EVALUATION OF SIMILARITY MEASURES FOR RECONSTRUCTION-BASED REGISTRATION IN IMAGE-GUIDED RADIOTHERAPY AND SURGERY

DARKO ŠKERL, M. SC., DEJAN TOMAŽEVIČ, PH.D., BOŠTJAN LIKAR, PH.D., AND FRANJO PERNUŠ, PH.D.
Department of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia

Purpose: A promising patient positioning technique is based on registering computed tomographic (CT) or magnetic resonance (MR) images to cone-beam CT images (CBCT). The extra radiation dose delivered to the patient can be substantially reduced by using fewer projections. This approach results in lower quality CBCT images. The purpose of this study is to evaluate a number of similarity measures (SMs) suitable for registration of CT or MR images to low-quality CBCTs.

Methods and Materials: Using the recently proposed evaluation protocol, we evaluated nine SMs with respect to pretreatment imaging modalities, number of two-dimensional (2D) images used for reconstruction, and number of reconstruction iterations. The image database consisted of 100 X-ray and corresponding CT and MR images of two vertebral columns.

Results: Using a higher number of 2D projections or reconstruction iterations results in higher accuracy and slightly lower robustness. The similarity measures that behaved the best also yielded the best registration results. The most appropriate similarity measure was the asymmetric multi-feature mutual information (AMMI).

Conclusions: The evaluation protocol proved to be a valuable tool for selecting the best similarity measure for the reconstruction-based registration. The results indicate that accurate and robust CT/CBCT or even MR/CBCT registrations are possible if the AMMI similarity measure is used. © 2006 Elsevier Inc.

Image-guided radiotherapy, Cone-beam CT, Image registration, Similarity measures.

INTRODUCTION

The precise localization of target and sensitive surrounding anatomic structures, at both the pretreatment and intratreatment stages, represents the most significant challenge to the implementation of three-dimensional (3D) conformal radiation therapy, intensity-modulated radiotherapy (IMRT), and minimally invasive surgery in routine clinical use. With the tremendous development of 3D imaging modalities such as computed tomography (CT), magnetic resonance (MR), and positron emission tomography (PET), finer and finer pretreatment localization of different anatomic structures has been made possible. Whereas pretreatment localization is based on image segmentation, intratreatment localization or verification of patient setup is based on registration of pre- and intratreatment information, which is usually the information encountered in images. Registration allows any 3D point defined in the pretreatment image to be located in a patient, treatment machine, or treatment room coordinate system. Modern planning and execution of radiotherapy or minimally invasive surgery are critically dependent on image guidance, which comprises image acquisition, segmentation, registration, and visualization (1–3).

Different registration strategies for online verification of patient setup in the treatment room have been used in the past, and new ones are constantly being proposed. A more recent and popular registration method in image-guided radiotherapy and surgery is based on registration of two-dimensional (2D) electronic portal images or X-ray images to digitally reconstructed radiographs (DRRs) (4, 5). DRRs are X-ray projection images computed for arbitrary angles of projection through the CT image (6, 7). The unknown pose of the pretreatment CT volume relative to the portal or fluoroscopic X-ray image set, defining the patient in the treatment room, may be automatically estimated by optimizing the similarity measure calculated from DRR and corresponding portal or fluoroscopic X-ray images (4, 8–11). Similarity measure is a function that associates a numeric value with a pair of images with the idea that a higher, or for some similarity measures lower, value indicates greater similarity. Similarity measure can be looked upon as a K-dimensional function, with K the number of parameters of the spatial...
transformation model. In the ideal case, the K-dimensional parametrical space, in which the value at each location corresponds to the value of the similarity measure for that transformation estimate, would contain a sharp maximum (minimum) with monotonically decreasing (increasing) values with distance away from the optimum. The range of transformations around the optimum for which the value of the similarity measure is a monotonic function of misregistration is called the capture range. The distance from an optimum, to which registration converged, to the true registration position is the registration accuracy. Some of the local optima may be very small, caused by either interpolation artifacts or by local good matches between voxel intensities. Robustness is the ability of a registration method to produce similar results on all trials, i.e., regardless of the starting position, implementation details, and differences in image contents. Ideally, a similarity measure, together with the optimization method, should provide accurate and robust registrations. Registrations that are based on similarity measures adjust the parameters of an appropriate spatial transformation model until the similarity measure reaches an optimum.

Unfortunately, the intensity-based registration of DRRs and corresponding portal or fluoroscopic X-ray images is slow because of time-consuming calculation of DRRs. To some extent, the problem may be reduced by calculating DRRs containing only structures of interest (9) or by implementing a faster DRR generation method (12). The other drawback of this method is that it is not best suitable for registration of MR to portal or fluoroscopic X-ray images, as there is practically no correlation between MR-based DRRs and X-ray projections (13). Probably the most important drawback of the DRR-based registration method is that by projecting a high-quality 3D CT image into 2D to generate a DRR image, valuable information needed for accurate and robust registration is lost.

