gradient-based similarity measures focus on structural information in images rather than raw intensity values. Since anatomical structures in X-ray images are often characterized by edges and gradients, these metrics can improve robustness to intensity inconsistencies.

4.3.1. Gradient Correlation

Gradient correlation evaluates the similarity between the spatial gradients of the projection image and the DRR. Let $\nabla I = (\partial I/\partial x, \partial I/\partial y)$ denote the image gradient. The gradient correlation metric can be expressed as

$$\mathrm{GC}(T) = \frac{\sum_{u,v} \nabla I_{obs}(u,v) \cdot \nabla I_{\mathrm{DRR}}(u,v;T)}{\sqrt{\sum_{u,v} ||\nabla I_{obs}||^2} \sqrt{\sum_{u,v} ||\nabla I_{\mathrm{DRR}}||^2}}. \tag{14}$$

Gradient-based approaches emphasize structural boundaries and are less sensitive to global intensity variations. These properties have made gradient correlation a popular choice in many classical 2D–3D registration frameworks [5].

4.4. Normalized Gradient Information (NGI)

Normalized gradient information (NGI) was introduced to further improve robustness to intensity inconsistencies between DRRs and fluoroscopic images. Instead of comparing gradient magnitudes, NGI evaluates the alignment between gradient directions.

Let $\mathbf{g}_1$ and $\mathbf{g}_2$ denote normalized gradient vectors of the two images. The NGI metric evaluates their directional similarity, which reduces the influence of gradient magnitude differences caused by varying imaging conditions.

NGI has been shown to provide improved robustness in challenging clinical scenarios involving low-contrast images and imaging artifacts. It has been widely adopted in DRR-based registration methods, particularly in applications involving fluoroscopic guidance [3].

 doi:10.20944/preprints202605.0431.v1

4.5. Phase-Based Similarity Metrics

Phase-based metrics have recently attracted attention due to their robustness to illumination and intensity variations. These methods represent images in the frequency or wavelet domain and compare phase information rather than intensity values.

One example is phase-based mutual information, which computes mutual information using phase representations extracted from complex wavelet transforms. In particular, the dual-tree complex wavelet transform (DTCWT) provides approximately shift-invariant and orientation-selective phase information that captures structural features of images.

Let $\phi(I)$ denote the phase representation of image I obtained from the wavelet transform. Phase-based mutual information can be expressed as

$$\mathrm{PMI} = \mathrm{MI}(\phi(I_{obs}), \phi(I_{\mathrm{DRR}})). \tag{15}$$

This representation has been shown to improve robustness to noise and intensity inconsistencies between projection images and DRRs [42]. By focusing on structural phase information, these methods can produce smoother optimization landscapes and improved convergence behavior.

4.6. Learned Similarity Metrics

Recent advances in deep learning have enabled the development of learned similarity metrics for image registration. Instead of designing hand-crafted similarity measures, these approaches train neural networks to estimate similarity directly from data.

In a typical formulation, a convolutional neural network (CNN) receives the observed X-ray image and the corresponding DRR as inputs and outputs a similarity score:

$$\mathcal{S}_\theta(I_{obs}, I_{\mathrm{DRR}}), \tag{16}$$

where θ denotes the network parameters. The network is trained to assign higher similarity scores to correctly aligned image pairs than to misaligned pairs.

Learning-based similarity metrics have been explored in several recent works on 2D–3D registration and differentiable rendering frameworks [12,41]. These approaches benefit from the ability of deep networks to learn complex cross-modal relationships between DRRs and real X-ray images.

However, learned similarity metrics require large training datasets and may exhibit limited generalization across different imaging devices or anatomical regions. Consequently, many modern systems combine learned metrics with physics-based models and classical optimization techniques.

4.7. Discussion

The choice of similarity metric has a profound impact on the performance of 2D–3D registration algorithms. Classical intensity-based metrics, such as SSD and NCC, are simple and computationally efficient, but they often struggle with cross-modal inconsistencies between DRRs and X-ray images, particularly in the presence of noise, scatter, and nonlinear intensity transformations. Information-theoretic measures, such as mutual information, provide greater flexibility for multimodal registration by modeling statistical dependence rather than direct intensity correspondence. However, in 2D–3D registration, their performance can degrade due to limited sampling in the joint intensity distribution and sensitivity to noise and interpolation artifacts.

Empirical studies have shown that metrics incorporating structural information are generally more robust in practical 2D–3D registration scenarios. In particular, Penney et al. [5] demonstrated that *pattern intensity* and *gradient difference* achieve the most reliable performance across challenging conditions, including the presence of soft-tissue structures and interventional instruments. In contrast, mutual information and cross-correlation were found to be more sensitive to intensity outliers and background variations, leading to higher failure rates. These findings highlight the importance of

 doi:10.20944/preprints202605.0431.v1

designing similarity measures that emphasize local structural consistency while remaining robust to intensity mismatches.

More recent approaches, including gradient-based, phase-based, and learned similarity metrics, continue this trend by focusing on invariant representations of anatomical structures. Gradient- and phase-based methods improve robustness by emphasizing edges and structural features, while learned metrics aim to capture complex cross-modal relationships directly from data. In practice, many state-of-the-art systems combine multiple cues, such as gradient information and intensity statistics, to balance robustness and accuracy.

Looking forward, the integration of differentiable rendering and learning-based similarity models is expected to further advance this area. These developments enable end-to-end optimization and data-driven metric design, offering the potential to overcome limitations of hand-crafted similarity measures while maintaining the physical consistency of image formation models.

5. Digitally Reconstructed Radiographs (DRR) Generation

Digitally reconstructed radiographs (DRRs) play a fundamental role in many rigid 2D–3D registration algorithms. In intensity-based frameworks, DRRs provide the bridge between volumetric images and projection images by simulating the X-ray image formation process from a three-dimensional (3D) volume under a candidate transformation. The accuracy, efficiency, and differentiability of DRR generation strongly influence the performance of the registration algorithm, since DRRs must typically be generated repeatedly during the optimization process.

Given a volumetric image containing attenuation coefficients $\mu(x)$ and a projection geometry defined by the imaging system, DRR generation simulates the X-ray intensity detected at each pixel by integrating attenuation values along the corresponding projection ray. According to the Beer–Lambert law, the detected X-ray intensity can be modeled as

$$I(u,v) = I_0 \exp\left(- \int_{\mathcal{R}(u,v)} \mu(x) dx \right), \tag{17}$$

where I_0 denotes the incident X-ray intensity and $\mathcal{R}(u,v)$ represents the ray path corresponding to detector pixel (u,v). In practice, most DRR generation methods compute the line integral of attenuation coefficients along each ray and ignore the exponential term, since the relative attenuation values are sufficient for similarity evaluation.

Over the years, several computational strategies have been proposed to efficiently approximate this line integral. In this section, we review the most widely used techniques, including ray-casting approaches, the Siddon ray-tracing algorithm, GPU-accelerated DRR generation, and recent differentiable rendering methods.

5.1. Ray-Casting Methods

Ray-casting methods represent one of the earliest approaches for generating DRRs. In these methods, a ray is cast from the X-ray source through the volume toward each detector pixel. The attenuation values along the ray are sampled at regular intervals, and the resulting samples are accumulated to approximate the line integral.

