Joint Learning of Appearance and Transformation for Predicting Brain MR Image Registration

Qian Wang[1,2], Minjeong Kim[1], Guorong Wu[1], and Dinggang Shen[1]

[1] Department of Radiology and BRIC, [2] Department of Computer Science
University of North Carolina at Chapel Hill, US
qianwang@cs.unc.edu, {mjkim,grwu,dgshen}@med.unc.edu

Abstract. We propose a new approach to register the subject image with the template by leveraging a set of training images that are pre-aligned to the template. We argue that, if voxels in the subject and the training images share similar local appearances and transformations, they may have common correspondence in the template. In this way, we learn the sparse representation of certain subject voxel to reveal several similar candidate voxels in the training images. Each selected training candidate can bridge the correspondence from the subject voxel to the template space, thus predicting the transformation associated with the subject voxel at the confidence level that relates to the learned sparse coefficient. Following this strategy, we first *predict* transformations at selected key points, and retain multiple predictions on each key point (instead of allowing a single correspondence only). Then, by utilizing all key points and their predictions with varying confidences, we adaptively *reconstruct* the dense transformation field that warps the subject to the template. For robustness and computation speed, we embed the *prediction-reconstruction* protocol above into a multi-resolution hierarchy. In the final, we efficiently refine our estimated transformation field via existing registration method. We apply our method to registering brain MR images, and conclude that the proposed method is competent to improve registration performances in terms of time cost as well as accuracy.

1 Introduction

Image registration has been intensively investigated and widely applied in medical image analysis during past decades. By normalizing individual images into the same space, researchers are able to conduct population-based analysis. A typical setting in processing brain MR images, for instance, often involves selecting template and then aligns each subject image to the template via pairwise registration. After registering all subject images, qualitative and quantitative analyses can be performed in the template space, e.g., to infer the population atlas or to measure group difference.

Pairwise registration between individual subject and the template is usually formulated as an optimization problem [1], with the objective function evaluating both the subject-template *similarity* and the *smoothness* of the non-rigid transformation field. However, this straightforward optimization scheme may suffer when applied to subject images that have significant anatomical differences to the template. Further, the independent registration can hardly take advantages of intrinsic similarity of subject

J.C. Gee et al. (Eds.): IPMI 2013, LNCS 7917, pp. 499–510, 2013.

images, although similar subjects share similar transformations to the template and can potentially help each other in registration.

In fact, performances of image registration can be greatly improved if information from other images in the population is well incorporated. For example, recent studies show very promising alignment of images in the groupwise manner [2]. Although it is difficult to directly warp a significantly different subject to the template in pairwise registration, the problem could become much easier when using other intermediate images as bridges [3-5]. That is, the input subject can deform first towards its nearby intermediate image, and then towards the template by borrowing the already established pathway from the intermediate to the template.

We here propose a novel approach to predict the transformation between a new *subject* image and the *template* by leveraging a set of images that are pre-registered with the template. Specifically, the prediction is achieved by joint learning of patch-based image appearances and transformations, which significantly differs from existing methods. For clarity, we term the images to help estimate the subject-template transformation as *training* images, and denote the transformation in the Lagrangian framework. In particular, we aim to estimate the field $\phi(\cdot)$ that warps the subject (as well as any training image) to the template. That is, for the grid point x in the template, $\phi(x)$ maps it to the subject space. Reversely, $\phi^{-1}(\cdot)$ projects from subject to template. The template voxel x and the subject voxel $\phi(x)$ are regarded as *correspondences* to each other.

The transformation associated with certain subject voxel can be predicted by proper candidate voxels in the training images, if they are correspondences. Fig. 1(a) shows an intuitive example, where voxels from the template, a training image, and the subject are enumerated in red boxes, respectively. For y in the training image, we locate its template correspondence at $x = \phi_{tr}^{-1}(y)$. Further, supposing $\hat{x}$ in the subject to be the correspondence of y (i.e., with similar local appearances in red circles), $\hat{x}$ should locate its template correspondence at $x = \phi_s^{-1}(\hat{x})$ as well. The established correspondence between the template voxel x and the subject voxel $\hat{x}$, bridged by the training voxel y, implies that $\phi_s^{-1}(\hat{x})$ can be predicted by the training field $\phi_{tr}^{-1}(y)$.

We will develop the **P**rediction-**R**econstruction (P-R) protocol to directly estimate $\phi_s(\cdot)$ instead of its inverse $\phi_s^{-1}(\cdot)$, and further apply it to registering brain MR images. For given subject voxel, we identify its correspondences in training images to bridge the subject-template correspondence. To this end, we learn the sparse representation of the patch-based appearance of the subject voxel by all possible correspondence candidate voxels in training images. The sparse learning reveals several training candidates, each of which predicts the local transformation at the confidence relating to the learned sparse coefficient. Therefore, we can *predict* multiple transformations on a set of selected key points, and adaptively *reconstruct* the dense transformation field by considering the confidence of each prediction. For the sake of robustness and computation speed, the P-R protocol is embedded in a multi-resolution hierarchy. Moreover, since the tentative transformation field conveys important voxel descriptions other than appearances, it also participates into the sparse learning to better predict transformations. In the end, we can refine our estimated transformation field via existing registration method efficiently (i.e., with limited iterations of optimization).

