

## A neural network-based 2D/3D image registration quality evaluator for pediatric patient setup in external beam radiotherapy

Jian Wu,<sup>1a</sup> Zhong Su,<sup>2</sup> and Zuofeng Li<sup>2</sup>

*Department of Radiation Oncology,<sup>1</sup> University of Florida, Gainesville, FL; University of Florida Health Proton Therapy Institute,<sup>2</sup> Jacksonville, FL, USA*  
jwu2@ufl.edu

Received 29 July, 2014; accepted 4 October, 2015

Our purpose was to develop a neural network-based registration quality evaluator (RQE) that can improve the 2D/3D image registration robustness for pediatric patient setup in external beam radiotherapy. Orthogonal daily setup X-ray images of six pediatric patients with brain tumors receiving proton therapy treatments were retrospectively registered with their treatment planning computed tomography (CT) images. A neural network-based pattern classifier was used to determine whether a registration solution was successful based on geometric features of the similarity measure values near the point-of-solution. Supervised training and test datasets were generated by rigidly registering a pair of orthogonal daily setup X-ray images to the treatment planning CT. The best solution for each registration task was selected from 50 optimizing attempts that differed only by the randomly generated initial transformation parameters. The distance from each individual solution to the best solution in the normalized parametrical space was compared to a user-defined error tolerance to determine whether that solution was acceptable. A supervised training was then used to train the RQE. Performance of the RQE was evaluated using test dataset consisting of registration results that were not used in training. The RQE was integrated with our in-house 2D/3D registration system and its performance was evaluated using the same patient dataset. With an optimized sampling step size (i.e., 5 mm) in the feature space, the RQE has the sensitivity and the specificity in the ranges of 0.865–0.964 and 0.797–0.990, respectively, when used to detect registration error with mean voxel displacement (MVD) greater than 1 mm. The trial-to-acceptance ratio of the integrated 2D/3D registration system, for all patients, is equal to 1.48. The final acceptance ratio is 92.4%. The proposed RQE can potentially be used in a 2D/3D rigid image registration system to improve the overall robustness by rejecting unsuccessful registration solutions. The RQE is not patient-specific, so a single RQE can be constructed and used for a particular application (e.g., the registration for images acquired on the same anatomical site). Implementation of the RQE in a 2D/3D registration system is clinically feasible.

PACS numbers: 87.57.nj, 87.85.dq, 87.55.Qr

Key words: image registration, neural network, image-guided radiation therapy, brain tumor, registration quality assurance

### I. INTRODUCTION

In external beam radiation therapy, registration of 2D X-ray projection images and 3D CT images has been used clinically for patient setup and motion mitigation.<sup>(1,2)</sup> Compared to 3D/3D registration-based patient setup strategies, the 2D/3D registration method does not require

<sup>a</sup> Corresponding author: Jian Wu, 2000 SW Archer Rd, P O Box 100385, Gainesville, FL 32610, USA; phone: (352) 265 0680; fax: (352) 265 7057; email: jwu2@ufl.edu

lengthy 3D image acquisition on the daily basis and, more importantly, reduces imaging dose to the patient. The stochastic risk of carcinogenesis caused by imaging dose is a major concern for pediatric patients because of their longer life expectancy. For patient setup, although a number of dose reduction strategies for the CBCT imaging have been proposed,<sup>(3)</sup> a 2D X-ray imaging technique may still be a better choice if the 2D registration yields solutions that are clinically acceptable. At the present time, CBCT imaging is not available in our proton therapy systems. Patient setup relies on an X-ray imaging system with orthogonal projection geometry and a 2D image matching technique. Development of a reliable automated 2D/3D registration system may potentially reduce patient setup time and improve setup accuracy.

2D/3D registration approaches for patient positioning have been investigated<sup>(4–8)</sup> and were also implemented in commercial systems such as the Brainlab ExacTrac X-ray imaging system (Brainlab AB, Feldkirchen, Germany), the Accuray CyberKnife Xsight system (Accuray, Sunnyvale, CA), and the MedCom VeriSuite system (MedCom GmbH, Darmstadt, Germany) for proton therapy systems. These systems align two X-ray (orthogonal or oblique) projection images to a treatment planning CT image and compute six transformation (three translational and three rotational) parameters. When used with a robotic patient couch that is able to move in five or six-degrees of freedom, the system can achieve better patient setup correction compared to traditional correction strategies that are based only on translations.

Most 2D/3D registration methods require computation of digitally reconstructed radiographs (DRR) from a 3D CT dataset. A few years ago the computation burden of this step was a major obstacle that prevented the clinical implementation of such systems. However, this problem has been overcome by advances of graphics card-based computation techniques.<sup>(7,9)</sup> The calculation time required for an intensity-based 2D/3D registration has been reduced from a few hours using CPUs<sup>(4)</sup> to several seconds using graphics processing units (GPUs). With the technology of graphics-based computation continuously evolving, the speed of registration is no longer a concern. System reliability in terms of registration accuracy and robustness for a variety of anatomical disease sites has become the primary challenge in clinic.

