Point Similarity Measures Based on MRF Modeling of Difference Images for Spline-based 2D-3D Rigid Registration of X-ray Fluoroscopy to CT images

Guoyan Zheng[1], Xuan Zhang[1], Slavica Jonić[2, 3], Philippe Thévenaz[2], Michael Unser[2], and Lutz-Peter Nolte[1]

[1] MEM Research Center, University of Bern, Stauffacherstrasse 78, CH-3014, Bern, Switzerland
Guoyan.Zheng@MEMcenter.unibe.ch
[2] Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), CH-1015, Lausanne VD, Switzerland
[3] Institut de Minéralogie et de Physique des Milieux Condensés, Université Pierre et Marie Curie, F-75015 Paris, France

Abstract. One of the main factors that affect the accuracy of intensity-based registration of two-dimensional (2D) X-ray fluoroscopy to three-dimensional (3D) CT data is the similarity measure, which is a criterion function that is used in the registration procedure for measuring the quality of image match. This paper presents a unifying framework for rationally deriving point similarity measures based on Markov random field (MRF) modeling of difference images which are obtained by comparing the reference fluoroscopic images with their associated digitally reconstructed radiographs (DRR's). The optimal solution is defined as the maximum a posterior (MAP) estimate of the MRF. Three novel point similarity measures derived from this framework are presented. They are evaluated using a phantom and a human cadaveric specimen. Combining any one of the newly proposed similarity measures with a previously introduced spline-based registration scheme, we develop a fast and accurate registration algorithm. We report their capture ranges, converging speeds, and registration accuracies.

1 Introduction

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. An extensive study of six similarity measures applied specifically to 2D-3D registration has been performed by Penney *et al.* [1]. The similarity measures considered by the authors were: normalized cross-correlation [2], entropy of the difference image [3], pattern intensity [4], mutual information [5], gradient correlation [6], and gradient difference [1]. Using the fiducial markers to get the "gold-standard" registration, the authors ranked these measures based on their accuracy and robustness. They found that pattern intensity was one of the two similarity measures that were able to register accurately and robustly, even when soft tissues and interventional

instruments were present in the X-ray images. Unfortunately, pattern intensity was designed by using some heuristic rules [4].

This work formulates a MRF model on the difference images obtained by comparing the input fluoroscopic images with their associated DRR's. The optimal solution is defined as the MAP estimated of the MRF. By using this unifying MAP-MRF framework, we can derive new point similarity measure in a rational way. The optimization 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) [7] and pattern intensity [4], 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 describes the derivation of point similarity measures based on MRF modeling of the difference images. Section 4 presents the experimental results, followed by conclusions in Section 5.

2 Spline-based 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 scheme [7, 8]. This scheme follows the computation framework of intensity-based methods. Given a set of X-ray images and a CT volume, it iteratively optimizes the six rigid-body parameters describing the orientation and the translation of the patient pose, by generating and comparing floating DRR's with the reference X-ray images using appropriate similarity measure. The differences between this method and other intensity-based methods lie in [7]: 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 Marquardt-Levenberg non-linear least-squares optimizer was adapted to a multi-resolution context. The registration was performed from the coarsest resolution until the finest one. The accuracy of this method depends on the chosen similarity measure. Previously, accuracy of approximately 1.4 $\pm$ 0.2 mm when SSD was used [7] has been reported.

3 Deriving point similarity measures based on MRF modeling of different 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 [9].

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 $p(D|T)$ that describes the distribution of the observed difference images D by comparing 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

$$p(T \mid D) \propto p(D \mid T)p(T) \tag{1}$$

4. Drawing inference based on the posterior distribution

3.1 Prior distribution

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. We treat all parameter configurations equally. Due to the Euler angle based parameterization of rotation in our approach, $p(T)$ are a uniform distribution. But it is possible to use this property to favor certain transformations when different parameterization forms such as quaternion are used.

