Reconstructing 3D Lung Shape from a Single 2D Image during the Deaeration Deformation Process using Model-based Data Augmentation

1st Shuqiong Wu
Graduate School of Informatics, Kyoto University, Kyoto, Japan
wusq@sys.i.kyoto-u.ac.jp

2nd Megumi Nakao
Graduate School of Informatics, Kyoto University, Kyoto, Japan
megumi@i.kyoto-u.ac.jp

3rd Junko Tokuno
Department of Thoracic Surgery, Kyoto University Hospital, Japan
jtokuno@kuhp.kyoto-u.ac.jp

4rd Toyofumi Chen-Yoshikawa
Department of Thoracic Surgery, Kyoto University Hospital, Japan
fengshic@kuhp.kyoto-u.ac.jp

5rd Tetsuya Matsuda
Graduate School of Informatics, Kyoto University, Kyoto, Japan
tetsu@i.kyoto-u.ac.jp

***Abstract*—Three-dimensional (3D) shape reconstruction is particularly important for computer assisted medical systems, especially in the case of lung surgeries, where large deaeration deformation occurs. Recently, 3D reconstruction methods based on machine learning techniques have achieved considerable success in computer vision. However, it is difficult to apply these approaches to the medical field, because the collection of a massive amount of clinic data for training is impractical. To solve this problem, this paper proposes a novel 3D shape reconstruction method that adopts both data augmentation techniques and convolutional neural networks. In the proposed method, a deformable statistical model of the 3D lungs is designed to augment various training data. As the experimental results demonstrate, even with a small database, the proposed method can realize 3D shape reconstruction for lungs during a deaeration deformation process from only one captured 2D image. Moreover, the proposed data augmentation technique can also be used in other fields where the training data are insufficient.**

***Index Terms*—CNN, deaeration deformation, machine learning, data augmentation, 3D shape reconstruction**

I. Introduction

In recent years, three-dimensional (3D) shape reconstruction has become widely used in computer-based surgical navigation because 3D models can provide an integral knowledge of the human inner body. Many approaches have been used to obtain the 3D information of internal organs in various situations. The most correct and simple method is to measure the 3D shapes by computed tomography (CT), magnetic resonance imaging (MRI), or other 3D imaging technologies. However, intraoperative imaging places an additional burden on both surgeons and patients. To solve this problem, some researchers computed the 3D information of intraoperative organs using preoperative CT volume data or a pre-known 3D model [1]–[5]. For example, Koo et al. realized the registration of a preoperative 3D model with an intraoperative laparoscopy image by adopting contour and shading features [4]. Similarly, Collins et al. deformed a pre-computed 3D model for fitting to 2D intraoperative laparoscopy videos [5]. These algorithms achieve reliable performance, and have already been applied to surgical navigation. Nevertheless, they are effective only when the deformation of the organ is small. In the case of thoracic surgeries for lung cancers, where deaeration deformation always occurs, these approaches, which are based on preoperative 3D models, lose accuracy.

To solve the problem caused by large deformation, stereoscopic or multiple intraoperative images have been used to implement 3D reconstruction [6]–[8]. Three-dimensional reconstruction based on multiple images depends on the disparities among these different images. This kind of reconstruction only considers intraoperative information. Thus, it is less sensitive to large deformation [7], [8]. Nonetheless, using only the visual information causes the performance to be easily influenced by image blur, poor illumination, and occlusion. Moreover, the reconstructed 3D shape is always partial because it is difficult to capture 2D images of the whole organ during a surgery.

In contrast to preoperative-model-based and multi-image-based approaches, machine-learning-based methods can reconstruct an integral 3D shape from a single image, as long as the amount of training data is sufficient [9]–[11]. These methods train artificial networks such as convolutional neural networks (CNN) using a large range of data. They are robust to large deformations, poor illumination, occlusion, and image blur. However, when these techniques are applied to 3D reconstruction in the medical field, a crucial problem is that the collection of various 2D-3D pair data is not realistic. To

This research was funded by the Japan Agency for Medical Research (AMED) and the Acceleration Transformative Research for Medical Innovation (ACT-M) Program. This study was performed in accordance with the animal research ethics committee, Kyoto University. We thank Kimberly Moravec, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

solve this problem, we propose a CNN-based algorithm that can reconstruct the 3D shape of lungs from a single 2D image. To guarantee the training of the CNN on a small database, a data augmentation technique is also proposed to create various training data. The contribution of this research is (i) it provides a promising solution for single-image-based 3D lung reconstruction; and (ii) it proposes a novel data augmentation technique to address the problem of data insufficiency for machine-learning-based algorithms in the medical field.

