AUTOMATED ANATOMICAL LANDMARK DETECTION ON DISTAL FEMUR SURFACE USING CONVOLUTIONAL NEURAL NETWORK

Dong Yang[1], *Shaoting Zhang*[2], *Zhennan Yan*[1], *Chaowei Tan*[1], *Kang Li*[3], *Dimitris Metaxas*[1]

[1] CBIM, Rutgers University, Piscataway, NJ, US
[2] Department of Computer Science, University of North Carolina at Charlotte, NC, US
[3] Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, US

ABSTRACT

Accurate localization of the anatomical landmarks on distal femur bone in the 3D medical images is very important for knee surgery planning and biomechanics analysis. However, the landmark identification process is often conducted manually or by using the inserted auxiliaries, which is time-consuming and lacks of accuracy. In this paper, an automatic localization method is proposed to determine positions of initial geometric landmarks on femur surface in the 3D MR images. Based on the results from the convolutional neural network (CNN) classifiers and shape statistics, we use the narrow-band graph cut optimization to achieve the 3D segmentation of femur surface. Finally, the anatomical landmarks are located on the femur according to the geometric cues of surface mesh. Experiments demonstrate that the proposed method is effective, efficient, and reliable to segment femur and locate the anatomical landmarks.

Index Terms— Deep learning, anatomical landmark detection, convolutional neural network, graph cut, mesh curvature

1. INTRODUCTION

Knee joint surgery, e.g. knee replacement, has been one of the most commonly performed surgeries since it was introduced in 1968 [1]. According to the Agency for Healthcare Research and Quality, more than 600,000 people accept knee surgeries every year in United States. By the year 2030, 3.48 million U.S. adults are estimated to undergo total knee replacement [2]. Accurate localization of 3D anatomical landmarks on the distal femur bone is vital to the success of these computer-aided surgeries. Also, the anatomical landmarks are important for biomechanical studies of bones and attached muscles (e.g. joint kinematics analysis).

During knee surgery procedures, the landmark localization process is often conducted manually or by using the inserted auxiliaries, such as markers, metallic pins [3]. However, these methods cannot guarantee a high degree of localization accuracy, especially in the views of 3D medical image. As a result, automatic bone landmark localization was introduced in the recent studies [4, 5, 6]. Although the automatic methods achieve high accuracy, they are not without their own shortcomings. For example, they are either dependent heavily on initial manual localization or likely lack of geometric distinctiveness in the prediction.

As shown in Figure 1, some landmarks, that carry geometrically distinct information, include Lateral Peak (LP), Lateral Epicondyle (LE), Lateral Distal Point (LDP), Medial Peak (MP), Adductor Magnus Tubercle (AMT), Medial Epicondyle (ME) and Medial Distal Point (MDP). In this study, we aim to locate these seven landmarks in the 3D medical images.

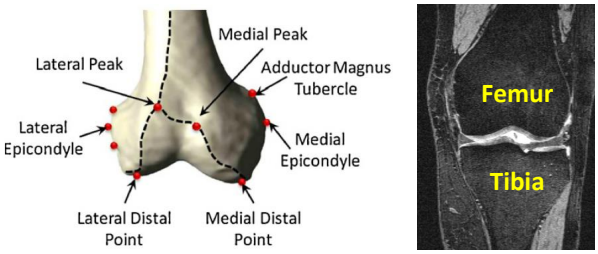

Fig. 1. Left: Anatomical landmarks on a distal femur (from [3]). Right: One slice of 3D knee joint MR image.

In this paper, we propose a novel framework to automatically locate femur landmarks from the 3D MR images using the convolutional neural network (CNN) and graph cut optimization. During the process of localization, both the global shape and local surface curvatures are taken into consideration, because they define the geometric features of landmarks. In what follows, we describe the methodology and experiments, discuss the results, and conclude with the future directions. Our study contributes to the practical application of the 3D medical image processing by improving the accuracy of landmark localization.

