

A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration

Graeme P Penney¹, Jürgen Weese², John A Little, Paul Desmedt³,
Derek LG Hill¹, and David J Hawkes¹

¹ Division of Radiological Sciences, UMDS, Guy's & St Thomas' Hospitals, London SE1 9RT, UK. {email G.Penney@umds.ac.uk}

² Philips Research Hamburg, Rontgenstraße 24-26, 22335 Hamburg, Germany.

³ EasyVision Advanced Development, Philips Medical Systems, Veenpluis 4-6, P.O. Box 10.000, 5680 DA Best, the Netherlands.

Abstract. A comparison of six similarity measures for use in intensity based 2D-3D image registration is presented. The accuracy of the similarity measures are compared to a “gold-standard” registration which has been accurately calculated using fiducial markers. The similarity measures are used to register a CT scan to a fluoroscopy image of a spine phantom. The registration is carried out within a region of interest in the fluoroscopy image which is user defined to contain a single vertebra. Many of the problems involved in this type of registration are caused by features which were not modelled by a phantom image alone. More realistic “gold standard” data sets were simulated using the phantom image with clinical image features overlaid. Results show that the introduction of soft tissue structures and interventional instruments into the phantom image can have a large effect on the performance of some similarity measures previously applied to 2D-3D image registration. Two measures were able to register accurately and robustly even when soft tissue structures and interventional instruments were present as differences between the images. These measures are called pattern intensity and gradient difference.

1 Introduction

Common modalities for guiding interventions are ultrasound or x-ray fluoroscopy. These modalities are “real-time” but only two dimensional, so they lack the spatial information contained in computed tomography (CT) and magnetic resonance (MR) images. There are also a number of important anatomical features which are not visualised well using these modalities, but can be observed using CT and/or MR. One method of allowing information from CT images to be used during interventional procedures is to register the CT scan to an intra-operative x-ray fluoroscopy image. A number of papers have described techniques to achieve this registration [1, 3, 4, 7, 8, 14]. Current techniques to achieve registration divide into those that match features, such as bony structures [1, 4, 7], and those that use image intensities directly [3, 8, 14]. The former relies on a pre-processing step to segment appropriate features and are therefore difficult to automate.

This paper compares six intensity based similarity measures to determine which is the most accurate and robust. Rigid body registrations are carried out between a CT scan and a fluoroscopy image of a spine phantom. The final registrations are compared to a “gold-standard” registration calculated using fiducial markers. More clinically realistic fluoroscopy images are simulated by overlaying structures segmented from clinical fluoroscopy images on to the fluoroscopy image of the spine phantom.

2 Comparing fluoroscopic images and DRRs

Digitally reconstructed radiographs (DRRs) are produced by casting rays through a CT volume and integrating the Hounsfield numbers along each ray. There are two main types of difference between DRRs and fluoroscopy images, those which are caused by changes in the imaged object and those due to differences in image formation.

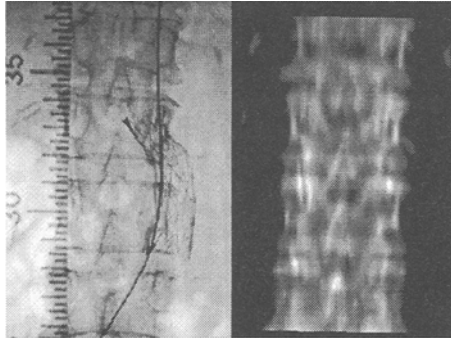


Fig. 1. Clinical fluoroscopy image (left) and digitally reconstructed radiograph (right) at registration.

2.1 Differences due to changes in imaged object

Overlying and underlying structures. We are assuming that the transformation between CT and fluoroscopy can be described by a perspective projection of rigid 3D motion. Soft tissue structures, however, can deform between the pre-operative CT image and interventional fluoroscopy. Because of this, we threshold the CT volume so no soft tissue structures are projected (see figure 1).

Interventional instruments in the field of view can create large differences between the images.

Spinal deformation can cause the vertebrae to be in different positions in the CT and fluoroscopy images.

