Learning Similarity Measure for Multi-Modal 3D Image Registration

Daewon Lee\textsuperscript{1}, Matthias Hofmann\textsuperscript{1,2}, Florian Steinke\textsuperscript{1,3}, Yasemin Altun\textsuperscript{1}, Nathan D. Cahill\textsuperscript{2,4}, and Bernhard Schölkopf\textsuperscript{1}

\textsuperscript{1}Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany
\textit{firstname.lastname@tuebingen.mpg.de}
\textsuperscript{2}Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK
\texttt{\{mh,cahill\}@robots.ox.ac.uk}
\textsuperscript{3}Siemens Corporate Technology, 81739 Munich, Germany
\textsuperscript{4}Research and Innovation, Carestream Health, Inc., Rochester, NY 14618, USA

Abstract

Multi-modal image registration is a challenging problem in medical imaging. The goal is to align anatomically identical structures; however, their appearance in images acquired with different imaging devices, such as CT or MR, may be very different. Registration algorithms generally deform one image, the floating image, such that it matches with a second, the reference image, by maximizing some similarity score between the deformed and the reference image. Instead of using a universal, but a priori fixed similarity criterion such as mutual information, we propose learning a similarity measure in a discriminative manner such that the reference and correctly deformed floating images receive high similarity scores. To this end, we develop an algorithm derived from max-margin structured output learning, and employ the learned similarity measure within a standard rigid registration algorithm. Compared to other approaches, our method adapts to the specific registration problem at hand and exploits correlations between neighboring pixels in the reference and the floating image. Empirical evaluation on CT-MR/PET-MR rigid registration tasks demonstrates that our approach yields robust performance and outperforms the state of the art methods for multi-modal medical image registration.

1. Introduction

Many medical imaging applications require multi-modal registration, or the precise spatial alignment of images of the same person taken with different scanning devices. This is an intrinsically difficult problem, since corresponding locations in the different images show different intensities, and often there is not a one-to-one mapping between the intensities in the two images. For example in MR-CT registration, black pixels in the MR image can correspond to either bone or air tissue, which have maximally distinct CT values. Thus, one cannot simply use photo-consistency as a similarity score for this task, not even after rescaling the image intensities.

The most popular approach in multi-modal image registration maximizes the mutual information (MI) of the images based on their joint intensity histogram, that is, the two dimensional histogram of the pixel intensities of corresponding point pairs [11]. This approach favors similarity scores that yield the most information about the intensity distribution at one location in the floating image given the intensity at the corresponding position in the reference image. While this is a plausible and very general assumption, it also discards much useful information. Considering again the example of MR-CT registration, a black pixel in MR does not tell us that the output pixel should have a uniquely defined intensity. In fact, we know precisely that the output pixel intensity in CT should be either one, in the case of bone, or zero, in the case of air. Such knowledge can be learned when an exact registration is known, and it can be used to improve existing registration algorithms.

The first successful approaches in learning similarity functions for medical image registration have been undertaken within a generative framework [5, 4, 8, 13]. Leventon \textit{et al.} [5] propose to estimate the underlying joint intensity distribution from registered example image pairs, and then to employ a maximum likelihood (ML) approach to define the alignment measure for new image pairs. Chung \textit{et al.} [4] minimize the Kullback-Leibler (KL) divergence between the learned joint intensity distribution and the joint distribution of the new images. Similarly, Sabuncu \textit{et al.} [8] use the entropic graph-based Jensen-Rényi (JR) divergence for the same minimization problem. One problem, however, with MI and all similarity measures based on sim-
Figure 1. A flowchart of the proposed registration method using the learned similarity function.
the size of overlapping regions [2]. Furthermore, we benchmark against the learned joint density based measure (LJD) [5]. These analyses show that our approach outperforms the state of the art methods for multi-modal medical image registration.

2. Learning the Similarity Function

In this section we describe the steps involved in learning the similarity measure for multi-modal image registration. Fig. 1 sketches the whole methodology, which consists of learning the similarity measure from a pre-registered set of images and applying the learned measure for registering new images.