An intensity-based 2D/3D registration algorithm aligns the acquired planar images with the computed DRRs by minimizing a cost function. Such an optimization process can be trapped in a local minimum due to the complexity of patient anatomy, image noise and artifacts, organ deformation, and the lack of monotonicity of the similarity function, which results in incorrect solutions. The frequency of such failures can depend on the initial conditions of 2D/3D registrations. Unsuccessful registration rate has been reported to be as high as 32% when initial shifts were 6 to 16 mm.<sup>(10)</sup> Traditionally, evaluation of the acceptability of a registration solution is carried out by visually comparing the overlapped source and target images. This method is subjective and inconsistent. Pattern classifier-based registration quality evaluation methods have been proposed by Wu and colleagues to identify local minima and avoid premature terminations, and have been successfully applied to 2D/2D<sup>(11)</sup> and 3D/3D<sup>(12)</sup> registration applications. Based on phantom and patient tests, they have demonstrated that the registration quality evaluator (RQE) could reject registration results with errors larger than a user-defined tolerance. Furthermore, based on a phantom study, Wu and colleagues<sup>(13)</sup> showed the potential of using the neural network-based RQE on 2D/3D registration applications. Although this pioneer study gave very favorable results, this approach has yet to be validated in a clinical setting.

For this paper, we used a similar method to the one described in Wu's phantom study in a clinical application where 2D/3D registration can be used to assist patient setup before daily radiation therapy. A neural network-based RQE was used to automatically identify bad registration solutions. Compared to the 3D/3D RQE, the construction of a 2D/3D RQE with good performance is more technically challenging. Although the similarity functions and the optimization algorithms used in 3D/3D registration can be used in 2D/3D registration with minimal modification, the failure pattern for 2D/3D registration is significantly different. A 3D/3D registration is typically either successful or failed badly (i.e., far away from its optimal point in the transformation parameter space) because the objective function is smooth and an

optimization converges easily to a solution around an optimum. The objective function for a rigid 2D/3D registration has much more local optima even in the close vicinity of the globally optimum. The use of a pattern classifier to identify those nonoptimal solutions near the optimal solution is, therefore, much more challenging.

## II. MATERIALS AND METHODS

### A. Patient data acquisition

In this study, we used the daily setup images of six randomly selected pediatric patients with ages ranging from 3 to 14. Patients selected for this study have previously consented to participation in our institutional IRB-approved protocol. All patients were diagnosed with brain tumors and have completed their proton therapy treatments in our institute. A pair of orthogonally projected kilovoltage (kV) X-ray images was acquired each day during the patient setup. The kV X-ray images, with  $1152 \times 1600$  pixels, have pixel sizes of  $0.18 \times 0.18 \text{ mm}^2$  and  $0.21 \times 0.21 \text{ mm}^2$  for panel A and panel B, respectively. Treatment planning CT images were acquired with the pixel size of  $0.75 \times 0.75 \text{ mm}^2$  in the axial plane and the slice thickness of 1 mm.

### B. 2D/3D image registration

The original X-ray images were first down-sampled at the ratio of 1:2 and then rigidly registered to their treatment planning CT images. Registrations between two X-ray images and the CT image were performed using an in-house GPU-based 2D/3D registration program. This program uses an iterative method to optimize the normalized mutual information (NMI) between the digitally reconstructed radiographs (DRRs) and the acquired X-ray images. The DRRs are computed in each iteration by ray-tracing a 3D CT image for a given set of transformation parameters using a graphics card. Each registration is based on the region of interest (ROI) as defined on the X-ray images. The room coordinates of the X-ray images were determined by image overlays of physical crosshair devices acquired simultaneously with the imaging subjects. The right–lateral and the posterior–anterior (PA) projection X-ray images of a selected patient are shown in Figs. 1(a) and 1(b), respectively. Figures 1(c) and 1(d) show the calculated final DRRs when optimization was completed. Figures 1(e) to 1(h) show the overlapped images of X-ray images and the edge images of DRRs. The edge images of DRRs were created by applying a Laplacian of Gaussian filter. Figures 1(e) and 1(f) show the images before registration and Fig. 1(g) and 1(h) the images after registration.

Each patient received 30 treatment fractions, thus there are a total of 30 unique registration tasks per patient. The rigid registrations computed three translational and three rotational parameters.

