Automatic Categorization of Anatomical Landmark-Local Appearances Based on Diffeomorphic Demons and Spectral Clustering for Constructing Detector Ensembles

Shouhei Hanaoka[1], Yoshitaka Masutani[1,2], Mitsutaka Nemoto[1], Yukihiro Nomura[1], Takeharu Yoshikawa[3], Naoto Hayashi[3], and Kuni Ohtomo[1,2]

[1] Department of Radiology,
[2] Division of Radiology and Biomedical Engineering, Graduate School of Medicine
[3] Department of Computational Diagnostic Radiology and Preventive Medicine,
The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
`hanaoka-tky@umin.ac.jp`

Abstract. A method for categorizing landmark-local appearances extracted from computed tomography (CT) datasets is presented. Anatomical landmarks in the human body inevitably have inter-individual variations that cause difficulty in automatic landmark detection processes. The goal of this study is to categorize subjects (i.e., training datasets) according to local shape variations of such a landmark so that each subgroup has less shape variation and thus the machine learning of each landmark detector is much easier. The similarity between each subject pair is measured based on the non-rigid registration result between them. These similarities are used by the spectral clustering process. After the clustering, all training datasets in each cluster, as well as synthesized intermediate images calculated from all subject-pairs in the cluster, are used to train the corresponding subgroup detector. All of these trained detectors compose a detector ensemble to detect the target landmark. Evaluation with clinical CT datasets showed great improvement in the detection performance.

Keywords: anatomical landmark, diffeomorphic demons, spectral clustering.

1   Introduction

Detection algorithms for anatomical landmark positions have a wide range of applications in medical image processing. For example, algorithms are frequently used to determine the initial condition of statistical shape models for the segmentation of various organs. A simple template matching technique has been frequently used in practice on the assumption that a landmark has sufficiently small inter-individual variations in its local appearance (i.e., the intensities of the voxels around the landmark point). However, many practically important anatomical landmarks in the human body, such as tips of bone structures or bifurcations of vessels, have considerably large inter-individual variations in appearance. Although difficult, detection of these

landmarks is a critical task in medical image processing, such as the precise segmentation of organs [1] or the identification of anatomical anomalies [2]. Therefore, a more reliable method for detecting such landmarks is desired.

It is especially challenging to develop a general but reliable method to detect diverse types of landmarks. One problem is that the intensity of voxels within the landmark-local appearance can greatly change according to the various shapes of the anatomical structure on which the landmark is defined. For example, the inter-individual variety in the length of the 12th rib tip significantly changes the voxel values of the local appearance (Fig. 1). To overcome this problem, a method that can interpolate the shapes—not appearances—between two given training datasets by synthetically augmenting the amount of the training datasets would be beneficial. Developing a shape interpolator for this purpose is one of the goals of this study.
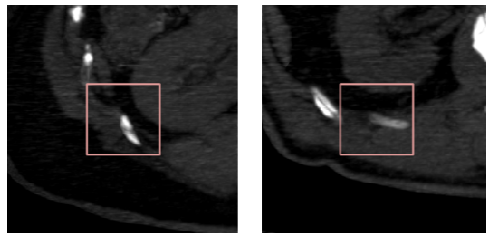

Fig. 1. An example of a landmark difficult to detect due to inter-individual variety (the tip of the 12th right rib in axial images from two datasets)

Another problem is that the shape of the anatomical structure on which a landmark is defined can be multimodal, which means, the structure has several subgroups of shape varieties. If the distribution of appearance has several local maxima (i.e., the distribution cannot be represented by a concave probability function), a significant problem arises if one tries to approximate it by a multidimensional Gaussian distribution (or any other concave distribution). Unfortunately, in our experience in defining and detecting landmarks, this problem is not uncommon (as shown in Fig. 3). In such a case, standard modeling with principal component analysis (PCA) is insufficient. However, at least two different approaches to solve this problem can be considered: 1) sub-categorizing the appearances of training datasets and dividing them into several subsets, or 2) applying machine learning techniques that do not need any assumption of concavity in the probability distribution. The second goal of this study is to develop a method to sub-categorize the appearances (the first of the two approaches).