978-1-4799-2374-8/15/$31.00 ©2015 IEEE

2. METHODOLOGY

The automatic landmark localization framework contains three consecutive steps. Firstly, we detect several landmarks with distinct surface curvatures from the 3D medical images to calculate rigid transformation. Secondly, we transform the mean shape of femur from the training pool to the initial shape of the new image, and then the initial shape is refined with the proper optimization to obtain the segmentation. This segmentation step is necessary since mesh curvatures obtained from segmentation determine the positions of anatomical landmarks. Finally, the anatomical landmarks are localized by both of their initial positions from segmentation and local surface geometry.

2.1. Landmark Detection

The initial landmarks are chosen manually from the mean mesh of the training pool, based on the geometric characteristics such as large absolute values of curvatures.

Many methods have been applied to detect 3D landmarks in medical images [7]. Two of the most commonly used methods are regression with random forest [8, 9], and classification with marginal space learning and probabilistic boosting tree [10, 11]. Although these two methods are able to achieve a reasonable degree of accuracy and efficiency, they require the subjective selection of features such as the Haar-like feature and steerable feature, which might be time consuming and still cannot guarantee the optimal performance of detection. Recently, researchers have introduced deep convolutional neural network (CNN) to the field of automatic feature selection [12, 13, 14, 15]. In the CNN, the feature map of each layer is computed by the convolution of the entire image with the same filter h_i. h_i is non-linear function (e.g. $\tanh$) of the weights W_k and the bias terms b_k. Then the feature map is

$$h_k(i,j) = \tanh[W_k * x(i,j) + b_k]. \quad (1)$$

Sharing the same filter at each layer would reduce the memory size and improve the performance. In the particular area of image classification, the CNN has achieved excellent performance. However, most of the current applications focus on 2D images, with few exceptions in the 3D medical image domain, mainly due to computational complexity.

In our study, we explore the use of CNN in 3D medical image processing in a more efficient way. We treat the localization problem as the binary classification. During the training, we convert all 3D images into three sets of 2D images with *X, Y, Z* axes respectively. For each landmark, 2D images are labeled as either positive or negative based on whether they contain this landmark. Therefore, a total number of $3n$ labels are created, with n equals to the number of initial landmarks. Accordingly, $3n$ CNN classifiers are trained with both images and their labels. During testing, the new 3D image is also sliced into 3 sets of 2D images as above. For each axis of each landmark, the corresponding classifier is applied to achieve a probability distribution of 2D images. The coordinate along one axis is determined by the index of image slice with the highest probability. Combining results from all three classifiers, the coordinates of each landmark are obtained.

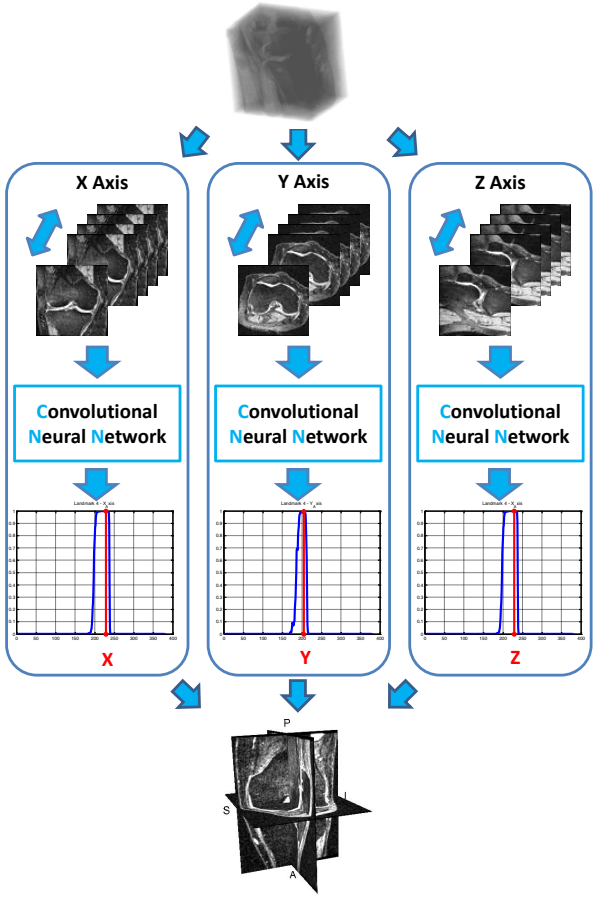

Fig. 2. The visual description of our method.