These drawbacks have motivated the development of novel registration methods, based on reconstruction of a 3D volume, called cone-beam CT (CBCT), from a series of 2D projection images, and registration of pretreatment CT or MR to CBCT images (14–18). By this approach, the similarity is computed between a pair of 3D images instead of between numerous 2D image pairs, which is the case for DRR-based registration. In radiation therapy, two approaches have been considered based on whether kilovoltage (kV) or mega-voltage (MV) beams are used to form the image (2, 19). For kV CBCT, a separate kV source and a second dedicated electronic portal imaging device, attached perpendicular to the treatment beam, have been proposed to acquire the projection images (15, 20). For MV CBCT, the MV source of the linac and electronic portal imaging device currently employed to acquire portal images are used for imaging (14, 17, 21). In image-guided minimally invasive surgery, a C-arm has been used to acquire fluoroscopic X-ray images for CBCT formation (18, 22). An important clinical issue of these techniques is the extra radiation dose delivered to the patient with each setup verification scan. To keep radiation to which the patient is exposed as low as possible, it is desired that CBCT images are reconstructed from as few projection images as possible. As a consequence, the quality of a 3D image reconstructed from a small number of projection images will be low. The similarity measure applied to automatically register the pretreatment CT or MR image to the CBCT image must therefore be able to cope with low image quality.

The similarity measure that nowadays enjoys the reputation of an accurate and robust multimodality image registration criterion is mutual information (14, 23, 24). However, even registrations based on mutual information will often fail if images contain insufficient information. Failure of registration is even more likely if an MR image is to be registered to the CBCT image. Probably the only way to cope with the insufficient information is to add spatial information, for instance in the form of intensity gradients, into the registration process (18).

We have conducted two experiments. To be able to choose the best similarity measure for a reconstruction-based registration using a small number of projection images, we have first analyzed the behavior of nine similarity measures for rigid registration of CT and MR images to CBCT images. The behavior of each similarity function was analyzed using publicly available spine phantom image data (25) and a recently proposed protocol for evaluation of similarity measures (26, 27). The protocol allows estimation of a similarity measure’s capture range, number and extent of local optima, and the accuracy and distinctiveness of the global optimum. Second, to show that the information obtained by the evaluation protocol in the first experiment is valuable for choosing the best similarity measure for reconstruction-based registration, we have tested the performance of the reconstruction-based registration method (18) using each of the nine similarity measures. The reconstruction-based registration method has been tested on the same data using standardized evaluation methodology (25) by which registration accuracy and robustness were assessed.

METHODS AND MATERIALS

Experimental data

The publicly available image database (http://www.isi.uu.nl/Research/Databases/) consisted of 2D X-ray images and corresponding 3D, 3D rotational X-ray (3DRX), CT, and MR images of two defrosted segments of vertebral bodies (25). The first vertebral column consisted of three thoracolumbar vertebrae bodies, whereas the second segment consisted of five thoracic vertebrae bodies. The X-ray images were acquired by a clinical floor-mounted 3DRX C-arm system (Integris BV5000; Philips Medical Systems, Best, The Netherlands). During an 8-s run of 180° around each phantom, two sets of 100 projection images were acquired and used to reconstruct two high-resolution 3D volumes using a filtered back-projection reconstruction technique (28). As the C-arm was calibrated, the projection geometry of the X-ray images with respect to the reconstructed volume was known. Therefore, the geometric relationship between a 3DRX volume and the 100
corresponding X-ray images was known. The CT-images of the two vertebral columns were acquired with a clinical 16-detector-row multislice CT scanner (MSCT; Philips Medical System, Best, The Netherlands). The MR images were obtained with a clinical 1.5-Tesla MR scanner (Gyroscan NT; Philips Medical System, Best, The Netherlands) using a sagittal 3D turbo spin echo acquisition and turbo factor of 29, repetition time/echo time of 1500 ms/90 ms. The “gold standard” 3D/3D registrations of 3DRX to corresponding CT or MR data were obtained by maximization of mutual information (24). Figure 1 shows two orthogonal X-ray image projections and a projection of CT and MR images of the first spine phantom. Figure 2 shows slices of CBCT images reconstructed from four (top) and eight (bottom) projections. The number of iterations of the reconstruction algorithm was 1, 3, 5, and 7 (left to right).

Cone-beam CT

To study the influence of the number of X-ray images on the quality of the CBCT image and, consequently, on the behavior of the similarity measures and quality of a reconstruction-based registration, we have reconstructed a number of CBCTs using different subsets of X-ray images out of the two sets of 100 X-ray images. Generally, the more X-ray images are used for reconstruction, the better is the reconstructed image. The iterative simultaneous algebraic reconstruction technique (SART) was used for reconstruction (29). SART, which is slow but simple to implement, generally provides good results in situations when a small number of projections is available or when projections are not uniformly distributed. Usually, just a few iterations are enough to obtain useful reconstruction results, whereas increasing the number of iterations may increase reconstruction artifacts, which are caused by computation errors that are amplified in each iteration. Increasing the number of reconstruction iterations also results in a more time-consuming reconstruction. Besides the number of images used for reconstruction, we have therefore also analyzed the impact of the number of iterations on the behavior of similarity measures.