Let $\mathbf{r}(s)$ denote the ray parameterized by distance s from the source. The DRR intensity at pixel (u,v) can be approximated as

$$I_{\mathrm{DRR}}(u,v) \approx \sum_{k=1}^{N} \mu(\mathbf{r}(s_k)) \Delta s, \tag{18}$$

where s_k represents discrete sampling locations along the ray and Δs is the sampling step size.

Ray-casting methods are conceptually simple and relatively easy to implement. However, their accuracy depends on the sampling resolution along each ray, and dense sampling can lead to high computational costs. Moreover, interpolation of voxel values is required when the sampling points do

 doi:10.20944/preprints202605.0431.v1

not coincide with voxel centers. Despite these limitations, ray casting remains widely used in early 2D–3D registration systems and serves as the foundation for many modern DRR implementations.

5.2. Siddon Ray-Tracing Algorithm

Digitally reconstructed radiographs (DRRs) are typically generated by simulating the propagation of X-ray rays through a volumetric image. As illustrated in Figure 4, rays are emitted from a virtual X-ray source and traverse the 3D volume toward the detector plane. Along each ray, the attenuation contributions of the intersected voxels are accumulated to produce the simulated projection intensity.

To improve computational efficiency and accuracy, Siddon proposed a fast algorithm for computing exact ray–voxel intersection lengths in a 3D grid [27]. Instead of sampling points along the ray, the Siddon algorithm calculates the distance traveled by the ray within each voxel and accumulates the corresponding attenuation contributions.

Let l_i denote the intersection length of the ray within voxel i, and μ_i denote the attenuation coefficient of that voxel. The DRR intensity can then be expressed as

$$I_{\mathrm{DRR}}(u, v) = \sum_i \mu_i l_i. \tag{19}$$

By computing the exact intersection lengths between the ray and the voxel grid, the Siddon algorithm avoids redundant sampling and provides an accurate geometric solution for ray traversal. As a result, it has become a widely used technique in both computed tomography reconstruction and DRR generation.

Despite its improved accuracy, Siddon's method remains computationally expensive when large volumes and high-resolution detectors are used. Since a DRR must be generated for each candidate transformation during registration, computational efficiency remains a critical consideration.

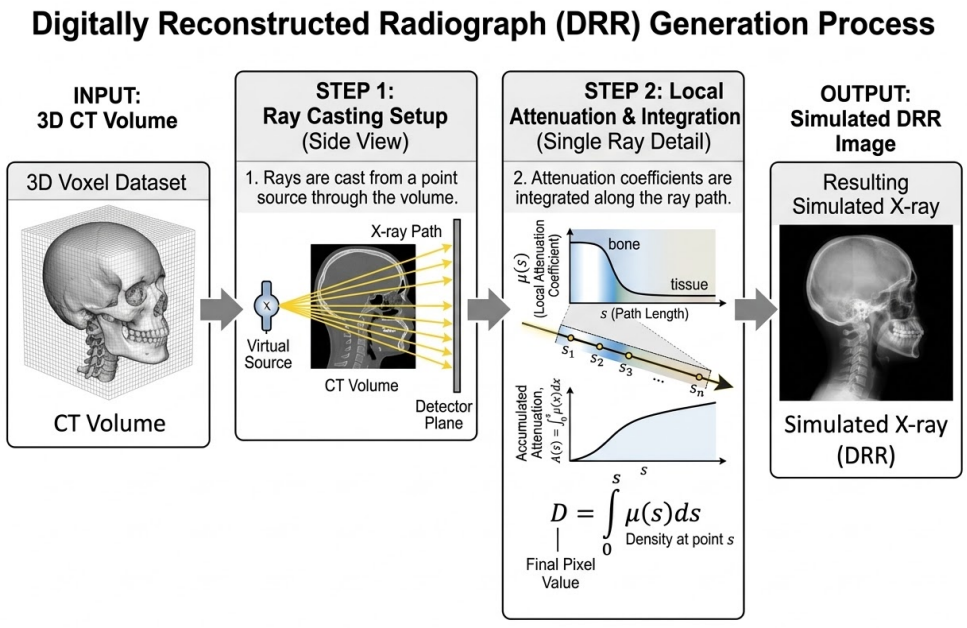


Figure 4. Illustration of digitally reconstructed radiograph (DRR) generation. Rays are cast from the virtual X-ray source through the 3D volume, and attenuation values are integrated along each ray to simulate the X-ray projection image.

5.3. GPU-Accelerated DRR Generation

With the increasing availability of programmable graphics processing units (GPUs), many DRR generation methods have been adapted to exploit parallel computing architectures. GPU-based

 doi:10.20944/preprints202605.0431.v1

DRR implementations can significantly accelerate ray integration by performing thousands of ray computations simultaneously.

In typical GPU-based implementations, the volume is stored as a three-dimensional texture, and hardware-accelerated interpolation is used to efficiently sample attenuation values along projection rays. Fragment shaders or compute kernels are then used to integrate attenuation values along the rays corresponding to detector pixels.

GPU acceleration has enabled real-time or near real-time DRR generation, which is essential for many clinical applications involving interactive registration or intraoperative navigation. Several studies have demonstrated substantial speed improvements over CPU-based implementations while maintaining comparable projection accuracy [43].

However, GPU implementations may introduce numerical differences due to interpolation schemes and floating-point precision. Careful implementation and validation are therefore necessary when using GPU-based DRR generation in medical applications.

5.4. Differentiable DRR Rendering

Recent advances in differentiable rendering have introduced a new paradigm for DRR generation in which gradients of the projection operation with respect to transformation parameters can be computed analytically or through automatic differentiation.

In differentiable DRR frameworks, the rendering process is formulated as a differentiable function

$$I_{\mathrm{DRR}} = f(V, T), \tag{20}$$

where V denotes the volumetric image and T represents the rigid transformation. By enabling gradient propagation through the rendering process, these methods allow gradient-based optimization or end-to-end training within modern deep learning frameworks.

DeepDRR is one such framework that combines physics-based X-ray simulation with machine learning techniques to generate realistic DRRs for training and evaluation of image-guided intervention algorithms [29]. The framework incorporates models of X-ray attenuation, scatter, and detector response to produce more realistic simulated images.

More recently, differentiable DRR frameworks such as DiffDRR have been developed to directly integrate DRR generation into optimization pipelines for 2D–3D registration [41,44]. These approaches enable efficient gradient-based optimization of pose parameters and have shown promising results in both classical and learning-based registration systems.

Differentiable rendering represents an important step toward bridging physics-based imaging models and data-driven learning approaches. By combining accurate forward models with gradient-based optimization, these methods offer new opportunities for improving registration robustness and computational efficiency.

5.5. Discussion

The generation of accurate and efficient DRRs remains a critical component of many 2D–3D registration algorithms. Classical ray-casting and Siddon-based methods provide physically meaningful projections but may be computationally expensive for iterative optimization frameworks. GPU-based acceleration has substantially improved computational efficiency and enabled real-time DRR generation in many applications. More recently, differentiable DRR rendering techniques have opened new possibilities for integrating physics-based image formation models with modern machine learning methods.

The choice of DRR generation technique therefore depends on the specific requirements of the application, including accuracy, computational speed, and compatibility with gradient-based optimization frameworks. As hardware and software technologies continue to evolve, DRR generation methods will likely play an increasingly important role in the development of next-generation 2D–3D registration systems.

 doi:10.20944/preprints202605.0431.v1

6. Optimization Strategies for 2D–3D Registration

Given the registration objective defined in Section 2,

$$\hat{T} = \arg\min_{T \in SE(3)} \mathcal{S}(I_{obs}, I_{\mathrm{DRR}}(T)), \tag{21}$$

the goal of the optimization algorithm is to identify the rigid transformation T that maximizes the similarity between the observed projection image and the simulated DRR. Since the transformation space is six-dimensional and the similarity landscape is often highly non-convex, the choice of optimization strategy plays a critical role in determining the robustness, accuracy, and computational efficiency of the registration process.