There are many benefits associated with conducting the detection in this way. First, this method tends to produce stable results, since the femur position is relatively stable in MR image without obvious translation, rotation, or scaling problems. Second, the power of CNN is fully utilized, because the 3D problems are converted into 2D problems. Previous studies using 3D detection methods ran on every voxel and the surrounding area. The computation complexity is $O(MNPK^3)$. *M, N*, and *P* are the dimensions of 3D images. *K* is the average side length of the 3D image cube. Compared to the previous methods, our method is better in a sense that the complexity is $O(3MNP)$. The last but not the least, feature selection is automatically solved without the bother of designing the feature space. Our method only requires the configuration of few system parameters, for instance, the number of hidden layers and nodes.

2.2. Femur Segmentation

We follow the conventional method to generate segmentation meshes within two steps. First, the mean shape is obtained by taking the average of the training meshes, and then is rigidly transformed into the new image as the initial segmentation. We use the predicted landmarks obtained from the previous steps and Procrustes analysis to calculate parameters of rigid transformation. Procrustes analysis provides a close solution for parameters (translation, rotation, scaling) of shape fitting.

Second, we use the narrow band graph cut optimization for mesh refinement. Such method guarantees the shape deformation within the neighborhood of initialization, which implies the shape constraints [16, 17, 18, 19]. The inner and outer bounds of the neighborhood are set by shrinking and inflating the initial meshes along the normal direction, resulting point-to-point correspondences of two bounds. More points are sampled along the line segments between the corresponding points of the inner and outer bounds. Graph is built by connecting the surrounding points, as shown in Figure 3. This graph method is preferred since the quantity of sampled points is controllable. Unlike the previous methods, this optimization does not run on voxel-level, and thus it performs efficiently without down-sampling. After graph cut optimization, the surface mesh is refined which contains only few negligible defects. The principle component analysis (PCA) is used for smoothing the final results.

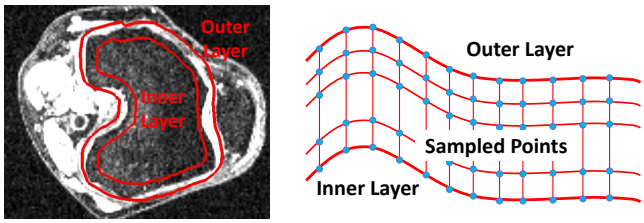

Fig. 3. Left: Inner and outer bounding layers. Right: Graph built between two layers.

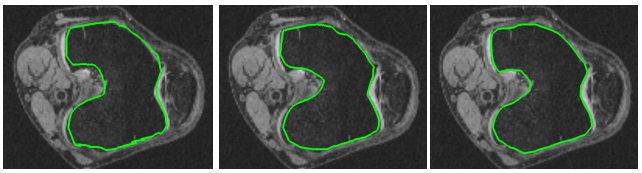

Fig. 4. Left: Initial mesh from rigid transformation. Middle: Segmentation after graph cut optimization. Right: PCA refined mesh.

2.3. Anatomical Landmark Localization

Since the segmentation mesh has point-wise correspondence with mean mesh, the anatomical landmark locations on mesh can be inferred by searching by the indices of those points on the mean mesh. However, the points with right indices cannot guarantee to be the true landmarks. Therefore we search for the ultimate landmark points within the neighbourhood of previous predicted points. The points with maximum absolute values of curvature are determined as the anatomical landmarks because of landmark definition in the beginning.

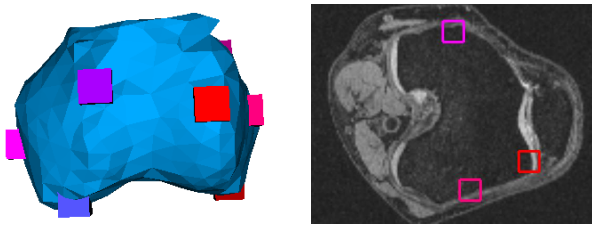

Fig. 5. Final detection results in 2D and 3D views (Blue is segmentation mesh, boxes with other colors represent different landmark position).