Similarity measure evaluation protocol

The evaluation protocol requires that images, typical for a specific registration task, and “gold standard” registration of these images are given, and that a spatial transformation model is selected (26, 27). The spatial transformation $T$ that is supposed to bring two images, a floating and a target image, into correspondence is assumed to be rigid and is therefore composed of three translational $(t_x, t_y, t_z)$ and three rotational $(\Theta_x, \Theta_y, \Theta_z)$ parameters. The six-dimensional parametrical space is first normalized so that equal changes of each of the six parameters in the normalized parametrical space will have approximately equal impact on the transformation magnitude. By normalizing the parametrical space, Euclidean metrics may be used to determine distances from the position at which the images are “best” aligned and where a similarity function should have its optimum. Let the origin $X_0$ of the six-dimensional parametrical space be at the known “gold standard” position and let $SM(X)$ be the value of a similarity measure for the spatial transformation defined by location $X; X = [x_1, \ldots, x_k]$ in this space. Similarity measure values $SM(X_{n,m})$ are defined for image pairs, with the target image at the origin $X_0$ and the floating image transformed from the origin to location $X_{n,m}$. Values $SM(X_{n,m}), n = 1, 2, \ldots, N; m = -M/2, \ldots, M/2,$ of a similarity measure are defined on $N$ lines probing the six-dimensional parametrical space and at $M + 1$ points evenly spaced along each line. Each of the $N$ lines is defined by a randomly selected starting position $X_{n,-M/2}$ at a distance $R$, $R = \|X_{n,M/2}\|$, from the origin and its mirror point $X_{n,M/2}$. To make the similarity measure invariant to the absolute scale, each original similarity measure $SM(X_{n,m})$ is normalized to the interval $[0, 1]$

$SM(X_{n,m}) = \frac{SM(X_{n,m}) - SM_{\text{min}}}{SM_{\text{max}} - SM_{\text{min}}} \quad (1)$

where $SM_{\text{min}}$ and $SM_{\text{max}}$ are the minimal and maximal values of $NM + 1$ similarity measure values before normalization, respectively. If a similarity function is such that its minimum is sought for by optimization, the similarity measure value $SM(X_{n,m})$ is changed to $-SM(X_{n,m})$. Let $X_{n,\text{opt}}$ be the position and $SM(X_{n,\text{opt}})$ the value of the global optimum of the similarity measure along line $n$, and let $X_{n,\text{loc}}$ be the position of the local optimum closest to $X_{n,\text{opt}}$.

The behavior of a similarity measure is assessed by five properties. All properties are statistical estimations, derived from the “gold standard” position, similarity measure values $SM(X_{n,m})$ and positive gradients $d_{n,m}$:

$$d_{n,m} = \begin{cases} 
SM(X_{n,m-1}) - SM(X_{n,m}) & \text{if } m < \text{opt} \text{ and } SM(X_{n,m-1}) > SM(X_{n,m}) \\
SM(X_{n,m+1}) - SM(X_{n,m}) & \text{if } m > \text{opt} \text{ and } SM(X_{n,m+1}) > SM(X_{n,m}) \\
0 & \text{otherwise}
\end{cases} \quad (2)$$
The five properties are:

1. Accuracy ACC of a similarity measure is defined as the root mean square (RMS) of distances $||X_{n, opt} - X||$ between the origin $X_0$ and global optima $X_{n, opt}$, $n = 1, 2, \ldots, N$:

$$ACC = \sqrt{\frac{1}{N} \sum_{n=1}^{N} ||X_{n, opt} - X_0||^2} \ [\text{mm}]$$ (3)

2. Distinctiveness of optimum DO is defined as a function of distance $r$, $r = s \cdot \delta_s$, from the optimum, where $\delta_s$, $\delta_s = 2R/M$ is the distance between two consecutive points along a line and $s, s = 1, 2, \ldots$ is the number of steps:

$$DO(r) = \frac{1}{2rN} \sum_{n=1}^{N} 2 \cdot SM(X_{n, opt}) - SM(X_{n, opt+1})$$

$$-SM(X_{n, opt+1}) [10^{-3}/\text{mm}]$$ (4)

3. Capture range CR is defined as the minimal distance between the position of optima $X_{n, opt}$ and the closest minimum $X_{n, loc}$:

$$CR = \min_{s} ||X_{n, opt} - X_{n, loc}|| [\text{mm}]$$ (5)

4. Number of minima NOM(r) is the sum of minima of the similarity measure within distance $r$ from each of the $N$ global optima, i.e., a cumulative number of minima as a function of distance $r$. The average number of minima per line NOM(r)/N is denoted by NOM.

