

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

Graeme P. Penney, Jürgen Weese, John A. Little, Paul Desmedt, Derek L. G. Hill, and David J. Hawkes*

Abstract—A comparison of six similarity measures for use in intensity-based two-dimensional–three-dimensional (2-D–3-D) 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 computed tomography (CT) scan of a spine phantom to a fluoroscopy image of the 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 modeled 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 2-D–3-D 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 were pattern intensity and gradient difference. Their registration accuracy, for all the rigid-body parameters except for the source to film translation, was within a root-mean-square (rms) error of 0.54 mm or degrees to the “gold-standard” values. No failures occurred while registering using these measures.

Index Terms—Gold-standard, image-guided interventions, image registration, similarity measures, two-dimensional–three-dimensional (2-D–3-D).

I. INTRODUCTION

THREE-dimensional (3-D) imaging modalities such as computed tomography (CT) and magnetic resonance (MR) imaging are widely used in clinical diagnosis and treatment planning, but at present their use as interventional imaging modalities has been limited. Common modalities for guiding interventions are ultrasound or X-ray fluoroscopy. These modalities are “real-time” but only two-dimensional (2-

D), so they lack the spatial information contained in CT and MR images. There are also a number of important anatomical features which are not visualized 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. These can be split into two main categories: feature-based methods and intensity-based methods. Feature-based methods [1], [7], [12] register on salient features that have been segmented from each image. They use a reduced amount of data which makes the algorithms fast, once the segmentation has been undertaken. Unfortunately, the segmentation stage is very hard to achieve automatically and errors in segmentation can lead to errors in the final registration. Intensity-based methods [6], [13], [22] compare the voxel and pixel values directly using a measure based on image statistics. These are typically slower, but require little or no segmentation.

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 were simulated by overlaying structures segmented from clinical fluoroscopy images on to the fluoroscopy image of the spine phantom.

II. COMPARING FLUOROSCOPIC IMAGES AND DRR'S

A. Production of DRR's

Digitally reconstructed radiographs (DRR's) are produced by casting rays through a CT volume (Fig. 1). Each of these rays will go through a number of voxels. If the Hounsfield numbers of these voxels are integrated along the ray and projected onto an imaging plane the resultant image should resemble a radiograph.

B. Differences Between Fluoroscopy Images and DRR's

There are two main types of difference between fluoroscopy images and DRR's, those which are caused by changes in the imaged object and those due to differences in image formation. Examples of the first set of differences are the presence of interventional instruments and soft-tissue deformation. These differences would still exist even if DRR's could be produced

Manuscript received December 12, 1997; revised July 28, 1998. This work was supported by Philips Medical Systems, EasyVision Advanced Development, Best, the Netherlands. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was A. Manduca. Asterisk indicates corresponding author.

G. P. Penney, J. A. Little, and D. L. G. Hill are with the Division of Radiological Sciences, UMDS, Guy's and St. Thomas' Hospitals, London SE1 9RT, U.K.

J. Weese is with Philips Research Hamburg, Röntgenstraße 24-26, 22335 Hamburg, Germany.

P. Desmedt is with EasyVision Advanced Development, Philips Medical Systems, Veenpluis 4-6, 5680 DA Best, the Netherlands.

*D. J. Hawkes is with the Division of Radiological Sciences, UMDS, Guy's and St. Thomas' Hospitals, London SE1 9RT, U.K. (e-mail: d.hawkes@ipg.umds.ac.uk).

Publisher Item Identifier S 0278-0062(98)08048-3.

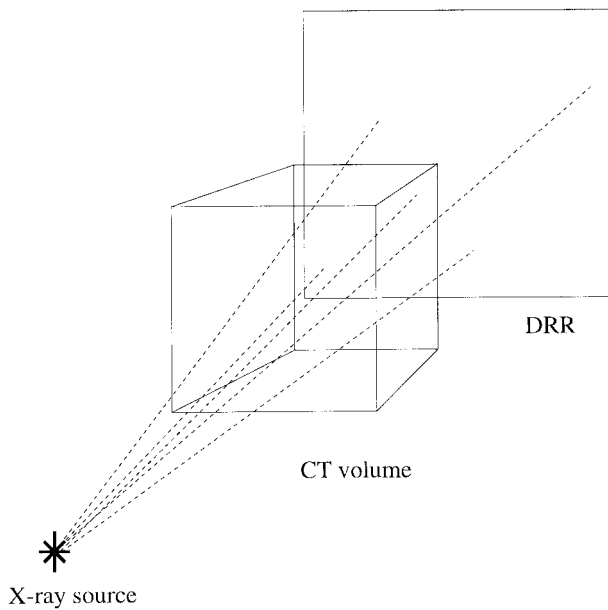


Fig. 1. How a DRR is created by casting rays through the CT volume.

to perfectly resemble a fluoroscopy image. The second set represent differences caused by the inherent dissimilarity between CT and fluoroscopic images. We have corrected for some of these differences, such as pincushion distortion correction and blurring the fluoroscopy to more accurately represent the CT resolution. No attempt has been made to correct for other differences such as the X-ray tube anode heel effect and effects due to differences in X-ray energies. A good similarity measure for 2-D-3-D image registration should be able to achieve accurate registrations despite the presence of these differences between the images.

1) Differences Due to Changes in Imaged Object:

a) Overlying and underlying structures: We are assuming that the transformation between CT and fluoroscopy can be described by a perspective projection of rigid 3-D motion. Soft-tissue structures, however, can deform between the preoperative CT image and interventional fluoroscopy. Because of this, we threshold the CT volume so no soft-tissue structures are projected. Therefore, the registration is carried out between a DRR which contains only bony detail and the fluoroscopy image which contains both bony detail and soft-tissue structures, as can be seen in Fig. 2.