The major contributions of this work include:

1. We investigate the appearance-transformation relationship to derive the hierarchy for predicting the transformation field in registration;
2. We apply sparse representation to local appearance/transformation and then propagate the sparse learning to update predictions of the transformation field;
3. We introduce an adaptive way to reconstruct the dense transformation field based on key points and their associated multiple predictions with varying confidences.

We will detail the proposed method in Section 2 and demonstrate its performances in Section 3. We will conclude this work with discussions in Section 4.

2 Method

We introduce the **P**rediction-**R**econstruction protocol in Section 2.1, which predicts transformations at selected key points and then reconstructs the dense field accordingly. The protocol is embedded into a hierarchical framework in Section 2.2, and specifically applied to brain MR image registration. In order to predict the transformation of each key point, we apply sparse representation for joint learning of appearances and transformations in Section 2.3. Finally, in Section 2.4, we show details on adaptive reconstruction of the dense transformation field.

2.1 The P-R Protocol

We aim to estimate the transformation field $\phi_s(\cdot)$ that warps the subject image to the template in the P-R protocol. For the template voxel x, we can predict its correspondence $\phi_s(x)$ in the subject by utilizing the correspondence between x in the template and $\phi_{tr}(x)$ in the training image. In Fig. 1(b), we assume the subject voxel $\hat{x}$ and the training voxel y to be a pair of correspondences. Then, the offset from $\phi_s(x)$ to $\hat{x}$ in the subject image should equal the offset from $\phi_{tr}(x)$ to y in the training image

$$\phi_s(x) - \hat{x} = \phi_{tr}(x) - y. \tag{1}$$

The relation above, proved in the appendix, implies that the subject transformation $\phi_s(x)$ and the training transformation $\phi_{tr}(x)$ are closely related if $\hat{x}$ and y are correspondences in terms of their similar appearances. Based on (1), we can *predict* transformations on selected key points in the template, and then adaptively *reconstruct* the dense transformation field $\phi_s(\cdot)$.

Predicting $\phi_s(x)$ in (1) requires several inputs that are addressed in the following. Specifically, we convert (1) to an incremental refinement model by letting $\hat{x} = \phi_s^{\tau_{l-1}}(x)$ that reflects the tentatively deformed location of the template voxel x at time τ_{l-1}. Then, the prediction at time τ_l, later than τ_{l-1}, is updated by

$$\phi_s^{\tau_l}(x) = \phi_s^{\tau_{l-1}}(x) + \phi_{tr}^{\tau_l}(x) - y, 0 \le \tau_{l-1} \le \tau_l. \tag{2}$$

This incremental refinement is also illustrated in Fig. 1(c), where blue and red dashed arrows denote transformations at τ_{l-1} and τ_l, respectively.

After fixing $\hat{x} = \phi_s^{\tau_{l-1}}(x)$, we further select the training image with $\phi_{tr}^{\tau_l}(\cdot)$ and determine the training voxel y as the correspondence to $\hat{x}$ for predicting $\phi_s^{\tau_l}(x)$. Multiple predictions usually exist for a single template voxel x. Given the two training images in Fig. 1(c), for instance, two predictions on $\phi_s^{\tau_l}(x)$ are available by choosing y_1 from the first training image and y_2 from the second. Moreover, the number of predictions on $\phi_s^{\tau_l}(x)$ can be much higher, since multiple correspondence candidates to $\phi_s^{\tau_{l-1}}(x)$ may emerge from even a single training image. To acquire only reliable predictions on $\phi_s^{\tau_l}(x)$, we apply sparse representation to qualify both transformation $\phi_{tr}^{\tau_l}(x)$ and candidates of y in the training images, with details provided in Section 2.3. The sparse learning also computes the confidence for each specific prediction attempted by an arbitrary combination of $\phi_{tr}^{\tau_l}(x)$ and y.