Up to now, a number of detection methods for multiple landmarks with high reliability have been reported. Seifert et al. [1] reported a framework for detecting landmarks that includes 1) key slice detectors, 2) landmark point detectors that utilize a probabilistic boosting tree with 3D Haar-like features, and 3) a belief propagation algorithm on a predefined inter-landmark graph network. In a previous paper [3], we reported another framework that consists of 1) single-landmark detectors with PCA-based appearance models, 2) false positive (FP) reduction with MadaBoost-based classifiers, and 3) final determination using prior knowledge on inter-landmark distances. Both of these frameworks use a boosting-based approach to handle a wide variety of landmark-local

appearances. In our previous approach, however, it was assumed that the multimodal distribution problem described above affected the framework, because the framework relies on a PCA-based appearance model in the first step.

In this study, we propose a novel method to 1) categorize the given subjects, i.e., training CT datasets, according to landmark-local shape variations of the anatomical structure, 2) synthesize an appearance (a local volume around the landmark) that has the intermediate anatomical shape between any given pair of training datasets, and 3) utilize them to improve the detection performance of landmark detectors through composing a "detector ensemble," which consists of detectors, one of which is designed for only a single subgroup of the local appearance of the target landmark. We used the detector proposed in our previous work [3] as a baseline and improved its performance by replacing the detector with a detector ensemble.

2   Methods

The proposed method consists of three components: 1) shape interpolation and similarity evaluation using the diffeomorphic demons registration algorithm [4], 2) clustering of the appearance variation by using the spectral clustering algorithm [5] and 3) training of subgroup-specific detectors (Fig. 2).

2.1   Shape Interpolation and Similarity Evaluation

2.1.1   Diffeomorphic Demons Registration between Each Appearance Pair

The diffeomorphic demons algorithm is a non-rigid registration algorithm that ensures diffeomorphism of the resulting deformation field [4]. We chose this algorithm because of its invertibility and relatively low computational cost. Invertibility is critical in our application for synthesizing the intermediate shape image.

Let $I_a(\mathbf{p})$, $I_b(\mathbf{p})$ be image functions of the two images to be registered. The vector $\mathbf{p}$ indicates an arbitrary point in the image. Suppose that a 3-D transformation s is defined by a displacement vector field $\mathbf{s}(\cdot)$ such that point $\mathbf{p} + \mathbf{s}(\mathbf{p})$ is warped to point $\mathbf{p}$. Then an image $I_a \circ s$, which is the deformed image of I_a by the transformation s, is defined as

$$(I_a \circ s)(\mathbf{p}) = I_a(\mathbf{p} + \mathbf{s}(\mathbf{p})). \tag{1}$$

We modified the diffeomorphic demons algorithm to be symmetrical for our purpose. (Vercauteren et al. introduced another solution using the Baker-Campbell-Hausdorff (BCH) approximation [4], but we chose this modification because of its simplicity.) The algorithm is as follows:

Algorithm 1 (Diffeomorphic demons with symmetric forces)

1. Initialize the speed image $\mathbf{u}(\mathbf{p})$ to be $(0,0,0)^{\mathrm{T}}$ at any point $\mathbf{p}$ in the image.
2. Calculate a pair of the symmetrical deformation field c_a, c_b by calculating the exponential (in the sense of Lie algebra theory, as described in [4]) of the speed image as follows:

$$c_a = \exp(\mathbf{u}), c_b = \exp(-\mathbf{u}) \tag{2}$$

3. For diffusion-like regularization, apply Gaussian smoothing with kernel size σ_{diff} to the vector fields $\mathbf{c}_a$ and $\mathbf{c}_b$. Let the smoothed results be $\mathbf{s}_a$ and $\mathbf{s}_b$.
4. Update the speed image by

$$\mathbf{u} \to \mathbf{u} - \frac{I_a \circ s_a - I_b \circ s_b}{\|\mathbf{J_p}\|^2 + |I_a \circ c_a - I_b \circ c_b|^2} \cdot \mathbf{J_p}^T \tag{3}$$

Here, $(\mathbf{p})$ is omitted and $\mathbf{J_p} = -\frac{1}{2}\left(\nabla_p^T(I_a \circ s_a) + \nabla_p^T(I_b \circ s_b)\right)$ is a symmetrized Jacobian of the current images.

5. For fluid-like regularization, apply Gaussian smoothing with kernel size σ_{fluid} to the speed vector fields $\mathbf{u}$.
6. Iterate steps 2 through 5 for a sufficient number of times.