2.1. Max margin structured prediction

In multi-modal image registration, we are interested in the task of inferring a spatial transformation \( T : \Omega_r \rightarrow \Omega_f \) for a reference image \( I_r : \Omega_r \rightarrow \mathbb{R} \) and its corresponding floating image \( I_f : \Omega_f \rightarrow \mathbb{R} \) image, where \( \Omega_r, \Omega_f \subset \mathbb{R}^d \) are the feasible position sets of the reference and floating images respectively. Given a similarity function \( s \) that quantifies the compatibility of aligned reference-floating image pairs, the optimal transformation of \( (I_r, I_f) \) is found by maximizing the similarity over all possible transformations,

\[
T^* = \text{argmax}_{T \in T} s(I_r, I_f \circ T).
\]

Our goal is to train a similarity function \( s \) over a sample of pre-aligned image pairs such that the empirical cost \( \Delta \) of misregistration, e.g. the target registration error [12], is minimized.

We assume that the similarity of two images decomposes into the similarities of local regions,

\[
s(I_r, I_f \circ T) = \sum_{p \in \Omega \subseteq \Omega_r} s(f(I_r, p), f(I_f, T(p))),
\]

where \( f \) extracts a vectorial description of the local surroundings of point \( p \) from the given image. In this paper we focus on rectangular image patches centered at \( p \), but other feature representations would be possible. One can extend this framework further by incorporate compatibility terms \( \tilde{s} \) for multiple patches of the floating image \( \tilde{s}(f(I_f, p), f(I_f, p')) \). For example, \( \tilde{s} \) can impose spatial consistency of neighboring patches after transformation. Since the standard datasets include a small number of registered full images, we restrict our attention to the formulation that ignores dependencies across floating image patches.

The optimal similarity function should give the highest score to the correctly aligned patch pairs and lower scores to the incorrectly aligned pairs. This is exactly the optimization goal of training a predictor that infers the floating image patch \( y = f(I_f, p') \) corresponding to a given reference image patch \( x = f(I_r, p) \) by maximizing \( s(x, y) \) over all the possible patches of \( I_f \). Let \( D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \) be a sample of correctly aligned image patches. We restrict the space of \( s \) to linear functions over some feature representation \( \phi \) that is defined on the joint input-output space,

\[
s(x, y; w) = \langle w, \phi(x, y) \rangle.
\]

We train the similarity function \( s \) by maximizing the minimum margin of the sample with respect to \( w \), where the margin is defined as \( \gamma(x, y; w) = s(x, y; w) - \max_{y \neq y_i} s(x, y'; w) \) and \( y \neq y_i \) is any structured object from the output space \( Y \) (the set of all patches from the floating image \( I_f \)). If we allow margin violations with linear penalties and we control the norm of \( w \), the optimization problem can be stated as a convex program [10]

\[
\min_{w, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i
\]

s.t. \( s(x_i, y_i; w) - \max_{y \in Y, y \neq y_i} s(x_i, y; w) \geq \Delta(y_i, y) - \xi_i \),

\[
\xi_i \geq 0, \forall i.
\]

Note that due to the non-linearity of the constraints, this is not a quadratic program (QP), but it can be converted into a QP by replacing each margin constraint with a set of linear penalty constraints that is defined on the joint input-output space, \( \phi \), and \( s \) be a sample of correctly aligned image patches. We restrict the number of constraints [2]. Furthermore, we benchmark against the learned joint density based measure (LJD) [5]. These analyses show that our approach outperforms the state of the art methods for multi-modal medical image registration.
where $\psi$ and output space, respectively. In our experiments, we solve the dual problem of (3) via kernel function explicitly induced by a kernel function defined over the joint input-output space via

$$k((x, y), (x', y')) = \langle \phi(x, y), \phi(x', y') \rangle = \psi(x, x') \cdot \psi'(y, y')$$

where $\psi$ and $\psi'$ denote the inner product kernel in input and output space, respectively. In our experiments, we use Gaussian kernels for $\psi$ and $\psi'$. In each iteration of Algorithm 1, we solve the dual problem of (3) via kernel functions and obtain the learned similarity function $s(x, y; w)$ expressed in terms of the dual variables $\alpha$ as

$$s(x, y; w) = \langle w, \phi(x, y) \rangle$$

$$s(x, y; w) = \sum_{i=1}^{n} \sum_{y \in S_i} \alpha_i \psi(x_i, x) \cdot (\psi'(y_i, y) - \psi'(\bar{y}, y)).$$