3. EXPERIMENTS

The dataset are the 3D MR images of knee joints from the Osteoarthritis Initiative (OAI) public dataset. The OAI knee MRI protocol provides imaging data on multiple articular structures and features relevant to knee OA that balances requirements for high image quality and consistency [20]. The dimension of each volume is $384 \times 384 \times 160$, and the resolution is $0.365 \times 0.365 \times 0.7mm^3$. Small noise and vague boundaries exist in the images as shown in figures. In total, 50 volumes of them are used in the experiment and annotated by experts. 40 volumes are utilized for training, 10 for testing. Initially the annotation data are labelled masks with a paint brush tool, they are converted into meshes for convenience.

During training, 3D volumes are sampled into 2D image slice sets along three axes X, Y, Z respectively for each landmark. In each set, the slices from different volumes, which contains the one of landmark, are labelled as positive samples Their neighboring slices in the range $\pm 2.5mm$ are also set as positive because they share very similar appearance. The negative samples are evenly selected from slices except positive samples. During sampling, because positive samples is considerably less than negative samples, more positive sample can be achieved by rotating initial positive slices with slight angle along axes from the volume and interpolating neighboring slices from them, which increases variance of training pool and robustness of detection.

In our experiment, the CNN contains two convolutional layers, two max-pooling layers, one hidden layer and one logistic regression layer. The first convolutional layer has 20 7×7 kernels, and the second has 40 7×7 kernels. Each convolutional layer is followed by a max-pooling layer with kernel size 2×2. The hidden layer is after the second max-pooling

Table 1. Quantitative comparison.

Landmark	Mean	STD	Max	Min
AMT	5.19	2.43	8.14	1.89
LP	4.64	2.23	7.57	0.89
MP	4.55	2.30	7.30	1.26
ME	4.79	2.91	9.16	0.77
LE	4.66	2.20	8.73	0.66
LDP	4.86	2.32	7.06	0.42
MPD	4.13	1.70	7.08	1.61

layer and it has 500 hidden neurons. The logistic regression comes at the end. For computational efficiency, image slices are down-sampled to 64×64 for both training and testing.

For testing, the input volume is also sampled in X, Y, Z axes as independent image slice sets. The corresponding CNN classifiers runs on every slice of these sets and return probability distributions along each axis for each landmark. The position with the highest response in distribution is set as landmark's coordinates along one axis. Therefore the coordinates are determined by combining with results from all the three image sets. Given initial landmark positions and their related positions in mean shape, Procrustes analysis provide close solution of transformation parameters (scaling, rotation, translation) for fitting mean shape in the volume. Next, the inner and outer bounding layers are generated by shrinking or inflating along the vertex normal directions for $5.0mm$. 20 points are evenly sampled in each line segment between the corresponding vertices of inner and outer bounds for building the graph. The final positions of the anatomical landmarks are searched in neighborhood with maximum distance $2.0mm$ after graph cut optimization.

Table 1 shows the evaluation results of anatomical landmark localization. The average time of training one CNN classifier is about 30 minutes. The total training time is directly dependent on the quantity of landmarks. The average time of testing is around 90 second, 80% of which is consumed by running the CNN classifiers. All programs run in MatLab and Python on a machine with a 2.66 GHz CPU and 4 GB memory.

4. CONCLUSION

In this paper, we proposed a novel framework to locate seven anatomical landmarks of the distal femur bone. Our approach is automatic and it combines both global shape information and local mesh curvatures. There are several directions for future research work. One possible direction is to enlarge training variance and do cascade detection for higher accuracy. This work can also be extended to locate other anatomical landmarks for other bones (e.g. tibia) or organs in medical images with difference modalities.

5. REFERENCES

[1] Mike D Van Manen, James Nace, and Michael A Mont, "Management of primary knee osteoarthritis and indications for total knee arthroplasty for general practitioners," *JAOA: Journal of the American Osteopathic Association*, vol. 112, no. 11, pp. 709–715, 2012.