This is a fully symmetric version of diffeomorphic demons which deforms each of two images simultaneously so as to fit to the other. Note that the deformation c_a is an inverted version of c_b, and *vice versa*, in the sense that $c_a \circ c_b = c_b \circ c_a = \text{Id}$ (the identity transformation). Finally, the intermediate image is calculated as a mean of two deformed images, or, $\frac{1}{2}\left((I_a \circ s_a)(\mathbf{p}) + (I_b \circ s_b)(\mathbf{p})\right)$.

2.1.2   Calculating the Similarity Measure between Each Appearance Pair

Prior to clustering the landmark-local appearances extracted from the given training datasets, the shape similarity measure must be calculated adequately. In this study, the shape similarity is estimated by the squared norm of the divergence of the deformation vector fields $\mathbf{s}_a$ and $\mathbf{s}_b$ calculated above. That is,

$$\text{Sim}(I_a, I_b) = \exp\left\{-\frac{1}{2\,\sigma_{sim}^2}\int_\Omega \frac{1}{2} \cdot (\|\nabla \cdot \mathbf{s}_a\|^2 + \|\nabla \cdot \mathbf{s}_b\|^2)dv\right\}, \tag{4}$$

where Ω is the domain of the landmark-local appearance and dv is the volume element. The parameter σ_{sim} was determined empirically in this study (the experimental data is omitted), as well as σ_{fluid} and σ_{diff} in 2.1.1.

2.2   Clustering the Landmark-Local Appearances

Spectral clustering methods are algorithms that cluster data samples using eigenvectors of matrices derived from the data. In [5], Ng et al. presented a simple spectral clustering algorithm and the theoretical background. The following is a brief description of the algorithm to cluster n given datasets into k clusters:

Algorithm 2 (Spectral clustering)

1. Form an affinity matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ as its factor A_{ij} equals to the similarity measure between the i-th and j-th appearances. The diagonal elements A_{ii} are set as zero.

2. Define a diagonal matrix $\mathbf{D}$ so that $D_{ii} = \sum_{j=1}^{n} A_{ij}$, and a matrix $\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$.

3. Find the k largest eigenvectors of $\mathbf{L}$. Let them be $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$. Form a matrix $\mathbf{X} = (\mathbf{x}_1\ \mathbf{x}_2\ \dots \mathbf{x}_k) \in \mathbb{R}^{n \times k}$.

4. Form the matrix $\mathbf{Y}$ from $\mathbf{X}$ by renormalizing each row of $\mathbf{X}$, i.e., $Y_{ij} = X_{i,j}/\sqrt{\sum_{j=1}^{n} X_{i,j}^2}$.

5. Treat each row of $\mathbf{Y}$ as a sample point in $\mathbb{R}^k$, then cluster these sample points into k clusters via the K-means algorithm.

Here, the number of cluster k is a parameter determined in advance. In this study, we chose k from 2 to 6 for each landmark by the automatic cluster number determination method proposed by Zelnik-Manor et al [6].
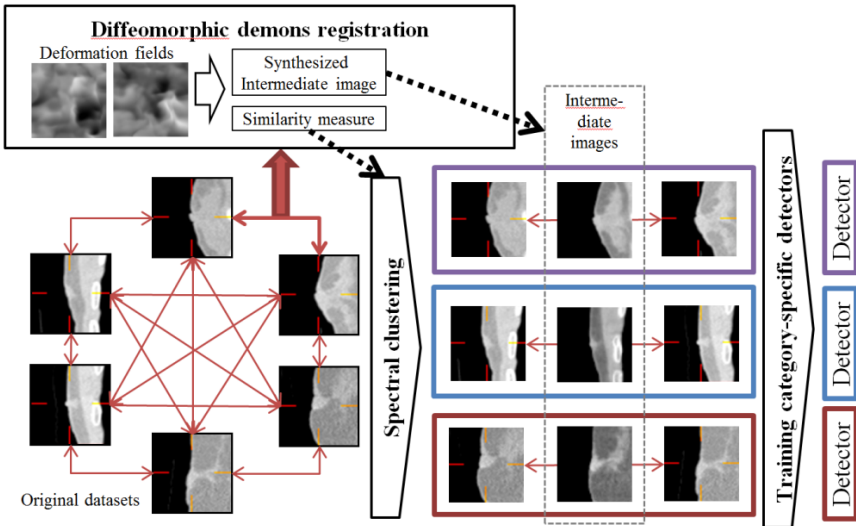

Fig. 2. A diagram of non-rigid registration, clustering and training of category-specific detectors. The images are sagittal cross-sections of the target landmark (left nipple).