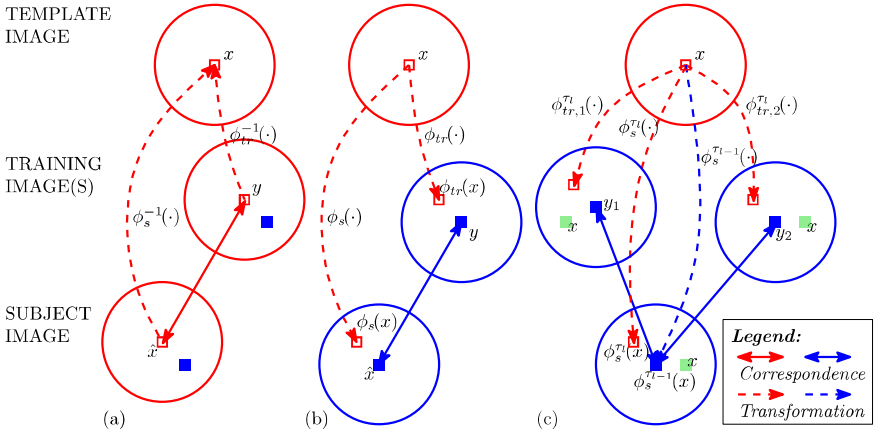

Fig. 1. Illustration of the predictable transformation: (a) The correspondence between the subject voxel $\hat{x}$ and the template voxel x is established due to the correspondence from $\hat{x}$ to y in the training image and from y to x as $x = \phi_{tr}^{-1}(y)$; (b) The template voxel x deforms to $\phi_s(x)$ in the subject, if $\hat{x}$ is the correspondence to y and the offset from $\phi_s(x)$ to $\hat{x}$ equals the offset from $\phi_{tr}(x)$ to y; (c) Multiple predictions are available given several candidates of y, and the prediction fits an incremental refinement model if $\hat{x}$ is assigned according to the tentative transformation.

The prediction is further applied to a set of key points that are automatically selected in the template space. Then, utilizing all key points scattered in the template space and their associated predictions, we are able to reconstruct the dense transformation field $\phi_s^{\tau_l}(\cdot)$ that warps the subject to the template. Note that the reconstruction is adaptive in order to account for multiple predictions (with varying confidences) of each key point, as detailed in Section 2.4.

2.2 The P-R Hierarchy in Brain MR Image Registration

We further embed the P-R protocol above into a hierarchical framework and apply it to registering bran MR images in this work. The hierarchy accounts for evolving resolutions, with the goal to provide better robustness. The transformation field output by the previous level works as initialization to the next level at higher resolution. In particular, after denoting the l-th level to finish at τ_l, the P-R hierarchy is as follows

```
Select a set of template key points 𝕏;
FOR each level l
  Select a subset of key points 𝕏_l ⊆ 𝕏;
  FOR each key point x ∈ 𝕏_l
    Predict φ_s^{τ_l}(x) following the rule in (2);
  END FOR
  Reconstruct the dense field φ_s^{τ_l}(·);
END FOR
```

The hierarchy above functions in the way similar to typical multi-resolution image registration methods. In particular, we use HAMMER [6] to align all training images to the template with recommended configurations (i.e., low-middle-high resolutions). The transformation fields of training images are then utilized at each level of our P-R hierarchy. The key points in the template are mostly sampled near the transitions of different brain tissues (i.e., white matter, grey matter, and cerebrospinal fluid), following the same strategy with [6]. The key points are abundant in contexture information and crucial to accurate alignment of neuroanatomical structures. Locations of all key points are available as part of the training data, thus resulting in no additional computation for a new subject. The subset of key points $\mathbb{X}_l$ in the l-th level is randomly sampled from $\mathbb{X}$, while the size of $\mathbb{X}_l$ enlarges when the level increases (i.e., 1.0e4 for the size of $\mathbb{X}_1$, 4.0e4 for $\mathbb{X}_2$, and 1.6e5 for $\mathbb{X}_3$ in the end). After exhausting all levels, the P-R hierarchy estimates the dense transformation field that registers the subject with the template. We can further refine the estimated transformation field efficiently, i.e., by feeding the field as initialization and running HAMMER with limited number of iterations at the high resolution only.

2.3 Predict Transformations via Joint Learning

The rule in (2) requires specific $\phi_{tr}^{\tau_l}(x)$ and y when attempting to predict $\phi_s^{\tau_l}(x)$. The selection of the two inputs determines the confidence of the resulted prediction. To this end, we evaluate the respective confidences of $\phi_{tr}^{\tau_l}(x)$ and y in predicting $\phi_s^{\tau_l}(x)$, and then define their product as the overall confidence of the resulted prediction. In particular, we *first* select several training images and assign specific confidences to their fields. *Then*, from those selected training images, we locate candidates of y that are correspondences to $\phi_s^{\tau_{l-1}}(x)$ in terms of local image appearances. The combination of certain $\phi_{tr}^{\tau_l}(x)$ and the candidate of y yields a prediction on $\phi_s^{\tau_l}(x)$, along with the respective confidence.