[2] Hans C Dreyer, Lisa A Strycker, Hilary A Senesac, Austin D Hocker, Keith Smolkowski, Steven N Shah, Brian A Jewett, et al., "Essential amino acid supplementation in patients following total knee arthroplasty," *The Journal of clinical investigation*, vol. 123, no. 11, pp. 4654–4666, 2013.

[3] K Subburaj, Bhallamudi Ravi, and Manish Agarwal, "Automated identification of anatomical landmarks on 3d bone models reconstructed from ct scan images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 5, pp. 359–368, 2009.

[4] Daniel Bystrov, Vladimir Pekar, Stewart Young, Sebastian PM Dries, Harald S Heese, and Arianne M van Muiswinkel, "Automated planning of mri scans of knee joints," in *Medical Imaging*. International Society for Optics and Photonics, 2007, pp. 65092Z–65092Z.

[5] Seung-Yeob Baek, Joon-Ho Wang, Insub Song, Kunwoo Lee, Jehee Lee, and Seungbum Koo, "Automated bone landmarks prediction on the femur using anatomical deformation technique," *Computer-Aided Design*, vol. 45, no. 2, pp. 505–510, 2013.

[6] Ning Xue, Michael Doellinger, Charles P Ho, Rachel K Surowiec, and Raphael Schwarz, "Automatic detection of anatomical landmarks on the knee joint using mri data," *Journal of Magnetic Resonance Imaging*, 2014.

[7] Yiqiang Zhan, Xiang Sean Zhou, Zhigang Peng, and Arun Krishnan, "Active scheduling of organ detection and segmentation in whole-body medical images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 313–321. Springer, 2008.

[8] Nathan Lay, Neil Birkbeck, Jingdan Zhang, and S Kevin Zhou, "Rapid multi-organ segmentation using context integration and discriminative models," in *Information Processing in Medical Imaging*. Springer, 2013, pp. 450–462.

[9] Dong Han, Yaozong Gao, Guorong Wu, Pew-Thian Yap, and Dinggang Shen, "Robust anatomical landmark detection for mr brain image registration," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pp. 186–193. Springer, 2014.

[10] Yefeng Zheng, Matthias John, Rui Liao, A Nottling, Jan Boese, Jörg Kempfert, Thomas Walther, Gernot Brockmann, and Dorin Comaniciu, "Automatic aorta segmentation and valve landmark detection in c-arm ct for transcatheter aortic valve implantation," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 12, pp. 2307–2321, 2012.

[11] Adrian Barbu, Michael Suehling, Xun Xu, David Liu, S Kevin Zhou, and Dorin Comaniciu, "Automatic detection and segmentation of axillary lymph nodes," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pp. 28–36. Springer, 2010.

[12] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[14] Dan Ciresan, Ueli Meier, and Jürgen Schmidhuber, "Multi-column deep neural networks for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3642–3649.

[15] Christian Szegedy, Alexander Toshev, and Dumitru Erhan, "Deep neural networks for object detection," in *Advances in Neural Information Processing Systems*, 2013, pp. 2553–2561.

[16] Yefeng Zheng, Dong Yang, Matthias John, and Dorin Comaniciu, "Multi-part modeling and segmentation of left atrium in c-arm ct for image-guided ablation of atrial fibrillation," 2013.

[17] Dong Yang, Yefeng Zheng, and Matthias John, "Graph cuts based left atrium segmentation refinement and right middle pulmonary vein extraction in c-arm ct," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2013, pp. 86693U–86693U.

[18] Yuri Y Boykov and M-P Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in nd images," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 1, pp. 105–112.

[19] Mustafa Gökhan Uzunbaş, Chao Chen, Shaoting Zhang, Kilian M Pohl, Kang Li, and Dimitris Metaxas, "Collaborative multi organ segmentation by integrating deformable and graphical models," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pp. 157–164. Springer, 2013.

[20] CG Peterfy, E Schneider, and M Nevitt, "The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee," *Osteoarthritis and Cartilage*, vol. 16, no. 12, pp. 1433–1441, 2008.