2.2 Differences in image formation

Different x-ray energies between the modalities.

Heel effect and non-uniformity of the image intensifier response.

Different resolutions of fluoroscopy images and CT images.

Truncation when rays cut through the top or bottom of the CT scanned volume. Such rays are incomplete and so are not compared to the fluoroscopy image.

Geometric distortion in the fluoroscopy image is corrected for using a suitable phantom and software [5].

3 Similarity measures

The following sections outline a number of similarity measures. Each similarity measure is used to compare a fluoroscopic image (intensity values I_{fl}) with a DRR (intensity values I_{DRR}). The position and orientation of the CT volume with respect to the

fluoroscopy set are defined by ten parameters $\mathbf{P} = (X, Y, Z, \theta_x, \theta_y, \theta_z, c_s, l_s, k_1, k_2)$, see figure 2.

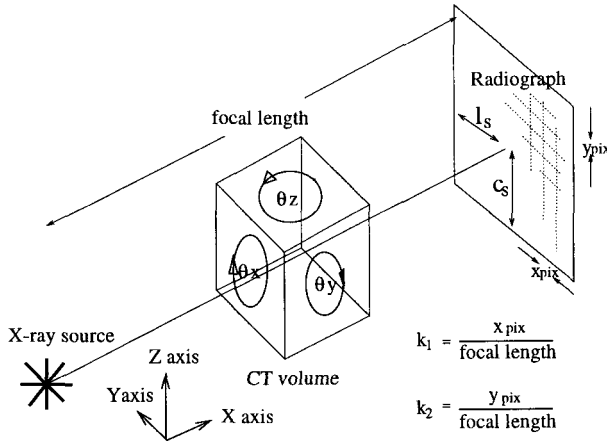


Fig. 2. Diagram showing the degrees of freedom in rigid body perspective projection. Six rigid body parameters, 3 translational X , Y and Z and 3 rotational θ_x , θ_y and θ_z . Four parameters c_s , l_s , k_1 and k_2 are associated with perspective projection: c_s and l_s are the coordinates on the film where the normal to the film goes through the x-ray source and k_1 and k_2 equal the pixel sizes x_{pix} and y_{pix} divided by the focal length.

3.1 Normalised Cross Correlation [8]

$$R = \frac{\sum_{(i,j) \in T} (I_{fl}(i,j) - \bar{I}_{fl})(I_{DRR}(i,j) - \bar{I}_{DRR})}{\sqrt{\sum_{(i,j) \in T} (I_{fl}(i,j) - \bar{I}_{fl})^2} \sqrt{\sum_{(i,j) \in T} (I_{DRR}(i,j) - \bar{I}_{DRR})^2}} \quad (1)$$

\bar{I}_{fl} and \bar{I}_{DRR} are the mean values of the images in the overlap region $(i,j) \in T$.

3.2 Entropy of the difference image [2]

$$I_{diff} = I_{fl} - s I_{DRR} \quad (2)$$

$$H(s) = - \sum_x p(x) \log p(x) \quad (3)$$

The entropy measure (H) described here operates on a single difference image (I_{diff}) which is created by subtracting the DRR from the fluoroscopy image using a suitable intensity scaling factor, s . A histogram is formed from the difference image and $p(x)$ denotes the probability of obtaining the pixel value x in I_{diff} .

3.3 Mutual information [9,11–13]

$$S = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (4)$$

Where $p(x)$ and $p(y)$ are the probability distributions in individual images and $p(x, y)$ is the joint probability distribution.

3.4 Gradient correlation [3, 8]

Horizontal and vertical Sobel templates were used to create gradient images dI_{fl}/di , dI_{fl}/dj and dI_{DRR}/di , dI_{DRR}/dj . Normalised cross correlation (equation 1) is then calculated between dI_{fl}/di and dI_{DRR}/di and between dI_{fl}/dj and dI_{DRR}/dj . The final value of this measure is the average of these normalised cross correlations.