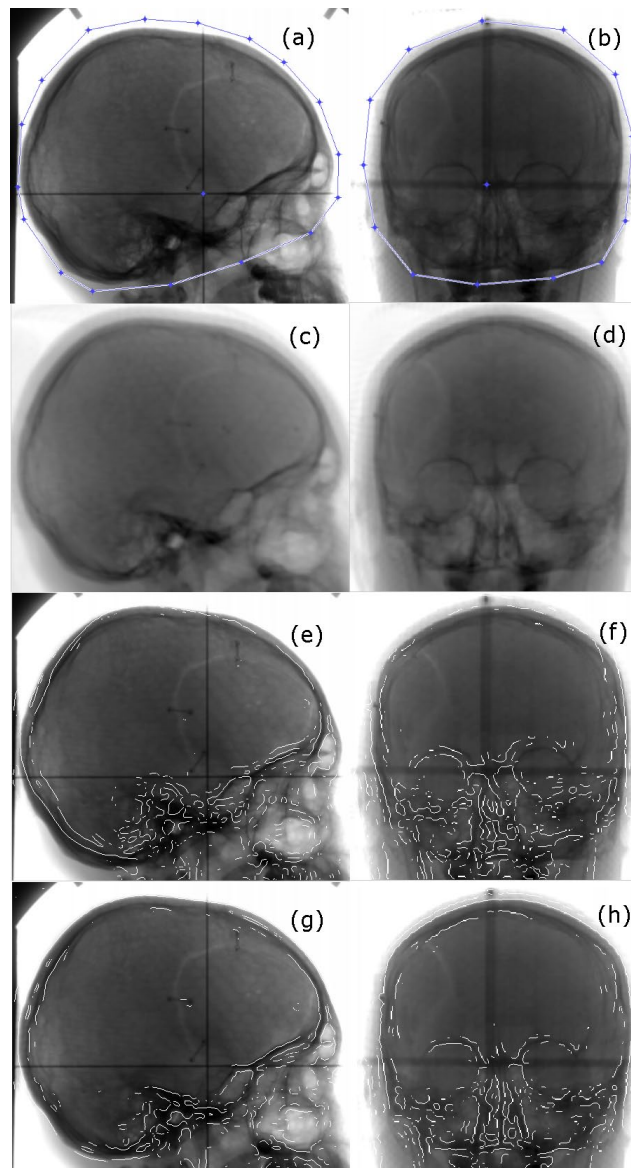


FIG. 1. The right-lateral view (first column) and the posterior–anterior view (second column) images of a selected pediatric patient. The setup verification X-ray images acquired just before treatment are shown in (a) and (b). The user-defined regions of interest are indicated by the line segments with blue vertices. The calculated final DRRs when optimization was completed are shown in (c) and (d). X-ray images and the edge images of DRRs are overlapped before the registration [(e) and (f)] and after the registration [(g) and (h)] to show the effects of image registration.

### C. Training and test data preparation

As described by Wu and Murphy,<sup>(12)</sup> a pattern classifier can be constructed and used to identify unsuccessful registration solutions that are caused by local minimum trapping or early termination. In this study, we applied this method to construct an RQE for 2D/3D registrations. Data preparation of the RQE involves: 1) sampling the parametrical space in the neighborhood of the point-of-solution, 2) computing a number of geometrical features from the sampled values, and 3) identifying the true category of a solution based on a user-defined error tolerance. We will describe these steps in details in the following sections.

### **C.1 Parametrical space sampling**

In the previous paper,<sup>(12)</sup> Wu and Murphy sampled the similarity measure values by varying one translational parameter at a time using two user-defined sampling step sizes (i.e., 1 mm and 5 mm). In this study, we sampled three rotational parameters in addition to the three translational parameters. This is because, for 2D/3D registrations, out-of-plane translations and rotations (with respect to the imaging planes) may change the similarity function value in different ways (e.g., different gradients) compared to the in-plane ones. The addition of rotational parameters will provide more information to the pattern classifier and potentially improve its performance. When we sampled the cost function, a normalized transform unit, the mean voxel displacement (MVD), was used for both translational parameters and rotational parameters in the parametrical space. Thus a unit change of a rotational or a translation parameter produces the same averaged voxel displacement within the ROI defined on the CT images. For the brain patients in this study, the ROIs on the CT images are rectangular cuboids that are approximately centered at the isocenter and include the whole brain. Those ROIs were manually defined by a user for each CT image. We investigated the impact of sampling step size on the performance of RQE by varying the step size from 1 mm to 30 mm with 5 mm increments.

### **C.2 Geometrical feature calculations**

The distinctiveness of optimum (DO) and the mirror symmetry (MS) have been used by Wu and Murphy<sup>(12)</sup> as geometrical features to evaluate their 3D/3D image registration solutions. The same geometrical features were used in this study, but were computed separately for the translational samples and rotational samples.

The DO is based on our assumption that a global minimum generally has a deeper valley compared to that for a local minimum. The MS is based on our assumption that a global minimum has a relatively symmetrical profile centered at its minimum. The profile of a local optimum, on the other hand, is usually more irregular because of its limited capture range. Calculations of both features require the similarity function be sampled along the translational and rotational major axes in the parametrical space for a given step size. For each axis, the sampled points were centered at the current registration solution.

As explained earlier, we computed DO and MS separately for translations and rotations, and denoted them as  $DO_{trans}$ ,  $DO_{rot}$ ,  $MS_{trans}$ , and  $MS_{rot}$ , respectively. These features were computed for each registration solution.

### **C.3 Determination of true categories**

We used the same method as proposed by Wu and Murphy<sup>(12)</sup> to determine the true category of a registration solution. In this method, a “gold standard” registration solution was obtained by taking the best solution over 50 registration attempts. These 50 registration attempts differ by randomly selected initial transformation parameters that are centered at the initial setup position (where the isocenters are aligned) and within the range of  $\pm 5$  mm and  $\pm 2^\circ$  in the parametrical space. The best solution has the minimal final cost function (i.e., negated NMI). This best solution was visually verified by using our in-house visualization software.