Soft-tissue structures usually have a lower attenuation coefficient than bony structures and therefore play a relatively small role in the intensity distribution in the final image. They are, however, present in the majority of pixels and so can form a large difference between the images. Soft-tissue structures tend to be larger in size and possess low spatial frequencies compared to bony structures. This characteristic can be used to filter out the effect of soft-tissue structures on the measure of alignment.

The threshold value was chosen by visual observation of the DRR's. A good threshold value should remove soft tissue but leave as much bony detail as possible. The DRR's in this paper were produced using a threshold value of 400 Hounsfield units. Initial results show that changes in the

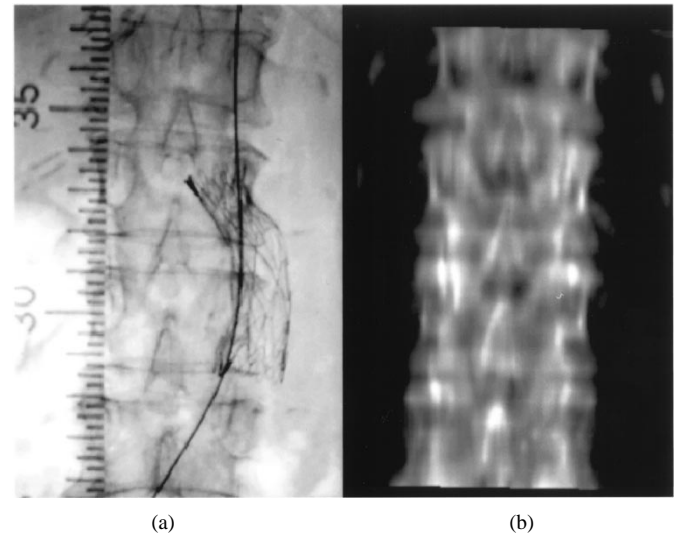


Fig. 2. (a) Clinical fluoroscopy image and (b) DRR's at registration.

threshold (by ± 50 Hounsfield units) has negligible effect on the final registrations.

b) Interventional instruments in the field-of-view: Interventional instruments are often made from materials with a high attenuation coefficient and therefore create large differences between the images. The difference is usually so large that these pixels contain very little useful information to aid registration. Similarity measures should ideally, therefore, be insensitive to these highly altered pixels. Fig. 2 shows three of these features: a ruler which is used to define positions during the procedure, a guidewire, and a deployed aortic stent.

c) Spinal deformation: The relative positions of the vertebrae may be different in the CT and fluoroscopy images. This may effect registration. One method to reduce the effect of deformation is to register on a single rigid structure, such as one vertebra [22]. Another method, not addressed further in this paper, is to correct for the deformation by warping one of the images [14].

2) Differences in Image Formation:

a) Different X-ray energies between the modalities: CT and fluoroscopy machines work at different effective energies; CT at about 80 keV, while fluoroscopy is typically between 50–60 keV. The linear attenuation coefficient of a material is dependent on X-ray energy. As a result, the attenuation coefficients measured by a CT scanner are different from the attenuation of a beam of X-rays used in fluoroscopy. At CT energies, the ratio of bone to soft-tissue attenuation is much lower than at radiographic energies. Therefore, if a DRR is produced from the whole CT volume, the image contrast from soft tissue will be comparable with the image contrast from bone, reducing the clarity of bony detail. Our approach avoids this problem by thresholding the CT volume to remove soft-tissue structures.

b) Heel effect and nonuniformity of the image intensifier response: The heel effect is caused by the anode emitting different intensities of X-rays depending on the angle between the incident electron, emitted X-ray, and anode target angle. The result is a gradual change in average intensity values in the

radiograph from the top to the base of the image, as shown by the darker top of the fluoroscopy image in Fig. 2. The image intensifier also has a nonuniform response, with images slightly brighter in the center of the field-of-view and darker at the periphery. These shading effects will not be present in a DRR, although they could be corrected for by calibrating the fluoroscopy set. The effect is similar to that of underlying and overlying tissue, but it is of a lower magnitude and more predictable.

c) *Different resolutions:* Fluoroscopy images and CT images typically have different resolutions. As a result, some structures in the DRR will appear more blurred (especially normal to the CT slice plane) than in the fluoroscopy image. By blurring the fluoroscopy image so its resolution resembles that of the DRR, the difference in resolutions can be reduced. To achieve this accurately is a very difficult task, as it not only depends on the resolution of both the CT and fluoroscopy images, but, as a perspective projection is involved, it also depends on the position of individual voxels within the CT scan relative to the pose of the fluoroscopy set. As a simple approximation, we smoothed the fluoroscopy image using a circular mask (diameter: five pixels). The mask was passed over the image and the central pixel set equal to the average value of all the pixels under the mask. This made the fluoroscopy resolution approximately consistent with the CT slice spacing.

d) *Truncation:* Truncation occurs when rays cut through the top or bottom of the CT scanned volume. Such rays are incomplete and so should not be compared to the fluoroscopy image.

e) *Distortion in the fluoroscopy image:* Pincushion distortion is caused by nonlinearities in the electron optics and the curved input face of the image intensifier. It can be corrected for using a suitable phantom and software [8], and this has been carried out for the fluoroscopy images used in this paper.

f) *Post processing of fluoroscopy images:* Clinical fluoroscopy images are routinely processed to increase contrast and suppress noise. These processes may alter the relationship between fluoroscopic and DRR intensity. In this investigation, we have attempted to simulate registrations to clinical images and, therefore, standard clinical settings were used to produce the fluoroscopy images.

III. 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 $\mathbf{P} = (X, Y, Z, \theta_x, \theta_y, \theta_z, c_s, l_s, k_1, \text{ and } k_2)$, see Fig. 3. The first six parameters are rigid-body translations (X, Y, Z) and rotations ($\theta_x, \theta_y, \theta_z$) of the CT volume. The last four parameters are concerned with the perspective projection geometry of the fluoroscopy set. The coordinates c_s and l_s represent the position 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.

