# Robust Anatomical Landmark Detection for MR Brain Image Registration

Dong Han, Yaozong Gao, Guorong Wu, Pew-Thian Yap, and Dinggang Shen

Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, NC, USA
dgshen@med.unc.edu

**Abstract.** Correspondence matching between MR brain images is often challenging due to large inter-subject structural variability. In this paper, we propose a novel landmark detection method for robust establishment of correspondences between subjects. Specifically, we first annotate distinctive landmarks in the training images. Then, we use regression forest to simultaneously learn (1) the optimal set of features to best characterize each landmark and (2) the non-linear mappings from local patch appearances of image points to their displacements towards each landmark. The learned regression forests are used as landmark detectors to predict the locations of these landmarks in new images. Since landmark detection is performed in the entire image domain, our method can cope with large anatomical variations among subjects. We evaluated our method by applying it to MR brain image registration. Experimental results indicate that by combining our method with existing registration method, obvious improvement in registration accuracy can be achieved.

## 1 Introduction

Accurate matching of anatomical structures is a key step in many medical image processing and analysis tasks [1]. However, anatomical structures can vary significantly across different individuals, thus posing huge challenges for accurate correspondence detection. Current methods typically perform local searching to determine anatomical correspondences between images [2,3]. Such approach is limited in two aspects: (1) Due to large inter-subject variability, even after affine registration, corresponding structures could be distant and exhibit different appearances, thus leading to inaccurate matching. (2) The similarity between corresponding points is often measured using predefined local image features that do not necessarily respect the local structural variations.

In this paper, we propose two ways to improve correspondence detection: (1) a global correspondence detection mechanism that can deal with large anatomical differences, and (2) development of feature descriptors that are unique and discriminative for correspondence detection. Specifically, a robust detector is learned for each landmark by using regression forest [4]. Note that a regression forest is an ensemble of randomly-trained binary decision trees that can map a complex input space to continuous output parameters [4]. In our framework, regression forest is used to simultaneously learn the optimal set of features that can

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 186–193, 2014.

best characterize the annotated landmarks and also a set of complex non-linear mappings from the local appearances of image points to their 3D displacements towards each landmark. The learned forests are then used to detect the corresponding landmarks in a new image based on the displacements predicted from each point in the image. Since landmark detection is performed in the entire image domain, our method is able to handle large structural variations. Moreover, since a unique detector is learned for each landmark, more discriminative features can be used to distinguish a landmark from other points in the image.

We evaluate our method by applying it to provide robust initial correspondences and subsequently the initial deformation field for MR brain image registration. The initial deformation field can be refined with a non-rigid registration method such as HAMMER [5]. Experimental results show that these learned landmark detectors are robust to large inter-subject anatomical variability. By combining these detectors with HAMMER, more accurate and robust results can be obtained, especially for the cases with large anatomical differences.

## 2 Methods

### 2.1 Multi-resolution Regression-Guided Landmark Detection

In this paper, landmark detection is formulated as a multivariate non-linear regression problem. Specifically, given a point $\boldsymbol{p} \in \mathbb{R}^3$ in the image, we want to predict the displacement $\boldsymbol{d} \in \mathbb{R}^3$ from $\boldsymbol{p}$ to the latent landmark $\boldsymbol{v} \in \mathbb{R}^3$ based on the local image appearance of $\boldsymbol{p}$. This highly non-linear mapping between $\boldsymbol{p}$ and $\boldsymbol{d}$ is learned by using regression forest, which belongs to random forest family and is specialized for non-linear regression problems. A regression forest consists of a set of binary decision trees. Each tree contains a number of split and leaf nodes. A split node is associated with a split function, which directs an incoming data item to either the left or right child based on a single feature and the learned threshold. A leaf node stores a statistical distribution of outputs of training data items that fall into it, which will be used to determine the output of an unseen data item. In the following, we will introduce (1) in the training stage how the regression forest is learned to capture non-linear relationship between $\boldsymbol{p}$ and $\boldsymbol{d}$, and (2) in the application stage how the learned regression forest can be used to predict the location of the corresponding landmark in a new image.