To confirm accuracy of the gold standard, 100 validation registrations were performed for all daily setup images of a randomly selected patient (i.e., Patient 1). Those 100 registrations are based on 100 random initial guesses that are centered on the best known alignment. The maximal deviations from the center are also limited to  $\pm 5$  mm and  $\pm 2^\circ$ . The differences between the new registration solutions and the gold standard are compared.

Once the “gold standard” solutions were established, a new round of 20 registration repetitions was carried out for each registration problem to generate the training and test dataset for the RQE. These 20 registration repetitions also differ by randomly selected initial transformation parameters that are centered at the initial setup position where the isocenters are aligned. This results in a total of 3600 registration solutions (i.e., 6 patients  $\times$  30 fractions/patient  $\times$  20 registrations/fraction) to be used to train and test the RQE. By balancing the number of successful



and unsuccessful registrations, the ranges of random perturbations to the initial transformation parameters were empirically chosen to be  $\pm 20$  mm and  $\pm 8^\circ$ .

The registration errors for these new solutions were quantified by calculating the MVD with respect to their “gold standard” solutions. As described by Skerl and colleagues,<sup>(14)</sup> the MVD is equivalent to the Euclidean distance from the point of the current solution to the point of the “gold standard” solution in the normalized parametrical space. So the MVD of a registration solution can be calculated conveniently and compared to a user-defined error tolerance to determine the acceptance of this solution. In this paper, a MVD tolerance of 1 mm was used. We believe this choice is reasonable since 1 SD of all successfully registration solutions is about 0.5 mm.

#### D. Training and validation of the neural network

The same two-layer, feed-forward neural network as proposed by Wu and Murphy<sup>(12)</sup> was used in this study. The network was implemented using the MATLAB Neural Network Toolbox (MathWorks, Inc., Natick, MA). There are four inputs to the neural network. The inputs are the computed geometrical features —  $DO_{trans}$ ,  $DO_{rot}$ ,  $MS_{trans}$ , and  $MS_{rot}$ . There is only one single numerical output ranging from 0 to 1. An output greater than 0.5 is considered to belong to the category of unsuccessful registrations. Hyperbolic tangent sigmoid transfer functions were used in both the hidden layer and the output layer. To avoid the problem of overfitting, effort has been taken to keep the network structure simple. We started with having 20 neurons in the hidden layer and gradually reducing the number of neurons. We found out the classifier still has very good performance when the number was reduced to two, but its performance deteriorated remarkably when there was only one neuron left in the hidden layer. Thus in our final neural network structure there are two neurons in the hidden layer. During the network construction, the scaled conjugate gradient optimization algorithm<sup>(15)</sup> was used to minimize the mean sum-of-squares of the network output error. This method is based on the conjugate gradient method, but was designed to avoid the time-consuming line search by combining with the model-trust region approach.

To overcome the problem of overtraining and improve generality, the available data were divided into three subsets. Firstly, the data generated from each individual patient was chosen in turns as the test dataset. This is so called “leave-one-out” scheme. Then the rest of the data from five other patients were randomly divided into the training dataset (80%) and the validation dataset (20%). The training set was used for computing the gradient and updating the network weights and biases. After each training iteration, the network was tested on the validation set. The test data had no effect on training and they provided an independent measure of network performance during and after training. Thus it provided us with a valid method to evaluate how the network was generalized to new patients. Two criteria, sensitivity and specificity, were used to evaluate the performance of the classifier.

#### E. Test of the integrated robust 2D/3D registration system

The RQE was integrated into our in-house, GPU-based 2D/3D registration system. The same patient dataset as mentioned above was used in this test. The RQE constructed with the first patient left out (arbitrarily chosen) were added to the 2D/3D registration system. Once a registration solution was reached, the RQE would determine whether the result was acceptable. If a solution was rejected, the registration method would restart with another trial, but using a different set of randomly generated initial transform parameters. In this paper, we set the maximal number of trials as five. If the RQE failure to accept a solution after the maximal number of trials has been reached, we consider the registration results were not reliable and a manual registration was required. Here we intend to estimate the efficiency of this integrated system rather than evaluate the accuracy and performance of the system. Consider two scenarios. 1) If a registration solution was accepted after too many registration trials, then the system would have little efficiency. 2) If all solutions were accepted on their first trial, then we would not need an RQE. It should be noted that, for the first case, it may not mean the RQE has bad performance.

It may just mean the registration method failed to find a good solution. The method described in section D above is a better way to evaluate the performance of the RQE itself.

### III. RESULTS

Figures 2(a) and 2(b) show the profiles of the cost function (i.e., negative NMI) along shifts in x-, y-, and z-axes and rotations about x-, y-, and z-axes from the point-of-solutions (i.e., the origins) in the parametrical space. The units of shifts and rotational deviations have been normalized to MVD. The profiles in Fig. 2(a) are centered on a typical good solution, where the curves on the positive side and the negative side are proximately symmetric. The profiles in Fig. 2(b) are centered on a typical bad solution, where some curves are not symmetric and the valleys for all curves are shallower than the corresponding ones in Fig. 2(a). This demonstrates why DO and MS can be used to describe those two features.