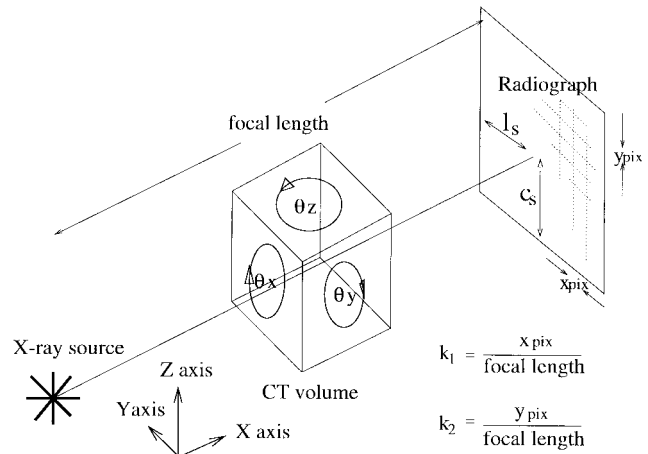


Fig. 3. Diagram showing the degrees of freedom in rigid-body perspective projection. Six rigid-body parameters, three translational X , Y , and Z , and three rotational θ_x , θ_y , and θ_z . Four parameters c_s , l_s , k_1 , and k_2 are associated with perspective projection: c_s and l_s are the coordinates 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.

A. Normalized Cross Correlation

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

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

In cross correlation, the contribution from a given pixel is strongly dependent on the intensity of the pixel, and so, a few large differences in intensity (such as may be caused by an interventional guidewire) can have a substantial effect on the similarity measure. This measure has been previously used in [13].

B. Entropy of the Difference Image

The entropy measure (H) described here operates on a single difference image (I_{dif}) which is created by subtracting the DRR from the fluoroscopy image using a suitable intensity scaling factor s [see (2)]

$$I_{dif} = I_{fl} - sI_{DRR} \quad (2)$$

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

A histogram is formed from the difference image and $p(x)$ [see (3)] denotes the probability of obtaining the pixel value x in I_{dif} . Because histograms are used, the same weight is given to each pixel regardless of its intensity. This should make the measure robust to differences between the images, which result in large differences in the intensity of a small number of pixels. This measure has been used previously in 2-D-2-D registration to correct for motion between images of a digital subtraction angiography (DSA) sequence [3].

To calculate I_{dif} , the optimal scaling value s must be determined. The method used here starts with s equal to zero and increases it in uniform steps until a minimum value of H is found. The size of the steps used in this paper were equal to the ratio of the intensity ranges in the two images divided by 100.

C. Mutual Information

Mutual information (S) or relative entropy is given by

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

where $p(x)$ and $p(y)$ are the probability distributions in individual images and $p(x,y)$ is the joint probability distribution. It has been found to be very effective in 3-D-3-D multimodality image registration between modalities such as MR and positron emission tomography (PET), or MR and CT [15], [19]–[21]. Mutual information does not assume a linear relationship between the pixel values of the two images, but instead assumes that the co-occurrence of the most probable values in the two images is maximized at registration. This is a useful property for 3-D-3-D multimodality image registration. In the case of 2-D-3-D image registration, the DRR has been created so that it resembles a fluoroscopy image and so, at registration, the relationship should be close to linear.

D. Gradient Correlation

Gradient-based measures initially transform I_{fl} and I_{DRR} by differentiation. Horizontal and vertical Sobel templates were used to create gradient images dI_{fl}/di and dI_{fl}/dj representing the derivative of the fluoroscopy intensity in the two orthogonal axes of the image. Normalized cross correlation (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 normalized cross correlations.

Gradient measures have the advantage that they filter out low spatial frequency differences between the images, such as those caused by soft tissue. They also concentrate the contributions to the similarity measure on edge information, which intuitively appears sensible. Because this measure uses cross correlation, it is expected to be effected by a few large differences in intensity. In particular, the presence of interventional instruments in the fluoroscopy image, especially after they have been enhanced by the Sobel templates, may effect the performance of this measure. This type of measure has been used previously in [6] and [13].

E. Pattern Intensity

Pattern intensity [22] operates on a difference image (I_{dif}) as explained in the entropy measure of Section III-B. If the images are registered then, when sI_{DRR} is subtracted from I_{fl} to give I_{dif} , the structure from the vertebrae will vanish and there will be a minimum number of structures or patterns within the difference image (I_{dif}). Pattern intensity considers a pixel to belong to a structure if it has a significantly different intensity value from its neighboring pixels. All the pixels within a radius r are considered in this calculation. The form

of (5) is such that as the number of structures tend to zero, the measure tends to a maximum value and as the number of structures increase the measure asymptotically tends to zero. A constant (σ) is used to weight the function, so that small deviations in intensity (such as those caused by noise) result in the measure remaining near to its maximum value. Due to the asymptotic nature of the measure, large differences in intensity have the same effect on the measure regardless of their magnitude. This makes the measure robust to large differences in pixel intensity

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

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

The regional nature of pattern intensity is able to reduce the effect of differences between the images which are on a larger scale than r . Differences with a low spatial frequency, such as soft-tissue structures are filtered out, leaving the high-frequency bony detail on which the registration is based. The values chosen for the constants were $\sigma = 10$ and $r = 3$ pixels, as used in [22], though r was increased to five pixels when coarse images (see Section IV-C) were used as this was found to increase the robustness of the measure.

F. Gradient Difference

As mentioned in Section III-D gradient-based measures are potentially insensitive to soft-tissue deformation, but gradient correlation is expected to be sensitive to thin line structures. Because of this we propose another gradient-based measure, gradient difference. It uses a difference image as explained in the entropy measure of Section III-B, though this time the difference image is calculated from gradient images (8). It employs the same $1/(1+x^2)$ form as pattern intensity which should make the measure robust to thin line structures

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

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

The gradients in (8) are described in Section III-D, and A_v and A_h are constants, which for these experiments were the variance of the respective gradient fluoroscopy image. The use of two images allows the similarity measure to compare the direction of gradients, as well as their magnitude.

