Hierarchical Multi-modal Image Registration by Learning Common Feature Representations

Hongkun Ge\textsuperscript{1,2}, Guorong Wu\textsuperscript{1}, Li Wang\textsuperscript{1}, Yaozong Gao\textsuperscript{1,2}, and Dinggang Shen\textsuperscript{1,2}

Dinggang Shen: dinggang_shen@med.unc.edu

\textsuperscript{1}Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

\textsuperscript{2}Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Abstract

Mutual information (MI) has been widely used for registering images with different modalities. Since most inter-modality registration methods simply estimate deformations in a local scale, but optimizing MI from the entire image, the estimated deformations for certain structures could be dominated by the surrounding unrelated structures. Also, since there often exist multiple structures in each image, the intensity correlation between two images could be complex and highly nonlinear, which makes global MI unable to precisely guide local image deformation. To solve these issues, we propose a hierarchical inter-modality registration method by robust feature matching. Specifically, we first select a small set of key points at salient image locations to drive the entire image registration. Since the original image features computed from different modalities are often difficult for direct comparison, we propose to learn their common feature representations by projecting them from their native feature spaces to a common space, where the correlations between corresponding features are maximized. Due to the large heterogeneity between two high-dimension feature distributions, we employ Kernel CCA (Canonical Correlation Analysis) to reveal such non-linear feature mappings. Then, our registration method can take advantage of the learned common features to reliably establish correspondences for key points from different modality images by robust feature matching. As more and more key points take part in the registration, our hierarchical feature-based image registration method can efficiently estimate the deformation pathway between two inter-modality images in a global to local manner. We have applied our proposed registration method to prostate CT and MR images, as well as the infant MR brain images in the first year of life. Experimental results show that our method can achieve more accurate registration results, compared to other state-of-the-art image registration methods.

1 Introduction

Deformable image registration plays a very important role in medical image analysis [1, 2]. According to the number of image modalities used in registration, the deformable registration methods can be categorized into two types: single-modal and multi-modal image

Correspondence to: Dinggang Shen, dinggang_shen@med.unc.edu.
registration. For the latter, mutual information (MI) [3] or normalized mutual information [4] are widely used, by assuming the existence of statistical relation between intensities of two (multi-modal) images under registration. In the last two decades, MI-based registration methods have achieved many successes in medical imaging area, such as for registration of CT and MR brain images [2].

However, MI-based image registration methods have the following limitations. (1) MI measurement is often estimated from the entire image in order to have sufficient number of intensities to estimate histogram. Since intensity correlation of entire images is often optimized by estimating local deformations point by point, the registration of small structures (e.g., tumor) could be dominated by surrounding large structures, unrelated structures or even background. (2) Since there often exist multiple structures in the images under registration, their intensity correlation could be highly nonlinear and complex, thus making global MI unable to precisely guide local image registration. This can be demonstrated by CT and MR prostate images shown in Fig. 1. Intensities of bladder (blue) and rectum (red) are similar in CT, but different in MR image, in which the intensities of bladder are much brighter than those of rectum. Obviously, it is difficult to find simple intensity correlation to characterize both bladder and rectum.

Feature-based image registration is one of possible approaches to overcome the above-mentioned issues in the conventional MI-based registration methods. This registration approach is often driven by the anatomical correspondences hierarchically established between two images, by using image features (e.g., intensities [5]) extracted from a neighborhood of each key point as a morphological signature. However, since image features computed from different modality images are often distributing differently in the feature space, it is difficult to measure feature similarities. As a result, it is not straightforward to apply the feature-based registration framework to multi-modal images by using their native features.

To solve this problem, we propose to learn the common feature representations for two different modality images via Kernel CCA [6]. Specifically, we extract image features from well-registered image pairs at each key point in a multi-resolution manner, for characterizing anatomical structures in different resolutions. For each resolution, we can apply Kernel CCA to find respective nonlinear mappings that maximize the correlation between two different modality features in common space. In this way, the anatomical correspondences between two multi-modal images can be quantitatively measured by feature similarities between the learned common feature representations.