3.5 Pattern intensity [14]

Pattern intensity [14] operates on a difference image (I_{diff}) as explained in the entropy measure of section 3.2. If the images are registered then, when $s I_{DRR}$ is subtracted from I_{fl} to give I_{diff} , the structure from the vertebrae will vanish and there will be a minimum number of structures or patterns within the difference image (I_{diff}). Pattern intensity is given by,

$$P_{r,\sigma}(s) = \sum_{i,j} \sum_{d^2 \leq r^2} \frac{\sigma^2}{\sigma^2 + (I_{diff}(i, j) - I_{diff}(v, w))^2} \quad (5)$$

$$d^2 = (i - v)^2 + (j - w)^2 \quad (6)$$

where σ is a constant to reduce the effect of noise. The values chosen for the constants were $\sigma = 10$ and $r = 3$ pixels, as used in [14], although r was increased to 5 pixels when coarse images (see section 4.2) were used as this was found to increase the robustness of the measure.

3.6 Gradient difference

Gradient difference uses a difference image as explained in the entropy measure of section 3.2, though this time the difference image is calculated from gradient images (equation 8). It employs the same $1/(1 + x^2)$ form as pattern intensity which should make the measure robust to thin line structures.

$$G(s) = \sum_{i,j} \frac{A_v}{A_v + (I_{diffV}(i, j))^2} + \sum_{i,j} \frac{A_h}{A_h + (I_{diffH}(i, j))^2} \quad (7)$$

$$I_{diffV}(i, j) = \frac{dI_{fl}}{di} - s \frac{dI_{DRR}}{di}, \quad I_{diffH}(i, j) = \frac{dI_{fl}}{dj} - s \frac{dI_{DRR}}{dj} \quad (8)$$

The gradients in equation 8 are described in the gradient correlation section 3.4, and A_v and A_h are constants, which for these experiments were the variance of the respective gradient fluoroscopy image.

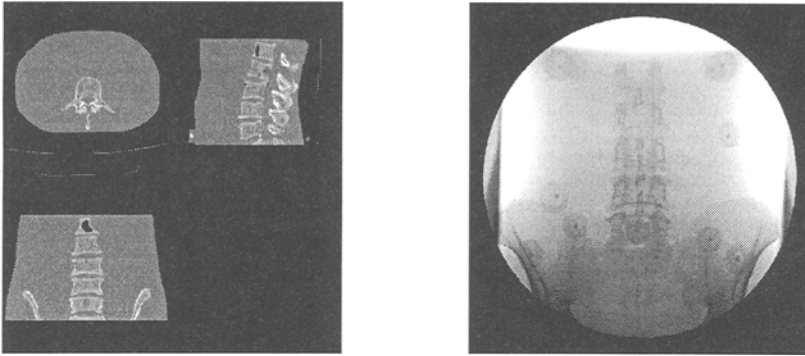


Fig. 3. Reformatted axial, sagittal and coronal slices from a CT image of a lumbar spine embedded in an acrylic matrix (left). Fluoroscopy image of the spine phantom (right).

4 Comparison using a gold standard

An experiment was carried out to investigate how accurately and robustly six similarity measures registered a CT scan of a spine phantom to a fluoroscopy image of the phantom. The phantom consisted of the five lumbar vertebrae and pelvis encased in acrylic which is approximately tissue equivalent at diagnostic x-ray energies. Registrations were compared to a “gold-standard” registration calculated using fiducial markers. Phantom images were used as it is an extremely difficult task to obtain an accurate “gold-standard” registration on clinical images. Many of the problems involved in our application are caused by image features which are not modelled by the phantom image alone. To assess how the similarity measures should perform on clinical images they were simulated by overlaying features segmented from clinical images onto the phantom image. Three clinical images were simulated by adding soft tissue, a stent (figure 4), and both soft tissue and a stent. No soft tissue structures were introduced to the CT volume as these would be removed during the thresholding stage (section 2.1).