Over the years, a wide range of optimization techniques have been applied to rigid 2D–3D registration. These approaches can broadly be divided into gradient-free methods, gradient-based methods, multi-resolution optimization schemes, stochastic optimization strategies, and more recent differentiable optimization frameworks based on Lie group parameterizations. Table 3 summarizes several representative optimization methods commonly used in the literature.

Table 3. Optimization methods commonly used in rigid 2D–3D registration.

Method	Type	Derivatives	Characteristics
Powell	Derivative-free	No	Robust but slower
CMA-ES	Evolutionary	No	Good global search
Gauss–Newton	Gradient-based	Yes	Fast near optimum
Levenberg–Marquardt	Gradient-based	Yes	Stable nonlinear optimization
Stochastic optimization	Random sampling	Optional	Good for large search spaces

6.1. Gradient-Free Optimization

Early 2D–3D registration systems often relied on gradient-free optimization methods because the similarity metrics and DRR generation process were not differentiable. These algorithms search the transformation space using only function evaluations without requiring analytic gradients.

6.1.1. Powell's Method

Powell's method is a derivative-free optimization algorithm that iteratively minimizes a function along a set of search directions. At each iteration, the algorithm performs a sequence of one-dimensional line searches to update the transformation parameters.

Powell's method has been widely used in classical 2D–3D registration frameworks due to its simplicity and relatively robust convergence properties [5]. However, its performance can be sensitive to initialization, and the number of required similarity evaluations can become large when the search space is high-dimensional.

6.1.2. Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

Evolutionary algorithms have also been applied to rigid registration problems. Among these, the covariance matrix adaptation evolution strategy (CMA-ES) has gained popularity due to its ability to efficiently explore non-convex search spaces.

CMA-ES maintains a probabilistic model of the search distribution and iteratively updates the covariance matrix based on successful candidate solutions. This adaptive mechanism allows the algorithm to capture correlations between transformation parameters and improve search efficiency.

In the context of 2D–3D registration, CMA-ES has demonstrated improved robustness compared to deterministic gradient-free methods, particularly in scenarios where the similarity landscape contains multiple local minima.

 doi:10.20944/preprints202605.0431.v1

6.2. Gradient-Based Optimization

When gradients of the similarity function can be computed or approximated, gradient-based optimization methods can significantly accelerate convergence.

6.2.1. Gauss–Newton Optimization

The Gauss–Newton method is commonly used to solve nonlinear least squares problems. In the context of image registration, it approximates the Hessian matrix using the Jacobian of the residuals.

Let $\mathbf{r}(T)$ denote the vector of residuals between the observed image and the DRR. The Gauss–Newton update step can be written as

$$T_{k+1} = T_k - (J^T J)^{-1} J^T \mathbf{r}(T_k), \tag{22}$$

where J denotes the Jacobian matrix of the residuals with respect to the transformation parameters.

Gauss–Newton optimization can provide rapid convergence near the optimal solution, but it requires reliable gradient information and may become unstable when the approximation of the Hessian matrix is poor.

6.2.2. Levenberg–Marquardt Optimization

The Levenberg–Marquardt algorithm combines the advantages of Gauss–Newton and gradient descent methods by introducing a damping term that improves numerical stability.

The update step can be expressed as

$$T_{k+1} = T_k - (J^T J + \lambda I)^{-1} J^T \mathbf{r}(T_k), \tag{23}$$

where λ is a damping parameter that controls the balance between gradient descent and Gauss–Newton updates.

Levenberg–Marquardt optimization has been successfully applied in many medical image registration problems due to its robustness and efficient convergence behavior.

6.3. Multi-Resolution Optimization

The similarity landscape of 2D–3D registration is often highly non-convex due to image noise, anatomical structures, and projection artifacts. Multi-resolution optimization strategies address this challenge by performing registration at multiple image scales.

In a typical multi-resolution framework, both the projection image and the DRR are represented as image pyramids with progressively increasing resolution. The registration process begins at a coarse resolution level where the similarity landscape is smoother and large-scale misalignments can be corrected. The estimated transformation is then used to initialize optimization at finer resolution levels.

This coarse-to-fine strategy significantly reduces the risk of convergence to local minima and improves overall registration robustness.

6.4. Stochastic Optimization

Stochastic optimization methods have also been explored to improve efficiency when evaluating similarity metrics over large images. Instead of computing the similarity over all pixels, these methods use randomly sampled subsets of pixels to approximate the objective function.

Examples include stochastic gradient descent (SGD) and Monte Carlo sampling strategies. By reducing the number of pixel evaluations, stochastic optimization can substantially accelerate the registration process while maintaining acceptable accuracy.

However, stochastic approaches may introduce additional variance in the optimization process and require careful tuning of sampling strategies.

 doi:10.20944/preprints202605.0431.v1

6.5. Differentiable SE(3) Optimization

Recent developments in differentiable rendering and deep learning have enabled optimization frameworks that operate directly on the Lie group $SE(3)$. In these approaches, the rigid transformation is not updated by independently perturbing Euler angles and Cartesian translations. Instead, it is parameterized in the tangent space $\mathfrak{se}(3)$ using a six-dimensional twist vector, which gives an analytically convenient local representation of rigid motion.

Let $\boldsymbol{\xi} = (\boldsymbol{\omega}^\top, \mathbf{v}^\top)^\top \in \mathbb{R}^6$ denote the twist coordinates, where $\boldsymbol{\omega} \in \mathbb{R}^3$ represents the infinitesimal rotation and $\mathbf{v} \in \mathbb{R}^3$ the infinitesimal translation. The corresponding matrix representation in the Lie algebra is

$$\hat{\boldsymbol{\xi}} = \begin{bmatrix} [\boldsymbol{\omega}]_\times & \mathbf{v} \\ \mathbf{0}^\top & 0 \end{bmatrix}, \tag{24}$$

where $[\boldsymbol{\omega}]_\times$ is the skew-symmetric matrix of $\boldsymbol{\omega}$. The rigid transformation is then recovered by the exponential map

$$T = \exp(\hat{\boldsymbol{\xi}}) \in SE(3). \tag{25}$$

This representation is attractive because optimization can be carried out in the Euclidean tangent space while preserving the manifold structure of rigid motions. In practice, a current estimate T_k is often updated using a left-multiplicative perturbation,

$$T_{k+1} = \exp(\widehat{\Delta\boldsymbol{\xi}})T_k, \tag{26}$$

where $\Delta\boldsymbol{\xi} \in \mathbb{R}^6$ is a small increment predicted by an optimizer. Compared with directly optimizing Euler angles, this avoids singular parameterizations, reduces the need for ad hoc angle wrapping, and better respects the geometry of the pose space [14,15].