3.2 Observation model

Given a realization of the prior distribution, the observation model $p(D|T)$ describes the conditional distribution of the observed difference images 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 MRF with respect to the rth order neighborhood system $N = \{N_{i,j}^r\}$ we can derive the energy function for the observation model as:

$$E(D \mid T) = \sum_{q=1}^{Q} [\alpha \sum_{i,j}^{I,J} V(d_{i,j}) + (1-\alpha) \sum_{i,j}^{I,J} \frac{1}{card(N_{i,j}^r)} \sum_{(i',j') \in N_{i,j}^r} V(d_{i,j}, d_{i',j'})] \tag{2}$$

where Q is the number of images and $I \times J$ is the size of each image. The first term is the potential function for single-pixel cliques and the second term is the potential function for all other cliques. $\alpha \in [0:1]$ weights the influence of these two terms. $card(N_{i,j}^r)$ means to compute the number of pixels in neighborhood $N_{i,j}^r$.

The selection of the potential functions in Eq. (2) is a critical issue in MRF modeling [9]. As pointed out below, its selection decides the form of similarity measure.

The computation of the difference images also plays an important role in the present framework. In [4], an adjustable scaling parameter was used to build the difference images. To eliminate this parameter, Jonić *et al.* [7] tries to normalize the intensity range of the input reference fluoroscopic images and that of the corresponding DRR's by removing their mean and then dividing by their standard deviation. In this paper, we use a similar method. But unlike in [7], where the mean and the standard deviation were computed from the complete region of interest (ROI), we compute them only using those pixels in the neighborhood $N_{i,j}^r$.

3.3 MAP estimate

The posterior conditional probability distribution is given by:

$$p(T \mid D) \propto \mathbf{exp}(-E(D \mid T)) \tag{3}$$

In search for the MAP estimate:

$$\hat{T} = \mathbf{arg\,max}_T \, p(T \mid D) \tag{4}$$

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

Sum-of-Squared-Difference (SSD): It can be derived from Eq. (2) by specifying $\alpha = 1$ and $V(d_{i,j}) = d_{i,j}^2$.

Pattern Intensity: the pattern intensity proposed in [4] is written in the form:

$$P_{r,\sigma} = \sum_{i,j} \frac{1}{card(N_{i,j}^r)} \sum_{(i',j') \in N_{i,j}^r} \frac{\sigma^2}{\sigma^2 + (d_{i',j'} - d_{i,j})^2} \tag{5}$$

where r and σ are two parameters to be experimentally determined. $N_{i,j}^r$ is a neighborhood with radius r. It can be derived from the present framework by specifying $\alpha = 0$ and using following pairwise clique potential function:

$$V(d_{i,j}, d_{i',j'}) = -\frac{1}{1 + \frac{(d_{i',j'} - d_{i,j})^2}{\sigma^2}} \tag{6}$$

where $d_{i',j'}$ is a pixel in the neighborhood $N_{i,j}^r$.

3.4 Deriving new point similarity measures

More generally, by choosing different neighborhood system and by specifying different clique potential functions that incorporates different *a priori* constraints, we can derive different new similarity measures.

Isotropic rth order neighborhood system and pairwise potential function with 1st order smoothness constraint (INrS1): It is defined using following equation:

$$\alpha \sum_{i,j}^{I,J} d_{i,j}^2 + (1-\alpha) \sum_{i,j}^{I,J} \frac{1}{card(N_{i,j}^r)} \sum_{(i',j') \in N_{i,j}^r} (d_{i',j'} - d_{i,j})^2 \tag{7}$$

It is actually a combination of **SSD** and a modified form of pattern intensity [10]. Following the suggestion in [4], we also choose r=3 pixels. From now on, we call this similarity measure **IN3S1.**

Two anisotropic similarity measures can be derived using following equation:

$$P_\alpha = \alpha \sum_{i,j}^{I,J} d_{i,j}^2 + (1-\alpha) \sum_{i,j}^{I,J} (d_{(i,j),x}^2 + d_{(i,j),y}^2) \quad (8)$$