5. Risk of nonconvergence RON(r) is the property that describes the behavior of a similarity measure around the $N$ global optima. It is defined as the average of positive gradients $d_{n,m}$ within distance $r$ from each of the $N$ global optima:

$$RON(r) = \frac{1}{2rN} \sum_{n=1}^{N} \sum_{m=opt-r}^{opt+r} d_{n,m} [10^{-6}/\text{mm}]$$ (6)

A large value of RON(r) indicates that a similarity measure has distinctive (deep) or broader local optima in which optimization may get trapped. RON(r) is shorter denoted as RON.

The better a similarity measure is, the smaller are the values of the accuracy, number of minima, and risk of nonconvergence and the larger the capture range and distinctiveness of optimum values. CR, NOM, and RON are the three properties that describe robustness. The better these values are, the more robust is the similarity measure. More details on the evaluation protocol can be found in the studies of Škerl et al. (26, 27) and in the online version at http://lit.fe.uni-lj.si/Evaluation.

Similarity measures

We have implemented and evaluated the following nine similarity measures:

1. Asymmetric gradient–based mutual information AMMI (18) allows one to incorporate multi-variable image features (e.g., image intensity, image intensity gradients) into a single information-based similarity measure. AMMI is based on the assumption that just one of the floating image features has an arbitrary distribution, which is estimated by intensity histo-

2. Symmetric gradient–based mutual information SMIMI (18) also allows one to incorporate multi-variable image features into a single information-based similarity measure, by assuming that one floating and one target image feature are arbitrarily distributed and can be estimated via histograms.

3. Mutual information MI (23, 24) is a single feature similarity measure, which measures the amount of information that one image contains about the other. The influence of the size of the overlapping region between the images during registration is taken into account.

4. Normalized mutual information NMI (30) also measures the amount of information between the images. The influence of the size of the overlapping region between the images during registration is smaller than for MI.

5. Entropy correlation coefficient ECC (24, 31) is a measure of statistical dependence between the images.

6. Joint entropy H (32, 33) measures the amount of information that we have in the combined images. The more similar the images are, the lower is the joint entropy compared with the sum of individual entropies. The important limitation of this measure is that it does not deal with the overlap problem.

7. Correlation ratio COR (34) is an asymmetric measure of functional dependence between the images.

8. Woods’ criterion WC (35), or partitioned intensity uniformity, first partitions the target image into isointensity sets (or histogram bins) based on voxel values. Then it optimizes the uniformity of the floating image’s voxel values within each bin.

9. Pearson’s correlation coefficient PCC is a symmetric measure of linear dependence between the images. The optimization of linear dependence limits the use of this measure to registration of mono-modal images.

RESULTS

In all the following experiments, the number of lines $N$ was set to 50, $R$ to 35 mm, $M$ to 140, $\delta_s$ to 0.5 mm, and $s$ to 1. All similarity measures were applied to overlapping voxels of the floating and target images. In all experiments, the CBCT image served as the target image and CT or MR VOIs as the floating images. Similarity measures were formulated on the 2D joint histogram or joint probability distribution of the intensities of the two images. The joint histogram was obtained using 64 bins and partial volume interpolation (24). As similar trends were observed for each of the eight vertebral images, we present results only for the first vertebrae of the first vertebral column.

The impact of pretreatment imaging modalities

First, we have analyzed how the similarity measures behaved if CT or MR images had been registered to the CBCT image. To analyze the similarity measures in CT/CBCT image registrations, the CBCT image was reconstructed from four X-ray images with one reconstruction iteration while for MR/CBCT registrations the CBCT image was reconstructed from eight X-ray images with three reconstruction iterations. More images and iterations were used for reconstructing the CBCT to be registered with MR
because these modalities differ more than the CT and CBCT. Table 1 shows the values of five properties of the nine similarity measures. The numbers in bold and italic correspond to the best and the worst similarity measure, respectively, for a specific criterion. In CT/CBCT registrations, AMMI was the most accurate, had the most distinctiveness of optimum, the smallest RON, high CR, and small NOM. PCC had the smallest NOM indicating that it is robust but was not as accurate as AMMI and SMMI, which on the other hand was not robust as indicated by the large values of NOM and RON. For MR/CBCT registrations AMMI again had the smallest RON but was outperformed in terms of accuracy and distinctiveness of optimum by the SMMI similarity measure. The other measures that proved to be robust were MI and NMI. As expected, because of differences in modalities almost all properties of all similarity measures were worse in cases of MR/CBCT image registrations.