When the registration objective is differentiable, gradients can be propagated through the rendering pipeline and expressed with respect to the twist parameters. For a similarity loss $\mathcal{L}(T) = \mathcal{S}(I_{obs}, I_{\mathrm{DRR}}(T))$, one can apply first-order updates of the form

$$\Delta\boldsymbol{\xi} = -\eta\, \nabla_{\boldsymbol{\xi}}\mathcal{L}, \tag{27}$$

or use second-order approximations in the tangent space. This makes $\mathfrak{se}(3)$ especially attractive for gradient-based pose refinement with automatic differentiation, since the Jacobian of the rendered image with respect to the pose parameters can be computed without breaking the rigid-body constraint.

Modern differentiable 2D–3D registration methods frequently combine this pose representation with differentiable DRR generators such as DiffDRR and with locality-aware or geodesic losses [15,41]. A common design is to use a neural network for coarse initialization and then perform test-time optimization in $\mathfrak{se}(3)$, allowing the system to retain the capture range of learning-based methods while recovering the precision of physics-based registration. In this setting, the optimization variable is no longer treated as a generic 6-vector, but as a tangent-space element whose update is mapped back to $SE(3)$ through the exponential map after every iteration.

An additional advantage of the Lie-theoretic formulation is that it supports geometrically meaningful loss functions. Instead of penalizing rotation and translation parameters independently in an arbitrary coordinate system, one can measure error using geodesic distances on $SE(3)$ or by combining image-space similarity with transformation- space regularization. This is particularly useful in medical image registration because depth ambiguity and anisotropic imaging geometry can make some directions in pose space much harder to estimate than others. By encoding updates in $\mathfrak{se}(3)$, these anisotropies can be handled more systematically in both optimization and learning- based refinement frameworks.

 doi:10.20944/preprints202605.0431.v1

6.6. Discussion

Optimization strategies are a critical component of rigid 2D–3D registration algorithms, as they determine how effectively the transformation parameters can be estimated in the presence of non-convex similarity landscapes. Gradient-free methods such as Powell's method and CMA-ES provide robustness in situations where gradient information is unavailable but may require a large number of similarity evaluations. Gradient-based methods, including Gauss–Newton and Levenberg–Marquardt optimization, can achieve faster convergence when reliable derivatives are available. Multi-resolution frameworks improve robustness by smoothing the optimization landscape at coarse scales, while stochastic methods reduce computational cost by approximating similarity evaluations.

More recently, differentiable optimization on the $SE(3)$ manifold has emerged as a powerful paradigm that integrates classical registration models with modern machine learning frameworks. These approaches enable efficient gradient-based optimization and open new possibilities for end-to-end training of registration systems.

7. Datasets and Evaluation Protocols

Reliable evaluation is essential for assessing the accuracy, robustness, and computational performance of rigid 2D–3D registration methods. Because these algorithms are typically designed for image-guided clinical applications, evaluation protocols must quantify both geometric alignment accuracy and practical runtime performance under realistic imaging conditions.

In practice, evaluating 2D–3D registration algorithms presents several challenges. Unlike many computer vision problems, the ground truth transformation between a preoperative volumetric image and an intraoperative X-ray image is rarely known exactly in clinical settings. As a result, researchers often rely on a combination of synthetic datasets, physical phantoms, and carefully controlled clinical experiments to establish reference transformations.

This section reviews commonly used datasets and evaluation metrics in rigid 2D–3D registration studies.

7.1. Evaluation Metrics

Several quantitative metrics have been proposed to measure the accuracy of estimated transformations. These metrics typically evaluate either the geometric alignment between anatomical structures or the error in the estimated camera pose.

7.1.1. Target Registration Error (TRE)

Target Registration Error (TRE) is one of the most widely used metrics in medical image registration. TRE measures the Euclidean distance between corresponding anatomical landmarks after registration.

Let $\mathbf{x}_i$ denote a set of reference 3D landmarks and $\mathbf{x}_i^{gt}$ denote their ground truth positions after projection. If $\hat{T}$ denotes the estimated transformation, TRE can be computed as

$$\text{TRE} = \frac{1}{N}\sum_{i=1}^{N}\left\|\Pi(\hat{T}\mathbf{x}_i) - \Pi(T_{gt}\mathbf{x}_i)\right\|, \tag{28}$$

where $\Pi(\cdot)$ represents the projection operator and T_{gt} denotes the ground truth transformation.

TRE directly measures the geometric alignment of anatomical structures and is therefore highly relevant for clinical applications. In many surgical navigation tasks, achieving TRE values below 1–2 mm is considered necessary for safe and accurate guidance.

7.1.2. Pose Error

Pose error measures the difference between the estimated rigid transformation and the ground truth transformation. This error is typically decomposed into rotational and translational components.

The translation error can be computed as

 doi:10.20944/preprints202605.0431.v1

$$e_t = \|t_{est} - t_{gt}\|, \tag{29}$$

where t_{est} and t_{gt} denote the estimated and ground truth translation vectors.

The rotational error can be measured using the angle between the estimated rotation matrix R_{est} and the ground truth rotation matrix R_{gt}:

$$e_r = \cos^{-1}\left(\frac{\text{trace}(R_{est}R_{gt}^{\top}) - 1}{2}\right). \tag{30}$$

Pose error provides a direct measure of transformation accuracy and is often used when ground truth transformations are available, particularly in simulated experiments.

7.1.3. Reprojection Error

Reprojection error evaluates how well the estimated transformation aligns projected 3D points with their corresponding locations in the 2D image.

Given a set of 3D points $\mathbf{X}_i$ and their corresponding 2D image observations $\mathbf{u}_i$, the reprojection error can be defined as

$$e_{proj} = \frac{1}{N}\sum_{i=1}^{N} \|\Pi(\hat{T}\mathbf{X}_i) - \mathbf{u}_i\|. \tag{31}$$

This metric is commonly used in camera calibration and pose estimation problems and is particularly useful when accurate landmark correspondences are available.

7.2. Datasets

Different types of datasets are commonly used to evaluate rigid 2D–3D registration algorithms. Each dataset category provides different trade-offs between ground-truth accuracy and imaging realism. Table 4 summarizes the most common dataset types used in the literature and highlights their main advantages and limitations.

Table 4. Summary of dataset types commonly used for evaluating rigid 2D–3D registration methods.

Dataset type	Ground truth	Main advantages	Main limitations
Synthetic / DRR-based	Exact	Fully controlled imaging conditions, known transformations, and easy generation of large test sets. Particularly useful for evaluating pose error, TRE, capture range, and robustness under systematic perturbations.	Simulated projections may not fully reproduce scatter, beam hardening, detector noise, occlusions, or other complexities of real X-ray imaging, which can lead to optimistic performance estimates.
Physical phantom	High	Real X-ray acquisition geometry with controlled experimental conditions. Fiducial markers or external tracking systems can provide accurate reference transformations and enable repeatable experiments.	Limited anatomical variability and imperfect representation of patient-specific soft-tissue and imaging characteristics.
Cadaveric	Moderate–high	More realistic anatomy and image appearance than phantoms while still allowing partially controlled acquisition setups. Useful for assessing robustness in anatomically realistic conditions.	Limited availability, small cohort sizes, and nontrivial estimation of reference poses.
Clinical	Approximate	Highest realism and strongest relevance to practical image-guided interventions. Captures true workflow variability, image artifacts, and patient anatomy.	Ground truth is difficult to obtain accurately, and dataset access may be limited by privacy, calibration, and workflow constraints.

