A Unifying MAP-MRF Framework for Deriving New Point Similarity Measures for Intensity-based 2D-3D Registration

Guoyan Zheng and Xuan Zhang
MEM Research Center, University of Bern, Stauffacherstrasse 78, CH-3014 Bern, Switzerland
Guoyan.Zheng@MEMcenter.unibe.ch

Abstract

Similarity measure is one of the main factors that affect the accuracy of intensity-based 2D-3D registration of X-ray fluoroscopy to CT images. This paper presents a unifying MAP-MRF framework for rationally deriving point similarity measures based on Bayes theorem. Three new similarity measures derived from this framework are presented and evaluated using a phantom and a human cadaveric specimen. Their behaviors are compared to other well-known similarity measures and the comparison results are reported. Combining any one of the new similarity measures with a previously introduced spline-based multi-resolution 2D-3D registration scheme, we develop a fast and accurate registration algorithm. We report their capture ranges, converging speeds, and registration accuracies.

1. Introduction

2D-3D registration of a set of two-dimensional (2D) X-ray images with a three-dimensional (3D) volume data has shown great potential in computer-assisted radiosurgery [1] and orthopaedic surgery [2]. It obviates the invasive procedure of the conventional registration methods, which match the actual location of implanted fiducial markers, anatomical landmarks, or a cloud of points on the surface of the anatomy to the corresponding points in the model extracted from the volume data. The reported techniques to achieve this registration can be split into two main categories: feature-based methods [3-4] and intensity-based [5-7] methods. Feature-based methods require a prerequisite segmentation stage which is error-prone and hard to achieve automatically. The errors in segmentation can lead to errors in the final registration. In contrast, intensity-based methods directly compare the fluoroscopy images with digitally reconstructed radiographs (DRR’s), which are obtained by simulating X-ray projection of a CT volume. No segmentation is required.

One of the main factors that affect the accuracy of intensity-based 2D-3D registration is the similarity measure, which is a criterion function that is used in the registration procedure for measuring the quality of image match. Penney et al. compare six similarity measures, based on fiducial markers, for registering fluoroscopy images to a 3D CT of a spine phantom [7]. They find that mutual information is the least accurate of the six similarity measures and that pattern intensity is one of the two similarity measures that are able to register accurately and robustly, even when soft tissues and interventional instruments were present in the X-ray images. Unfortunately, pattern intensity is designed directly by using some heuristic rules [6].

This work formulates a Markov Random Filed (MRF) model for 2D-3D registration based on Bayesian framework. The optimal solution is defined as the maximum a posteriori (MAP) estimate of the MRF. By using this unifying MAP-MRF framework, we can derive new point similarity measure in a rational way, in contrast to heuristic methods. The minimization of each individual similarity measure derived from this framework leads to optimal registration. We point out that two previously published similarity measures, i.e., sum-of-squared-difference (SSD) [5] and modified pattern intensity [2], can be also derived from this framework.

The remainder of this paper is organized as follows. Section 2 briefly introduces the 2D-3D registration scheme used in this paper. Section 3 presents the basic concept of Markov Random Field. Section 4 describes the derivation. Section 4 presents the experimental results, followed by conclusions in section 5.

2. 2D-3D registration scheme

The 2D-3D registration scheme used in this paper is based on a recently introduced spline-based multi-resolution 2D-3D registration method [2, 5]. This 2D-
3D registration method follows the same computation framework as other intensity-based methods. Given a set of X-ray fluoroscopies and CT images, it iteratively optimizes the six rigid-body parameters describing the orientation and the translation of the patient pose, by generating and comparing DRR’s with the X-ray images using appropriate similarity measure. The differences between this method and other intensity-based methods lie in: 1) a cubic-splines data model was used to compute the multi-resolution data pyramids for both CT volume and X-ray images, the DRR’s, as well as the gradient and the Hessian of the cost function; 2) a Levenberg-Marquardt non-linear least-squares optimizer was adapted to a multi-resolution context. The registration was performed from coarsest resolution until finer one. The accuracy of this method depends on the chosen similarity measure. Previously, accuracy of approximately 1.4 ± 0.2 mm when SSD is used [5] and accuracy of 0.8 ± 0.1 mm when modified pattern intensity is used [2] have been reported.