The best solutions from 100 validation registrations differ no more than  $0.16^\circ$  in rotation and 0.09 mm in translation compared to the gold standard. The average absolute differences between the best solutions and the gold standard are  $0.04^\circ$ ,  $0.03^\circ$ , and  $0.04^\circ$  in rotation along x-, y-, and

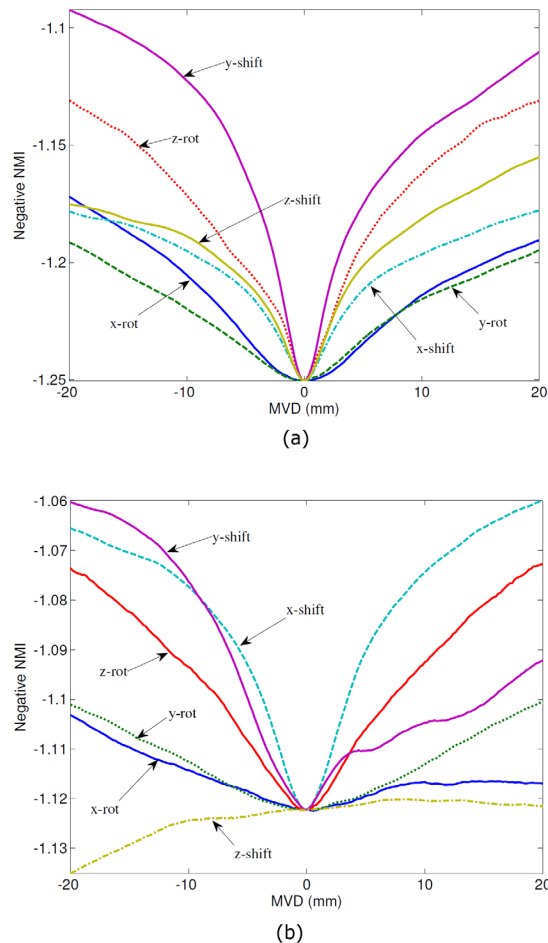


FIG. 2. Profiles of the cost function (i.e., negative NMI) along shifts in x-, y-, and z-axes and rotations about x-, y-, and z-axes from the point-of-solutions (i.e., the origins). The units of shifts and rotational deviations have been normalized to MVD. The profiles of a typical good solution and a bad solution are shown in (a) and (b), respectively.

z-axes, respectively. The differences for shifts are equal to 0.02 mm along all three major axes. Owing to the consistency of the registration solutions and the fact that all best solutions have been visual-checked, we have confidence that the gold standard is accurate and highly reliable.

The scatter plots of all data used in training, validation, and testing with 5 mm sampling size are shown in Figs. 3(a) and 3(b). There are 3,600 data points (i.e., 6 patients  $\times$  30 fractions/patient  $\times$  20 registrations/fraction) in each plot. Since the “leave-one-out” scheme was used, 3,000 data points were used to construct each classifier. In Fig. 3(a), the distinctiveness of optimum for translations ( $DO_{\text{trans}}$ ) was plotted against the distinctiveness of optimum for rotations ( $DO_{\text{rot}}$ ). In Fig. 3(b), the mirror symmetry for translations ( $MS_{\text{trans}}$ ) was plotted against the mirror symmetry for rotations ( $MS_{\text{rot}}$ ). The successful solutions are shown as blue circles and the unsuccessful solutions are shown as red crosses. As illustrated in 2D plots, there is no boundary to separate those two categories completely. It is also reasonable to assume that a simple boundary is not likely to be found in the 4D feature space.

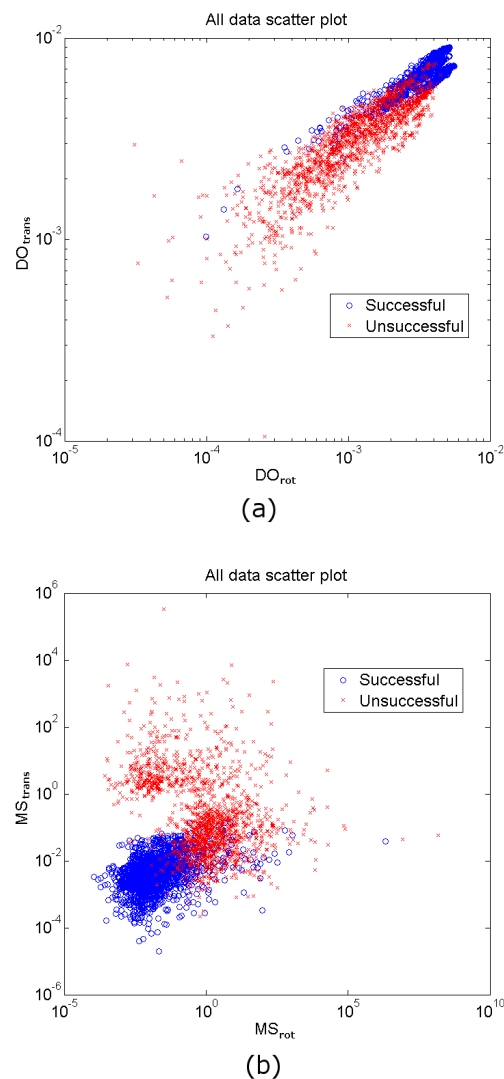


FIG. 3. Scatter plots of (a)  $DO_{\text{trans}}$  v.s.  $DO_{\text{rot}}$  and (b)  $MS_{\text{trans}}$  v.s.  $MS_{\text{rot}}$  of all data with 5 mm sampling step size. The successful solutions are shown as blue circles and the unsuccessful solutions as red crosses.