To further confirm the properties of similarity measures, we have conducted a registration-based experiment. The number of images used for reconstruction and the number of iterations were the same as in the first experiment. Registrations incorporating different similarity measures were evaluated using the standardized methodology for 3D/2D registration evaluation (25). The evaluation methodology uses the mean target registration error (mTRE) to measure the distance of a VOI’s position from “gold standard” before and after registration. The positions of all image elements in a VOI were used as target points. To evaluate the robustness of the CT or MR to CBCT image registration method, the methodology requires that for each VOI and corresponding CBCT image a large number of registrations are performed from different starting positions. Van de Kraats et al. (25) provided 200 starting positions for each of the 8 VOI. The starting positions were randomly generated around the “gold standard” position in such a way that the distance from “gold standard” measured by mTRE was uniformly distributed in the interval of 0 to 20 mm, with 10 positions in each of the 1-mm-wide subintervals. Powell optimization scheme was used as the search strategy. Each registration was considered successful if mTRE after registration was lower than the predefined clinically relevant threshold, which had been set to 2 mm. The registration error was defined as mTRE of all successful registrations. The results in Table 2 are highly correlated with the results in Table 1. The success rate of COR, WC, and PCC similarity measure respectively, for a specific criterion.

| Table 1. Properties of nine similarity measures in CT/CBCT and MR/CBCT registrations |
|-------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                                   | CT/CBCT |     |     |     |     |     |     |     |     |     |     |
|                                   | ACC     | DO  | CR  | NOM | RON | ACC     | DO  | CR  | NOM | RON |     |
| AMMI                               | 0.11    | 26.68 | 25.6 | 0.14 | 6.8 | 0.96    | 7.00 | 14.41 | 0.26 | 31.1 |     |
| SMMI                               | 0.24    | 26.01 | 18.1 | 1.58 | 659.3 | 14.17  | 10.43 | 1.72 | 1486.0 |     |
| MI                                 | 0.79    | 8.43  | 25.6 | 0.12 | 26.8 | 1.24    | 3.15 | 24.84 | 0.18 | 86.2 |     |
| NMI                                | 0.82    | 9.07  | 25.6 | 0.12 | 27.1 | 1.20    | 3.37 | 24.84 | 0.14 | 80.6 |     |
| ECC                                | 0.82    | 7.04  | 23.3 | 0.28 | 110.8 | 1.20    | 1.93 | 24.84 | 0.18 | 137.4 |     |
| H                                  | 0.16    | 1.36  | 27.9 | 0.10 | 14.2 | 1.24    | 0.90 | 0.50 | 0.24 | 50.0 |     |
| WC                                 | 1.17    | 2.15  | 23.3 | 0.50 | 437.3 | 0.96    | 1.30 | 0.50 | 0.66 | 827.1 |     |
| PCC                                | 0.61    | 4.92  | 27.1 | 0.10 | 15.0 | 4.52    | 1.40 | 0.50 | 0.48 | 113.5 |     |

**Abbreviations:** CT = computed tomography; CBCT = cone-beam CT; MR = magnetic resonance; AMMI = asymmetric multi-feature mutual information; ACC = accuracy; DO = distinctiveness of optimum; CR = capture range; NOM = number of minima; RON = risk of nonconvergence; SMMI = symmetric gradient– based mutual information; MI = mutual information; NMI = normalized mutual information; ECC = entropy correlation coefficient; H = joint entropy; COR = correlation ratio; WC = Woods’ criterion; PCC = Pearson’s correlation coefficient.

Numbers in bold and italic correspond to the best and worst similarity measure, respectively, for a specific criterion.

| Table 2. Mean target registration errors and success rates of registrations using different similarity measures and modalities |
|-------------------------------------|-----|-----|-----|-----|-----|-----|-----|
|                                   | CT/CBCT |     |     |     |     |     |     |
|                                   | mTRE (mm) | SR (%) | mTRE (mm) | SR (%) |
| AMMI                               | 0.3347 | 96.0 | 1.2897 | 91.5 |
| SMMI                               | 0.6643 | 88.0 | 0.4070 | 76.5 |
| MI                                 | 1.3413 | 94.5 | 1.8480 | 82.0 |
| NMI                                | 1.4069 | 94.5 | 1.8303 | 47.5 |
| ECC                                | 1.3517 | 96.0 | 1.8369 | 83.0 |
| H                                  | 1.6328 | 89.5 | 1.8200 | 33.5 |
| COR                                | 1.7421 | 96.5 | >*     | 0    |
| WC                                 | 1.9676 | 39.5 | >*     | 0    |
| PCC                                | 1.2692 | 92.5 | >*     | 0    |
| Combination                        | 0.6471 | 93.0 | 0.4149 | 93.0 |

**Abbreviations:** CT = computed tomography; CBCT = cone-beam CT; MR = magnetic resonance; mTRE = mean target registration error; SR = success rate; AMMI = asymmetric multi-feature mutual information; SMMI = symmetric gradient–based mutual information; MI = mutual information; NMI = normalized mutual information; ECC = entropy correlation coefficient; H = joint entropy; COR = correlation ratio; WC = Woods’ criterion; PCC = Pearson’s correlation coefficient. * > indicates that TREs were above 2 mm in all 200 registrations.
larity measures applied in MR/CBCT registration was 0% as TRES were above 2 mm in all 200 registrations. This is indicated by a symbol “>” in Table 2. This could be predicted from results in Table 1, as COR, WC, and PCC were neither accurate nor robust. The results in Table 1 also well predicted that AMMI was accurate and robust in CT/CBCT registrations and that SMMI was the most accurate measure for MR/CBCT image registrations.