3. Markov random field

Markov random field theory is a branch of probability theory for analyzing the spatial or contextual dependencies of physical phenomena. It has been used extensively in image restoration, segmentation, object recognition and matching [8]. In what follows, the basic concepts of the MRF are reviewed for completeness. For rigorous expositions one may refer to [8].

Let \( L = \{(i,j): 1 \leq i \leq M, 1 \leq j \leq N\} \) be an \( M \times N \) integer lattice; then \( D = \{D_{i,j}, (i,j) \in L\} \) denotes a family of random variables, i.e., a random field, defined on \( L \). Each image \( d \) can be viewed as a discrete sample realization of \( D \) assuming the intensity value on each pixel site. \( d = \{d_{i,j}, d_{i,j}, ..., d_{M,N}\} \) is referred to as a configuration of \( D \), the complete set of all the configuration is denoted as \( \mathcal{D} \).

Definition 1: A neighborhood system for \( L \) is defined as \( N = \{N_{i,j}, (i,j) \in L\} \), where \( N_{i,j} \) is the set of sites on the neighborhood of \((i,j)\), whose definition is as follows:
\[
N_{i,j} = \{(i',j') \mid (i',j') \in L, (i',j') \neq (i,j), |(i',j') - (i,j)| \leq r\}
\]
where \( r \) is a positive integer. An \( N \) is called \( n \)th order neighborhood system if \( r \) assumes the value of \( n \).

Definition 2: A clique \( c \) is a subset of \( L \) for which every pair of sites is a neighbor. Single pixels are also considered cliques. The set of all cliques on an integer lattice is denoted by \( \mathcal{C} \).

Definition 3: A random field \( D \) is a MRF with respect to the neighborhood system \( N \) if and only if:
\[
a. \quad P(d) > 0, \forall d \in \mathcal{D} \quad (2)
\]
\[
b. \quad P(d_{i,j} \mid d_{-i,j}) = P(d_{i,j} \mid d_{-i,j}) \quad (3)
\]

Definition 4: \( D \) is a Gibbs Random Field (GRF) with respect to \( (w.r.t.) \) the neighborhood system \( N \) if and only if:
\[
P(d) = \frac{1}{z} e^{-E(d)} \quad (4)
\]
where \( z \) is a normalization constant called the partition function and \( E(d) \) is the energy function of the form:
\[
E(d) = \sum_{c \in \mathcal{C}} V_c(d) \quad (5)
\]
where \( V_c \) is called the clique potential. Generally, \( V_c \) is a function of the cliques around the site under consideration.

Hammersley-Clifford theorem [8]: \( D \) is an MRF on \( L \) w.r.t. \( N \) if and only if \( D \) is a GRF on \( L \) w.r.t. \( N \).

4. Deriving point similarity measures based on MRF modeling of difference images

To find an optimal registration transformation we cast the problem into a Bayesian framework of MAP-MRF estimate. We thus follow the four steps of the MAP-MRF estimate.

1. Construction of a prior probability distribution \( p(T) \) for the registration transformation \( T \) matching the reference X-ray images to the floating DRR’s.
2. Formulation of an observation model \( p(D \mid T) \) that describes the distribution of the observed difference images \( D \) between the reference X-ray images and the floating DRR’s given any particular realization of the prior distribution.
3. Combination of the prior and the observation model into the posterior distribution by Bayes theorem
\[
p(T \mid D) \propto p(D \mid T)p(T) \quad (6)
\]
4. Drawing inference based on the posterior distribution.

4.1. Prior distributions

One advantage of formulating 2D-3D registration according to Bayesian framework is that we are able to specify a prior distribution for each configuration of registration parameter space. In this paper, we don’t take advantage of this property. But it is possible to use this property to design an energy term that is able to apply geometric regularization in a non-rigid 2D-3D registration, which prevents unrealistic deformations.
4.2. Observation model

Given a realization of the prior distribution, the observation model $p(D|T)$ describes the conditional distribution of the observed difference image $D$. By specifying an observation model we may favor a transformation that establishes matching between regions of similar properties. By modeling the difference image $D$ as a GRF with respect to the neighborhood system $N$ we can derive the energy function for the observation model as:

$$
\alpha \sum_{i,j} V(d_{ij}) + (1-\alpha) \sum_{i,j} \frac{1}{\text{card}(N_{ij})} \sum_{\delta \in \delta_{ij}} V(d_{ij}, d_{\delta_{ij}}) \quad (7)
$$

where the first term is the potential function for single-pixel cliques and the second term is the potential function for all other cliques defined by neighborhood system $N$. $\alpha \in [0;1]$ weights the influence of these two terms. $\text{card}()$ means to compute the number of pixels in neighborhood $N'_{ij}$.