**Algorithm 1** Cutting plane algorithm

| 1: | Input: $D = \{(x_1, y_1), \ldots, (x_n, y_n)\} $ |
| 2: | $S_i \leftarrow \emptyset$ for all $i = 1, \ldots, n$ |
| 3: | Define $\mathcal{Y}_i, \forall i$ as described in Section 2.3 |
| 4: | repeat |
| 5: | for $i = 1, \ldots, n$ do |
| 6: | $H(y) = \Delta(y_i, y) - s(x_i, y_i; w) + s(x_i, y; w)$ |
| 7: | compute $\hat{y} = \text{argmax}_{y \in \mathcal{Y}} H(y)$ |
| 8: | compute $\xi_i = \max \{0, \max_{y \in S_i} \} \}$ |
| 9: | if $H(\hat{y}) > \xi_i + \epsilon$ then |
| 10: | $S_i \leftarrow S_i \cup \{\hat{y}\}$ |
| 11: | Solve the dual of (3) over $S, S = \cup_i S_i$ |
| 12: | end if |
| 13: | end for |
| 14: | until no $S_i$ has changed during iteration |

all correct pairs from all incorrect pairs are significantly harder than the constraints of (3) and are possibly infeasible. The constraints of (3) impose separation for each training instance but not across all data. Furthermore, (3) naturally incorporates the error function, $\Delta$, into the optimization problem, which is not the case for the binary formulation. In our experiments, 0/1 loss is used for $\Delta$ function.

2.2. Joint Kernel Map

The choice of feature representation $\phi(x, y)$ should reflect the correlations between the components of the input and output variables. We consider feature maps that are implicitly induced by a kernel function defined over the joint input-output space via

$$k((x, y), (x', y')) = \langle \phi(x, y), \phi(x', y') \rangle = \psi(x, x') \cdot \psi'(y, y')$$

where $\psi$ and $\psi'$ denote the inner product kernel in input and output space, respectively. In our experiments, we use Gaussian kernels for $\psi$ and $\psi'$. In each iteration of Algorithm 1, we solve the dual problem of (3) via kernel functions and obtain the learned similarity function $s(x, y; w)$ expressed in terms of the dual variables $\alpha$ as

$$s(x, y; w) = \langle w, \phi(x, y) \rangle$$

$$s(x, y; w) = \sum_{i=1}^{n} \sum_{y \in S_i} \alpha_i \psi(x_i, x) \cdot (\psi'(y_i, y) - \psi'(\bar{y}, y)).$$

2.3. Output Space $\mathcal{Y}_i \subseteq \mathcal{Y}$

When iteratively optimizing problem (3) using Algorithm 1, we have to find the most violated constraint at each iteration (see line 7 of Algorithm 1). This is computationally problematic if the whole output space $\mathcal{Y}$ is searched through exhaustively, since the number of patches in the floating image may be large. A practical approach to solve this problem is to search only over a reduced set $\mathcal{Y}_i \subset \mathcal{Y}$ in the $i$-th iteration, where $\mathcal{Y}_i$ includes neighboring patches of training output patch $y_i$ in the floating image. By assuming that the neighborhood size is larger than the maximum shift of the center point of patch $y_i$ for the optimal transformation $T^*$, our restricted set $\mathcal{Y}_i$ remains plausible for registration purposes.

In some image registration problems, the correct floating image patch for a local reference image patch can be ambiguous in the vicinity of the true positions. If the cost function $\Delta$ does not capture this ambiguity, as in the case of 0/1 loss or target registration error, it is crucial to avoid imposing margin constraints for these floating image patches, as these can lead to infeasibility of the optimization problem. To state this problem more formally, let $(x_i, y_i)$ and $(x_j, y_j)$ be two correctly aligned image patch pairs in $D$. If $x_i$ and $x_j$ are very similar, i.e., $x_i \approx x_j$, and $y_j$ is in the neighborhood of $y_i$, i.e., $y_j \in \mathcal{Y}_i$, then the constraints (4) require

$$s(x_i, y_i; w) > s(x_i, y; w),$$

$$s(x_j, y_j; w) > s(x_j, y; w).$$

Plugging $x_i$ for $x_j$ in (6) and combining with (5) yields the infeasibility problem mentioned above. We overcome this problem by explicitly removing all patches $y$ such that $x \approx x_i$ from $\mathcal{Y}_i$ when $\Delta$ function is 0/1 loss. We define $x \approx x_i$ to be true if the Euclidean distance between the intensities of $x$ and $x_i$ is smaller than some small positive threshold value.