The following subsections discuss these dataset categories in more detail and describe how they are typically used in experimental evaluation.

 doi:10.20944/preprints202605.0431.v1

7.2.1. Synthetic Datasets

In rigid 2D–3D registration, synthetic evaluation datasets are most commonly constructed by generating digitally reconstructed radiographs (DRRs) from CT volumes under known projection geometries and poses, rather than by relying on a single universally adopted public synthetic benchmark. In recent literature, commonly used sources for such synthetic data include the CT volumes from DeepFluoro in pelvic registration, the Ljubljana dataset in vascular applications, and large public CT collections such as CTSpine1K for spine registration [45–47]. Large pelvic CT collections such as CTPelvic1K may also be used to generate synthetic training images at scale for learning-based methods [48]. More recently, differentiable rendering frameworks have further popularized patient-specific synthetic dataset generation, in which large numbers of DRRs are sampled on the fly from a subject's preoperative CT using randomized poses and projection parameters [49].

7.2.2. Clinical and Phantom Datasets

Clinical and phantom datasets play complementary roles in the evaluation of rigid 2D–3D registration methods. Clinical datasets provide the highest degree of realism because they contain images acquired during real diagnostic or interventional procedures, whereas phantom datasets offer greater experimental control and often more accurate reference pose estimation.

Among named public clinical datasets, one of the most important examples is the **Ljubljana 3D–2D-GS-CA** dataset, which contains 3D-DSA, 2D-DSA, and 2D-MAX images from 10 patients with cerebrovascular pathologies together with gold-standard registrations obtained from fiducial-marker alignment [46]. This dataset is one of the few publicly available clinical benchmarks for quantitative evaluation of 2D–3D registration algorithms.

In contrast, phantom-based evaluation is very common in the literature, but phantom datasets are often institution-specific rather than shared as standard public benchmarks. Many classical studies evaluate registration methods on custom spine phantoms, skull phantoms, or plastic pelvis phantoms under calibrated fluoroscopic imaging setups [5,50]. Phantom experiments are especially valuable because they preserve real X-ray acquisition geometry while allowing controlled imaging conditions and relatively accurate ground-truth measurement using fiducials, calibration objects, or external tracking systems.

Overall, clinical datasets are essential for demonstrating practical relevance under real workflow conditions, while phantom datasets remain important for repeatable technical validation. However, the limited availability of shared public phantom and clinical benchmarks continues to make cross-paper comparison difficult in rigid 2D–3D registration.

7.3. Evaluation Protocols

In addition to accuracy metrics, evaluation protocols often consider several practical performance measures.

Capture range. The capture range refers to the maximum initial misalignment from which the registration algorithm can successfully converge to the correct solution.

Robustness. Robustness is typically evaluated by running the algorithm on a large number of randomly perturbed initial poses and measuring the success rate of convergence.

Computational efficiency. Runtime performance is an important consideration for clinical applications, particularly in intraoperative settings where real-time feedback is required.

Comprehensive evaluation studies typically report several of these metrics simultaneously in order to provide a balanced assessment of registration accuracy, robustness, and computational efficiency.

7.4. Discussion

The evaluation of rigid 2D–3D registration methods remains an active area of research due to the difficulty of obtaining reliable ground truth transformations in clinical data. Synthetic datasets provide precise control over experimental conditions but may fail to capture the complexity of real imaging

 doi:10.20944/preprints202605.0431.v1

scenarios. Clinical and phantom datasets offer greater realism but often suffer from limited ground truth accuracy.

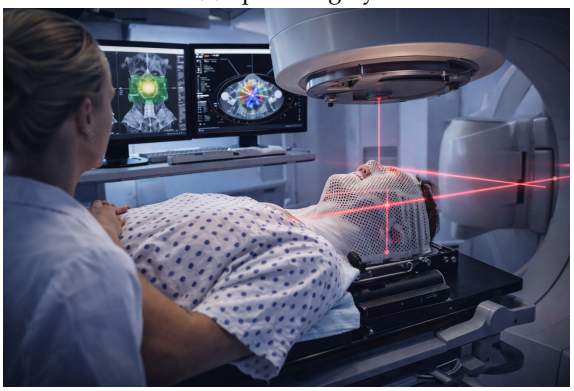
For this reason, many recent studies adopt hybrid evaluation protocols that combine synthetic experiments for quantitative accuracy assessment with clinical experiments for validation under realistic conditions. Establishing standardized datasets and benchmarking protocols remains an important direction for future research in 2D–3D registration.

8. Applications

Rigid 2D–3D registration plays an important role in many image-guided clinical procedures where intraoperative X-ray images must be aligned with preoperative volumetric scans. Representative applications include orthopedic surgery, spine navigation, radiation therapy, and interventional radiology. Figure 5 illustrates several typical clinical scenarios in which 2D–3D registration techniques are used [3,5].

(a) Spine surgery

(b) Orthopedic navigation



(c) Radiation therapy

(d) Interventional radiology

Figure 5. Representative clinical applications of rigid 2D–3D registration. These techniques are widely used in image-guided spine surgery, orthopedic navigation, radiation therapy patient positioning, and interventional radiology procedures.

8.1. Orthopedic Surgery

Orthopedic procedures frequently rely on fluoroscopic imaging for intraoperative guidance. However, fluoroscopic images provide only limited depth information and often require repeated imaging from multiple viewing angles. Rigid 2D–3D registration can address this limitation by aligning preoperative CT volumes with intraoperative fluoroscopic images, thereby enabling three-dimensional anatomical visualization during surgery.

This capability has been used in applications such as fracture reduction, implant placement, and joint replacement procedures. For example, Penney et al. [5] investigated the use of similarity measures for 2D–3D registration in orthopedic applications. Later studies further improved robustness and accuracy through gradient-based similarity metrics and optimized DRR generation methods

 doi:10.20944/preprints202605.0431.v1

[33]. These approaches allow preoperative anatomical models to be overlaid onto intraoperative fluoroscopic images, providing surgeons with improved spatial guidance.

8.2. Spine Navigation

Spinal surgery is another important application domain for rigid 2D–3D registration. Procedures such as pedicle screw insertion require high accuracy due to the proximity of critical anatomical structures including the spinal cord and nerve roots. In many navigation systems, preoperative CT scans are used to construct three-dimensional models of the vertebrae.

Rigid 2D–3D registration methods align these preoperative models with intraoperative fluoroscopic images to provide accurate visualization of vertebral anatomy during surgery [3,9,51]. Such systems have been shown to improve surgical accuracy while reducing the need for repeated fluoroscopic imaging, which may lower radiation exposure for both patients and medical staff.

8.3. Radiation Therapy

In radiation therapy, accurate patient positioning is essential to ensure that radiation doses are delivered precisely to the tumor while minimizing exposure to surrounding healthy tissues. Image-guided radiation therapy (IGRT) systems frequently use X-ray projection images to verify patient alignment before treatment.

Rigid 2D–3D registration methods can align planning CT images with portal images or cone-beam projections acquired during treatment [52,53]. This alignment enables automated patient positioning correction and improves treatment accuracy. Such techniques are particularly important in high-precision radiotherapy approaches such as stereotactic body radiation therapy (SBRT), where small positioning errors can significantly affect treatment outcomes.

8.4. Interventional Radiology

In interventional radiology, minimally invasive procedures are performed under fluoroscopic guidance. Examples include vascular interventions, cardiac catheterization, and tumor embolization procedures. Although fluoroscopy provides real-time imaging, it often lacks sufficient anatomical detail.

Rigid 2D–3D registration can align preoperative CT or MRI scans with intraoperative fluoroscopic images, enabling three-dimensional anatomical models to be overlaid on live X-ray images [29]. This technique improves navigation during complex interventions and can reduce procedure time by providing better anatomical context.