IV. 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

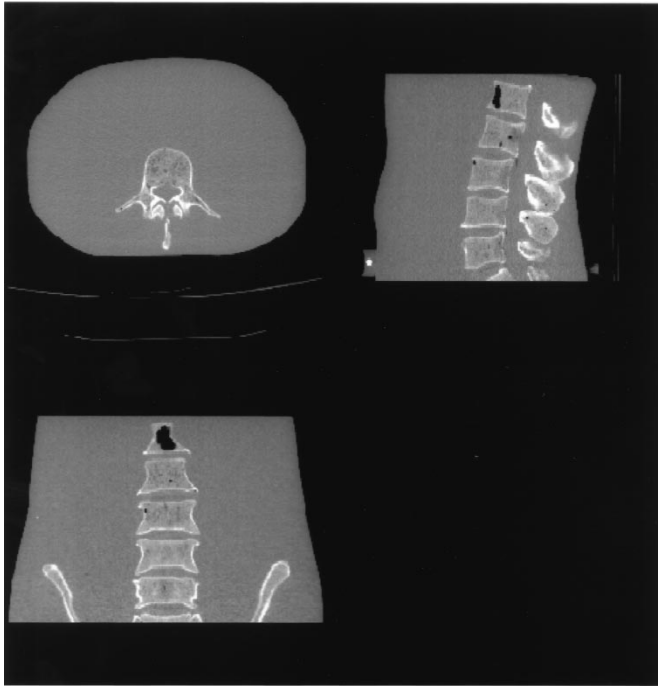


Fig. 4. Reformatted axial (top left), sagittal (top right) and coronal (bottom left) slices from a CT image of a lumbar spine embedded in an acrylic matrix. Note one of the fiducial markers is visible in the sagittal slice.

with 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 modeled by the phantom image alone. To assess how the similarity measures should perform on clinical images, we combine the phantom image with image features delineated from clinical images.

A. Calculation of the “Gold-Standard” Registration

The “gold-standard” registration has been found using 12 fiducial markers (5-mm-diameter aluminum ball bearings cast in acrylic resin). A CT scan was acquired using a Philips TOMOSCAN SR 7000, which had voxel sizes $1.094 \times 1.094 \times 1.5$ mm and image dimensions $320 \times 320 \times 119$ voxels. A fluoroscopy image was taken on a Philips multi DIAGNOST 3, which had image size 1024×1024 pixels. These images can be seen in Figs. 4 and 5. A distortion-correction phantom and software were used to correct for pincushion distortion in the fluoroscopy image [8], and a cubic phantom of known dimensions was used to correct for any geometric scaling errors in the CT scan [11].

An intensity weighted method was used to find the center of the fiducial markers in the CT scan. This method is capable of finding the centroid of a marker to sub-voxel accuracy [2], [4]. It was not possible to use this method in the fluoroscopy image due to underlying and overlying intensity fluctuations, and so, the following method was used. A pixel within the marker was manually picked (e.g., point p in Fig. 6) and lines were drawn out from this pixel at 45° intervals (e.g., line AB). Each of these lines was taken in turn and a graph plotted of intensity

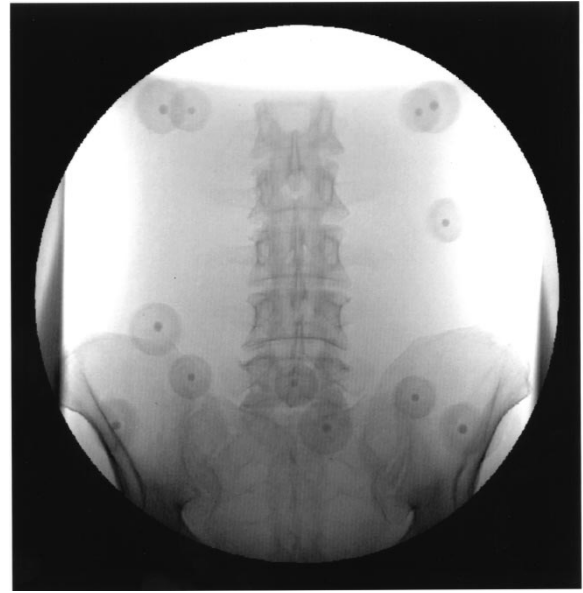


Fig. 5. Fluoroscopy image of the spine phantom. Note the attenuation from the tissue equivalent acrylic in which the spine is embedded and the spherical fiducial markers.

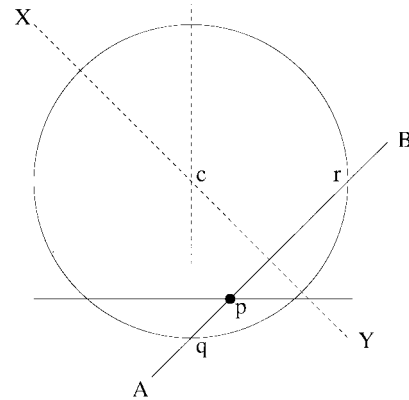


Fig. 6. Diagram to show how the center of the fiducial markers were found in the fluoroscopy image. Point p manually picked. Lines drawn out at 45° intervals (e.g., AB). Intercepts between AB and edge of marker found q and r . Normal to AB drawn equidistant to q and r ; line XY . Line XY should go through center c . Triangulation was used to obtain an average value for the position of c .

as a function of distance. This function was differentiated to calculate the positions of steepest upslope and downslope. These positions were refined by calculating a second-order derivative to find the positions of the points of inflection. These represent points q and r , which are near to the edge of the marker. A normal (line XY) was then drawn through the line which was equidistant from these points. This normal should go through c , the center of the marker. Normals are found for each of the four lines at 45° to each other and their intersection points found. The center of the marker is taken to be the average of these intersection points.