In the training stage, the training dataset consists of a set of linearly aligned MR brain images $\{I_j\}$. Each image is associated with the annotated location of a corresponding landmark $\{\boldsymbol{v}_j \in \mathbb{R}^3\}$. For each image $I_j$, we sample a set of training image points $\{\boldsymbol{p}_j^i \in \mathbb{R}^3\}$ in a spherical region $\Omega_j$ centered at $\boldsymbol{v}_j$. The output of each $\boldsymbol{p}_j^i$ is calculated as $\boldsymbol{d}_j^i = \boldsymbol{v}_j - \boldsymbol{p}_j^i$. Here, we use $\mathfrak{P}$ and $\mathfrak{D}$ to denote the training points from all images and their corresponding outputs, respectively. The features used to characterize each point are Haar-like features [4], i.e., displaced intensity patch differences. When training a specific tree, we first randomly generate a set of Haar-like features to form a specific feature pool $\boldsymbol{F} = \{f_k\}$. Then, we push $\mathfrak{P}$ and $\mathfrak{D}$ through the tree starting at the root.

At each split node $n$, we randomly generate one set of thresholds $\boldsymbol{T}_k$ for each feature $f_k$ based on the distribution of the feature responses $f_k(\mathfrak{P}_n)$, where $\mathfrak{P}_n$ denotes the training points reaching node $n$. Then, we perform exhaustive search to find the best feature $f_n^*$ and threshold $t_n^*$ that maximize the information gain $\mathrm{G}_n(f,t) = \mathrm{E}(\mathfrak{D}_n) - w_L\mathrm{E}(\mathfrak{D}_n^L) - w_R\mathrm{E}(\mathfrak{D}_n^R)$ after splitting, where $\mathfrak{D}_n$ denotes the set of outputs of training points reaching node $n$, $\mathfrak{D}_n^L = \{\boldsymbol{d} \in \mathfrak{D}_n | f(\boldsymbol{d}) \leq t\}$ and $\mathfrak{D}_n^R = \{\boldsymbol{d} \in \mathfrak{D}_n | f(\boldsymbol{d}) > t\}$ are the output sets for the left and right children after splitting, $w_L = |\mathfrak{D}_n^L|/|\mathfrak{D}_n|$, $w_R = |\mathfrak{D}_n^R|/|\mathfrak{D}_n|$, $\mathrm{E}(\mathfrak{D}_n) = \mathrm{Trace}(\mathrm{Cov}(\mathfrak{D}_n))$ computes the consistency of $\mathfrak{D}_n$, Trace denotes matrix trace, and Cov denotes covariance matrix. The best $f_n^*$ and $t_n^*$ are saved in node $n$ and will be used in the application stage. For each leaf node, we can save the posterior probability $g(\boldsymbol{d}|\boldsymbol{p})$ by summarizing the outputs of all training points reaching this leaf node. Here, we assume that the outputs in each leaf node obey Gaussian distribution. In this case, only mean $\bar{\boldsymbol{d}}$ and covariance $C$ of the outputs need to be saved. The training of the entire tree continues until all training paths reach the leaves.