Confidence of $\phi_{tr}^{\tau_l}(x)$

The correspondence detected in brain MR images is meaningful only if restricted within a limited range such that $\left\|\phi_s^{\tau_{l-1}}(x) - y\right\|$ is small. The latent intimacy between $\phi_s^{\tau_{l-1}}(x)$ and y encourages us to select the training image with the field $\phi_{tr}^{\tau_l}(x)$ highly resembling $\phi_s^{\tau_l}(x)$ according to (2). The importance in selecting $\phi_{tr}^{\tau_l}(\cdot)$ can also be observed in Fig. 1(c). The two candidates y_1 and y_2, from two respective training images, yield the same predictions for $\phi_s^{\tau_l}(x)$. However, the training field $\phi_{tr,2}^{\tau_l}(x)$ in the right is more similar to $\phi_s^{\tau_l}(x)$ than the left field $\phi_{tr,1}^{\tau_l}(x)$, in reference to the coordinate of x in green. The correspondence between $\phi_s^{\tau_{l-1}}(x)$ and y_1 thus can only be detected by searching in a much wider area, with more computation, as $\left\|\phi_s^{\tau_{l-1}}(x) - y_1\right\| > \left\|\phi_s^{\tau_{l-1}}(x) - y_2\right\|$. Therefore, we conclude that $\phi_{tr,2}(x)$ is superior to $\phi_{tr,1}(x)$ in predicting $\phi_s^{\tau_l}(x)$ due to the relatively high similarity between $\phi_{tr,2}(x)$ and $\phi_s(x)$. Consequently, we relate the confidence of $\phi_{tr}^{\tau_l}(x)$ in predicting $\phi_s^{\tau_l}(x)$ as their in-between similarity, since more similar $\phi_{tr}^{\tau_l}(x)$ can better approximate $\phi_s^{\tau_l}(x)$ in (2).

In order to determine the confidence of $\phi_{tr}^{\tau_l}(x)$ from individual training image, we investigate the sparse representation of $\phi_s^{\tau_l}(x)$ in terms of all possible training fields at x. The coefficient in representing $\phi_s^{\tau_l}(x)$ indicates the similarity between $\phi_s^{\tau_l}(x)$ and $\phi_{tr}^{\tau_l}(x)$ from a specific training image [7], thus capturing the confidence of $\phi_{tr}^{\tau_l}(x)$ in predicting $\phi_s^{\tau_l}(x)$. The confidence of $\phi_{tr}^{\tau_l}(x)$, however, can hardly be estimated directly since $\phi_s^{\tau_l}(x)$ is not yet predicted. To this end, we evaluate the similarity of $\phi_{tr}^{\tau_{l-1}}(x)$ and $\phi_s^{\tau_{l-1}}(x)$ as an alternative, by assuming that the transformation fields at different levels are mildly changing. To facilitate sparse representation [7] for the evaluation of the confidence of each training field, we denote the set of training images as $\{I_{tr,i}|i = 1, \cdots M\}$ and their transformations as $\{\phi_{tr,i}^{\tau_l}(\cdot)|i = 1, \cdots M\}$. The local patch of the transformation field ϕ centered at x is then vectorized to the column vector $\boldsymbol{\psi}$. The patch for the subject transformation $\boldsymbol{\psi}_s$ can thus be represented as $\boldsymbol{\psi}_s = \boldsymbol{\Psi}_{tr}\mathbf{u}$ by solving

$$\mathbf{u} = \arg\min_{\mathbf{u}} \|\boldsymbol{\psi}_s - \boldsymbol{\Psi}_{tr}\mathbf{u}\|^2 + \alpha\|\mathbf{u}\|_1, \tag{3}$$

where $\mathbf{u} = [u_1, \cdots, u_i, \cdots, u_M]^T$, $\boldsymbol{\Psi}_{tr} = \left[\boldsymbol{\psi}_{tr,1} \cdots, \boldsymbol{\psi}_{tr,i}, \cdots \boldsymbol{\psi}_{tr,M}\right]$, and $\forall u_i \geq 0$. The matrix $\boldsymbol{\Psi}_{tr}$ is the dictionary wrapping up all patches of training fields. The vector $\mathbf{u}$ records the coefficients to linearly represent $\phi_s(x)$, where the superscript τ_{l-1} indicating the level is intentionally omitted for short. The non-negative scalar α imposes the sparseness of $\mathbf{u}$ by penalizing its L_1-norm $\|\mathbf{u}\|_1$. The derived u_i, measuring the *similarity* of $\boldsymbol{\psi}_{tr,i}$ and $\boldsymbol{\psi}_s$ [7], indicates the *confidence* to predict $\phi_s(x)$ by $\phi_{tr,i}(x)$.