Once the position of these markers had been found, a “gold-standard” registration matrix was calculated using a singular value decomposition method [9], [17], [18]. The matrix was decomposed into ten perspective projection parameters P using a method described by Ganapathy [5].

TABLE I
EXPECTED STANDARD DEVIATION OF “GOLD-STANDARD” PARAMETERS
CALCULATED USING 990 SETS OF PARAMETERS WHICH, WHEN USED
TO PROJECT 3-D POINTS TO TWO DIMENSIONS, GIVE AN AVERAGE
RMS ERROR OF 1.2 PIXELS IN THE POSITION OF THE 2-D POINTS

	rigid body 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

Observed target registration error (TRE_o), as described in Maurer *et al.* [16], was calculated in the following way. Two fiducial markers were removed from the set of 12 so they could be used as target points. A matrix was calculated using the remaining markers and used to project the 3-D positions of the two target points into two dimensions. The distance between these projected 2-D positions and the 2-D target point positions found from the fluoroscopy image were calculated. A distance was calculated for each possible set of ten points from 12, resulting in 66 distances. The rms value of these distances was 1.2 pixels, which corresponds to an error of 0.44 mm at the imaging plane (the input to the image intensifier).

To relate the 2-D error TRE_o to the rigid-body parameters ($\theta_x, \theta_y, \theta_z, X, Y, Z$), a numerical simulation was carried out. The “gold-standard” 3-D points were projected into two dimensions to form a “perfect” set of 3-D and 2-D points. Two target points were removed from the set of 12 and a random Gaussian distributed error (standard deviation ε pixels) was added to the 2-D positions of the remaining ten points to simulate true target registration error (TRE_t) [16]. A new perspective projection matrix was calculated using an iterative scheme which altered the rigid-body parameters to minimize the rms distance between the 3-D points projected into two dimensions, and the 2-D points with random error added. The perspective parameters (k_1, k_2, c_s, l_s) were kept fixed at their “gold-standard” values. The rms error between the target points projected into two dimensions using the new matrix and their “perfect” 2-D positions was calculated. This rms error represents the TRE_t of the new matrix and hence also the TRE_t of the new set of rigid-body parameters. This process was repeated 15 times for each possible set of ten points from 12, resulting in 990 values of TRE_t from which an average was calculated. By experiment it was found that $\varepsilon = 1.35$ pixels produced an average TRE_t of 1.2 pixels, which is equal to the TRE_o of the “gold-standard” matrix. This experiment resulted in 990 sets of rigid-body parameters which, when used to project points from three dimensions to two dimensions, have an average TRE_t of 1.2 pixels. The standard deviations of these 990 sets of parameters are shown in Table I. They give a good indication of the expected error in each of the “gold-standard” rigid-body parameters.

The TRE_o is made up from TRE_t and a target localization error (TLE). If TLE is the same in each image space and uncorrelated, then $(TRE_o)^2 \approx (TRE_t)^2 + 2(TLE)^2$ [16]. In the above, we have set $TLE = 0$ and have, therefore, calculated an upper bound for TRE_t .

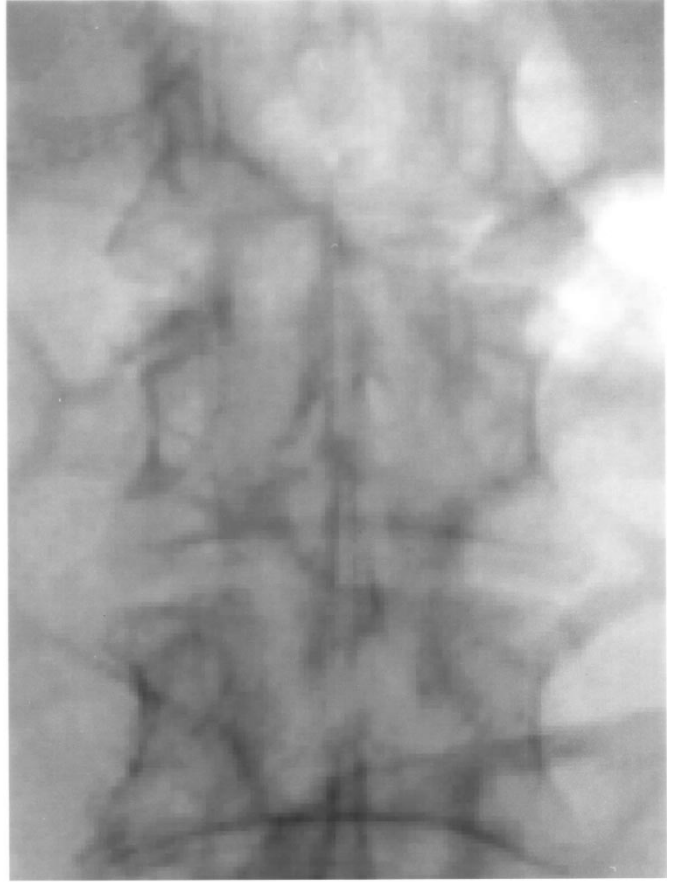


Fig. 7. Enlarged portion of the fluoroscopy image with soft-tissue structures overlaid.

B. Simulating Clinical Images

Clinical fluoroscopy images were simulated by overlaying features segmented from clinical images onto the phantom image. Three clinical images were simulated by adding soft-tissue, a stent, 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 II-B.1).

1) *Overlying Soft Tissue* (Fig. 7): a region that contained a variety of soft-tissue structures was segmented from a clinical fluoroscopy image. The major soft-tissue structure visible was the small bowel, which produced some large variations in images intensity due to the presence of gas. The segmented soft tissue was overlaid on top of the original spine phantom fluoroscopy image by adding the pixel intensities of both images together, using a weighting factor ($w = 0.7$), $I_{combined} = I_{phantom} + wI_{soft\ tissue}$. The weighting factor was necessary to achieve a visually realistic ratio between the contributions from bone and soft tissue. It was chosen by visual inspection; comparing the simulated image with clinical images.