In the application stage, when a new image comes, we can use the learned forest to predict the displacement from any image point $\boldsymbol{p}$ to the latent landmark. Specifically, for each tree in the forest, we push $\boldsymbol{p}$ through it starting at the root. At each split node $n$, we compute the feature response $f_n^*(\boldsymbol{p})$ and compare it with threshold $t_n^*$. If $f_n^*(\boldsymbol{p}) \leq t_n^*$, $\boldsymbol{p}$ is passed to the left child; otherwise, it is passed to the right child. When $\boldsymbol{p}$ reaches a leaf, the saved $\bar{\boldsymbol{d}}$ in this leaf is regarded as the prediction result of this tree. The final prediction $\widehat{\boldsymbol{d}}$ of a forest is computed as the average over predictions of all trees in the forest. When performing landmark detection, we can adopt the idea of majority voting [4]. Specifically, for each $\boldsymbol{p}$ in the image, we can use the forest to estimate $\widehat{\boldsymbol{d}}$. Then, a vote is given to the point nearest to $\boldsymbol{p} + \widehat{\boldsymbol{d}}$. After voting from all image points, the point that receives the most votes is regarded as the landmark location. However, the major limitation of such voting strategy is that when large anatomical variations exist near the landmark, image points far away from the landmark will not be informative about that and their votes will be less reliable, which may eventually lead to false landmark detection. *To make the prediction more robust to anatomical structure variations, a "point jumping" strategy is adopted in our method.* The basic idea is that instead of directly using $\boldsymbol{p} + \widehat{\boldsymbol{d}}$ as the predicted landmark location [6,7], we repeatedly use the forest to estimate a pathway from $\boldsymbol{p}$ to the landmark, and during this procedure the landmark prediction is iteratively refined. Specifically, given the current location of $\boldsymbol{p}$, we can estimate its displacement $\widehat{\boldsymbol{d}}$. Then, we let $\boldsymbol{p}$ jump to the new location $\boldsymbol{p} + \widehat{\boldsymbol{d}}$. After several such jumps, the final location of $\boldsymbol{p}$ is regarded as the predicted landmark location from $\boldsymbol{p}$. Given a set of image points, we can use "point jumping" to obtain one predicted landmark location from each point. Then, the predicted location with the smallest $||\widehat{\boldsymbol{d}}||$ is regarded as the landmark location. Because the prediction from each point is iteratively refined via "point jumping", a small set of sampled points is generally enough for successful landmark detection, which makes this approach very efficient.

To further improve the robustness of our landmark detection method, we adopt *multi-resolution strategies* in both training and application stages. Specifically, in the training stage, we train one regression forest for each resolution. In the coarsest resolution, we sample the training points all over the image in order to ensure the robustness of the detector; in the finer resolutions, we only sample the training points near the annotated landmark to increase the specificity of the detector. During the application stage, in the coarsest resolution we sparsely sample a set of points from the entire image. Using the "point jumping" on these sampled points, we can obtain an initial guess of the landmark location. In a finer resolution, we take the predicted location in the previous resolution as the initialization and then sample points only near the initialization, thus improving the specificity of landmark detection. In this way, the detected landmark will be gradually refined from coarse to fine resolutions.

### 2.2 Application of the Proposed Method to Image Registration

We evaluate our landmark detection method by applying it to MR brain image registration. Ideally, to achieve this goal, we can learn a large number of detectors that are densely distributed in the whole brain. However, for typical MR brain images, thousands of landmark detectors are needed to model sufficiently the high-dimensional deformation field between two images, which is computationally infeasible. To resolve this issue, we choose an alternative way. Specifically, we learn a relatively small set of landmark detectors (located at distinctive brain regions) to provide robust initial correspondences between two images. Then, a dense deformation field can be interpolated from the initial correspondences on the landmarks, and is used to initialize the deformable image registration. Since the most important and challenging correspondences can be well established by our landmark detectors, the subsequent refinement of deformation field becomes much easier than the direct registration between two images. In this way, the conventional image registration can be divided into two sub-tasks, i.e., (1) accurate initial correspondence establishment by robust landmark detectors and (2) efficient deformation refinement by an existing image registration method. Next, we will describe the landmark-based initialization method in detail.

For detector training, given a group of training MR brain images, we annotate a small set of corresponding landmarks in each image. These landmarks should cover the entire brain, in order to effectively drive the whole deformation field. After landmark annotation, a unique detector is learned for each landmark by using the method described in Subsection 2.1.