2.4. Selecting training and test image patches

For training our similarity measure for multi-modal registration, we need a training set $D$ of well-aligned image patches. Although one could simply extract patches from all positions of a set of registered training image pairs $I_r, I_f$, this can yield a very large dataset that is not practical to work with. Moreover, we can only expect the patch-wise similarity measure to be informative in regions that have some image contrast and are not uniformly coloured. Otherwise, we can always shift the patches relative to each other without changing the similarity score $s(f(I_r, p), f(I_f, T(p)))$. Thus, during both the training and testing steps, our similarity score is only needed for a subset of the image space.

We therefore define a restricted region $\Omega := \{p \mid \|I_r(p)\| > \theta, p \in \Omega_r\}$, where $\nabla I_r(p)$ denotes the norm of the image gradient at $p$ and $\theta$ is a threshold parameter. This selection of the training and testing steps implies that our similarity score $s$ and the floating image is only computed for patches in $\Omega$.

189
resulting registration algorithms will mainly focus on aligning anatomical boundaries. This is a plausible assumption in cases where images from different modalities fundamentally depict the same anatomical structures, as in CT and MR images [6]. When other modalities such as PET or SPECT are used, this assumption may be violated; however, a learned similarity measure may still be useful as long as one modality contains structural information.

Moreover, using patches that are very close together in \( \Omega \) will always yield the same similarity scores. Such patch pairs thus contribute neither to the training nor to the testing step. Instead, they simply increase the computation time. We therefore constraint the positions in \( \Omega \) to also have at least some minimal distance from each other. The resulting positions \( \Omega \) for one example image pair are shown in Fig. 2.

3. Validating the learned similarity measure separately from multi-modal registration

In this section, we first examine the learned local similarity measure between patches (2), and then we show some properties of the induced image-wise criterion (1). The description of the data sets and experimental setup for learning the similarity function is given in Sections 4.1 and 4.2, respectively.

3.1. Local similarity measure

To illustrate the learned similarity function between patches, we plot its values for a MR-CT example in Fig. 3. We show a reference MR image of a human head in (a), and the corresponding CT image in (b). To show the validity of the learning function \( s \) as a local similarity measure, we pick one position \( x_1 \) in the reference MR image, and compute the scores \( s(x_1, y) \) for all patches \( y \) of the CT image that are within a box of the size of the maximal expected shift. The results are color-coded in (d). While MI and NMI are typically computed between two whole images, they can also be computed between two local image patches. In order to compare NMI against our proposed local approach, we show such similarity scores for patch-wise NMI in (c).

A good similarity measure should be uniquely maximised for the correct match \( (x_1, y_1) \). While this goal is achieved by the learned similarity measure, the NMI has two maxima in the vicinity of the true match; this can easily lead to misregistration. Another shortcoming of NMI can be seen by examining the matches \( (x_1, y_2) \) and \( (x_1, y_3) \) more closely. \( y_2 \) looks relatively similar to \( y_1 \), whereas the appearance of \( y_3 \) is quite different to \( y_1 \). Nevertheless, NMI gives a better score to the pair \( (x_1, y_3) \) than to \( (x_1, y_2) \). This counterintuitive behaviour is not seen for our learning based approach, which may thus be more easily interpretable.

3.2. Image-wise similarity measure

To evaluate the combined image-wise similarity function (1), we conducted the following synthetic experiments. We took a correctly aligned CT-MR image pair, and translated and rotated the CT image in different directions while the MR image was kept fixed. For each CT-MR image pair resulting from such a transformation, we computed our learned similarity score of (1) as well as NMI. NMI was calculated both patch-wise and image-wise (Patch-wise NMI is the sum of single patch pair NMI values, similar to (1), whereas, image-wise NMI means a NMI value between the whole image pair). The results are shown in Fig. 4. Note that the obtained graphs for the learned similarity score are smoother than those for patch-wise NMI. They also show that the learned similarity score has a unique local max-
imum at zero, which is the true transformation. In contrast, the patch-wise NMI curves show multiple local maxima, which will lead any local optimization algorithm to stop short of the true global maximum. On the other hand, image-wise NMI shows comparable smooth graphs, but the global maximum for the image-wise NMI scores is obtained at positions $+1^\circ, -6\text{mm}, +0\text{mm}$ for the rotation and translations, respectively. Thus, image-wise NMI would not yield to a precise alignment of the two datasets.