To investigate the effect of the sampling step size on the performance of the RQE, the RQEs were constructed and tested separately for various step sizes ranging from 1 mm to 30 mm with 5 mm increment. As described in the previous section, each time, data from one of the six patients were chosen as the test data. The RQE was constructed using data from the rest of the patients for each given step size. Performances of the RQE in terms of sensitivity and specificity as a function of sampling step size are plotted in Figs. 4(a) and 4(b), respectively. The sensitivities for all test patients are similarly high with the sampling step size between 5 mm and 25 mm. The specificities are very close for all patients in the entire test range except for Patient 2, whose specificity peaks at 5 mm step size. The sensitivity and the specificity of all patients with 5 mm step size are tabulated in Table 1.

The registration results of the integrated 2D/3D registration system are shown in Table 2. The trial-to-acceptance ratio (TAR), which is defined as the ratio of the total number of trials to the number of acceptance, for all patients, is equal to 1.48. This indicates the integrated system has about 48% overhead compared to the original system. The final acceptance ratio (FAR), which is defined as the number of accepted solutions to the number of registration attempts (i.e., number of acceptances + number of final rejections), is 92.4%.

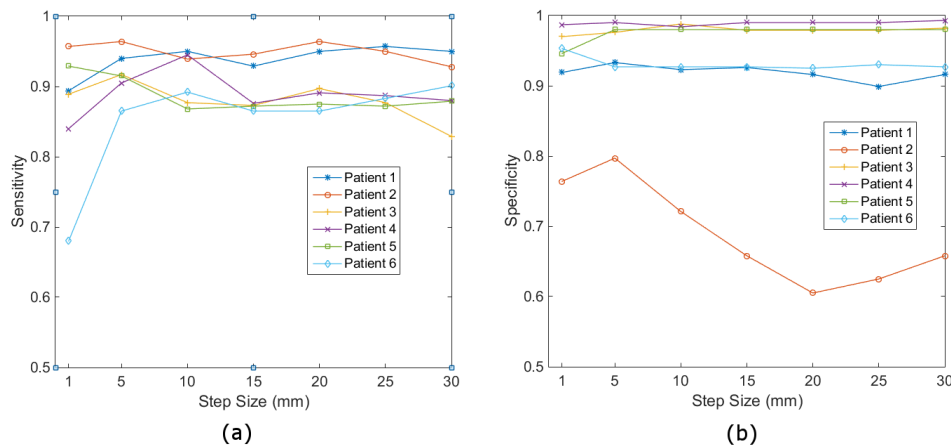


FIG. 4. Sensitivity (a) and specificity (b) of the classifier trained with various sampling step sizes evaluated using test datasets.

TABLE 1. Sensitivity and specificity of the classifier for test datasets with 5 mm sampling step size.

	<i>Patient 1</i>	<i>Patient 2</i>	<i>Patient 3</i>	<i>Patient 4</i>	<i>Patient 5</i>	<i>Patient 6</i>
Sensitivity	0.940	0.964	0.917	0.905	0.915	0.865
Specificity	0.933	0.797	0.976	0.990	0.980	0.927

TABLE 2. Registration results for the integrated robust 2D/3D registration system.

	<i>Patient 1</i>	<i>Patient 2</i>	<i>Patient 3</i>	<i>Patient 4</i>	<i>Patient 5</i>	<i>Patient 6</i>	<i>Total</i>
# of Acceptance	317	190	320	310	320	290	1747
# of Trials	359	850	334	388	349	298	2578
# of Final Rejections	3	130	0	10	0	0	143

#### IV. DISCUSSION

Due to the similarity between the 2D/3D registration and the 3D/3D registration, this paper used the same training data generation, RQE generation, and performance evaluation method as previously reported.<sup>(12)</sup> However, we added features for rotational deviations in our 2D/3D RQE instead of utilizing only translational features. This is because, presumably, object translations and rotations about different axes may change the cost function at difference rates. This may not be a problem for the 3D/3D RQE because the cost function was calculated from 3D voxels. But the cost function for 2D/3D RQE is based on 2D pixels on two orthogonally projected images. Depending on their axes, a translation or a rotation can be in-plane or out-of-plane with respect to their projection image planes, which would result in different gradients in their cost function profiles. Generally speaking, including more nonredundant features in the classifier would potentially improve its performance with the cost of increased computation complexity, which is not significant in our case.