Confidence of y

Given the *i*-th training image with the field $\phi_{tr,i}^{\tau_l}(\cdot)$, we aim to locate candidate voxels of y that are correspondences of $\phi_s^{\tau_{l-1}}(x)$. The search for y can be conducted as

typical *correspondence detection*, by evaluating the appearance similarity between $\phi_s^{\tau_{l-1}}(x)$ and the candidate of y in the nearby. Higher similarity measure indicates better reliability of the detected correspondence, thus implying higher confidence of the prediction. To this end, we learn the sparse representation of the patch-based appearance of $\phi_s^{\tau_{l-1}}(x)$ based on candidates of y from all training images. The coefficients computed in sparse learning, also measuring the similarities between $\phi_s^{\tau_{l-1}}(x)$ and individual training candidates [7], tell the confidences in predicting $\phi_s^{\tau_l}(x)$ according to different candidates of y.

We denote y_{ij} for the *j*-th candidate of y from the *i*-th training image. All candidates y_{ij} for y, which satisfy $\left\|\phi_s^{\tau_{l-1}}(x) - y_{ij}\right\| \le \rho_l$ with ρ_l the maximal magnitude of a reasonable correspondence, are enumerated. We then denote $\boldsymbol{\theta}_s$ for the intensity patch centered at $\phi_s^{\tau_{l-1}}(x)$ and $\boldsymbol{\theta}_{ij}$ for the patch at y_{ij}. All training patches are concatenated in the dictionary $\boldsymbol{\Theta}_{tr} = \left[\cdots, \boldsymbol{\theta}_{1j}, \cdots, \boldsymbol{\theta}_{ij}, \cdots, \boldsymbol{\theta}_{Mj}, \cdots\right]$. Consequently, the coefficient vector $\mathbf{v}$ to represent $\boldsymbol{\theta}_s$ based on the dictionary matrix $\boldsymbol{\Theta}_{tr}$ can be obtained by solving

$$\begin{aligned} &\mathbf{v} = \arg\min_{\mathbf{v}} \|\boldsymbol{\theta}_s - \boldsymbol{\Theta}_{tr}\mathbf{v}\|^2 + \beta\|\mathbf{v}\|_1, \\ &\text{s.t.} \quad \mathbf{v} = \left[\cdots, v_{1j}, \cdots, v_{ij}, \cdots, v_{Mj}, \cdots\right]^T, \forall v_{ij} \ge 0. \end{aligned} \tag{4}$$

The non-negative constant β, similar to α in (3), encourages deriving $\boldsymbol{\theta}_s$ by a sparse linear representation of column basis in $\boldsymbol{\Theta}_{tr}$. Note that the training candidate $\boldsymbol{\theta}_{ij}$ can be automatically excluded from $\boldsymbol{\Theta}_{tr}$ if the confidence of $\phi_{tr,i}^{\tau_l}(x)$, or u_i obtained in the step above, happens to be zero. That is, the previous sparse learning upon training fields helps reduce the size of $\boldsymbol{\Theta}_{tr}$, thus the optimization in (4) can be expedited significantly.

Any arbitrary combination of $\phi_{tr,i}^{\tau_l}(x)$ and y_{ij} yields an attempt to predict $\phi_s^{\tau_l}(x)$. In particular, we define the confidence w_{ij} for the attempt as the product of confidences of $\phi_{tr,i}^{\tau_l}(x)$ and y_{ij}, or $w_{ij} = u_i v_{ij}$. The sparsity enforced in selecting $\phi_{tr}^{\tau_l}(x)$ and y results in multiple, but limited number of, predictions with non-zero confidences. In this way, we **1)** avoid local minima if only acquiring a single but incorrect prediction for the key point; **2)** suppress a majority of predictions of low reliability. We further normalize the confidences of each key point by $w_{ij} \leftarrow w_{ij} / \sum \left\| w_{ij} \right\|$, to impose equal priors of all key points.

2.4 Reconstruct Dense Transformation Field

In the next, we reconstruct the dense transformation field to fit the multiple predictions of all key points. We turn to radial basis function (RBF) to represent the field. Suppose the RBF kernel function is $G(\cdot)$ and γ_x the RBF coefficient vector for the key point $x \in \mathbb{X}$, the dense field at the arbitrary location x' is then computed by

$$\phi(x') = \sum\nolimits_{x \in \mathbb{X}} G(\|x' - x\|)\gamma_x. \tag{5}$$

We further define the kernel matrix $\mathbf{G}$, in which the entry at the m-th row and the n-th column is calculated by feeding the distance between the m-th and the n-th key points to the kernel function $G(\cdot)$. If only a single prediction was ever attempted for each key point, the residuals for the dense field to fit the predicted transformations of all key points could be easily computed as $\|\mathbf{\Phi} - \mathbf{G\Gamma}\|^2$, while the predicted transformation (in row vector form) of the m-th key point is recorded in the m-th row of $\mathbf{\Phi}$ and its RBF coefficient in the m-th row of $\mathbf{\Gamma}$ accordingly.