9. Clinical Translation and Deployment Roadmap

Although many rigid 2D–3D registration algorithms achieve strong results in controlled research studies, translation into routine clinical practice requires much more than improved image similarity or lower TRE. A clinically deployable system must fit within a procedural workflow, operate within strict latency constraints, remain reliable in the presence of calibration errors and artifacts, and satisfy quality, safety, and regulatory requirements. Figure 6 summarizes a typical pathway from algorithm development to real-world deployment.

 doi:10.20944/preprints202605.0431.v1

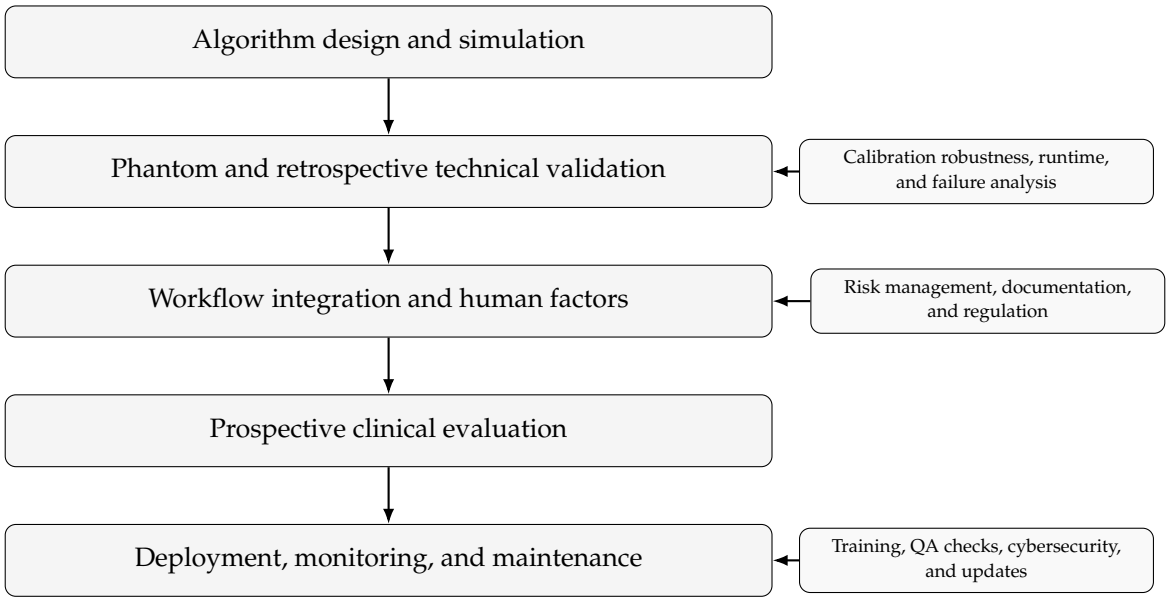


Figure 6. Typical pathway from a research prototype to clinical deployment for rigid 2D–3D registration systems. Translation requires not only geometric accuracy but also workflow compatibility, low latency, risk management, quality assurance, and post-deployment monitoring.

9.1. Technical Validation Beyond Accuracy

The first stage of translation is usually technical validation on simulated data, phantoms, cadavers, or retrospective clinical cases. At this stage, geometric metrics such as TRE, pose error, and reprojection error remain essential, but they are not sufficient on their own. Investigators must also characterize capture range, calibration sensitivity, robustness to occlusions and metal artifacts, and failure modes under poor initialization. For clinical deployment, reporting only mean accuracy is often inadequate; reproducibility, worst-case behavior, and uncertainty estimates are equally important.

9.2. Workflow Integration and Real-Time Performance

A registration method that is accurate in isolation may still fail to be clinically useful if it disrupts operating-room workflow. Practical systems must define when registration is triggered, who verifies the result, how the method interacts with existing navigation or C-arm hardware, and what happens when registration fails. Runtime is particularly important: in many interventional settings, the total latency budget includes image transfer, preprocessing, DRR generation, optimization, visualization, and user confirmation. Consequently, methods that achieve excellent accuracy but require repeated manual intervention or multi-minute computation may have limited clinical utility despite strong offline benchmarks.

9.3. Regulatory, Safety, and Quality Requirements

From a deployment perspective, 2D–3D registration software must be treated as part of a safety-critical system. This introduces requirements that are often outside the scope of algorithm papers, including software verification and validation, traceability, cybersecurity, logging, version control, and risk management for known failure modes. In addition, a clinically deployed system must support routine quality assurance, including imaging geometry checks, calibration verification, and safeguards against silent registration failure. Learning-based components add further challenges because model updates, domain shift, and data provenance must be controlled in a way that is compatible with clinical governance.

9.4. Prospective Evaluation and Post-Deployment Monitoring

The final stages of translation involve prospective testing in the intended use environment and continued monitoring after deployment. Prospective studies should evaluate not only registration

 doi:10.20944/preprints202605.0431.v1

accuracy but also procedure time, rate of manual corrections, radiation burden, operator trust, and downstream clinical outcomes. Once deployed, performance drift can arise from hardware changes, protocol updates, or previously unseen anatomy and implants. For this reason, robust monitoring, fallback strategies, and clear criteria for human override are central to safe long-term use.

10. Challenges and Future Directions

Despite significant progress in rigid 2D–3D registration research, several challenges remain that limit the widespread clinical adoption of these methods. Addressing these challenges is an important direction for future research.

10.1. Dose Variations

Fluoroscopic images acquired during clinical procedures often exhibit large variations in radiation dose and imaging conditions. Differences in exposure settings can lead to substantial changes in image contrast, noise levels, and intensity distributions. These variations can degrade the performance of intensity-based similarity metrics and complicate the registration process [3]. Developing similarity metrics and feature representations that are robust to dose variations remains an important research challenge.

10.2. Occlusion and Metal Artifacts

Surgical instruments, implants, and other metallic objects can introduce strong artifacts in X-ray images. Metal implants may produce streak artifacts and occlusions that obscure anatomical structures, making reliable registration more difficult.

Several approaches have been proposed to address these challenges, including robust similarity metrics, outlier rejection strategies, and segmentation-based masking techniques [33]. However, handling severe occlusions and artifacts remains an open research problem.

10.3. Generalization of Learning-Based Methods

Recent learning-based approaches have demonstrated promising performance in rigid 2D–3D registration tasks [12,41]. These methods leverage deep neural networks to estimate transformation parameters or learn cross-modal similarity metrics directly from data.

However, learning-based models often struggle to generalize across different imaging devices, anatomical regions, and clinical environments. Improving the robustness and generalization ability of learning-based registration methods is therefore an important direction for future research[54][55][56].

10.4. Real-Time Constraints

Many clinical applications require real-time or near real-time registration performance. For example, intraoperative navigation systems must provide immediate feedback to surgeons during procedures.

Achieving high registration accuracy while maintaining low latency remains a challenging problem. Advances in GPU acceleration, efficient DRR generation, and fast optimization algorithms have significantly improved performance in recent years [29], but further improvements are still needed for real-time clinical deployment.

10.5. Multi-View Registration

Many modern imaging systems can acquire multiple projection images from different viewing angles. Incorporating information from multiple views can significantly improve registration accuracy and reduce ambiguities caused by overlapping anatomical structures[57].