II. Proposed Algorithm

In this section, we first explain the model-based data augmentation method; and then we describe the CNN designed for 3D lung shape reconstruction.

A. Statistical 3D model

In the last two decades, statistical shape models (SSD) have been widely used for 3D medical image segmentation [12], [13]. By matching a labelled statistical shape model with a new 3D image, the area of interest of the new image can be segmented automatically [12]. A statistical model is always computed from a group of sample data. We suppose there is a set of samples $\{M_1, M_2, ..., M_n\}$, where all M_i are aligned 3D point clouds obtained from CT volume data. Then, the principal component analysis (PCA) method is used to reduce the number of dimensions of variation among these samples. Finally, the statistical model can be represented by [12]

$$M = \overline{M} + \sum_{k=1}^{C} P_k b_k \tag{1}$$

where $\overline{M}$ is the mean of the samples, C is the number of dimensions, P is the eigenvectors of the covariance matrix, and b is the vector of weight parameters. Each sample can be represented by (1) using its own parameter vector b [12].

In this study, we assume that the training database only contains eight samples. As is well-known, the shape and size of lungs vary significantly case by case. Thereby, a statistical shape model of the eight cases may be biased because the variation of the training data is too small. However, the deaeration processing of different cases may share some common pattern. Therefore, we propose a new model that we name statistical displacement model (SDM), instead of the traditional SSD for the data augmentation. The displacement measures the deformation during the deaeration process. We use M_i^p and M_i^q to respectively denote the inflated and deflated lungs for Case i ($i \in \{1, 2, ..., 8\}$) in this study, and then $D_i = M_i^p - M_i^q$, where D_i is the displacement of deaeration deformation of Case i. Finally, the SDM is represented by

$$D = \overline{D} + \sum_{j=1}^{\Phi} \rho_j \varphi_j, \tag{2}$$

where $\overline{D}$ is the mean displacement of the eight cases, Φ is the number of dimensions after PCA, ρ is the eigenvector of the covariance matrix, and φ denotes the weight parameters. In this study, Φ is set to be 2, where the cumulative contribution rate reached 98%. Then we increase or decrease φ step by step to create different deaeration displacements D. Finally,

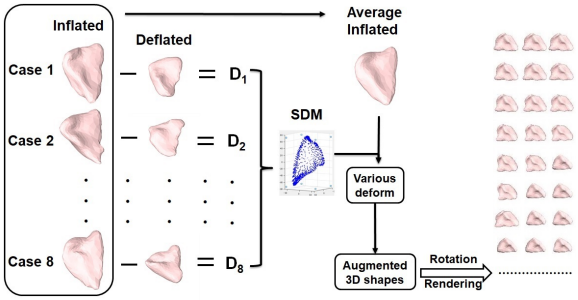

Fig. 1. Schematic flow of the proposed data augmentation method

k different displacements are applied to the average inflated lung $\overline{M^p}$, and k deflated 3D shapes $\overline{M_k^q}$ will be augmented. Moreover, the proposed algorithm can produce different size of 3D shape between $\overline{M^p}$ and $\overline{M_k^q}$ by interpolation.

B. Data augmentation

To realize single-image-based 3D shape reconstruction, 2D–3D pair data are required for training, i.e., a 2D image and its corresponding 3D shape is regarded as one pair of data. Using the CT data of both the inflated and deflated lungs of eight cases, a total of 16 3D shape models were built (eight inflated and eight deflated). As we mentioned above, first we need to augment 3D shapes for learning. We apply various displacements to the average inflated 3D model $\overline{M^p}$, to create various deflated 3D shapes, which are then added to the 16 shapes as training data. For each 3D shape of training data, 2D images are rendered from n different viewpoints. Thus, n pairs of 2D–3D data are augmented from one 3D shape model. This idea is inspired by PointNet [9], which uses rendered 2D images from multiple viewpoints for training. Figure 1 shows the structure of the proposed data augmentation.