2.3   Training of Baseline Detectors and Detector Ensembles

In this study, 50 training datasets were used to train the detector for each landmark. The positions of landmarks were input by a board-certificated radiologist.

The baseline detector. As previously noted, we used the landmark detection framework presented in [3] for evaluation of the proposed method. A detector is composed of two processes, A) PCA-based candidate detection and B) MadaBoost-based likelihood estimation. The detector outputs a list of candidate points with their estimated likelihoods. More details are available in [3]. Note that the proposed method is not

specific to our detector but can be applied to any detection system designed to output a list of candidates with likelihoods, because our algorithm only does clustering and augmenting of the training datasets prior to detector training.

The detector ensemble. After the local appearances of each target landmark from all training datasets were divided into clusters (described in 2.2.), each subgroup (=cluster) detector was trained in the same manner as that of the baseline detector. In the training for candidate detection, not only the original datasets but also the synthesized intermediate images in the cluster were used as the training datasets.

2.4   Experimental Settings

The proposed method was evaluated with 13 landmarks: inferior tip of the sternum, tip of the coccyx, tips of the left/right 12th ribs, umbilicus, superior tips of the right/left kidneys, first bifurcation of the intrahepatic portal vein, inferior tip of the liver, roots of the celiac and superior mesenteric arteries, and right/left nipples. Most of the landmarks were difficult to detect by the baseline detector.

The evaluation was performed with 10 CT datasets (other than the training datasets) without intravenous contrast agent administration. The voxel size of all datasets was $0.977 \times 0.977 \times 1.250$ mm.

In the real landmark detection process, the outputs of the subgroup detectors within any ensemble were integrated as follows. 1) All subgroup detectors, as well as the baseline detector, were applied individually. 2) The resulting candidate lists were concatenated and sorted by likelihood. 3) Finally, only the 100 candidates with the largest likelihoods were used and the others were discarded. For comparison, the outputs of each baseline detector were also truncated to have 100 candidates.

The performance of each detector (baseline or ensemble) was evaluated by a receiver operating characteristic (ROC) analysis by changing the cut-off threshold of the candidate likelihood. The performance was scored by the criterion of the area under the curve (AUC). Each detection result was judged as correct if the detected point was within the range of 2 cm from the manually inputted gold standard point.

3   Experimental Results and Discussions

The results are shown in Table 1. Examples of the clustering results and the ROC curves are shown in Figs. 3 and 4. The AUCs were improved in 11 landmarks (84.6%). The average and standard deviation of the improvements was 0.13 ± 0.12. Remarkable improvements were seen in landmarks even when corresponding baseline detectors showed very poor performances. Therefore, we believe that our method successfully clustered and augmented the training datasets with multimodal variation.

In a related study, Heckemann et al. [7] proposed an automatic segmentation method by non-rigidly registering multiple labeled images (atlases) into the given unseen

image and integrating the results to form the final result. This method showed excellent results and seems to be effective for handling multimodal variation. However, compared to their multi-atlas approach, our approach has the benefit of less computational cost in processing unseen images, in return for a larger cost of registration in the training process. Therefore, we believe that our approach can be an alternative choice for handling the multimodal inter-individual variety of organ shapes.
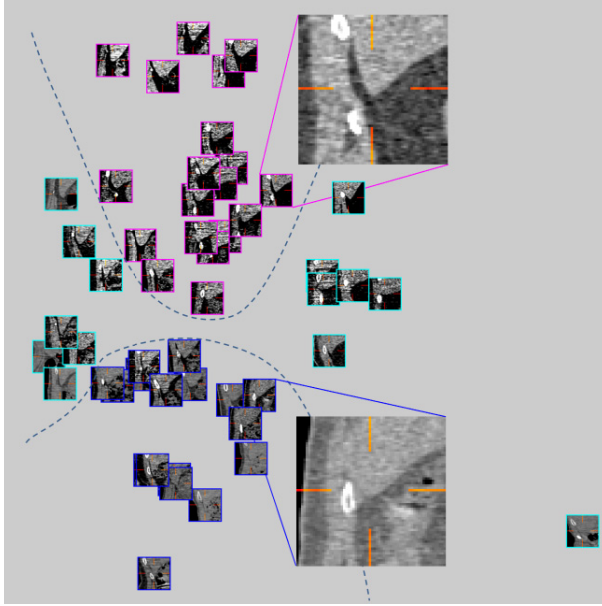