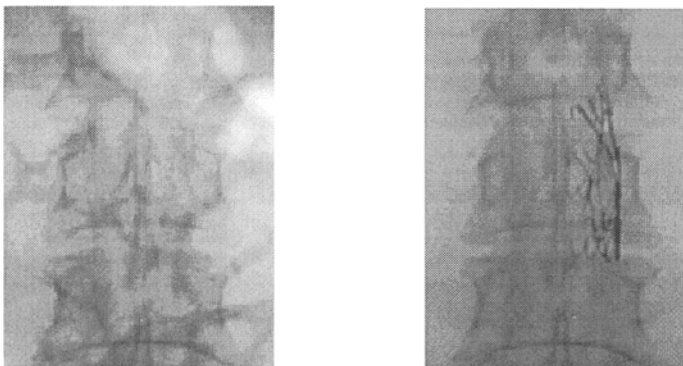


Fig. 4. Enlarged portion of the fluoroscopy image; (left) with soft tissue structures overlaid, (right) with interventional stent overlaid.

4.1 Calculation of the “gold-standard” registration

The “gold-standard” registration has been found using twelve fiducial markers (5mm diameter aluminium ball bearings cast in acrylic resin). A CT scan was acquired using a Philips TOMOSCAN SR 7000, which had voxel sizes 1.094x1.094x1.5mm and image dimensions 320x320x119 voxels. A fluoroscopy image was taken on a Philips multi DIAGNOST 3, which had image size 1024x1024 pixels. These images can be seen in figure 3. A distortion correction phantom and software were used to correct for pincushion distortion in the fluoroscopy image [5] and a cubic phantom of known dimensions was used to correct for any geometric scaling errors in the CT scan [6]. The “gold-standard” registration was calculated and its expected accuracy is given in table 1 (see [10] for a full description of the method used).

Table 1. Expected error in “gold-standard” parameters.

	rotation (degrees)			translation (mm)		
	θ_x	θ_y	θ_z	X	Y	Z
SD	0.09	0.09	0.09	1.34	0.15	0.15

4.2 The Registration Algorithm

The registration was carried out on the L3 vertebra. The search space for the algorithm was the six rigid body degrees of freedom. The parameters associated with perspective projection (c_s , l_s , k_1 and k_2) were held fixed at their “gold standard” values.

The search strategy alters each of the degrees of freedom in turn by $\pm\delta\mathbf{P}$ and calculates a new value of the similarity measure. When all the degrees of freedom have been checked a weighted movement is made in the directions which improve the similarity measure, with the parameter showing the greatest improvement being given the greatest weight. A multi-resolution approach is used. The algorithm starts at a coarse resolution (fluoroscopy image reduced, by blurring, to a 128x128 pixel image) and a large step size ($\delta\mathbf{P}$ set equal to 4 mm or 4°) and ends at a finer resolution (256x256) and small step size ($\delta\mathbf{P} = 0.5$ mm or 0.5°).

4.3 Assessing accuracy and robustness

Each similarity measure was used to register a CT scan of the spine phantom to four different fluoroscopy images of the phantom. The starting estimates for registration were the “gold standard” value plus or minus $\Delta\mathbf{P}$ (see table 2). There are sixty four possible combinations of “gold-standard” plus or minus $\Delta\mathbf{P}$ which are at the corners of a six dimensional hyper-cuboid.

Table 2. The displacement ($\Delta\mathbf{P}$) of the starting positions from “gold-standard” for each of the six rigid body parameters.

	rotation (degrees)			translation (mm)		
	θ_x	θ_y	θ_z	X	Y	Z
$\Delta\mathbf{P}$	3.4	7.6	7.8	50.8	3.6	2.4

5 Results

Our results are presented in four tables, one for each of the different fluoroscopy images. They show the RMS error between “gold-standard” values and final registration positions (excluding failures) and the number of failed registrations. A failure was defined as when the final registration position, in any of the six degrees of freedom was further from “gold-standard” than ΔP as shown in table 2. The average registration took 74 seconds on a Sun Ultra 30 (300 MHz).

Table 3. RMS error values in rigid body parameters for registrations to fluoroscopy image of spine phantom with no added structures.