Multi-view 2D–3D registration methods jointly optimize the pose parameters across several projection images [3]. Although these approaches can provide more robust solutions, they also increase computational complexity and require accurate calibration of imaging geometry.

 doi:10.20944/preprints202605.0431.v1

11. Conclusion

Rigid 2D–3D registration is a fundamental problem in medical image computing that enables the integration of preoperative volumetric images with intraoperative projection imaging. Over the past two decades, numerous approaches have been proposed, ranging from classical intensity-based optimization methods to modern learning-based frameworks.

In this survey, we presented a comprehensive overview of rigid 2D–3D registration techniques. We first introduced a unified mathematical formulation of the problem and described the imaging geometry underlying X-ray projection. We then organized existing methods into several methodological categories and reviewed key algorithmic components, including similarity metrics, DRR generation methods, and optimization strategies. Furthermore, we summarized commonly used datasets and evaluation protocols and discussed important clinical applications of these methods.

Despite substantial progress, several challenges remain, including robustness to imaging artifacts, generalization of learning-based approaches, and real-time computational requirements. Continued advances in differentiable rendering, deep learning, and high-performance computing are expected to further improve the accuracy and efficiency of future 2D–3D registration systems.

We hope that this survey provides a useful reference for researchers and clinicians working in image-guided interventions and helps guide future developments in rigid 2D–3D registration.

References

1. Maintz, J.B.A.; Viergever, M.A. A survey of medical image registration. *Medical Image Analysis* **1998**, *2*, 1–36.
2. Hill, D.L.G.; Batchelor, P.G.; Holden, M.; Hawkes, D.J. Medical image registration. *Physics in Medicine and Biology* **2001**, *46*, R1–R45.
3. Markelj, P.; Tomaževič, D.; Likar, B.; Pernuš, F. A Review of 3D/2D Registration Methods for Image-Guided Interventions. *Medical Image Analysis* **2012**, *16*, 642–661.
4. Navab, N.; Bani-Hashemi, A.; Mitschke, M.; et al. 3D reconstruction from two fluoroscopic views: Applications to spine surgery. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 1998, pp. 476–483.
5. Penney, G.P.; Weese, J.; Little, J.A.; Desmedt, P.; Hill, D.L.G.; Hawkes, D.J. A Comparison of Similarity Measures for Use in 2-D–3-D Medical Image Registration. *IEEE Transactions on Medical Imaging* **1998**, *17*, 586–595. https://doi.org/10.1109/42.730404.
6. Penney, G.P.; Edwards, P.J.; King, A.P.; Hawkes, D.J. A framework for registration of 2D and 3D images for image-guided interventions. *Medical Image Analysis* **2001**, *5*, 79–98.
7. Shirato, H.; Shimizu, S.; Kunieda, T.; Kitamura, K.; et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *International Journal of Radiation Oncology Biology Physics* **2000**, *48*, 1187–1195.
8. Groher, M.; Zikic, D.; Navab, N. Comparison of 2D/3D registration methods for image-guided spine surgery. *Medical Image Analysis* **2009**, *13*, 884–899.
9. Birkfellner, W.; Figl, M.; Huber, K.; Watzinger, F.; Wanschitz, F.; Hummel, J.; Ewers, R.; Bergmann, H. Rigid 2D–3D Registration for Image-Guided Surgery. *Medical Physics* **2005**, *32*, 235–245. https://doi.org/10.1118/1.1828289.
10. Mitchell, S.C.; Penney, G.P.; King, A.P.; Hawkes, D.J. Multimodality image registration by maximization of mutual information. *Medical Image Analysis* **2001**, *5*, 79–98.
11. Viola, P.; Wells, W. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision* **1997**, *24*, 137–154.
12. Miao, S.; Wang, Z.J.; Zheng, Y.; Liao, R. A CNN Regression Approach for Real-Time 2D/3D Registration. *IEEE Transactions on Medical Imaging* **2016**, *35*, 1352–1363. https://doi.org/10.1109/TMI.2016.2521800.
13. Liao, R.; Miao, S.; de Tournemire, P.; Grbic, S.; Kamen, A. Deep learning based 2D/3D registration. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2017, pp. 137–145.
14. Unberath, M.; Zaech, J.; Lee, S.C.; Bier, B.; Fotouhi, J.; Armand, M.; Navab, N. DeepDRR: A catalyst for machine learning in fluoroscopy-guided procedures. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2018, pp. 98–106.
15. Gopalakrishnan, V.; Golland, P. DiffDRR: A Differentiable X-ray Renderer for 2D–3D Registration. *arXiv preprint* **2021**, *arXiv:2109.11028*.

 doi:10.20944/preprints202605.0431.v1

16. Zheng, G. 2D-3D registration for image-guided interventions. *Computerized Medical Imaging and Graphics* **2009**, *33*, 559–565.
17. van Herk, M.; Remeijer, P.; Rasch, C.; Lebesque, J.V. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology, Biology, Physics* **2000**, *47*, 1121–1135. https://doi.org/10.1016/S0360-3016(00)00518-6.
18. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2 ed.; Cambridge University Press, 2003.
19. Kak, A.C.; Slaney, M. *Principles of Computerized Tomographic Imaging*; SIAM, 2001.
20. Li, L.; Aubert, B.; Kemper, P.; Plumley, A. Intrinsic Tolerance in C-Arm Imaging: How Extrinsic Re-optimization Preserves 3D Reconstruction Accuracy, 2026, [arXiv:cs.CV/2603.14031].
21. Murray, R.M.; Li, Z.; Sastry, S.S. *A Mathematical Introduction to Robotic Manipulation*; CRC Press, 1994.
22. Ma, Y.; Soatto, S.; Košecká, J.; Sastry, S.S. *An Invitation to 3-D Vision: From Images to Geometric Models*; Springer, 2010.
23. Pennec, X. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision* **2006**, *25*, 127–154.
24. Otake, Y.; Wang, A.; Siewerdsen, J.H. An efficient GPU-based framework for intensity-based 2D–3D registration. *Medical Physics* **2012**, *39*, 3156–3169.
25. Bushberg, J.T.; Seibert, J.A.; Leidholdt, E.M.; Boone, J.M. *The Essential Physics of Medical Imaging*, 3 ed.; Lippincott Williams & Wilkins, 2011.
26. Zhou, C.; Faruqui, S.H.A.; An, D.; Patel, A.; Abdalla, R.N.; Hurley, M.C.; Shaibani, A.; Potts, M.B.; Jahromi, B.S.; Ansari, S.A.; et al. Single-View Fluoroscopic X-Ray Pose Estimation: A Comparison of Alternative Loss Functions and Volumetric Scene Representations. *Journal of Imaging Informatics in Medicine* **2025**, *38*, 2600–2612. Epub 2024 Dec 13, https://doi.org/10.1007/s10278-024-01354-w.
27. Siddon, R.L. Fast Calculation of the Exact Radiological Path for a Three-Dimensional CT Array. *Medical Physics* **1985**, *12*, 252–255.
28. Göksel, O.; et al. GPU-accelerated 2D–3D registration for real-time image-guided interventions. *IEEE Transactions on Medical Imaging* **2013**, *32*, 1707–1720.
29. Unberath, M.; Zaech, J.N.; Lee, S.C.; Bier, B.; Fotouhi, J.; Armand, M.; Navab, N. DeepDRR: A Catalyst for Machine Learning in Fluoroscopy-Guided Procedures. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2018, Lecture Notes in Computer Science.
30. Gopalakrishnan, V.; et al. DiffDRR: Dynamic differentiable digitally reconstructed radiographs. *arXiv preprint arXiv:2208.12737* **2022**.
31. Pluim, J.P.W.; Maintz, J.B.A.; Viergever, M.A. Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging* **2003**, *22*, 986–1004.
32. Roche, A.; Malandain, G.; Pennec, X.; Ayache, N. Rigid registration of 3-D ultrasound with MR images: A new approach combining intensity and gradient information. In Proceedings of the IEEE Transactions / MICCAI-era registration literature reference, 1998.
33. Wein, W.; Khamene, A.; Sauer, F.; Navab, N. Robust 3D–2D Registration in Fluoroscopy. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2008, Vol. 5241, *Lecture Notes in Computer Science*, pp. 1000–1008.
34. Maes, F. Multimodality Image Registration by Maximization of Mutual Information. *IEEE Transactions on Medical Imaging* **1997**, *16*, 187–198.
35. Grangeat, P. Mathematical framework of cone beam 3D reconstruction via the first derivative of the Radon transform. In Proceedings of the Mathematical Methods in Tomography; Herman, G.T.; Louis, A.K.; Natterer, F., Eds., Berlin, Heidelberg, 1991; pp. 66–97. https://doi.org/10.1007/BFb0084509.
36. Frysch, R.; Pfeiffer, T.; Rose, G. A novel approach to 2D/3D registration of X-ray images using Grangeat's relation. *Medical Image Analysis* **2021**, *67*, 101815. https://doi.org/10.1016/j.media.2020.101815.
37. Zhou, C.; Cha, T.; Peng, Y.; Li, G. Transfer learning from an artificial radiograph-landmark dataset for registration of the anatomic skull model to dual fluoroscopic X-ray images. *Computers in Biology and Medicine* **2021**, *138*, 104923. https://doi.org/10.1016/j.compbiomed.2021.104923.
38. Gopalakrishnan, V.; Golland, P. DiffDRR: Differentiable Digitally Reconstructed Radiographs. *arXiv preprint* **2022**, *arXiv:2208.12737*.
39. Gao, C.; Liu, X.; Gu, W.; Killeen, B.; Armand, M.; Taylor, R.; Unberath, M. Generalizing Spatial Transformers to Projective Geometry with Applications to 2D/3D Registration. In Proceedings of the MICCAI, 2020, [2003.10987].