4.3. MAP estimate

The posterior conditional probability distribution is given by:

$$
p(T|D) \propto \exp(-E(d)) \quad (8)
$$

In search for the MAP estimate

$$
\hat{T} = \arg \max_T p(T|D) \quad (9)
$$

To illustrate how to derive similarity measures using the presented framework, two examples of previously published similarity measures are given as follows.

**Sum-of-Squared-Difference (SSD):** SSD can be derived from eq. (7) by specifying $\alpha = 1$ and $V(d_{ij}) = d_{ij}^2$. 

**Pattern Intensity:** the original pattern intensity proposed in [8] is written in the form

$$
P_{r,\sigma} = \sum_{i,j} \frac{1}{\text{card}(N'_{ij})} \sum_{\delta \in \delta_{ij}} (d_{ij}^2 + (d_{\delta_{ij}} - d_{ij})^2) \quad (10)
$$

where $r$ and $\sigma$ are two parameters to be experimentally determined. $N'_{ij}$ is a neighborhood with radius $r$.

As long as the parameters $r$ and $\sigma$ are determined, maximizing eq. (10) is equivalent to minimizing

$$
\sum_{i,j} \frac{1}{\text{card}(N'_{ij})} \sum_{\delta \in \delta_{ij}} (d_{ij}^2 - d_{\delta_{ij}}^2)^2 \quad (11)
$$

where $(d_{ij}^2 - d_{\delta_{ij}}^2)$ is the clique potential function for a pairwise-interaction model [9], which is equivalent to applying a first order smoothness constraint to a neighborhood $N'_{i}$ of order $r$ around site $t = (i,j)$. Numbering and order coding of the neighborhood up to order five is shown in Figure 1. Note that $N'_{i} \supseteq N'_{1}$.

![Figure 1. Numbering and order coding of neighborhood structure for a pairwise interaction model. Left: the neighborhood of $d_{i}$ up to order five; Right: the location of site $d_{i,4}$ in the neighborhood system](image)

4.4. Deriving new similarity measures

More generally, by choosing different neighborhood system and by specifying different clique potential functions to apply different orders of smoothness constraint on the difference images, we can derive different new similarity measures.

**Isotropic rth order neighborhood system with 1st order smoothness constraint (INrS1):** INrS1 is defined using following equation.

$$
\alpha \sum_{i,j} d_{ij}^2 + (1-\alpha) \sum_{i,j} \frac{1}{\text{card}(N'_{ij})} \sum_{\delta \in \delta_{ij}} (d_{\delta_{ij}} - d_{ij})^2 \quad (12)
$$

It is actually a combination of SSD and LSNPI. Two anisotropic similarity measures can be derived using following equation:

$$
P_a = \alpha \sum_{i,j} d_{ij}^2 + (1-\alpha) \sum_{i,j} (d_{(i,j)}^2 + d_{(i,j)}^2) \quad (13)
$$

where $d_{(i,j)} = \frac{\partial d}{\partial x}(i,j); d_{(i,j)} = \frac{\partial d}{\partial y}(i,j)$ is the first derivatives of the difference image $D$.