Due to multiple predictions of each key point, we further expand $\mathbf{\Phi}$ and introduce the confidence matrix $\mathbf{W}$ to fitting. Suppose the p-th row of $\mathbf{\Phi}$ records the prediction for the m-th key point with the confidence w_{ij}, we set the entry of $\mathbf{W}$ at the p-th row and the m-th column as w_{ij} and set all other entries in the p-th row as zero. The overall residuals then become $\|\mathbf{\Phi} - \mathbf{WG\Gamma}\|^2$.

Smoothness regularization is essentially important to the reconstruction of the dense field. To this end, the kernel functions $G(\cdot)$ is usually designed as low-pass filter [8]. Further, if $\mathbf{G}$ is positive definite, the regularization can be attained by solving $\mathbf{\Gamma} = \arg\min_{\mathbf{\Gamma}} \|\mathbf{\Phi} - \mathbf{WG\Gamma}\|^2 + \lambda \mathrm{tr}(\mathbf{\Gamma}^T \mathbf{G\Gamma})$ [9], where λ controls the strength of the smoothness constraint. To generate the dense transformation field, the RBF coefficients in $\mathbf{\Gamma}$ are solvable in the following

$$(\mathbf{G} + \lambda(\mathbf{W}^T\mathbf{W})^{-1})\mathbf{\Gamma} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{\Phi}. \tag{6}$$

Here, $\mathbf{W}^T\mathbf{W}$ is a diagonal matrix, where the m-th diagonal entry equals the sum of squares of the confidences for all predictions upon the m-th key point.

The kernel $G(\cdot)$ is designed such that $\mathbf{G}$ is positive definite and $G(\cdot)$ has low-pass response. Abundant choices of RBF kernels are available, i.e., the thin plate splines (TPS) [10, 11] with polynomial decay in frequency domain. Most RBF kernels, however, are globally supported, leading to dense matrix $\mathbf{G}$ and suffering from numerical instability. As a remedy, we use the compactly supported kernel [12] in this work

$$G(\|x' - x\|) = \left(1 - \frac{\|x' - x\|}{c}\right)^2 \cdot \exp\left(-\frac{\|x' - x\|^2}{2\sigma^2}\right), \forall \|x' - x\| \le c. \tag{7}$$

The kernel function $G(\cdot)$ cuts to zero if beyond the compact support, or $\|x' - x\| > c$. The resulted matrix $\mathbf{G}$ is thus sparse and further benefits solving (6).

To tackle the concern on the optimal scale of the kernel, we apply multi-kernel strategy to recursively reconstruct the transformation field [13]. In particular, we fix σ in (7) and adjust c to derive a set of RBF kernels $\{G_h\}$. The size of the compact support for G_h, or c_h, satisfies to $c_{h-1} = 2c_h$. We start reconstruction by applying G_1 to (6). The residuals after G_{h-1} are further reconstructed by G_h. The recursion iterates until that the overall residual, or $\|\mathbf{\Phi} - \mathbf{WG\Gamma}\|^2$, is tiny enough.

3 Experimental Results

We apply the proposed P-R hierarchy to NIREP datasets to verify its time cost and accuracy in estimating the transformation fields. The NIREP datasets contain 16

images, each of which is labeled by 32 ROIs. All images are resampled to the isotropic size of 256×256×256 and pre-processed (including bias correction, skull-stripping, etc.). We then randomly select one image (i.e., the fourth image) as the template and align all other images to the template in affine registration (i.e., using FLIRT). The estimation of the rest non-rigid transformation between each subject and the template is the focus of our study. All experiments are performed on a machine with an Intel Core i5 CPU (3.1GHz, single thread only) and 8G RAM.

We apply HAMMER to carefully register all images to the template, in order to acquire transformation fields as training data. The code of HAMMER is freely available through NITRC[1]. We follow the recommended settings for HAMMER and specify 50 iterations to each of the low, middle, and high resolutions. The outputs of HAMMER are comparable to [14], assuring the quality of the training data.

Time Cost

In order to predict the transformation for a certain subject image by our method, we utilize all other 14 images and their transformations from HAMMER for training. A leave-one-out cross validation is conducted for each subject image. The predicted transformation usually needs further refinement (i.e., applying HAMMER for a limited number of iterations at the high resolution only). In particular, we designate 0 (Setting 1), 10 (Setting 2), and 50 (Setting 3) iterations of refinement, respectively, for the proposed method. Note that Setting 1 produces results with no refinement via registration, while refinement in Setting 3 is the same with the high resolution of standard HAMMER. The time cost for each setting, compared with HAMMER, is summarized in Table 1.