Fig. 3. An example of landmark appearance variation (inferior tip of the liver). Shown are the coronal cross-sections, mapped according to the dimension reduction result by a Laplacian eigenmap. The curves indicate our spectral clustering result. Note that appearances in the upper subgroup have an obvious fat pad between the liver and the abdominal wall muscles, while those in the lower subgroup have only a trace amount of fat pad.
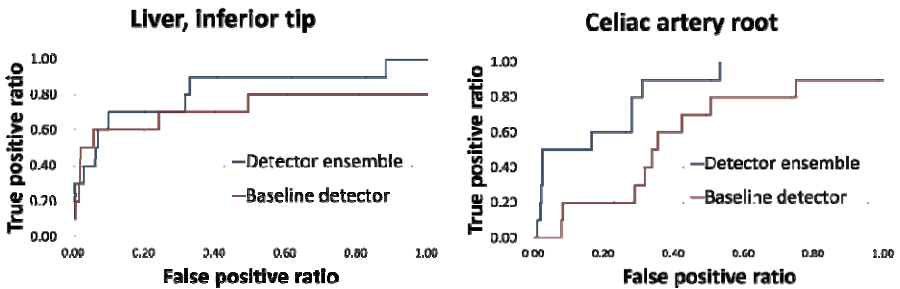

Fig. 4. Two examples of ROC curves (inferior tip of the liver and root of the celiac artery)

Table 1. The AUC (area under the curve) results of the baseline detector and the detector classifier

LM name	Sternum, inf. tip	Tip of coccyx	Tip of rt. 12th rib	Tip of lt. 12th rib	Umbilicus	Rt. Kidney, sup. tip	Lt. Kidney, sup. tip	Bifurcation of IHPV	Liver, inf. tip	Celiac artery root	SMA root	Rt. nipple	Lt. nipple	Average
Baseline	0.31	1.00	0.80	0.61	0.75	0.80	0.66	0.57	0.72	0.59	0.66	0.31	0.07	0.60
Ensemble	0.46	1.00	0.78	0.66	0.82	0.77	0.69	0.70	0.82	0.84	0.81	0.60	0.48	0.73
Improve-ment	0.15	0.00	-0.02	0.05	0.07	-0.03	0.03	0.13	0.10	0.25	0.15	0.29	0.41	0.13

4   Conclusion

A method for clustering and augmenting landmark-local appearances prior to training of anatomical landmark detectors was presented. Ensembles of cluster-specific detectors showed large improvements in detection performance, even when detection by the baseline detector was significantly difficult.

References

1. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Samei, E., Hsieh, J. (eds.) Medical Imaging 2009: Physics of Medical Imaging. Proceedings of the SPIE, vol. 7258, pp. 725902–725902-8 (2009)
2. Hanaoka, S., Masutani, Y., Nemoto, M., Nomura, Y., Yoshikawa, T., Hayashi, N., Yoshioka, N., Ohtomo, K.: Probabilistic Modeling of Landmark Distances and Structure for Anomaly-proof Landmark Detection. In: Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy, pp. 159–169 (2011)
3. Nemoto, M., Masutani, Y., Hanaoka, S., Nomura, Y., Yoshikawa, T., Hayashi, N., Yoshioka, N., Ohtomo, K.: A unified framework for concurrent detection of anatomical landmarks for medical image understanding. In: Proc. SPIE, vol. 7962, p. 79623E (2011)
4. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. Neuroimage 45(1) suppl. 1, S61–S72 (2009)
5. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems, vol. 14, pp. 849–856 (2001)
6. Zelnik-Manor, L., Perona, P.: Self-Tuning Spectral Clustering. In: Advances in Neural Information Processing Systems, vol. 17, pp. 1601–1608 (2004)
7. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33(1), 115–126 (2006)