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

Anisotropic 4-neighborhood system and potential functions with first order smoothness constraint (AN4S1): It computes the first derivative in Eq. (8) using 4-neightborhood system with following convolution masks:

- $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$ for the determination of $d_{(i,j),x}$ and
- $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T$ for the determination of $d_{(i,j),y}$

Anisotropic 8-neighborhood system and potential functions with first order smoothness constraint (AN8S1): It also computes the first derivative in Eq. (8) using 4-neightborhood system but with following convolution masks:

- $\begin{bmatrix} -(1/3) & 0 & (1/3) \\ -(1/3) & 0 & (1/3) \\ -(1/3) & 0 & (1/3) \end{bmatrix}$ for the determination of $d_{(i,j),x}$ and
- $\begin{bmatrix} -(1/3) & -(1/3) & -(1/3) \\ 0 & 0 & 0 \\ (1/3) & (1/3) & (1/3) \end{bmatrix}$ for the determination of $d_{(i,j),y}$

4 Experiments

A phantom and a human cadaveric spine specimen together with their ground truths were used in our experiments. Both phantom and cadaveric specimen were scanned by a GE LightSpeed Ultra CT scanner (GE Healthcare, Chalfont St. Giles, United Kingdom) with same intra-slice solution (0.36 mm x 0.36 mm) but with different inter-slice thickness, 1.25 mm for the phantom and 2.5 mm for the cadaveric specimen, which resulted in volume dataset of size 512x512x93 volxels for phantom and 512x512x72 for the cadaveric specimen, respectively. The 2D projection images of both phantom and cadaveric specimen were acquired from a Siemens ISO-C C-arm (Siemens AG, Erlangen, Germany). They are calibrated and undistorted with custom-made software with high accuracy. 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 cadaveric specimen to simulate a practical situation in image-guided therapy

The ground truths were obtained by implanting fiducial markers. Both phantom and cadaveric specimen were equipped with infrared light emitting diodes (LEDs) markers to establish a patient coordinate system (P-COS) and was tracked using an optoelectronic position sensor (Northern Digital OptoTrak 3020). The actual locations of fiducial markers were digitized in P-COS using an optoelectronically tracked pointer and were matched to the corresponding points in CT volume dataset. The ground truths were then obtained

using singular value decomposition with an accuracy of 0.52 mm for phantom and 0.65 mm for cadaver, respectively.

For all three newly derived similarity measures, the parameter α was chosen as 0.5. Each time, two nearly orthogonal C-arm images from the corresponding dataset were used for the experiments described below.

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. Though mutual information was ranked as least accurate in [1], other group [11, 12] later found that it performed reasonably well. The results were given in Figure 1. 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 present MAP-MRF framework showed a superior behavior compared to other two well-known similarity measures. More specially, 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. It is also evident that the behavior of mutual information is better than that of **SSD**.

Combining any one of the similarity measures with the 2D-3D registration scheme described in Section 2, we developed a 2D-3D registration algorithm. The second experiment was designed to evaluate their capture ranges, converging steps, and registration accuracies of these registration algorithms. Based on the investigation results obtained in the first experiment, we only performed this experiment on the human cadaveric specimen dataset to compare the three newly derived similarity measures. 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 2. The results on registration accuracies are shown in Table I. It was found that **IN3S1** had larger capture range than other two similarity measures but it was also less accurate and required more steps to be converged.

5 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.