2) *An Interventional Stent* (Fig. 8): an image of a stent was segmented from a clinical image and then overlaid ($w = 1.0$) on top of the spine phantom fluoroscopy image. The objective was to create an overlap with a large number of sharp edges, which is characteristic of a number of interventional instruments e.g., guidewires or needles.

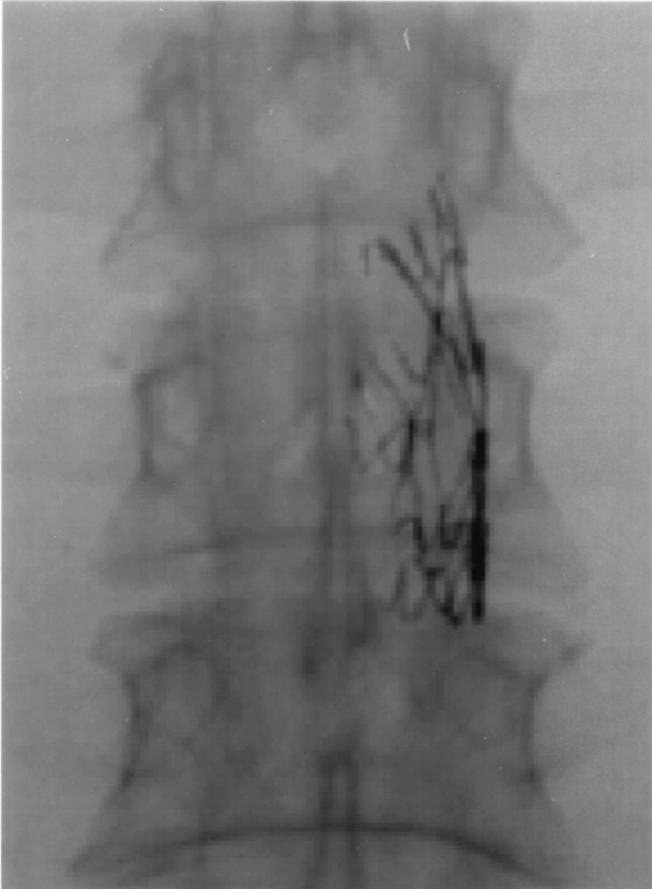


Fig. 8. Enlarged portion of the fluoroscopy image, with interventional stent overlaid.

The third image was a combination of the above two images with both soft tissue and a stent overlaid.

C. The Registration Algorithm

We have found that when registering clinical images it is advantageous to register on a specific region-of-interest (ROI) within the fluoroscopy image, rather than the whole image. Instead of casting rays through the CT volume to every pixel in the DRR, they are only projected to pixels within the ROI. This has a number of advantages; first, it greatly speeds up the registration, as DRR's are only produced, and similarity measures only need to operate, within this region. Second, as the registration is only carried out in the ROI, rather than the whole image, it should be more accurate within this region. Third, if the ROI contains a single rigid bony structure, then deformation is not an issue. In this paper, a ROI was manually drawn around the L3 vertebra and all the registrations were carried out in this region.

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. Some similarity measures (entropy, pattern intensity, and gradient difference) necessarily calculate the intensity scaling between the two images (s). This introduces another parameter into the optimization.

The search strategy uses the current parameters (\mathbf{P}) to form a DRR, which is then compared to the fluoroscopy image using a similarity measure to obtain a value (V_0). The rigid-body parameters of \mathbf{P} are then altered one at a time by $\pm\omega\mathbf{P}\delta\mathbf{P}$, a new DRR created, and a new value of the similarity measure calculated (V_1, V_2, \dots, V_{12}). The weighting factor $\omega\mathbf{P}$ equals unity for all the rigid-body parameters except for translations in the X direction. Because movements along X have a much smaller effect on the DRR than movements in the other degrees-of-freedom, ω_X was set to four. The values V_0, V_1, \dots, V_{12} indicate which parameters currently have a local optimal value and which parameters could be changed to improve the value of the similarity measure. A movement was made in these latter parameters which was weighted by the size of the improvement in the similarity measure. This was repeated until all six parameters were at an optimum value ($V_0 < V_i, i = 1 \rightarrow 12$). At this point, $\delta\mathbf{P}$ was reduced and the process restarted.

A multiresolution approach is used. The fluoroscopy image was reduced, by blurring, to 128×128 pixel images, and $\delta\mathbf{P}$ set equal to 4 mm or 4° . When an optimum value was found in all six parameters, a finer resolution was used (256×256 pixel images) and $\delta\mathbf{P}$ was halved to 2 mm or 2° . When subsequent optimum values were found in all six parameters, $\delta\mathbf{P}$ was halved and the resolution is kept at 256×256 pixel images. This was repeated for another two steps until $\delta\mathbf{P}$ equalled 0.5 mm or 0.5° and an optimum value had been found in all six parameters. This multiresolution approach reduced processing time and helped avoid local minima.

The optimization scheme can effect the performance of a similarity measure. If a measure behaves smoothly within the search space, then it is unlikely to be penalized by the choice of optimization scheme, while a measure which produces a large number of local minima will require a robust optimization scheme to reach registration. A similar optimization scheme to the one described in this paper has been found to be robust for 3-D-3-D registration [19], [20]. We, therefore, believe this type of optimization scheme to be a good choice for this study. Further reductions in computing time were achieved by reducing the size of the CT volume to a region which just contains the spine. The amount of data used in the registration was not effected as DRR's were produced using only CT voxels containing bone.