When registering two new MR brain images, landmark detection is first performed in each image by using the learned landmark detectors. Then, sparse correspondences can be established between the two images based on the detected corresponding landmark pairs. Next, thin plate spline (TPS) method can be used to interpolate an initial dense deformation field from these sparse correspondences [8]. The generated initial deformation field can be further refined by existing deformable registration algorithms.

## 3 Experimental Results

### 3.1 Image Dataset and Parameter Setting

For evaluation, 62 MR adult brain images were selected from the ANDI dataset with large inter-subject variability. The dimensions of these images are $256 \times 256 \times 256$, and the voxel size is $1 \times 1 \times 1$mm. All images were first linearly aligned by using FLIRT in FSL package [9]. Then, 44 images were selected as the training set and the other 18 images as the testing set. To learn a landmark detector, the training of regression forest was performed on 3 image resolutions. For each resolution, 10 trees were trained. To train a single tree, 6,000 points were sampled from each training image and 1,500 Haar-like features were randomly selected. At each split node, 1,000 thresholds were randomly generated for each feature. The maximum tree depth was 12. In this setting, training one tree costs about 15 minutes. The detection of a single landmark costs about 0.5 second.

### 3.2 Validation of the Proposed Method

In this subsection, the performance of the proposed method is evaluated by training detectors for a few specified landmarks and comparing the detection results with the manually annotated landmark locations in the testing images. Specifically, in this experiment, we manually annotated 6 specific landmarks on the corners/edges of the ventricular regions in both training and testing images. Fig. 1 shows the 6 annotated landmarks in one training image. Then, one detector was trained for each of the 6 landmarks.

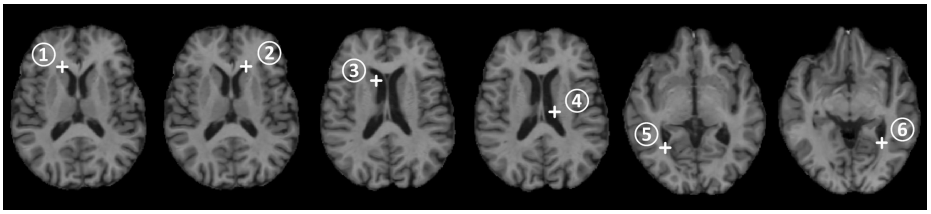


**Fig. 1.** The six manually annotated landmarks on the corners/edges of the ventricular region in one training image. The white cross marks denote the landmark locations.

After training, the detectors were used to detect the corresponding landmarks in each testing image. Then, the detection results were compared with the annotated landmark locations. To better evaluate our method, we compared it with the voting-based method. Table 1 presents the detection errors for the 6 landmarks from both methods. For the first four landmarks, both methods achieve reasonable detection accuracy with good stability. However, for landmarks 5 and 6, the performance of the voting-based method is far worse. This is mainly because the inter-subject variability of these two landmarks is much larger than

**Table 1.** The detection errors for the six landmarks from both voting-based method and our proposed method, respectively. Both mean errors and corresponding standard deviations are shown. (Unit: mm)

| No. | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Voting-Based | $2.69 \pm 1.14$ | $1.94 \pm 1.15$ | $2.36 \pm 1.21$ | $1.38 \pm 0.93$ | $8.60 \pm 9.25$ | $5.38 \pm 4.94$ |
| Proposed | $2.62 \pm 1.26$ | $2.40 \pm 1.08$ | $1.73 \pm 1.29$ | $1.48 \pm 0.77$ | $2.34 \pm 1.53$ | $2.77 \pm 1.24$ |

that of the other four landmarks. In this case, the prediction from the voting-based method is more likely to be dominated by the wrong votes. However, our method is much more robust to anatomical variability among individuals.

### 3.3 Evaluation of the Proposed Method in Image Registration

In this experiment, the proposed method was further evaluated by applying it to assist MR brain image registration, where a set of landmark detectors were learned to provide a good initial deformation field between the two images. Further refinement of the remaining deformation field was performed by the HAMMER algorithm. In this experiment, each MR brain image was further segmented into white matter (WM), gray matter (GM), ventricle (VN), and cerebrospinal fluid (CSF) regions by FAST in FSL package [10]. Here, we regard these segmentations as the ground truth for evaluating the accuracy of the registration results by comparing the tissue overlap ratio.