We observe that the proposed method can efficiently predict the dense transformation field in 5.0min by average (Setting 1). With 10 iterations of refinement in Setting 2, the overall time cost rises to 8.6min, which is still significant lower (64.5% less) than direct registration via HAMMER (24.2min in average). With 50 iterations of refinement (Setting 3), which is equal to the high-resolution setting of HAMMER, our method costs comparable time (22.1min in average). Since our method also achieves reasonable accuracy in estimating the transformation field as shown in the next, we conclude that the proposed method can potentially save computation time in deforming the subject to the template.

Table 1. Average time costs of HAMMER and different settings of the proposed method (unit: minute). The number in parenthesis counts iterations of registration.

		Stages of prediction/registration				*Overall time cost*
		Low level	*Mid level*	*High level*	*Refinement*	
HAMMER		1.7 (50)	5.3 (50)	17.2 (50)	NA	**24.2**
Proposed Method	*Setting 1*	0.2	0.9	3.9	NA	**5.0**
	Setting 2				3.6 (10)	**8.6**
	Setting 3				17.1 (50)	**22.1**

[1] http://www.nitrc.org/projects/hammerwml/

Accuracy of Estimated Transformation

The accuracy of the estimated transformation is evaluated in terms of the spatial alignment of neuroanatomical structures. To this end, after warping the ROIs of the subject image to the template, we calculate the Dice overlap ratio between all pairs of corresponding ROIs. We further compute the average Dice ratio, as well as the standard deviation, associated with each ROI and plot the results in Fig. 2. The left panel of Fig. 2 shows the Dice ratios for 16 ROIs in the left hemisphere, while the right panel is for the other 16 ROIs in the right hemisphere.

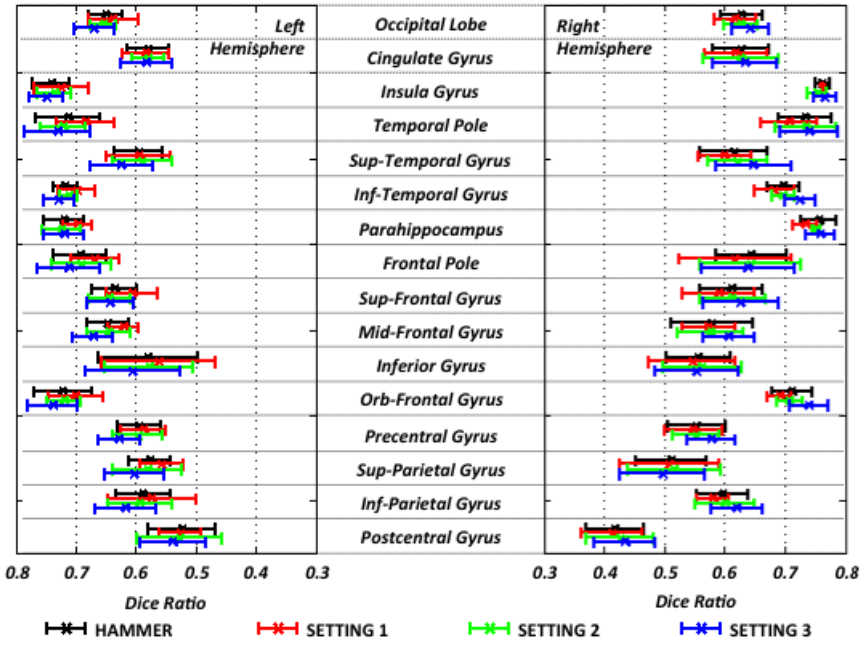

Fig. 2. The average Dice overlap ratios, as well as standard deviations, associated with 32 ROIs in NIREP datasets yield by HAMMER and the three different settings of the proposed method. The left panel shows scores for 16 ROIs in the left hemisphere, while the right panel is for the other ROIs in the right hemisphere.

Without any refinement to the proposed method (Setting 1, red bars in Fig. 2), the predicted transformation leads to an overall Dice ratio 2.01% lower than HAMMER (black bars). However, compared to the output of the middle resolution of HAMMER, our method yields 2.65% higher Dice ratio. The predicted transformation thus provides better initialization to the high-resolution refinement. Moreover, the Dice ratio can be rapidly improved by using only 10 iterations of refinement as in Setting 2 (green bars), which scores only 0.03% lower than HAMMER in the final. Considering the time cost of Setting 2, we conclude that *the proposed method is superior as it can achieve comparable accuracy in registration more efficiently*.

The accuracy in estimating transformations can be further enhanced in Setting 3, which utilizes 50 iterations of refinement. In particular, the average Dice ratio for Setting 3 is 1.87% higher than HAMMER, while its time cost is close to (or slightly less than) the counterpart. We further compute the Dice ratios of white matter and grey matter, as Setting 3 scores 4.05% and 4.28% higher than HAMMER, respectively. As the result, we claim that *our method achieves more accurate transformation fields given computation resources comparable to conventional registration method.*

4 Discussion

We have proposed an efficient approach to predict the transformation field for registering a new subject to the template. The prediction relies on the high correlation between image appearances and transformations. Thus, the sparse learning upon appearances can propagate to the prediction of transformations. After predicting transformations for selected landmarks, the dense field can be reconstructed adaptively. Compared with conventional registration method, the proposed method convincingly improves both time cost and accuracy of the estimated transformation field.