We have also analyzed the outcome of CT and MR to CBCT registrations using a combination of two similarity measures. We have started the CT/CBCT registrations with the PCC similarity measure, which, as Table 1 indicates, is supposed to be very robust. After reaching the optimum, PCC was replaced by the SMMI similarity measure. The same experiment was performed for MR/CBCT registrations, except that COR was used instead of PCC as the initial similarity measure as suggested by results in Table 1. Optimization of the COR similarity measure was not very successful in terms of the mTRE (success rate) because in all 200 cases the optima had been further away from the “gold standard” than the predefined threshold of a successful registration. Because, according to the results in Table 1, COR was robust but very inaccurate, it could be used first to find the initial position from which the registration with the SMMI similarity measure, which is very accurate but not so robust, is run. Indeed, the combination of COR and SMMI, as Table 2 indicates, resulted in a very accurate and robust registration. Results of this experiment, shown in the last row of Table 2, indicate that with a combination of similarity measures a much higher success rate can be achieved with SMMI without compromising the accuracy of registrations obtained with SMMI.

The impact of the number of 2D images used for reconstruction

In this experiment, the properties of the similarity measures were evaluated with respect to the number of projections used for reconstruction. CBCT images were reconstructed from 4, 8, 16, 32, and 96 projections using one reconstruction iteration, if intended for registration with CT, and three reconstruction iterations for registration with MR images. The results shown in Fig. 3 indicate that the accuracy and distinctiveness of optimum of all similarity measures improved when more images had been used for reconstruction. However, with more images the measures became slightly less robust, as indicated by the increase of NOM and RON values. This can be explained with more details presented in 3D images reconstructed from more 2D projections. Because with more images the accuracy improved and the robustness slightly deteriorated, the optimal number of images used for reconstruction is between 8 and 16 for CT and around 16 for MR images. Again, the similarity measures behaved better in CT/CBCT than in MR/CBCT registrations. The number of 2D projections had a greater impact on MR/CBCT than on CT/CBCT registrations. This was most probably due to larger differences between the MR and CBCT modalities.

The impact of the number of reconstruction iterations

In this experiment we varied the number of iterations from 1 to 7 and analyzed the impact of iterations on the behavior of similarity measures. The CBCT image was reconstructed from 4 and 8 projection images for registration with CT and MR, respectively. Results in Fig. 4 indicate that the behavior of the similarity measures did not improve when more than three iterations had been used to reconstruct the CBCT image. The number of iterations had almost no impact on the robustness and DO, whereas the accuracy of some of the measures, like H, WC, and COR, improved if three instead of one or two iterations had been used. For MR/CBCT registration, the similarity measures were the most accurate if three or more iterations of SART had been used while the robustness slightly deteriorated with more iterations. Considering the obtained results and corresponding computational demands, between 3 and 5 reconstruction iterations seem to be the most appropriate.

DISCUSSION

The reconstruction of a CBCT from kV or MV beams and its registration to a CT or MR image has been recently proposed in radiation therapy and minimally invasive image-guided surgery for patient setup verification (14–18). The advantages of CBCT over planar megavoltage radiographs are numerous. CBCT images are three-dimensional, they have better soft tissue contrast, and they are easier to compare with planning CT images (1). By registering two 3D images, one avoids losing valuable 3D information needed for accurate and robust registration. This is the case with DRR-based registrations where a high-quality 3D CT image is projected into 2D to generate a DRR image. An important clinical issue of reconstruction-based registration is the extra radiation dose delivered to the patient. The total dose required to perform kV or MV CBCT is approximately the product of the number of projection images and the dose delivered for each image. In the study of Pouliot et al. (14), the MV CBCT image was reconstructed from 180 projection images and the radiation delivered to the patient was 15 cGy. Jaffray et al. (15) reconstructed a MV and kV CBCT from 90 MV and 195 kV projections of a head phantom, respectively. The central-axis doses delivered in the MV and kV CBCT imaging procedures were 138 cGy and 3.1 cGy, respectively. In a later study by Jaffray et al. (36), a kV CBCT of an anthropomorphic phantom was reconstructed from 321 projections. By decreasing the dose below 2 cGy, CBCT could be performed daily and the dose to the patient would not significantly increase over the current standard. The dose can be reduced either by developing more sensitive flat-panel X-ray detectors or by reconstructing images from fewer projections. The latter approach would result in CBCT images of lower quality. Therefore, if a similarity measure based registration is to be used to register a low-quality CBCT image with the treatment planning CT image, the similarity measure should be able to cope with low image quality. Given the images to be registered and the
spatial transformation model, the quality of registration mainly depends on the similarity measure and the optimization method.