Similarity measure	rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Cross correlation	0.41	1.09	0.41	7.2	0.55	0.74	0
Entropy	0.32	0.72	0.36	10.4	0.64	0.46	0
Mutual Information	0.79	1.90	0.65	20.2	0.65	1.05	25
Gradient correlation	0.31	0.49	0.19	4.7	0.43	0.34	0
Pattern Intensity	0.28	0.49	0.26	4.1	0.45	0.34	0
Gradient difference	0.27	0.48	0.17	4.0	0.42	0.32	0

As can be seen from table 3 all six similarity measures, except for mutual information, performed well in registering to the fluoroscopy image of the spine phantom.

Table 4 shows how accurately the similarity measures registered when soft tissue structures were present. Soft tissue caused a large failure rate in mutual information (86%) and a fairly large failure rate in entropy (30%). The cross correlation measure failed relatively few times (8%), but the overall errors increased, particularly in X. Pattern intensity, gradient correlation and gradient difference were effected very little by the presence of soft tissue.

Table 4. RMS error values in rigid body parameters for registrations to fluoroscopy image of spine phantom with soft tissue structures overlayed.

Similarity measure	rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Cross correlation	1.64	1.63	0.41	13.2	0.86	1.06	8
Entropy	0.74	1.02	0.40	8.9	1.30	0.73	30
Mutual Information	2.00	2.43	0.57	21.0	0.97	1.42	86
Gradient correlation	0.51	0.49	0.27	6.9	0.39	0.39	0
Pattern Intensity	0.36	0.53	0.24	5.2	0.43	0.36	0
Gradient difference	0.37	0.36	0.21	6.5	0.40	0.39	0

When registrations to the fluoroscopy image with stent overlaid (table 5) were attempted two of the similarity measures, cross correlation and mutual information, failed to register a number of times and showed large errors when they did register (2.5° in θ_y and 22mm in X). Gradient correlation failed 3% of the time. Entropy was largely unaffected, with just a small general decrease in final accuracy. This result is expected, as the histogram used to compute entropy will be largely unaffected by a small number

of pixels with a large difference in intensity. Pattern intensity and gradient difference performed best, with very little change in overall registration error.

Table 5. RMS error values in rigid body parameters for registrations to fluoroscopy image of spine phantom with stent overlayed.

Similarity measure	rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Cross correlation	1.40	2.34	1.11	23.7	0.48	1.46	33
Entropy	0.49	0.86	0.72	14.8	0.69	0.62	0
Mutual Information	0.98	2.37	1.93	21.0	0.55	1.58	48
Gradient correlation	0.23	0.75	0.23	8.5	0.27	0.59	3
Pattern Intensity	0.32	0.39	0.28	4.9	0.42	0.36	0
Gradient difference	0.27	0.49	0.22	3.5	0.43	0.33	0

On registering to the image which contained both soft tissue and a stent (table 6), cross correlation, entropy and mutual information all failed a large number of times and when registrations were deemed valid, they were not accurate. Gradient correlation failed to register 5% of the time, though successful registrations were accurate. Pattern intensity and gradient difference both achieved accurate registrations to this image with no failures. This investigation has only used one pair of CT and fluoroscopy images of a rigid phantom. Different sets of images, in particular different views, may yield different results.

Table 6. RMS error values in rigid body parameters for registrations to fluoroscopy image of spine phantom with soft tissue and stent overlayed.

Similarity measure	rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Cross correlation	2.33	2.25	0.89	27.2	0.29	1.39	45
Entropy	1.36	1.69	1.87	16.9	1.52	1.02	53
Mutual Information	2.67	3.94	2.94	30.3	1.34	5.07	95
Gradient correlation	0.48	0.49	0.42	12.1	0.25	0.50	5
Pattern Intensity	0.31	0.41	0.43	10.7	0.36	0.47	0
Gradient difference	0.28	0.45	0.26	6.7	0.37	0.47	0

6 Conclusions

This paper has compared the accuracy and robustness of six 2D-3D registration algorithms which used different intensity based similarity measures. Final registration positions were compared to an accurate “gold standard” registration found using fiducial markers. Clinical images were simulated by overlaying structures segmented from clinical fluoroscopy images onto a fluoroscopy image of a spine phantom. We have shown that the introduction of soft tissue structures and interventional instruments into the phantom image can have a large effect on the performance of some similarity measures previously applied to 2D-3D image registration. Correlation measures can be effected by thin line structures, such as an interventional stent, which introduces pixels which have a large difference in intensity. Entropy type measures are insensitive to thin line structures, but fail when soft tissue structures create slowly varying changes in background intensity. For a measure to work well with medical images it must be able to

register accurately when soft tissue structures and thin line structures are present as differences between the images. Two of the measure described in this paper have achieved this: pattern intensity and gradient difference.