C. CNN-based 3D reconstruction

In this research, a CNN is used to learn the relationship between a single 2D image and its corresponding 3D point cloud. The design of the CNN is shown in Fig. 2, which is modified from VGG19. VGG19 is proved effective in processing images [15]. The Encoding in Fig. 2 denotes one convolution layer, one batch normalization layer, and one Relu layer. To speed up the training, we reduce the number of filters in each convolution layer. The batch normalization layer is

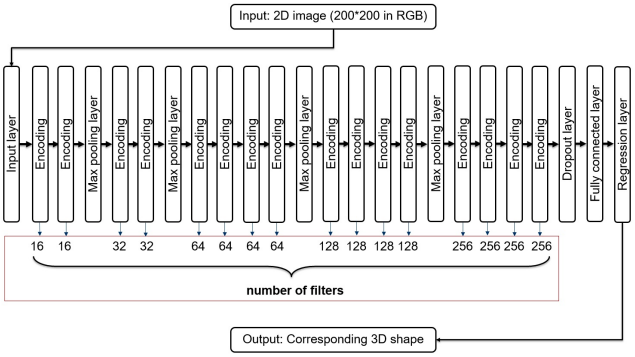

Fig. 2. Design of the CNN for 3D shape reconstruction from single image

added after each convolution layer so that all values can be normalized. Also, a dropout layer is added to restrain overfitting, and a regression layer is added to predict the 3D positions of each point in the point cloud. The input is a 200×200 pixel RGB image, and the output is a 3D point cloud. Note that, in the training, the input 2D images are totally different samples, but the output 3D shapes are the same for those images that are rendered from the same 3D model.

The CNN here learns the connection between a single 2D image and its corresponding 3D shape, and it is likely that, it also learns the disparity among those 2D images rendered from the same 3D model; because they are different inputs but have the same output. For simplicity, we assume that there are three groups of parameters in the CNN. The first two groups stand for the X and Y axes, and the last one represents the Z axis (depth). The first two groups can be easily deduced using the 2D input image. With respect to the last group, the CNN may select parameters randomly when it learned only one 2D-3D pair. However, these parameters are refined when it learns more 2D inputs who share the same 3D shape, because it needs to adjust the last group of parameters to produce the same output. In summary, the CNN learns how to adjust a CT-based 3D model to fit any 2D image captured from various viewpoints. Then in the test phase, the CNN automatically tunes its parameters to fit a new 2D image. In this manner, the 3D information can be computed from only one 2D image.

III. Experiment

In this section, we evaluate the effectiveness of the proposed algorithm. A total of ten cases of in vivo lung CT images were measured from the left lungs of ten Beagle dogs with two bronchial pressures (14 and 2 cmH2O) at the Institute of Laboratory Animals, Kyoto University, Japan. Eight cases were used for training, and the left two were used for testing. All the cases were aligned before training by a registration approach [14]. The training takes about 5 min using the Parallel Computing Toolbox (GPU) of MATLAB. The CPU and memory of the desktop we used are i7-6700 @ 3.40 GHz 3.41 GHz and 32 GB, respectively.

We assume that $T = 0$ represents the moment when the lung is inflated, and $T = 1$ represents the moment when the lung is totally deflated. Then $T \in [0, 1]$ can represent any moment during a deaeration deformation. In the training phase, the original 16 3D models (eight inflated and eight delated), plus 81 augmented 3D models from moment $T = 0.5$ are used. For each 3D model, the number of viewpoints n is set to 108. Thereby, a total of $(16 + 81) \times 108 = 10476$ pairs of 2D–3D data were used to train the CNN. In the test phase, to simplify the evaluation, we used the test 3D model at moments $T = 0$, $T = 0.5$, and $T = 1$ to perform the evaluation. The test images were synthesized by rendering from the test 3D model. A total of 2 cases $\times$ 3 moments $\times$ 108 viewpoints $= 648$ 2D images were tested. The RMSE between the ground truth and reconstructed result is computed by Iterative closest point method [16]. We compare our method with the shape from shading (SFS) algorithm. Figure 3 shows the results, where

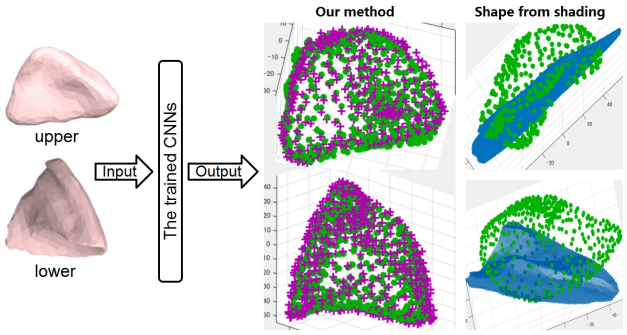

Fig. 3. One example of 3D shape reconstruction using the proposed method

the green dots show the ground truth, the blue dots show the results of SFS, and the purple dots show our results. Using our method, the RMSE of the upper lung is 2.59 mm, and that of the lower lung is 3.08 mm. SFS can hardly obtain the correct depth information. Its RMSEs of the upper and lower lungs are 8.61 mm and 7.55 mm, which are much larger than those of the proposed approach.