**Anisotropic 4-neighborhood system with first order smoothness constraint (A4NS1):** A4NS1 uses eq. (13) as the similarity measure and computes the first derivative using 4-neighborhood system with following convolution masks:

- $[-1 \ 0 \ 1]$ for the determination of $d_{(i,j)}^{x}$ and
- $[-1 \ 0 \ 1]^T$ for the determination of $d_{(i,j)}^{y}$

**Anisotropic 8-neighborhood system with first order smoothness constraint (A8NS1):** A8NS1 uses eq. (13) as the similarity measure and computes the
first derivative using 4-neighborhood system with following convolution masks:

- \[
\begin{bmatrix}
-1 & 0 & 1 \\
-1 & 0 & 1 \\
-1 & -1 & -1 \\
0 & 0 & 1 \\
1 & 1 & 1 \\
\end{bmatrix}
\]

for the determination of \( d_{i,j} \) and

- \[
\begin{bmatrix}
-1 & 0 & 1 \\
-1 & 0 & 1 \\
-1 & -1 & -1 \\
0 & 0 & 1 \\
1 & 1 & 1 \\
\end{bmatrix}
\]

for the determination of \( d_{i,j} \).

5. Experiments

A phantom and a human cadaveric spine specimen together with their ground truth were used in our experiments. The ground truth was obtained by implanting fiducial markers. The phantom was custom-made to simulate a good condition. In contrast, projections of interventional instruments were present in the X-ray images of the human cadaveric specimen to simulate a practical situation in image-guided therapy, as shown in Figure 2.

For all three newly derived similarity measures, the parameter \( \alpha \) was chosen as 0.5. For the similarity measure \( \text{INrS}1 \), the order of the neighborhood was chosen as \( r = 3 \). From now on, we call this similarity measure as \( \text{IN3S}1 \).

Figure 2. Experimental datasets. Left two images: Phantom and its C-arm image; right two images: volume rendering results of the CT data of the cadaveric specimen and its C-arm image presented with the projection of the interventional instruments (delineated by a black box).

Figure 3. Behavior of different similarity measures. Top row: cut through the minimum of different similarity measures on phantom data; bottom row: cut through the minimum of different similarity measures on cadaveric spine specimen data. The ordinate shows the value of different similarity measures (they are all normalized to the range [0.0, 1.0]), which are given as functions of each individual parameter varied in the range of \([-15^\circ, 15^\circ]\) or \([-15\text{mm}, 15\text{mm}]\) away from the associated ground truth (from left to right: (1) 1st column: rotation around X axis; (2) 2nd column: rotation around Y axis; (3) 3rd column: rotation around Z axis; (4) 4th column: translation along X axis; (5) 5th column: translation along Y axis; (6) 6th column: translation along Z axis). Zero in each abscissa means the ground truth for that individual parameter.
The first experiment was designed to compare the behaviors of the newly derived similarity measures to those of the published similarity measures such as SSD and mutual information. We cut through the minimums of all similarity measures by varying each one of the six registration parameters in the range of [-15°, 15°] or [-15mm, 15mm] away the associated ground truth of two datasets. The results were given in Figure 3. It was found that all similarity measures had similar behavior when tested on the phantom data but different behavior when tested on the cadaveric data. Those similarity measures derived from the presented MAP-MRF framework showed a superior behavior compared to other two published similarity measures. More specifically, the curves for the newly derived similarity measures have clear minima and are smoother, which is an important property to take the advantage of our 2D-3D registration scheme, which uses a gradient-based optimization technique to speed up the procedure.

Combining any one of the newly derived similarity measures with the 2D-3D registration scheme described in Section 2, we developed a fast and accurate 2D-3D registration algorithm. The second experiment was designed to evaluate the capture ranges, converging steps, and registration accuracies of these registration algorithms. This experiment was only performed on the human cadaveric specimen dataset. For this purpose, we perturbed the ground truth transformation by randomly varying each registration parameter in the range of [-2°, 2°] or [-2mm, 2mm] to get 100 positions, and then another 100 positions in the range of [-4°, 4°] or [-4mm, 4mm], and so on until the final range of [-12°, 12°] or [-12mm, 12mm]. We then performed our registrations and counted how many times they converged for each range (when the target registration error (TRE) measured on those fiducial markers was less than 1.5 mm). The capture range was defined when there was at least 95% successful rate. The experimental results on capture ranges and converging steps are given in Figure 4. The results on registration accuracies are shown in Table I.

### 6. Conclusions

In this paper, we introduced a unifying MAP-MRF framework to derive novel point similarity measures for 2D-3D registration of X-ray fluoroscopy to CT images. The derived novel point similarity measures had been evaluated using phantom and cadaver and the results showed that they provided satisfactory 2D-3D registration accuracy, even when interventional instruments were present.

### 7. References