In summary, these experiments give a first, strong hint that the learned similarity score yields a more accurate and more stable criterion for registration tasks than entropy-based similarity measures. In the next section, we will further validate this claim on a set of real multi-modal registration problems.

4. Validating the learned similarity measure within multi-modal registration

In order to evaluate the effectiveness of the learned similarity function in real applications, we conducted rigid registration experiments on a set of clinical brain image volumes comparing the performance of our learned similarity measure with the state-of-the-art alternatives.

4.1. RIRE dataset

In our experiments, we used CT, MR-T2, and PET image volumes from the Retrospective Image Registration Evaluation (RIRE) Project [12]. Note that all the MR-T2 images have been rectified by the RIRE project. The dataset consists of a training set for learning a similarity function and a test set for evaluating the registration performance. The RIRE project provides the ground truth transformation for one patient (pt-00), which we use for building a pair of pre-aligned training images. The test images are from seven different patients. Since no PET images are available for two of the seven patients, we only use five patients’ images for PET to MR registration (pt-01, pt-02, pt-05, pt-06, and pt-07), but all seven for CT to MR registration. The physical voxel size is $0.65 \times 0.65 \times 4 \text{ mm}^3$ for CT, $1.25 \times 1.25 \times 4 \text{ mm}^3$ for MR, and $2.59 \times 2.59 \times 8 \text{ mm}^3$ for PET images. Axial views of the CT, MR, and PET image volume of pt-01 are shown in Fig. 5.

To evaluate the accuracy of registration results for the various similarity measures, we use the target registration error (TRE) [12]. For each patient, the RIRE project has defined a set of volume of interests (VOIs) which are anatomically meaningful. TRE is the Euclidean distance between the VOI center in the reference image and its corresponding location in the deformed floating image. To obtain the TREs, we submit our transformation for each test image pair to the RIRE website, which computes the TREs and posts them online.

4.2. Experimental setup

We applied Algorithm 1 to RIRE dataset. The hyper-parameters (width of the Gaussian kernel, $\sigma$, and threshold, $\theta$, for the magnitude of image gradients) were selected by 5-fold cross validation among all combinations of values on a finite grid. The candidate parameter values were evaluated via their average TREs for the training patient’s VOIs given registrations computed with the respective parameters. The optimal parameters are reported in Table 1.

In order to obtain a fast and robust registration, we used a multi-resolution approach [11]. We resampled all images isotropically to $6 \times 6 \times 6 \text{ mm}^3$ for the coarse resolution (level 2), and $3 \times 3 \times 3 \text{ mm}^3$ for the finer resolution (level 1), respectively. The patch size is fixed as $30 \times 30 \times 18 \text{ mm}^3$ for all resolutions and modalities. Since generalisation of $s$ over different resolutions cannot be expected, we trained a separate similarity function for each image resolution.

The proposed learned similarity measure (Learned) is compared with several entropy-based measures: the mutual information (MI), normalized MI (NMI), entropy correlation coefficient (ECC), cumulative residual entropy correlation coefficient (CRECC), their modified overlap invariant measures (MMI, MECC, MCREECC) [2], and the learned joint density-based measure (LJD) [5]. Note that LJD estimates a joint density from the training image pair using the mixture of Gaussian method. For all of the comparison measures, we estimated densities (and joint densities) with histograms containing 64 (and 64x64) bins, and we accumulated samples using linear (bilinear) interpolation of histogram bins. We used the registration implementation of

<table>
<thead>
<tr>
<th>Modality</th>
<th>CT to MR</th>
<th>PET to MR</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter</td>
<td>level 2</td>
<td>level 1</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>2.0</td>
<td>4.0</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>$</td>
<td>\Omega</td>
<td>$</td>
</tr>
</tbody>
</table>

Table 1. Learning parameters as determined via cross-validation. $|\Omega|$ denotes the number of training patch pairs.

1TRE statistics for various registration methods are available in [http://www.insight-journal.org/rire/view_results.php]
Figure 4. Image-wise similarity functions (1) for an artificial CT-MR matching task, where the CT image is transformed as in the caption of the table while the MR image is fixed. Top to bottom row represent the learned similarity function values, patch-wise NMI, and image-wise NMI, respectively.