D. 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 (described in Section IV-B). The starting estimates for registration were the "gold-standard" value $\pm\Delta\mathbf{P}$ (see Table II). There are 64 possible combinations of "gold-standard" $\pm\Delta\mathbf{P}$ which are at the corners of a six-dimensional hypercuboid. These values were chosen to simulate how the algorithm operates when registering clinical images which have no "gold-standard" registration. In clinical practice, a starting estimate could be found using knowledge of patient position and imaging geometry, and by picking corresponding anatomical points in the images. It is much easier to obtain

an accurate estimate of some parameters, such as the in-plane translations (Y and Z) compared to other parameters, especially the out-of-plane translation (X). The values of ΔP are two times the estimated accuracy of a 2-D-3-D registration carried out by manually picking corresponding anatomical points in the two images. The following experiment was carried out to estimate how accurately this type of algorithm could register. Six anatomical points were picked on the spine and projected into two dimensions using the “gold-standard” matrix. A random Gaussian distributed error was added to the position of these points to simulate observer error. The size of the 3-D error (3-mm standard deviation) was based on the largest rms error found in a reproducibility experiment to find landmarks in a CT volume described in Hill *et al.* [10]. A similar experiment for 2-D errors was not found in the literature and so the 2-D error was roughly estimated to have a standard deviation of six pixels. An iterative scheme was used to alter the rigid-body perspective parameters to minimize the rms distance between the new 3-D points projected into two dimensions and the new 2-D points. The perspective parameters (k_1, k_2, c_s, l_s) were kept fixed at their “gold-standard” values. This was carried out 150 times and a mean and standard deviation of the final rigid-body parameters calculated. The values of ΔP were set to two times this standard deviation. It is very unlikely for all the starting positions to be out by two times the standard deviation (0.000 002% chance) and so this represents a rigorous test for the algorithm.

For the method to work, it is important that at the start of registration, some corresponding features are at least partially overlapping. One method of achieving this is to pick a single corresponding point in the DRR and in the fluoroscopy image and overlay these points. The starting positions for registration were modified by simulating this single-point-picking exercise. The “gold-standard” matrix was used to calculate a corresponding 2-D and 3-D point at the top of the L3 spinous process. Before registration begins, this 3-D point is cast into two dimensions using a matrix calculated from the initial parameters, a random Gaussian distributed error (standard deviation six pixels) is added to the position of this point and to the “gold-standard” 2-D point (to simulate observer error) and they are then overlaid by altering the in-plane translational parameters (Y and Z).

V. 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 II. The average registration took 74 s on a Sun Ultra 30 (300 MHz).

As can be seen from Table III, all six similarity measures performed well in registering to the fluoroscopy image of the spine phantom. Mutual information can be seen to be the worst of the measures.

TABLE II
THE DISPLACEMENT (ΔP) OF THE STARTING POSITIONS FROM “GOLD-STANDARD” FOR EACH OF THE SIX RIGID-BODY PARAMETERS

	Displacement in rigid body parameters					
	rotation (degrees)			translation (mm)		
	θ_x	θ_y	θ_z	X	Y	Z
ΔP	3.4	7.6	7.8	50.8	3.6	2.4

TABLE III
THE RMS ERROR VALUES IN RIGID-BODY PARAMETERS FOR REGISTRATIONS TO FLUOROSCOPY IMAGE OF SPINE PHANTOM WITH NO ADDED STRUCTURES

Similarity measure	RMS error in rigid body parameters rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Starting positions	3.40	7.60	7.80	50.8	1.03	0.52	-
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

TABLE IV
THE RMS ERROR VALUES IN RIGID-BODY PARAMETERS FOR REGISTRATIONS TO FLUOROSCOPY IMAGE OF SPINE PHANTOM WITH SOFT-TISSUE STRUCTURES OVERLAYED

Similarity measure	RMS error in rigid body parameters rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Starting positions	3.40	7.60	7.80	50.8	1.03	0.52	-
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

Table IV 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.

When registrations to the fluoroscopy image with stent overlaid (Table V) were attempted, two of the similarity measures—cross correlation and mutual information—failed to register a number of times and showed large errors when

TABLE V

THE RMS ERROR VALUES IN RIGID-BODY PARAMETERS FOR REGISTRATIONS TO FLUOROSCOPY IMAGE OF SPINE PHANTOM WITH STENT OVERLAYED

Similarity measure	RMS error in rigid body parameters rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Starting positions	3.40	7.60	7.80	50.8	1.03	0.52	-
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

TABLE VI

THE RMS ERROR VALUES IN RIGID-BODY PARAMETERS FOR REGISTRATION TO FLUOROSCOPY IMAGE OF SPINE PHANTOM WITH SOFT TISSUE AND STENT OVERLAYED

Similarity measure	RMS error in rigid body parameters rotation (degrees)			translation (mm)			No. fail, (%)
	θ_x	θ_y	θ_z	X	Y	Z	
Starting positions	3.40	7.60	7.80	50.8	1.03	0.52	-
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

they did register (greater than 2.34° in θ_y and 21 mm 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.

On registering to the image which contained both soft tissue and a stent (Table VI), cross correlation, entropy, and mutual information all failed a large number of times and when registrations were deemed valid, they were not accurate. The standard deviations of the successful final registration positions using entropy and cross correlation were also calculated. It is interesting that these standard deviation values are fairly small (all below 1° or mm except for the X translation), therefore, these measures are precise despite being inaccurate. This shows the value of basing the comparison on an accurate independent "gold-standard" rather than using the precision of results. 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.

VI. DISCUSSION

The least accurate of the six similarity measures was mutual information. Mutual information has shown itself to be very useful in 3-D multimodality registration but, with 3-D images, there are usually a large number of voxels to populate the 2-D histogram. In the case of 2-D-3-D registration there are far fewer pixels. Mutual information may therefore be failing as the 2-D histogram is relatively sparsely populated compared to the 1-D histogram used by entropy of the difference image. To investigate this a further experiment was carried out. A larger ROI was drawn in the fluoroscopy image (to also include the L2 and L4 vertebrae) which effectively trebled the number of pixels used. Sixty-four registrations were carried out using mutual information and the larger ROI. This resulted in a more accurate registration (similar to gradient correlation) to the plain fluoroscopy image. But a large number of failures (73%) still occurred when registering to the combined soft tissue and stent image.