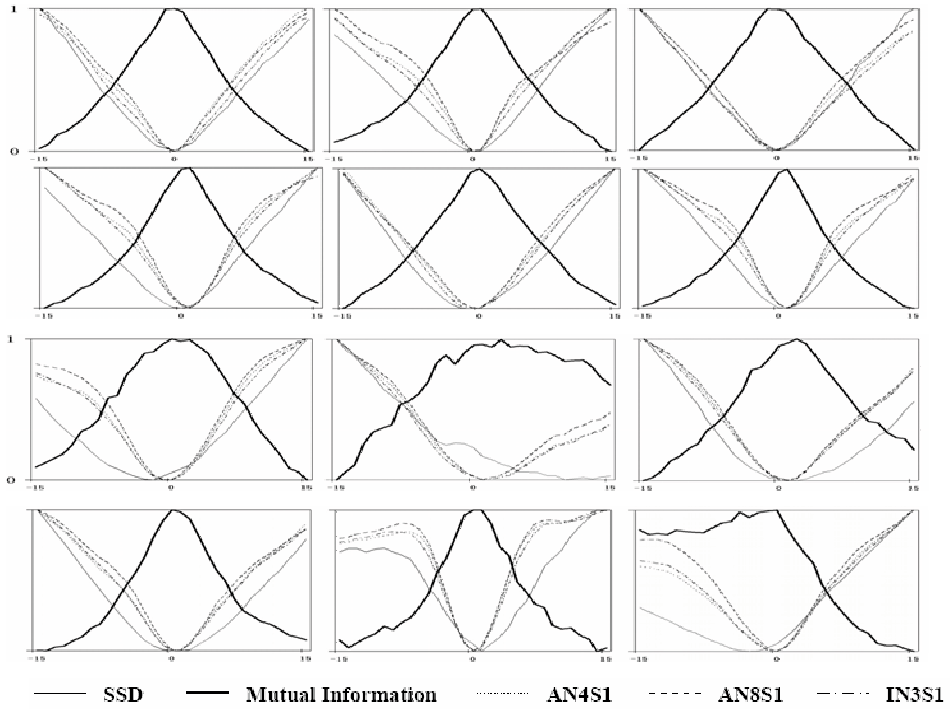

Fig. 1. Behavior of different similarity measures. Cut through the minimum of different similarity measures on the phantom data (the 1st and 2nd rows) as well as on the cadaveric spine specimen (the 3rd and 4th rows). The ordinate shows the value of different similarity measures (they are normalized to the range [0.0, 1.0]), which are given as functions of each rigid transformation parameter in the range of [-15°, 15°] or [-15 mm, 15 mm] away from the its ground truth ((1) 1st column of the 1st and 3rd rows: X rotation; (2) 2nd column of the 1st and 3rd rows: Y rotation; (3) 3rd column of the 1st and 3rd rows: Z rotation; (4) 1st column of the 2nd and 4th rows: X translation; (5) 2nd column of the 2nd and 4th rows: Y translation; (6) 3rd column of the 2nd and 4th rows: Z translation). Zero in each abscissa means the ground truth for that individual parameter, obtained by paired point matching based on fiducial markers.

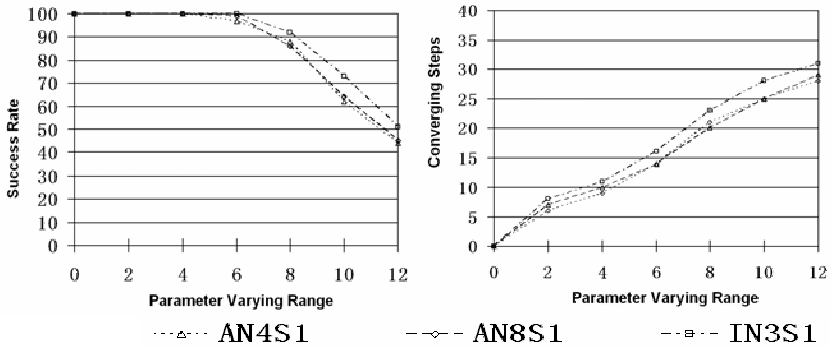

Fig. 2. Experimental results of capture ranges (left) and converging steps (right)