In the application (i.e., image registration) stage, these learned common features can be used for developing a hierarchical feature-based multi-modal image registration method. Specifically, given two new images, we first extract multi-resolution features in their own spaces. Then, we apply Kernel CCA mapping functions in each resolution to transform features into the common space. These common features can be used for correspondence detection. To improve the robustness of correspondence detection and subsequent registration, we initially select only a small number of key points with distinctive features in both images (i.e., salient points at image boundaries and corners), and let them drive
correspondence detection by matching the learned common features. After tentatively determining correspondences with these key points, we use thin-plate spline (TPS) to interpolate dense deformation field. To further refine the registration results, we gradually add more and more key points to guide image registration. By using hierarchical strategy, the registration performance can be iteratively improved.

We evaluate the performance of our proposed method for the cases of registering CT and MR prostate images, as well as the first-year infant brain MR images. Compared with several state-of-the-art methods such as MI-based registration method (Elastix [7]) and SyN, our method achieves more accurate registration results, in terms of ROI (Region of Interest) overlap degrees.

2 Method

2.1 Learning Common Feature Representations via Kernel CCA

Suppose we have \( T \) pairs of aligned training images, \( \{(I_s^{(1)}, I_s^{(2)})\}_{s=1, \ldots, T} \), where we use superscript to denote image modality. At each point \( v \), we consider using intensities of its surrounding image patch as the native image features, i.e., \( f_s^{(1)}(v) \) and \( f_s^{(2)}(v) \) represent the native features from images \( I_s^{(1)} \) and \( I_s^{(2)} \), respectively. Patch size is fixed here, but it’s straightforward to extend our learning procedure in a multi-resolution manner as we will explain later. We apply a random sampling procedure to obtain \( M \) pairs of native features from \( T \) pairs of training images. Note that each pair of native image features is extracted at the same anatomical location of two aligned modality images. Furthermore, we reshape native features of each point \( v \) into a column vector and thus construct two feature matrices:

\[
X = \begin{bmatrix} x_1, \ldots, x_i, \ldots, x_M \end{bmatrix}^T \quad \text{and} \quad Y = \begin{bmatrix} y_1, \ldots, y_i, \ldots, y_M \end{bmatrix}^T
\]

where the column vectors \( x_i = f_s^{(1)}(v) \) and \( y_i = f_s^{(2)}(v) \). By assuming \( L \) as the total number of image points in each image patch, both \( X \) and \( Y \) are \( M \times L \) matrices. Next step is to learn the common feature representations from \( X \) and \( Y \).

Although CCA is widely used to find the linear mapping, we resort to use Kernel CCA since we believe that the relationship between native image features of two modalities is highly nonlinear. Note that Kernel CCA shares the same idea as SVM that maps the data to a high-dimensional space by kernel functions, which is more flexible to deal with complex mapping than linear CCA. Without loss of generality, we use Gaussian kernel \( \phi \) in this paper, where \( \phi(a, b) = \exp(-\|a - b\|^2/2\sigma^2) \) denotes the Gaussian kernel applied to any two vectors \( a \) and \( b \).

First, we compute the kernel matrices for feature matrices \( X \) and \( Y \) by:

\[
K_X = [\phi(x_i, x_j)]_{M \times M}, \quad i, j = 1, \ldots, M
\]

\[
K_Y = [\phi(y_i, y_j)]_{M \times M}, \quad i, j = 1, \ldots, M
\]
Kernel CCA aims to determine the canonical vectors in the mapped feature space, which can be regarded as the basis vectors for projecting the native image features into a common space. The canonical vectors for \( K_X \) is defined as \( \alpha_{M \times D} = [\alpha_1, ..., \alpha_d, ..., \alpha_D] \), where each \( \alpha_d \) is a \( M \)-dimensional column vector and \( D \) is the minimum rank of matrices \( K_X \) and \( K_Y \), i.e., \( D = \min \left( \text{rank} \left( K_X \right), \text{rank} \left( K_Y \right) \right) \). Similarly, we define the canonical vectors for \( K_Y \) as \( \beta_{M \times D} = [\beta_1, ..., \beta_d, ..., \beta_D] \). The optimization function of Kernel CCA is:

\[
\arg\max_{\alpha_d, \beta_d}(K_X \alpha_d)^T(K_Y \beta_d), \quad d = 1, \ldots, D
\]

\[
s.t. \quad \alpha_d \cdot K_X \alpha_d = 1, \quad \beta_d \cdot K_Y \beta_d = 1.
\]

We sequentially solve each pair of \( \alpha_d \) and \( \beta_d \) by (1) maximizing the correlation between \( K_X \alpha_d \) and \( K_Y \beta_d \) across all \( M \) training pairs, and (2) requiring \( \alpha_d \) to be orthogonal to all previous computed canonical vectors \( \alpha_1, \ldots, \alpha_{d-1} \), and also \( \beta_d \) to be orthogonal to all previous computed canonical vectors \( \beta_1, \ldots, \beta_{d-1} \). Partial Gram-Schmidt Orthogonalization method can be used here to find the optimal \( \alpha \) and \( \beta \) in Eq. 3 [5].