Each of the numerous similarity measures introduced in the past has different properties, and is thus differently sensitive to the imaging modality, image content, sampling,
interpolation, histogram binning, partial image overlap, and image degradation, such as noise, intensity inhomogeneities, and geometrical distortions. To be able to choose the most appropriate similarity measure and its implementation, as well as the optimization method, it is often desirable to have some *a priori* information on the behavior of the similarity function with regard to the aforementioned factors. A thorough analysis of the similarity function at every point of the parametrical space is impossible because the parametrical space is too large to be analyzed even in the

Fig. 4. Five properties (top to bottom) of similarity measures as a function of the number of iterations for computed tomography/cone-beam computed tomography (CT/CBCT) (left column) and magnetic resonance/cone-beam computed tomography (MR/CBCT) (right column) registrations. The risk of nonconvergence $RON$ for symmetric gradient–based mutual information (SMMI) and Woods’ criterion (WC) is out of scale for both computed tomography/cone-beam computed tomography and magnetic resonance/cone-beam computed tomography registrations. For magnetic resonance to cone-beam computed tomography registration, the accuracy $ACC$ of Pearson’s correlation coefficient (PCC) was out of scale so it is not shown in the charts. Abbreviations as in legend to Fig. 3.
simplest case of rigid registration of 2D images. Limited information on the similarity measure may be obtained either indirectly from the accuracy and robustness of the registration itself or more directly, by drawing plots, showing their behavior when one image is systematically translated from or rotated around the “gold standard” registration position. Unfortunately, the information obtained by the latter approach is qualitative. To obtain more a priori and quantitative information on the behavior of a similarity measure, we have recently developed an optimization-independent similarity measure evaluation protocol (26, 27). By this protocol, a similarity measure’s capture range, number and extent of local optima, and the accuracy and distinctiveness of the global optimum are estimated independently of optimization. We believe that the proposed properties provide valuable quantitative information on the behavior of a similarity measure that can help researchers in deciding which similarity measure and which optimization protocol to use in a given application. However, it is clear that all possible behaviors of a similarity measure cannot be assessed, neither by these five features nor by any other ones. The first three properties characterize the robustness of a similarity measure. If within the space of possible geometrical transformations a similarity function is smooth, then it is less likely that registration will be influenced by the choice of the optimization scheme. The number of minima (NOM) and risk of nonconvergence (RON) provide complementary information on convergence. A high value of NOM alone does not necessarily mean that a cost function is bad. The local optima can be very small and the optimization may thus not end in one of them. However, NOM and RON together estimate not only the number but also the extent of local optima and therefore provide more information on convergence. The number of local optima and risk of nonconvergence might thus prove useful in selecting an appropriate optimization scheme. The third property that is potentially useful for optimization implementation is distinctiveness of the optimum (DO). DO(r) estimates the average change of a similarity measure near the global optimum, i.e., the behavior of a similarity measure in the neighborhood, defined by distance r, of the global optimum. When approaching the global optimum, a similarity measure may rise (fall) steeply and then, after reaching the optimum fall (rise) quickly as well, causing a distinctive (sharp) peak at the global optimum. Alternatively, it can rise and fall slowly, indicating an indistinctive (flat) optimum. DO(r) might help in selecting the stopping criterion of the optimization method.

Our definition of the capture range (CR) is rather strict. Because it is defined as the distance from the global optimum to the closest local optimum, regardless of its extent, it is actually the worst-case capture range. In practice, the optimization will probably not converge to a small local optimum. Besides, an optimization procedure will most probably not proceed along a line and the capture range may thus be larger than the CR obtained by the evaluation protocol. The capture range could be defined in many other ways. For instance, CR could be related to the distance r where RON(r) reaches a certain value or to the distance r where RON(r) abruptly rises.

In this article, we have evaluated nine similarity measures used for rigid body registration of pretreatment CT and MR to intertreatment CBCT images. To be as fair as possible, all similarity measures were formulated on the 2D histogram or joint probability distribution of the two images to be registered. In this way the impact of interpolation and histogram creation did not bias the results. We have evaluated the behavior of each of the nine similarity measures using the publicly available spine phantom CT, MR, and X-ray projection images for which “gold standard” registrations were available (25). The results of the analysis of similarity measures are thus important in the context of CT or MR to kV CBCT registrations (15, 18, 36).

Because CT images have excellent geometrical accuracy and provide electron density information directly, CT is nowadays the main imaging modality for treatment planning and patient setup verification. Results in Table 1 indicate that for CT to kV CBCT registrations the recently introduced AMMI similarity measure (18) is the most suitable as it was the most accurate and robust and had the most distinctive optimum of all the nine similarity measures analyzed. The AMMI similarity measure is based on image intensities and image intensity gradients of both images that are to be registered. The authors of AMMI believe that image intensities ensure larger capturing ranges whereas image intensity gradients preserve the accuracy of registration methods that use gradient features only (18). The information on the behavior of similarity measures provided by the evaluation protocol correlated well with the accuracy and robustness of CT/CBCT registrations run from 200 starting positions. The starting positions were uniformly distributed in the interval of 0 to 20 mm around the “gold standard” registration position. With AMMI, a registration accuracy of 0.33 mm and 96% success rate has been achieved if the CBCT had been reconstructed from only four projections and with one reconstruction iteration.