The “gold standard” of each registration task was determined by 50 registration trials that differed by their initial transformation parameters. We believe this is a better method than an exhaustive search method. The results of our 50 optimization attempts show majority of solutions are clustered in a small region within  $0.2^\circ$  and 0.2 mm (1 SD) around the mean solution when a few outliers (about 10) are excluded. This means we have confidence that the best solution we found is very close to its true global optimum. We also think further improvement on the accuracy is not necessary because the discrepancy between the global optimum determined by the mutual information and the true best alignment transform parameter has been reported to be about 1 mm.<sup>(16)</sup> Moreover, an exhaustive search method is not practical. Our optimization based method only needs to evaluate cost function about  $50 \times 300$  times for each registration task. Given six patients and 30 fractions per patient, the computation for “gold standard” using GPU takes about one day. However if an exhaustive search method is used, to cover even a small range such as  $\pm 0.2^\circ$  and  $\pm 0.2$  mm, we need at least 20 sampling points for each degree of freedom (DOF). In the parameter space with six DOF, that would result in 64 million points. Using the same hardware and computation method, the calculation would take more than 11 years.

The performance of the classifier (e.g., sensitivity and specificity) depends on the threshold value that is used to categorize the neural network output. The optimal value for this threshold could be determined if the costs of making both the false positive and negative errors and the frequencies of making those errors were known.<sup>(11)</sup> The choices of those parameters are application- and user-dependent and beyond the scope of this paper. In this study, a threshold value of 0.5 was used, which is based on the arbitrary assumption that the costs of making both errors are equal.

The optimal sampling size for the RQE in our study has been identified to be 5 mm. It should be noted that this value depends on many factors that could change the geometrical properties of the similarity function. Those factors may include treatment site, ROI definition, choice of similarity measure, DRR calculation method, and the imaging parameters such as pixel/voxel spacing, contrast, number of gray levels, imaging artifacts, and the noise level. However, for a specific clinical application, most of those factors are fixed. If we assume a relatively stable imaging system and have a protocol of ROI definition for creating the ROIs of the approximately same size and anatomical location among different patients, the optimal sampling size may be obtained per treatment site.

In this study, the NMI alone was used as the similarity measure and its geometrical properties are used by the classifier. It would be interesting to see if using a different similarity measure or using multiple measures simultaneously would improve the performance of the RQE.

The TAR (i.e., 1.48) we obtained indicates the RQE did effectively reject unsuccessful registration solutions. It also shows the integrated system requires about 48% more computation time compared to the original system. As mention before, this is not a major concern with the advance of the graphics card computation techniques. Of all the registration attempts, 92.4% of

cases had solutions that were finally accepted by the RQE before the maximal number of trials reached. Of all the rejected solutions, 90.9% originate from registering images for Patient 2. This is most likely due to the difference between the definitions of the ROI. A typical lateral projection X-ray image of Patient 2 is shown in Fig. 5(a) and its corresponding one for other patients shown in Fig. 5(b). A small portion of the posterior part of the cranial image was clipped out of the ROI for Patient 2 because of the selection of the imaging isocenter for this particular patient and the limited field-of-view (FOV) of the imaging system. For a scattering beam proton therapy system, the FOV of the X-ray imaging system corresponds to the choice of the treatment head (i.e., the snout size). Pediatric patients with brain tumor are typically treated with 25 cm snout. However, Patient 2 was treated with a  $30 \times 40$  cm snout, which is very rare in our clinic. This results in a reduced FOV in the anterior–posterior direction in the lateral view. If this particular case is excluded from our dataset, the TAR will be reduced to 1.11 (1 SD = 0.09) and FAR increased to 99.2%. With such a low TAR and high FAR, the RQE would greatly increase the robustness of the integrated 2D/3D registration system, without sacrificing much of the efficiency of the system. The user should be aware of the sensitivity of the classifier efficiency to the variations of the ROI definition. Efforts should be made to ensure the imaging content inside ROI when the classifier is constructed is the same as when the classifier is used. Clipping of the imaging content either by a user or an imaging system could adversely affect the efficiency of the classifier.

Our results are based on very limited patient data and a single treatment site. More accurate statistical evaluation would require the test on more patient data. The applications on other treatment sites that have non-rigid deformations with larger magnitude would be more challenging but also interesting.

It would also be interesting to adopt a scheme to adaptively update the RQE using the registrations results when it is used clinically.<sup>(17,18)</sup> The output of the RQE can be checked either online or offline by users. Erroneous decisions by the RQE can be corrected and fed back to the classifier to improve its future performance.

Two other more challenging, but interesting, extensions of this work are the applications of the RQE on multi-modality and/or deformable 2D/3D image registrations.<sup>(19–22)</sup> The key issue is to find features that are good indicators of a successful registration. For example, a simple intensity-based similarity measure is not adequate for a deformable registration because a registration solution with high similarity measure value does not necessarily indicate the deformation is physically meaningful. The features must also be robust so that they do not vary greatly among different patients.

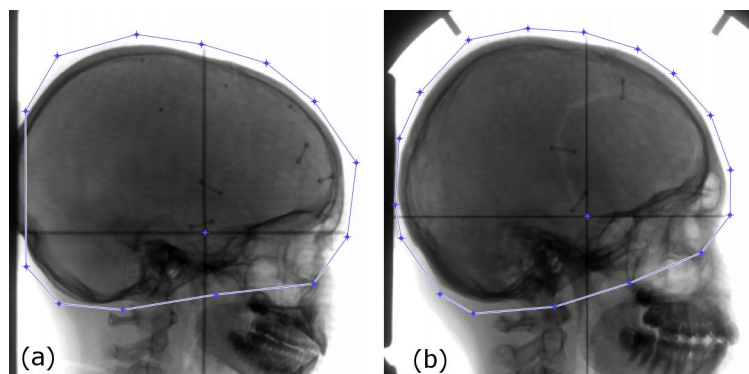


FIG. 5. A typical right–lateral projection X-ray image of (a) Patient 2 and (b) a patient other than Patient 2.