In the testing stage, we first compute the native image features from each testing image. Suppose that we obtain totally \( N \) native image features from each of two different modality images under registration. Following the same procedure in the training stage, we stack the native image features into two matrices \( P = [p_i]_{N \times L} \) and \( Q = [q_i]_{N \times L} \). Then we can get the following kernel matrices:

\[
K_{P_X} = [\phi(p_i, x_j)]_{N \times M}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, M
\]

\[
K_{Q_Y} = [\phi(q_i, y_j)]_{N \times M}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, M
\]

After that, it is straightforward to transform the native image features \( P \) and \( Q \) to the common space by using the learned canonical vectors, \( \hat{P} = K_{P_X} \alpha = K_{Q_Y} \beta \), where \( \hat{P} = [\hat{p}_i]_{N \times D} \) and \( \hat{Q} = [\hat{q}_i]_{N \times D} \) are common feature presentations for two different modality images. Note, if \( p_i \) and \( q_j \) are located at corresponding structure, \( p_i \) and \( q_j \) are supposed to be similar after mapping, though \( p_i \) and \( q_j \) could be very different.

**Multi-resolution Implementation**—To improve discrimination power of features, we learn the common features in a multi-resolution manner. Specifically, we use different Gaussian kernels to smooth each original image. Thus, although we fix the patch size in extracting native features, we can capture not only local but also global features. We apply Kernel CCA in each resolution independently to obtain the respective common feature representations.

### 2.2 Hierarchical Feature-Based Multi-modal Image Registration

Based on the learned common feature representations, we propose a hierarchical feature-based multi-modal image registration method, by iterating the following 3 steps until convergence (or reaching a certain number of iterations).

\[\text{Mach Learn Med Imaging, Author manuscript; available in PMC 2016 February 22.}\]
1) **Key Point Selection**—We first average and normalize gradient magnitude values over the entire image. These values are roughly considered as the importance of each point in registration. Based on this importance map, a small set of key points with higher importance values can be sampled. It is clear from Fig. 2(a) that the key points are more concentrated at rich-edge areas, where importance values are high. Furthermore, we employ a hierarchical mechanism for selecting key points. Fig. 2(a) shows the selected key points from different stages. Purple points are selected in the initial stages, while red and green voxels are gradually selected and added in the subsequent stages after relaxing the key-point selection criterion.

2) **Robust Feature Matching**—For each key point, we search for its corresponding point in another image by comparing feature similarity. Here, we use the Euclidian distance between common features as the similarity measure. Since we learn the common feature representations in a multi-resolution manner, we can detect correspondences from coarse to fine scales. A typical feature matching result is shown in Fig. 2, where one CT image point (indicated by a red cross in Fig. 2(b)) is compared w.r.t. all points in MR image (Fig. 2(c)). Fig. 2(e)–(g) show the respective similarity maps by using high-, middle-, and low-resolution features, where red and blue denote high and low similarities, respectively. The integrated similarity using multi-resolution features is shown in Fig. 2(h). Fig. 2(d) illustrates the similarity map using the linear CCA features. It is clear that (1) the learned common features by Kernel CCA can accurately guide feature matching for finding the matched anatomical structure while features by linear CCA are not able to do so; and (2) multi-resolution feature representations have more discrimination power than any of single-resolution features and thus can make the image registration more robust to superficial matches.

3) **Hierarchical Estimation of Deformation Field**—Given the tentatively estimated correspondences on key points, we consider the selected key points as the control points in the entire image domain, and then apply TPS to interpolate the dense deformation field. Thus, all image points can have their deformations immediately. Note that we here allow only the key points steering the entire registration, while other points only following the deformations of nearby key points. Since the key points can establish correspondences more reliably than others, our hierarchical deformation estimation mechanism can make our registration method robust against noise. After two images are approximately registered, we gradually add more and more key points to join the registration by sampling more key points.