We also compared the CNN trained using the proposed data augmentation with a CNN trained only using real data. Figure 4 shows one example of the comparison. The blue curves in the figure show the RMSEs of a CNN that is trained only by 16 models (eight inflated and eight deflated). The red curves show the RMSEs of a CNN that is trained by the proposed data augmentation. For each moment T, 108 2D images from different viewpoints were tested. These results show that the data augmentation actually reduces the RMSEs of the CNN, which makes training based on small size databases possible. Tables 1 and 2 summarize all the comparison results for the upper and lower lungs, respectively. In this study, 5-fold cross-validation is used for the evaluation, i.e., eight cases are used for the training and the remaining two are used for testing. The two tables show that the proposed data augmentation approach reduces the RMSEs significantly. This demonstrates that the proposed method is effective for single-image-based 3D reconstruction even though the training data would be insufficient for conventional machine learning methods.

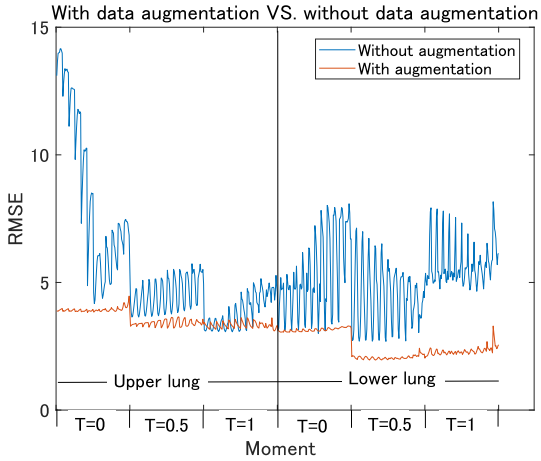

Fig. 4. Comparison result of the CNNs for Case 5

TABLE I
COMPARISON OF THE RESULTS OBTAINED WITH AND WITHOUT DATA AUGMENTATION FOR THE UPPER LUNG

Cases	**Without augmentation**		**With augmentation**	
(mm)	mean-RMSE	max-RMSE	mean-RMSE	max-RMSE
Case 1	4.01	5.67	2.93	4.18
Case 2	5.01	9.72	3.80	4.89
Case 3	5.82	11.89	4.08	5.22
Case 4	4.62	8.25	3.03	3.84
Case 5	5.69	14.16	3.56	4.47
Case 6	4.70	12.27	3.47	4.46
Case 7	4.36	5.86	3.80	4.65
Case 8	5.87	9.48	4.50	5.19
Case 9	5.19	7.38	3.75	4.65
Case 10	4.83	9.60	3.74	4.58
Average	5.01	9.43	3.67	4.61
Improved	—	—	26.75%	51.11%

TABLE II
COMPARISON OF THE RESULTS OBTAINED WITH AND WITHOUT DATA AUGMENTATION FOR THE LOWER LUNG

Cases	**Without augmentation**		**With augmentation**	
(mm)	mean-RMSE	max-RMSE	mean-RMSE	max-RMSE
Case 1	4.68	8.73	2.84	4.36
Case 2	5.14	10.57	3.95	4.86
Case 3	5.92	13.48	2.98	4.37
Case 4	3.63	6.40	2.22	2.78
Case 5	5.17	8.16	2.51	3.30
Case 6	4.66	8.23	3.54	4.86
Case 7	4.53	6.52	3.29	4.24
Case 8	4.76	7.12	3.49	4.43
Case 9	3.80	8.22	2.76	4.31
Case 10	4.52	6.98	3.43	4.08
Average	4.68	8.44	3.10	4.16
Improved	—	—	33.76%	50.71%

IV. DISCUSSION

In lung surgeries, especially in minimally invasive surgeries, only part of the organ can be captured via imaging. If these partial 2D images are used for 3D reconstruction, only a partial 3D shape can be reconstructed. Machine-learning-based techniques seem to be promising solutions for the above problem, because they can learn the model information beforehand. Learning-based 3D shape reconstruction is more like a fitting process than a building process. In this section, we use 2D images that capture partial organ to test the trained CNN. Figure 5 shows the reconstruction result. The green dots show the ground truth and the purple dots show the test result. The RMSEs of the two examples are 4.50 mm and 4.13 mm, respectively. Note that the training data for this CNN do not contain any 2D images of partial organs. Nonetheless, the CNN still can reconstruct a 3D shape that is close to the ground truth.