## V. CONCLUSIONS

Our patient study has demonstrated the proposed neural network-based RQE had fairly good performance when used with the NMI in identifying unsuccessfully 2D/3D registrations for daily patient setup. This technique can be used to improve the accuracy and the robustness of a patient setup system that relies on automated rigid 2D/3D registrations.

## REFERENCES

1. Bednarz G, Machtay M, Werner-Wasik M, et al. Report on a randomized trial comparing two forms of immobilization of the head for fractionated stereotactic radiotherapy. *Med Phys*. 2009;36(1):12–17.
2. Cho B, Suh Y, Dieterich S, Keall PJ. A monoscopic method for real-time tumour tracking using combined occasional x-ray imaging and continuous respiratory monitoring. *Phys Med Biol*. 2008;53(11):2837–55.
3. Yan H, Cervino L, Jia X, Jiang SB. A comprehensive study on the relationship between the image quality and imaging dose in low-dose cone beam CT. *Phys Med Biol*. 2012;57(7):2063–80.
4. Kim J, Fessler JA, Lam KL, Balter JM, Ten Haken RK. A feasibility study of mutual information based setup error estimation for radiotherapy. *Med Phys*. 2001;28(12):2507–17.
5. Clippe S, Sarrut D, Malet C, Miguët S, Ginestet C, Carrie C. Patient setup error measurement using 3D intensity-based image registration techniques. *Int J Radiat Oncol Biol Phys*. 2003;56(1):259–65.
6. Rohlfing T, Russakoff DB, Denzler J, Mori K, Maurer CR, Jr. Progressive attenuation fields: fast 2D-3D image registration without precomputation. *Med Phys*. 2005;32(9):2870–80.
7. Ino F, Gomita J, Kawasaki Y, Hagihara K. A GPGPU approach for accelerating 2-D/3-D rigid registration of medical images. Presented at the Fourth International Symposium on Parallel and Distributed Processing and Applications (ISPA'2006), Sorrento, Italy, 2006. Berlin, Germany: Springer-Verlag; 2006.
8. Steininger P, Neuner M, Weichenberger H, et al. Auto-masked 2D/3D image registration and its validation with clinical cone-beam computed tomography. *Phys Med Biol*. 2012;57(13):4277–92.
9. Spoerk J, Gendrin C, Weber C, et al. High-performance GPU-based rendering for real-time, rigid 2D/3D-image registration and motion prediction in radiation oncology. *Z Med Phys*. 2012;22(1):13–20.
10. Russakoff DB, Rohlfing T, Mori K, et al. Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2D-3D image registration. *IEEE Trans Med Imaging*. 2005;24(11):1441–54.
11. Wu J and Samant SS. Novel image registration quality evaluator (RQE) with an implementation for automated patient positioning in cranial radiation therapy. *Med Phys*. 2007;34(6):2099–112.
12. Wu J and Murphy MJ. A neural network based 3D/3D image registration quality evaluator for the head-and-neck patient setup in the absence of a ground truth. *Med Phys*. 2010;37(11):5756–64.
13. Wu J, Murphy M, Samant S. A neural network based registration quality evaluator for 2D-3D image registrations [abstract]. *Med Phys*. 2010;37(6):3128.
14. Skerl D, Likar B, Pernus F. A protocol for evaluation of similarity measures for rigid registration. *IEEE Trans Med Imaging*. 2006;25(6):779–91.
15. Møller MF. A scaled conjugate-gradient algorithm for fast supervised learning. *Neural Networks*. 1993;6(4):525–33.
16. Wu J, Kim M, Peters J, Chung H, Samant SS. Evaluation of similarity measures for use in the intensity-based rigid 2D-3D registration for patient positioning in radiotherapy. *Med Phys*. 2009;36(12):5391–403.
17. Yan H, Yin FF, Zhu GP, Ajlouni M, Kim JH. Adaptive prediction of internal target motion using external marker motion: a technical study. *Phys Med Biol*. 2006;51(1):31–44.
18. Murphy MJ and Pokhrel D. Optimization of an adaptive neural network to predict breathing. *Med Phys*. 2009;36(1):40–47.
19. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*. 1999;18(8):712–21.
20. Hopp T, Dietzel M, Baltzer PA, et al. Automatic multimodal 2D/3D breast image registration using biomechanical FEM models and intensity-based optimization. *Med Image Anal*. 2013;17(2):209–18.
21. Naini A, Patel RV, Samani A. CT-enhanced ultrasound image of a totally deflated lung for image-guided minimally invasive tumor ablation procedures. *IEEE Trans Biomed Eng*. 2010;57(10):2627–30.
22. Naini AS, Pierce G, Lee T, Patel RV, Samani A. CT image construction of a totally deflated lung using deformable model extrapolation. *Med Phys*. 2011;38(2):872–83.