3 Experiments

3.1 MR/CT Prostate Image Registration

In this section, we apply our hierarchical multi-modal image registration method for MR and CT prostate images used in radiation therapy. Since MR image has much better contrast in soft tissues, it is relatively easy to manually label the prostate in MR image. Thus, the key step in radiation therapy is to transfer the prostate label from MR image to CT image by MR/CT image registration. We divide the whole dataset to 10 groups and applied 5-fold
cross-validation (where, in each fold, 8 groups for training and another 2 groups for testing). Each group contains both CT and MR prostate images from the same patient with size $153 \times 193 \times 50$ and resolution $1 \times 1 \times 1 \text{mm}^3$.

In the training stage, we carefully register MR image to CT image with human intervention, for each pair of training images. To obtain middle- and low-resolution images, we apply Gaussian smoothing on the original image with kernel sigma 1.5 and 2.5, respectively. For each training patient, we randomly choose 2,000 image patches with size $7 \times 7 \times 5$ for the original, middle- and low-resolution images. Unless otherwise mentioned, the same parameters are used in the following experiments. Here, we compare with the MI-based registration method in Elastix software package.

Typical registration results by MI-based method and our proposed feature-based registration method are shown in Fig. 3. The yellow contours in four images are prostate/rectum boundaries obtained from same CT image as reference. The red contours denote corresponding structures in original MR image before registration (Fig. 3(b)) and two registered MR images by Elastix (Fig. 3(c)) and our method (Fig. 3(d)), respectively. Apparently, contours on the registered MR image by our method (Fig. 3(d)) are much closer to reference contours of CT image than by MI-based method, suggesting more accurate registration by our method. For quantitative evaluation, the mean and variance of prostate DICE ratio of Elastix is $0.8249 \pm 0.0735$, while proposed method achieves $0.8718 \pm 0.0458$, which is almost 5.7% improvement over Elastix.

### 3.2 Image Registration for Infant Brain MR Images in the First Year of Life

Accurate registration of infant brain images is very important in many brain development studies [8]. Due to rapid maturation and myelination of brain tissues in the first year of life, the intensity contrast of gray and white matter undergoes dramatic changes. As a result, the conventional image registration methods, even working well for adult brains, have limited performance for infant brain images in the first year of life.

In this section, we demonstrate the registration performance of our proposed method on the most challenging registration task, i.e., registering the 6-month-old infant brain images whose image contrasts between white matter (WM) and gray matter (GM) are the poorest in the first year [9] (Fig. 4(a)). In detail, we implement the registration between 6-month-old and 12-month-old brain images across different subjects, in which 6-month-old images are used as reference space. The dataset includes 11 subjects and each subject contains both 6-month-old and 12-month-old MR T1 images with size $256 \times 256 \times 198$ and resolution $1 \times 1 \times 1 \text{mm}^3$. Specifically, the first step is to learn common feature representations between 6-month-old and 12-month-old images. In each cross validation, 9 subjects are chosen as training images, in which 6-month-old and 12-month-old images of the same subjects are carefully registered via their segmented images, while other remaining two subjects serve as the testing subjects. Specifically, the 6-month-old image from one testing subject acts as the target image, while the 12-month-old image from another testing subject acts as the moving image. The whole process above (of our cross validation) is repeated 10 times for evaluation.
Since we have the manual tissue segmentations (WM, GM, and CSF (Cerebral-Spinal Fluid)) and also manual segmentation of hippocampus, we can quantitatively evaluate the registration performance based on both tissue and ROI overlap ratios. Table 1 presents the statistics of tissue DICE and ROI overlap ratio, compared with state-of-the-art Elastix (MI-based), diffeomorphic Demons, and SyN. As we can see, our proposed method achieves the best results over other counterpart registration methods. It is worth noting that our method improves DICE ratio by 2.56% over the second best method, which is also statistically significant under paired t-test (p<0.05). To further evaluate the performance of feature representations learned by Kernel CCA, we directly use native features, i.e., image intensities, instead of Kernel CCA features in our proposed method for comparison. As we can see, our proposed method with Kernel CCA achieves higher DICE ratio than that without Kernel CCA, which demonstrates the capability of using Kernel CCA features for guiding registration. Fig. 4 shows typical registration results by 5 different methods. Again, it is apparent that our proposed method with Kernel CCA (Fig. 4(c)) produces the most accurate results.