[2] for all similarity functions.

4.3. Results

The experimental results are shown in Fig.6, with numerical values are given in Table 2. The presented values are statistics computed from the TREs of all VOIs from all test patients.

For MR-CT registrations, our learned measure outperforms all standard measures, yielding the lowest mean and median TRE among all measures. A MR-CT registration can be judged successful if the TRE value is smaller than 4 mm, which is the largest voxel dimension of the respective image pairs; otherwise, it should be considered a misregistration [4]. In Table 2, one can see that for the learned measure the maximum TRE is smaller than 4 mm, implying that all of VOIs of the test patients were successfully registered. On the other hand, the maximum TREs for the other similarity measures are all larger than 4 mm, which means they failed to register some VOIs successfully.

Concerning PET-MR registration, our proposed similarity measure also leads to registrations for which the worst case TRE is still smaller than 8 mm (the maximum voxel dimension of the PET images) and shows much tighter quartile range compared to the other measures as shown in Fig.6. However, the median performance is not significantly better than the other measures except MI which performs much worse. This might be due to the low resolution and the high noise levels in the PET images, which renders the PET patches much less informative.

5. Conclusions and Future Works

In this paper, we have shown a method to learn a similarity measure for multi-modal 3D image registration. In contrast to universal similarity measures such as mutual information, our learned score can be adapted optimally for a given task. Furthermore, the new method also allows to exploit structural information contained in neighbourhoods around a voxel of interest. These two effects are achieved through applying a modified version of max-margin structured-output learning methods to this problem. The algorithm makes use of joint kernels for the input and the output space which provide an efficient way of capturing the statistics within and between the respective image modalities to be registered, through the implicit use of an infinite-dimensional feature space representation. Experimental comparison on CT-MR and PET-MR registrations on brain image volumes from the RIRE project shows that our learned similarity measure outperforms other similarity measures in terms of the robustness and accuracy. In our future work, we plan to investigate the use of more sophisticated feature functions, different joint kernel maps, and different $\Delta$ functions. We also plan to apply this approach to non-rigid registration.
<table>
<thead>
<tr>
<th>Modality</th>
<th>Statistics</th>
<th>MI</th>
<th>MMI</th>
<th>NMI</th>
<th>ECC</th>
<th>MECC</th>
<th>CRECC</th>
<th>MCRECC</th>
<th>LJD</th>
<th>Learned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>2.08</td>
<td>2.47</td>
<td>4.51</td>
<td>3.48</td>
<td>6.17</td>
<td>3.97</td>
<td>3.87</td>
<td>3.23</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>1.98</td>
<td>1.99</td>
<td>2.89</td>
<td>2.62</td>
<td>4.50</td>
<td>2.86</td>
<td>2.90</td>
<td>2.41</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>5.15</td>
<td>6.10</td>
<td>12.50</td>
<td>9.38</td>
<td>17.74</td>
<td>10.16</td>
<td>10.07</td>
<td>6.32</td>
<td>3.32</td>
</tr>
<tr>
<td>PET/MR</td>
<td>Mean</td>
<td>8.19</td>
<td>3.00</td>
<td>3.20</td>
<td>3.10</td>
<td>2.96</td>
<td>3.34</td>
<td>3.29</td>
<td>3.07</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>5.16</td>
<td>2.37</td>
<td>2.64</td>
<td>2.54</td>
<td>2.40</td>
<td>3.01</td>
<td>2.95</td>
<td>2.56</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>37.18</td>
<td>7.71</td>
<td>7.57</td>
<td>7.65</td>
<td>7.50</td>
<td>7.53</td>
<td>7.47</td>
<td>7.56</td>
<td>4.81</td>
</tr>
</tbody>
</table>

Table 2. Statistics of VOI TREs (in mm) across all test patients’ image volumes.

Figure 6. Box and whisker plot of the TREs for the RIRE data and different similarity measures. The box has lines at the lower quartile, median, and upper quartile values. Whiskers extend from each end of the box to the adjacent values in the data; the most extreme values within 1.5 times the interquartile range from the ends of the box. Outliers are data with values beyond the ends of the whiskers. Outliers are displayed with a red + sign.

Acknowledgements

The images and the standard transformation(s) were provided as part of the project “Retrospective Image Registration Evaluation”, National Institutes of Health, No. 8R01EB002124-03, PI J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN.

References