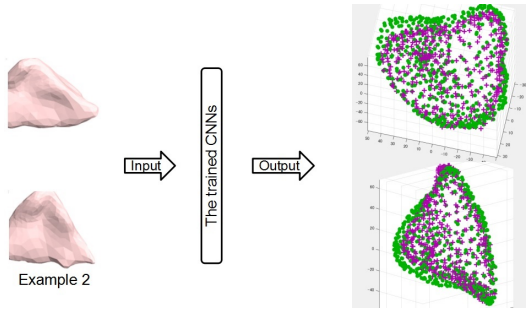

Fig. 5. Examples of reconstruction results from partial 2D images

V. CONCLUSION

This paper first analyzed the problem of 3D reconstruction during a lung deaeration deformation, and then proposed a CNN-based 3D shape reconstruction algorithm. The proposed method can reconstruct a 3D shape from only one 2D image. The training database contains only eight samples. However, the trained CNN is still effective, as demonstrated by the experimental results. In the current research, we only tested synthesized 2D images. In future work, we plan to use 2D endoscopic images or 2D captures for the evaluation. In that case, some image processing and synthesis techniques are required to reduce the discrepancy between the endoscopic images (or 2D captures) and the rendered images.

REFERENCES

[1] J. Hallet et al., "Systematic review of the use of pre-operative simulation and navigation for hepatectomy: current status and future perspectives," *J. Hepatobiliary Pancreat. Sci.* , vol. 22, no. 5, pp. 353–362, 2015.

[2] W. Zhang et al., "Application of preoperative registration and automatic tracking technique for image-guided maxillofacial surgery," *Comput. Assist. Surg.*, vol. 21, no. 1, pp. 137–142, 2016.

[3] M. Nakao, et al., "Physics-based interactive volume manipulation for sharing surgical process," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 809–816, 2010.

[4] B. Koo et al., "Deformable registration of a preoperative 3D liver volume to a laparoscopy image using contour and shading cues," *in Int. Conf. Med. Image. Comput. Comput. Assist. Interv. (MICCAI)* , 2017, pp. 326–334.

[5] T. Collins et al., "Robust, real-time, dense and deformable 3D organ tracking in laparoscopic videos," *in Int. Conf. Med. Image. Comput. Comput. Assist. Interv. (MICCAI)*, 2016, pp. 404–412.

[6] J. Lin et al., "Tissue surface reconstruction aided by local normal information using a self-calibrated endoscopic structured light system," *in Int. Conf. Med. Image. Comput. Comput. Assist. Interv. (MICCAI)*, 2015, pp. 405–412.

[7] D. Sun et al., "Surface reconstruction from tracked endoscopic video using the structure from motion approach," *in Int. Work. Med. Imaging Virt. Real.*, 2013, pp. 127–135.

[8] Q. Zhao et al., "The Endoscopogram: A 3D model reconstructed from endoscopic video frames," *in Int. Conf. Med. Image. Comput. Comput. Assist. Interv. (MICCAI)*, 2016, pp. 439–447.

[9] H. Fan et al., "A point set generation network for 3D object reconstruction from a single image," *in CVPR*, 2017, DOI:10.1109/CVPR.2017.264.

[10] CH. Lin et al., "Learning efficient point cloud generation for dense 3D object reconstruction," *in AAAI*, 2018 (To appear).

[11] N. Wang et al., "Pixel2Mesh: Generating 3D mesh models from single RGB images," *arXiv*, https://arxiv.org/abs/1804.01654, 2018.

[12] T. Heimann et al., "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.*, vol. 13, no. 4, pp. 543–563, 2009.

[13] J. Krger et al., "Statistical appearance models based on probabilistic correspondences," *Med. Image Anal.*, vol. 37, pp. 146–159, 2017.

[14] A. Saito et al., "Deformation estimation of elastic bodies using multiple silhouette images for endoscopic image augmentation,"*in IEEE Int. Symp. Mixed Augment. Real. (ISMAR)*, 2015, pp. 170–171.

[15] O. Russakovsky et al., "ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis. (IJCV)*, Vol. 115, no. 3, pp. 211–252, 2015.

[16] P. Bergstrm et al., "Robust registration of point sets using iteratively reweighted least squares, *Comput. Optimization and Applications*, vol. 58, no. 3, pp. 543–561, 2014.