4 Conclusion

We have presented a novel hierarchical feature-based multi-modal image registration method. To address the significant difference between native image features of multimodal images, we employ Kernel CCA to learn common feature representations for maximizing the statistical correlation of transformed native features in the common space. By using these common features, we further develop a hierarchical multi-modal image registration procedure through robust feature matching. We have shown promising results of our method in registering CT and MR prostate images, as well as the infant MR brain images in the first year of life.

References

1. Maintz J, Viergever M. A survey of medical image registration. Medical Image Analysis. 1998; 2:1–36. [PubMed: 10638851]
2. Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging. 2003; 22:986–1004. [PubMed: 12906253]
3. Wells I, William M, Viola P, Atsumi H, Nakajima S, Kikinis R. Multi-modal volume registration by maximization of mutual information. Medical Image Analysis. 1996; 1:35–51. [PubMed: 9873920]
4. Knops ZF, Maintz JBA, Viergever MA, Pluim JPW. Normalized mutual information based registration using k-means clustering and shading correction. Medical Image Analysis. 2006; 10:432–439. [PubMed: 16111913]
5. Wu G, Kim M, Wang Q, Shen D. S-HAMMER: Hierarchical Attribute-Guided, Symmetric Diffeomorphic Registration for MR Brain Images. Human Brain Mapping. 2014; 35
6. Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical Correlation Analysis: An Overview with Application to Learning Methods. Neural Computation. 2004; 16:2639–2664. [PubMed: 15516276]
7. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity based medical image registration. IEEE Transaction on Medical Imaging. 2010; 29:196–205.
8. Li G, Wang L, Shi F, Lyall AE, Lin W, Gilmore JH, Shen D. Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. The Journal of Neuroscience. 2014; 34:4228–4238. [PubMed: 24647943]
9. Wang L, Shi F, Gao Y, Li G, Gilmore JH, Lin W, Shen D. Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation. NeuroImage. 2014; 89:152–164. [PubMed: 24291615]
Fig. 1. Example CT and MR prostate images. Red and blue contours are rectum and bladder respectively.
Fig. 2.
Illustration of key points selection and feature matching in identifying the correspondence of a CT point in the MR prostate image. The similarity maps using linear CCA feature, high-, middle-, low-, and multi-resolution kernel CCA features are shown in (d)–(h), respectively.
Fig. 3.
Example of CT and MR registration by our method and MI-based method. (a) CT image used as reference; (b) Original MR image; (c) Registered MR image by MI-based method in Elastix; (d) Registered MR image by our method.
Fig. 4.
Typical registration results by 5 different methods. (a) A 6-month-old image. (b) A 12-month-old image. (c–g) Five registration results of warping 12-month-old image to 6-month-old image by our proposed method with/without Kernel CCA (c–d), SyN (e), Demons (f), and Ela-tix (g), respectively. Yellow contours in (b–g) denote the hippocampal boundaries of 12-month-old image in the original (b) and warped (c–g) spaces, while red contours in all images indicate the same hippocampal boundaries of 6-month-old image in different spaces.
Table 1
Comparison of tissue DICE and hippocampus overlap ratio by 5 different methods.

| Method                  | WM (±)   | GM (±)   | CSF (±)  | Hippocampus (±) |
|-------------------------|----------|----------|----------|-----------------|
| Elastix (MI-based)      | 63.24 ± 2.34 | 72.34 ± 1.24 | 41.98 ± 5.82 | 47.03 ± 6.38    |
| Demons                  | 63.08 ± 2.65 | 70.96 ± 1.70 | 46.47 ± 4.42 | 58.33 ± 7.65    |
| SyN                     | 65.03 ± 2.61 | 73.79 ± 1.53 | 49.17 ± 4.20 | 59.65 ± 7.49    |
| Proposed without KCCA   | 65.37 ± 1.96 | 72.15 ± 1.41 | 42.96 ± 2.22 | 55.27 ± 7.93    |
| Proposed with KCCA      | 67.59 ± 2.54 | 75.07 ± 1.67 | 50.02 ± 6.10 | 61.30 ± 5.88    |