In this experiment, 1,350 landmarks were annotated that covered the whole brain. After training all the landmark detectors, these detectors were used to establish rough initial correspondences between the two images. Here, one testing image was selected as the fixed image, and all other testing images were warped to it. Registration results from the landmark-based initialization, HAMMER, and the combination of these two methods were obtained and compared. The Dice ratios [11] (and their standard deviations) of these methods, computed for CSF, VN, GM and WM, are reported in Table 2. It can be observed that using landmark-based initialization increases the Dice ratios (and decreases the standard deviations) for all tissue types. This indicates that the correspondences established by the landmark detectors are accurate and robust. Refinement using HAMMER further improves the registration accuracy to a level that is better than the original HAMMER. To better illustrate the advantage of our approach, we select 5 images from the testing set that are substantially different in appearance from the fixed image. Fig. 2 presents the registration results for the 5 images. From Fig. 2 we can see that after landmark-based initialization, the inter-subject anatomical differences of these images are reduced significantly, which makes further refinement much easier than directly registering two images from scratch. The improvements in registration accuracy from the Landmark+HAMMER approach can be easily observed from these results.

**Table 2.** The Dice ratios and the corresponding standard deviations of CSF, VN, GM and WM regions before and after registration by using (1) the landmark detectors only, (2) the Landmark+HAMMER approach, and (3) the original HAMMER algorithm

| Region | Before Reg. | Landmark | Landmark+HAMMER | HAMMER |
|---|---|---|---|---|
| CSF | $40.3\% \pm 2.32\%$ | $48.6\% \pm 2.28\%$ | $62.7\% \pm 1.95\%$ | $61.1\% \pm 2.16\%$ |
| VN | $54.7\% \pm 10.5\%$ | $78.6\% \pm 3.36\%$ | $85.9\% \pm 1.51\%$ | $76.8\% \pm 11.1\%$ |
| GM | $45.2\% \pm 2.06\%$ | $50.4\% \pm 1.87\%$ | $66.1\% \pm 1.58\%$ | $65.1\% \pm 1.85\%$ |
| WM | $60.0\% \pm 2.65\%$ | $67.0\% \pm 1.39\%$ | $79.9\% \pm 1.28\%$ | $77.8\% \pm 2.09\%$ |