Another possible reason for the poor performance of mutual information is that it does not assume a linear relationship between the intensities in the two images. The DRR has been produced so that this relationship should be close to linear. Measures that assume a linear relationship make implicit use of this extra information, which may account for their greater accuracy.

Gradient correlation has been shown to be sensitive to thin line structures. To test this further, an experiment was carried out, whereby the weighting of the stent was doubled ($w = 2$) when the stent image was formed. This gave the stent the range of intensity which could be expected in an undeployed stent, or a measurement catheter. Registrations were then carried out using gradient correlation, pattern intensity, and gradient difference. Pattern intensity and gradient difference were largely unaffected by this increase in stent intensity, while gradient correlation failed 39% of the time and the accuracy was poor in the X direction (20 mm rms error).

This paper has concentrated on the use of intensity-based algorithms. A comparison with feature-based methods has not been carried out and so conclusions on their relative merits cannot be drawn. The performance of feature-based algorithms on the images used in this paper would greatly depend on how these algorithms segment 2-D features. If their 2-D segmentation method is invariant to the presence of soft-tissue structures and interventional instruments, then registration accuracy should not be effected. However, if the accuracy of 2-D feature selection is reduced by the presence of soft-tissue structures or interventional instruments, then registration accuracy may be effected. 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.

VII. CONCLUSIONS

This paper has compared the accuracy and robustness of six 2-D-3-D 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 2-D-3-D image registration. Correlation measures can be effected by thin line structures, such as an interventional stent, which introduces pixels that 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 types of measure described in this paper can achieve this: pattern intensity and gradient difference. They both account for thin line structures by having a function of the form $1/(1+x^2)$ though they differ in how they reduce the effect of soft-tissue structures. Pattern intensity uses a region r within which there is little variation in soft-tissue intensity and gradient difference relies on there being many more strong gradients due to bony features than due to soft-tissue structures and air.

ACKNOWLEDGMENT

The authors would like to thank Philips Medical Systems, EasyVision Advanced Development, for providing software. They would particularly like to thank A. van Eeuwijk for his help with distortion correction and F. Gerritsen for his encouragements and constructive criticisms. They would also like to thank Prof. A. Adam and Dr. M. Cowling for their clinical input and, finally, 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*, 1995, pp. 275–286.
- [2] C. B. Bose and I. Amir, “Design of fiducials for accurate registration using machine vision,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1196–1200, 1990.
- [3] T. M. Buzug, J. Weese, C. Fassnacht, and C. Lorenz, “Image registration: Convex weighting functions for histogram-based similarity measures,” in *Proc. CVRMed/MRCAS*, J. Troccaz, E. Grimson, and R. Mösges, Eds. Berlin, Germany: Springer-Verlag, 1997, pp. 203–212.
- [4] G. Chiorboli and G. P. Vecchi, “Comments on ‘Design of fiducials for accurate registration using machine vision’,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, pp. 1330–1332, 1993.
- [5] S. Ganapathy, “Decomposition of transformation matrices for robot vision,” *Pattern Recogn. Lett.* 2, 1984, pp. 401–412.
- [6] L. M. G. Brown and T. E. Boulton, “Registration of planar film radiographs with computed tomography,” in *Proc. MMBIA*, 1996, pp. 42–51.
- [7] 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 Med. Imag.*, 1998, pp. 81–94.
- [8] 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),” *Int. J. Card. Imag.*, vol. 6, pp. 39–45, 1990.
- [9] D. J. Hawkes, C. B. Mol, and A. C. F. Colchester, “The accurate 3D reconstruction of the geometric configuration of vascular trees from X-ray recordings,” in *Physics and Engineering of Medical Imaging*, R. Guzzardi, Ed., NATO ASI. Dordrecht, The Netherlands: Martinus Nijhoff, 1987, pp. 250–258.
- [10] D. L. G. Hill, D. J. Hawkes, Z. Hussain, S. E. M. Green, C. F. Ruff, and G. P. Robinson, “Accurate combination of CT and MR data of the head: Validation and applications in surgical and therapy planning,” *Computerized Med. Imag., Graphics*, vol. 17, no. 4/5, pp. 357–363, 1993.
- [11] 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.*, vol. 22, pp. 317–323, 1998.
- [12] 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*, vol. 1570, pp. 322–336, 1991.
- [13] 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.*, vol. 21, no. 11, pp. 1749–1760, 1994.
- [14] J. A. Little, D. L. G. Hill, and D. J. Hawkes, “Deformations incorporating rigid structures,” *Comput. Vision, Image Understanding*, vol. 66, pp. 223–232, 1997.
- [15] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Seutens, “Multimodality image registration by maximization of mutual information,” *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–198, Apr. 1997.
- [16] C. R. Maurer, Jr., G. B. Aboutanos, B. M. Dawant, R. J. Maciunas, and J. M. Fitzpatrick, “Registration of 3-D images using weighted geometrical features,” *IEEE Trans. Med. Imag.*, vol. 15, pp. 836–849, 1996.
- [17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [18] A. Rougée, C. Picard, C. Ponchut, and Y. Troussset, “Geometric calibration of X-ray imaging chains for three-dimensional reconstruction,” *Computerized Med. Imag., Graphics*, vol. 17, no. 4/5, pp. 295–300, 1993.
- [19] C. Studholme, D. L. G. Hill, and D. J. Hawkes, “Automated 3D registration of MR and CT images of the head,” *Med. Image Anal.*, vol. 1, no. 2, pp. 163–175, 1996.
- [20] ———, “Automated 3D registration of MR and PET brain images by multi-resolution optimization of voxel similarity measures,” *Med. Phys.*, vol. 24, pp. 25–35, 1997.
- [21] P. Viola and W. M. Wells, “Alignment by maximization of mutual information,” in *Proc. 5th Int. Conf. on Computer Vision (ICCV’95)*, 1995, pp. 16–23.
- [22] 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*, 1997, pp. 119–128.