When predicting the transformation for a certain key point, we first select training images via sparse learning of transformations, and then locate correspondence candidate voxels from selected training images only. This sequential solution effectively lowers the number of training candidates involved in correspondence detection. Moreover, with even more training images, the speed performance of our method is not expected to deteriorate, as the learning on transformations is fast and it regulates the complexity of the learning on appearances.

The P-R hierarchy provides a robust way to estimate the final dense transformation field. As level increases, more key points predict their transformations and the reconstructed field is obviously more accurate. Due to initialized transformation field at high levels, the search range in correspondence detection can gradually reduce compared with low levels. As the result, the size of the appearance dictionary matrix $\boldsymbol{\Theta}_{tr}$ decreases, thus obtaining better speed performance for the proposed method.

5 Appendix – The Rule in Predicting Transformations

Fig. 1(b) explains the rule in predicting transformations. Following the transformation field $\phi_{tr}(\cdot)$ that deforms a certain training image to the template, the template voxel x identifies its correspondence $\phi_{tr}(x)$ in the training image. Then, our task is to determine $\phi_s(\cdot)$ that assigns x to its correspondence $\phi_s(x)$ in the subject. We define the perturbation δx for x in the template and specifically let $\phi_s(x + \delta x) = \hat{x}$. We presume that voxels $\hat{x}$ in the subject and y in the training image form a pair of correspondences. The correspondence between $(x + \delta x)$ in the template and y in the training image, bridged by $\hat{x}$ in the subject, can immediately be established as $\phi_{tr}(x + \delta x) = y$. By expanding both transformations in their Taylor series, we have

$$\begin{aligned} \hat{x} &= \phi_s(x+\delta x) = \phi_s(x) + \nabla\phi_s(x)\delta x + \mathcal{O}(\delta x^T \delta x), \\ y &= \phi_{tr}(x+\delta x) = \phi_{tr}(x) + \nabla\phi_{tr}(x)\delta x + \mathcal{O}(\delta x^T \delta x). \end{aligned} \tag{8}$$

An obvious solution of $\phi_s(x)$ to the equations above is

$$\phi_s(x+\delta x) = \phi_{tr}(x+\delta x) + \hat{x} - y, \forall \delta x, \tag{9}$$

such that $\nabla^z\phi_s(x) \equiv \nabla^z\phi_{tr}(x)$ for any $z \in \mathbb{Z}^+$. It leads to (1) when δx vanishes.

Acknowledgement. This paper was supported in part by NIH grants (EB006733, EB008374, EB009634, MH088520, AG041721, and MH100217).

References

1. Rueckert, D., Schnabel, J.A.: Medical Image Registration. In: Deserno, T.M. (ed.) Biomedical Image Processing, pp. 131–154. Springer, Heidelberg (2011)
2. Wu, G., Jia, H., Wang, Q., Shi, F., Yap, P.-T., Shen, D.: Emergence of Groupwise Registration in MR Brain Study. In: Liang, H., Bronzino, J.D., Peterson, D.R. (eds.) Biosignal Processing: Principles and Practices (2012)
3. Jia, H., Yap, P.T., Shen, D.: Iterative multi-atlas-based multi-image segmentation with tree-based registration. NeuroImage 59, 422–430 (2012)
4. Kim, M., Wu, G., Yap, P.-T., Shen, D.: A General Fast Registration Framework by Learning Deformation-Appearance Correlation. IEEE Transactions on Image Processing 21, 1823–1833 (2012)
5. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D.: LEAP: Learning embeddings for atlas propagation. NeuroImage 49, 1316–1325 (2010)
6. Shen, D., Davatzikos, C.: HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging 21, 1421–1439 (2002)
7. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S.: Sparse Representation for Computer Vision and Pattern Recognition. Proceedings of the IEEE 98, 1031–1044 (2010)
8. Myronenko, A., Song, X.: Point Set Registration: Coherent Point Drift. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 2262–2275 (2010)
9. Girosi, F., Jones, M., Poggio, T.: Regularization Theory and Neural Networks Architectures. Neural Computation 7, 219–269 (1995)
10. Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 567–585 (1989)
11. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding 89, 114–141 (2003)
12. Genton, M.G.: Classes of kernels for machine learning: a statistics perspective. The Journal of Machine Learning Research 2, 299–312 (2002)
13. Floater, M.S., Iske, A.: Multistep scattered data interpolation using compactly supported radial basis functions. Journal of Computational and Applied Mathematics 73, 65–78 (1996)
14. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46, 786–802 (2009)