Table 1. Results of registration accuracies

Ranges	State	Target Registration Error (TRE) (mm)		
		AN4S1	AN8S1	IN3S1
[-2°, 2°] and [-2mm, 2mm]	Initial	2.3 ± 0.7	2.3 ± 0.7	2.3 ± 0.7
	Success Registration	0.7 ± 0.2	0.7 ± 0.1	0.8 ± 0.1
[-4°, 4°] and [-4mm, 4mm]	Initial	4.6 ± 1.4	4.6 ± 1.4	4.6 ± 1.4
	Success Registration	0.7 ± 0.2	0.8 ± 0.1	0.8 ± 0.1
[-6°, 6°] and [-6mm, 6mm]	Initial	6.9 ± 2.2	6.9 ± 2.2	6.9 ± 2.1
	Success Registration	0.7 ± 0.1	0.8 ± 0.1	0.8 ± 0.1
[-8°, 8°] and [-8mm, 8mm]	Initial	8.9 ± 2.8	8.9 ± 2.8	8.9 ± 2.8
	Success Registration	0.7 ± 0.2	0.8 ± 0.1	0.8 ± 0.1
[-10°, 10°] and [-10mm, 10mm]	Initial	10.4 ± 3.6	10.4 ± 3.6	10.4± 3.6
	Success Registration	0.7 ± 0.2	0.8 ± 0.1	0.9 ±0 .2
[-12°, 12°] and [-12mm, 12mm]	Initial	11.6 ± 3.9	11.6 ± 3.9	11.6 ± 3.9
	Success Registration	0.7 ± 0.2	0.8 ± 0.1	0.9 ± 0.2

References

1. Penney G.P., Weese J., Little J.A., Desmedt P., Hill D.L.G., and Hawkes D.J.: A comparison of similarity measures for use in 2D-3D medical image registration. IEEE T Med Imaging, Vol. 17, No. 4, (1998) 586-595
2. Lemieux L., Jagoe R., Fish D.R., Kitchen N.D., and Thomas D.G.T.: A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. Med Phys, Vol. 21, No. 11, (1994) 1749-1760
3. Buzug T.M., Weese J., Fassnacht C., and Lorenz C.: Image registration: convex weighting functions for histogram-based similarity measures. Lecture Notes in Computer Science, Vol. 1205, Springer-Verlag, Berlin Heidelberg New York (1997) 203-212
4. Weese J., Buzug T.M., Lorenz C., Fassnacht C.: An approach to 2D/3D registration of a vertebra in 2D x-ray fluoroscopies with 3D CT images. Lecture Notes in Computer Science, Vol. 1205, Springer-Verlag, Berlin Heidelberg New York (1997) 119-128
5. Maes F., Collignon A., Vandermeulen D., Marchal G., and Suetens P. : Multi-modality image registration by maximization of mutual information. IEEE T Med Imaging, Vol. 16, No. 2, (1997) 187-198
6. Brown L.M.G. and Boult T.E.: Registration of planar film radiographs with computed tomography. IEEE Proceedings of MMBIA (1996) 42-51
7. Jonić S., Thévenaz P., Zheng G., Nolte L.-P., and Unser M.: Spline-based registration of a 3-D CT to a set of C-arm images. Submitted to Internation Journal of Biomedical Imaging.
8. Jonić S., Thévenaz P., and Unser M.: Multiresolution-based registration of a volume to a set of its projection. Proceedings of the SPIE International Symposium on Medical Imaging: Image Processing (MI' 03), San Diego CA, USA, (2003), Vol. 5032, Part II, pp. 1049-1052
9. Li S.Z.: Markov random field modeling in computer vision. Springer-Verlag, Berlin Heidelberg New York, (1995).
10. Zheng G., Zhang X., and Nolte L.-P.: Assessing spline-based multi-resolution 2D-3D image registration for practical use in surgical guidance. Lecture Notes in Computer Science, Vol. 3150, (2004) 294-301
11. Zöllei L., Grimson E., Norbash A., and Wells W.: 2D-3D rigid registration of X-ray fluoroscopy and CT images using mutual information and sparsely sampled histogram estimators. Proceedings of CVPR'01, Volume 2, (2001) 679-703.
12. Russakoff D.B., Rohlfind T., and Maurer C.R. Jr.: Fast intensity-based 2D-3D image registration of clinical data using lighting fields. Proceedings of ICCV'03, (2003) 416-422.