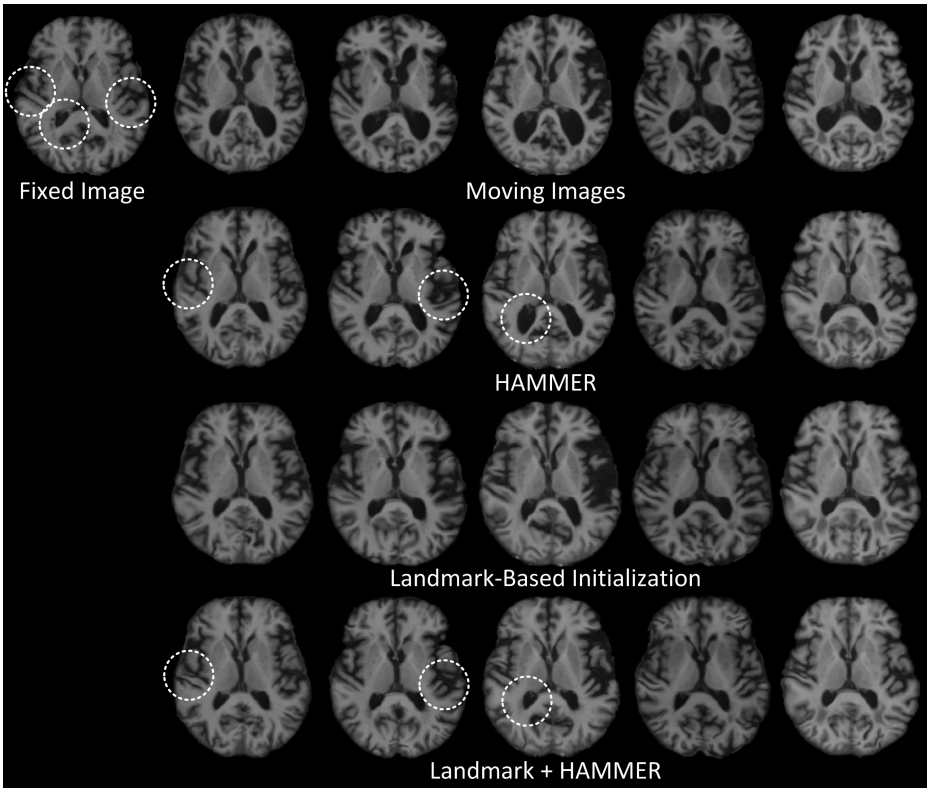


**Fig. 2.** Registration results on 5 testing images that are substantially different in appearance from the fixed image. Row 1: The fixed image and the original moving images. Row 2: Registration results given by HAMMER alone. Row 3: Registration results given by landmark-based initialization. Row 4: Registration results given by further refinement using HAMMER.

## 4 Conclusion

In this paper, we analyze the limitations of conventional anatomical correspondence detection methods, and then propose a multi-resolution regression-guided

landmark detection method to overcome these limitations. In our method, regression forest is employed to learn the complex non-linear mappings from local patch appearances of image points to their 3D displacements towards the landmarks. The learned regression forest is used to predict the locations of the corresponding landmarks in new images. We evaluate the proposed landmark detection method by applying it to assist MR brain image registration. Experimental results show that, with the help of reliable initial correspondences established by the proposed method, structural differences between two images can be reduced significantly. By refining the initial correspondences with the HAMMER algorithm, the final registration results are more accurate and robust than those given by HAMMER alone. Our future research will include (1) further improving landmark detection accuracy and (2) incorporating our landmark detection method with other non-rigid registration algorithms to further evaluate the advantage of our method in helping initialize image registration algorithms.

## References

1. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic Segmentation of MR Images of the Developing Newborn Brain. Med. Image Anal. 9, 457–466 (2005)
2. Shen, D.: Fast Image Registration by Hierarchical Soft Correspondence Detection. Pattern Recogn. 42, 954–961 (2009)
3. Wu, G., Kim, M., Wang, Q., Shen, D.: S-HAMMER: Hierarchical Attribute-Guided, Symmetric Diffeomorphic Registration for MR Brain Images. Hum. Brain Mapp. 35, 1044–1060 (2014)
4. Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K.: Regression Forests for Efficient Anatomy Detection and Localization in Computed Tomography Scans. Med. Image Anal. 17, 1293–1303 (2013)
5. Shen, D., Davatzikos, C.: HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration. IEEE Trans. Med. Imaging 21, 1421–1439 (2002)
6. Zhou, S.K., Comaniciu, D.: Shape Regression Machine. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 13–25. Springer, Heidelberg (2007)
7. Zhou, S.K., Zhou, J., Comaniciu, D.: A Boosting Regression Approach to Medical Anatomy Detection. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2007)
8. Bookstein, F.L.: Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Trans. Pattern Anal. Mach. Intell. 11, 567–585 (1989)
9. Jenkinson, M., Smith, S.M.: A Global Optimisation Method for Robust Affine Registration of Brain Images. Med. Image Anal. 5, 143–156 (2001)
10. Zhang, Y., Brady, M., Smith, S.: Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. IEEE Trans. Med. Imaging 20, 45–57 (2001)
11. Van Rijsbergen, C.J.: Information Retrieval, 2nd edn. Butterworth-Heinemann, London (1979)