Magnetic resonance imaging (MRI) provides superior image quality for soft tissue segmentation over CT and is used, for instance, for tumor and surrounding risk tissue segmentation in radiotherapy planning of brain tumors. However, the lack of electron density information, image distortion leading to geometrical inaccuracies, and the differences between intensities of MR and portal images have precluded the more widespread use of MR images in radiotherapy treatment planning and patient setup verification. A solution for both accurate MR-based segmentation and CT-based dose calculation is acquisition of both CT and MR images and their registration. The use of MRI alone for treatment planning and patient setup verification would remove any MR/CT registration errors and would reduce treatment cost by avoiding redundant CT scans and save patient, staff, and machine time (37). Beavis et al. (38) reported on brain tumor radiotherapy plans by assigning homogeneous electron density to the entire MR image. Chen et al. (37) showed that no clinically significant differ-
enences were found between the MRI-based and CT-based treatment plans using the same beam arrangements, dose constraints, and optimization parameters. Lee et al. (39) compared dose distributions created on the homogeneous density and bulk-density assigned MR images to original CT images of the prostate. They have observed negligible differences in dose distribution between radiotherapy plans using bone+water CT number bulk-assigned image and original CT. Bone segmentation which is required for MR planning can be done manually or automatically. For patient setup verification using MR images, registration methods have been proposed that are based on MR DRRs (37, 40). In this paper, we have analyzed the similarity measures that would be appropriate for MR/CBCT registrations for patient setup verification. Results in Table 1 indicate that SMMI was the most accurate but, on the other hand, the least robust similarity measure. As in CT/CBCT registrations AMMI was the most robust because it had the smallest risk of nonconvergence and a small number of local optima. The information on the behavior of similarity measures provided by the evaluation protocol again well correlated with the accuracy and robustness of MR/CBCT registrations run from 200 starting positions. In MR/CBCT registration the highest success rate (93%) and high accuracy (0.4 mm) were achieved with a combination of similarity measures. The registration was started with COR and, after reaching the optimum, continued with the SMMI similarity measure. That CT/CBCT registrations are more accurate and robust than MR/CBCT registrations was expected because of the larger differences between MR and CBCT modalities.

We have also analyzed the impact of the number of 2D projections that CBCT was reconstructed from on the behavior of the similarity measures. The accuracy and distinctiveness of the optimum improved significantly when CBCT had been reconstructed from more projections. This has been observed in both CT and MR to CBCT registrations. Unfortunately, the properties that describe robustness did not improve; they even slightly deteriorated. According to the results presented in Fig. 3, the optimal number of projections for CT and MR to CBCT registrations is approximately 16. Even better registration results can be expected if the CT or MR images are first registered to the CBCT image reconstructed from four or eight projections. The optimum of this registration will not be accurate, but it can be used to initialize further CT or MR registrations with the CBCT image reconstructed from 16 projections. More reconstruction iterations led to higher accuracy. The other properties of the nine similarity measures were only slightly affected by the higher number of iterations. Unfortunately, increasing the number of reconstruction iterations also results in a more time-consuming reconstruction procedure. The time needed for reconstruction depends on the size of the reconstructed volume, resolution of the projected image, number of projections, and number of reconstruction iterations. The time needed to reconstruct an image increases linearly with the number of iterations and projections. In our case, the reconstruction took approximately 5 s per projection and iteration on a 2.8 GHz Pentium IV personal computer. Using two iterations would thus increase the time to approximately 10 s.

CONCLUSION

Reconstruction-based registration for patient setup verification and the use of MR images for radiotherapy planning and setup verification are rather new and little explored. To keep the additional dose that a patient receives in such a scenario as low as possible, the CBCT has to be reconstructed from a small number of projection images, either using kV or MV beams. We have analyzed the behavior of several similarity measures used for registration of CT and MR to CBCT images reconstructed with different reconstruction parameters. With the evaluation protocol, we were able to find the best reconstruction setting and similarity measure. We have shown which of the presented similarity measures performs best for this kind of registration and for the images studied. It is, however, not clear that the similarity measure that performed best on images of the spine would also be the best for other body sites, such as the pelvis or head, as the behavior of a similarity measure depends not only on the imaging modality but also on image content and image degrading effects. We have also shown that a combination of different similarity measures can lead to a more robust registration. Considering the high accuracy and robustness of CT/CBCT and even MR/CBCT registrations using an appropriate similarity measure, the clinical use of such a registration for patient setup verification is worth further investigation.

REFERENCES