 doi:10.20944/preprints202605.0431.v1

40. Xiong, J.; Deng, R.; Yue, J.; Lu, S.; Guo, J.; Lionts, M.; Yao, T.; Cui, C.; Zhu, J.; Qu, C.; et al. ZeroReg3D: a zero-shot registration pipeline for 3D consecutive histopathology image reconstruction. *Journal of Medical Imaging* **2025**, *12*, 044002–044002.
41. Gao, C.; Unberath, M. DiffPose: Differentiable Pose Estimation for 2D–3D Registration. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2022, Lecture Notes in Computer Science.
42. Abugharbieh, R.; Bab-Hadiashar, A.; Jiang, D.B.D. Registration of 2D to 3D joint images using phase-based mutual information. In Proceedings of the Proceedings of SPIE Medical Imaging, 2007, Vol. 6512, p. 651209. https://doi.org/10.1117/12.709118.
43. Dorgham, O.M.; Laycock, S.D.; Fisher, M.H. GPU-Accelerated Generation of Digitally Reconstructed Radiographs for 2-D/3-D Image Registration. *IEEE Transactions on Biomedical Engineering* **2012**, *59*, 2595–2603. https://doi.org/10.1109/TBME.2012.2207898.
44. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM* **2021**, *65*, 99–106. https://doi.org/10.1145/3503250.
45. Grupp, R.B.; Unberath, M.; Gao, C.; Hegeman, R.A.; Murphy, R.J.; Alexander, C.P.; Otake, Y.; McArthur, B.A.; Armand, M.; Taylor, R.H. Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2D/3D registration. *International Journal of Computer Assisted Radiology and Surgery* **2020**, *15*, 759–769.
46. Mitrović, U.; Špiclin, Ž.; Likar, B.; Pernuš, F. On generating the gold standard data for evaluation of 3D–2D registration of cerebral angiograms. *Medical Image Analysis* **2013**, *17*, 911–921.
47. Deng, Y.; et al. CTSpine1K: A large-scale dataset for spinal vertebrae segmentation in computed tomography. *arXiv preprint arXiv:2105.14711* **2021**.
48. Peng, C.; et al. Deep adaptive collaborative learning for automated pelvic CT segmentation. *Medical Image Analysis* **2021**.
49. Gopalakrishnan, V.; et al. Intraoperative 2D/3D Image Registration via Differentiable X-ray Rendering. In Proceedings of the CVPR, 2024.
50. Zöllei, L.; Grimson, W.E.L.; Norbash, A.; Wells, W.M. 2D-3D rigid registration of X-ray fluoroscopy and CT images using mutual information and sparsely sampled histogram estimators. In Proceedings of the CVPR, 2001.
51. Suh, Y.; Li, L.; Plumley, A.; Zhou, C.; Moyer, D.; Kang, K. Pedicle Screw Pairing and Registration for Screw Pose Estimation from Dual C-arm Images Using CAD Models. *arXiv preprint* **2025**, *arXiv:2511.05702*.
52. Weese, J.; Schmitt, O.; Buzug, T.M.; Georgi, J.C. Automatic Patient Positioning in Radiotherapy. *Medical Physics* **1999**, *26*, 1221–1230. https://doi.org/10.1118/1.598635.
53. Khamene, A.; Bloch, P.; Wein, W.; Svatos, M.; Sauer, F. Automatic Registration of Portal Images and Volumetric CT for Patient Positioning in Radiation Therapy. *Medical Image Analysis* **2006**, *10*, 96–112. https://doi.org/10.1016/j.media.2005.06.002.
54. Ghosh, T.; Sheikhi, F.; Guo, J.; Singh, Y.; Younis, K.; Kuanar, S.; Faghani, S.; Farina, E.M.J.d.M.; Huo, Y.; Maleki, F. Foundation Models for Volumetric Medical Imaging: Opportunities, Challenges, and Future Directions. *Electronics* **2026**, *15*, 1245.
55. Qu, G.; Zhou, Z.; Calhoun, V.D.; Zhang, A.; Wang, Y.P. Integrated brain connectivity analysis with fMRI, DTI, and sMRI powered by interpretable graph neural networks. *Medical Image Analysis* **2025**, *103*, 103570. https://doi.org/https://doi.org/10.1016/j.media.2025.103570.
56. Qu, G.; Hu, W.; Xiao, L.; Wang, Y.P. A graph deep learning model for the classification of groups with different IQ using resting state fMRI. In Proceedings of the Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging; Krol, A.; Gimi, B.S., Eds. International Society for Optics and Photonics, SPIE, 2020, Vol. 11317, p. 113170A. https://doi.org/10.1117/12.2549274.
57. Zhu, Y.; Xiong, J.; Zhu, J.; Deng, R.; Zhao, S.; Wang, Y.; Wang, Y.; Qu, C.; Lu, Z.; Yang, Y.; et al. 3D Banff Lesion Scoring for Kidney Transplant Pathology: Feasibility and Utility for Volumetric Quantification. In Proceedings of the Medical Imaging with Deep Learning, 2026.