Acknowledgements

We would like to thank Philips Medical Systems, EasyVision Advanced Development for funding this work and providing software, particular thanks go to Alexander van Eeuwijk for his help with distortion correction and Frans Gerritsen for his encouragements and constructive criticisms. We would also like to thank Professor Andy Adam and Dr. Mark Cowling for their clinical input and to thank the radiographers at Guy's Hospital.

References

1. F. Betting and J. Feldmar. 3D-2D projective registration of anatomical surfaces with their projections. In *Proc. Information processing in medical imaging*, pages 275–286, 1995.
2. T.M. Buzug, J. Weese, C. Fassnacht, and C. Lorenz. Image registration: Convex weighting functions for histogram-based similarity measures. In J. Troccaz, E. Grimson, and R. Mösges, editors, *Proc. CVRMed/MRCAS*, pages 203–212. Berlin, Germany: Springer-Verlag, 1997.
3. L.M. Gottesfeld Brown and T.E. Boulton. Registration of planar film radiographs with computed tomography. In *Proc. MMBIA*, pages 42–51, 1996.
4. A. Guézic, P. Kazanzides, B. Williamson, R.H. Taylor, and D. Lord. Anatomy based registration of CT-scan and X-ray fluoroscopy data for intra-operative guidance of a surgical robot. In *Proc. SPIE Medical Imaging*, 1998.
5. P. Haaker, E. Klotz, R. Koppe, and R. Linde. Real-time distortion correction of digital x-ray II/TV-systems: an application example for digital flashing tomosynthesis (DFTS). *International Journal of cardiac imaging*, 6:39–45, 1990.
6. D. L. G. Hill, C. R. Maurer, Jr., C. Studholme, J. M. Fitzpatrick, and D. J. Hawkes. Correcting scaling errors in tomographic images using a nine degree of freedom registration algorithm. *J. Comput. Assist. Tomogr.*, 22:317–323, 1998.
7. S. Lavallée, R. Szeliski, and L. Brunie. Matching 3-D smooth surfaces with their 2-D projections using 3-D distance maps. In *Proc. SPIE Geometric methods in computer vision.*, volume 1570, pages 322–336, 1991.
8. L. Lemieux, R. Jagoe, D.R. Fish, N.D. Kitchen, and D.G.T. Thomas. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Med. Phys.*, 21(11):1749–1760, 1994.
9. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Seutens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
10. G.P. Penney, J. Weese, J.L. Little, Desmedt P., D.L.G. Hill, and D.J. Hawkes. A comparison of similarity measures for use in 2D-3D medical image registration. *IEEE Transactions on Medical Imaging*, 1998 (Provisionally accepted).
11. C. Studholme, D.L.G. Hill, and D.J. Hawkes. Automated 3D registration of MR and CT images of the head. *Medical Image Analysis*, 1(2):163–175, 1996.
12. C. Studholme, D.L.G. Hill, and D.J. Hawkes. Automated 3D registration of MR and PET brain images by multi-resolution optimisation of voxel similarity measures. *Medical Physics*, 24:25–35, 1997.
13. P. Viola and W. M. Wells. Alignment by maximization of mutual information. In *Proc. 5th International Conference on Computer Vision (ICCV'95)*, pages 16–23, 1995.
14. J. Weese, T.M. Buzug, C. Lorenz, and C. Fassnacht. An approach to 2D/3D registration of a vertebra in 2D x-ray fluoroscopies with 3D CT images. In *Proc. CVRMed/MRCAS*, pages 119–128, 1997.