

# Automated segmentation of mammary gland regions in non-contrast torso CT images based on probabilistic atlas

X. Zhou <sup>\*a</sup>, M. Kan <sup>a</sup>, T. Hara <sup>a</sup>, H. Fujita <sup>a</sup>, K. Sugisaki <sup>b</sup>, R. Yokoyama <sup>b</sup>, G. Lee <sup>a</sup>, and H. Hoshi <sup>b</sup>

<sup>a</sup> Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Sciences, Graduate School of Medicine, Gifu University, Gifu, Japan;

<sup>b</sup> Department of Radiology, Gifu University Graduate School of Medicine and University Hospital, Gifu, Japan

## ABSTRACT

The identification of mammary gland regions is a necessary processing step during the anatomical structure recognition of human body and can be expected to provide the useful information for breast tumor diagnosis. This paper proposes a fully-automated scheme for segmenting the mammary gland regions in non-contrast torso CT images. This scheme calculates the probability for each voxel belonging to the mammary gland or other regions (for example pectoralis major muscles) in CT images and decides the mammary gland regions automatically. The probability is estimated from the location of the mammary gland and pectoralis major muscles in CT images. The location (named as a probabilistic atlas) is investigated from the pre-segmentation results in a number of different CT scans and the CT number distribution is approximated using a Gaussian function. We applied this scheme to 66 patient cases (female, age: 40-80) and evaluated the accuracy by using the coincidence rate between the segmented result and gold standard that is generated manually by a radiologist for each CT case. The mean value of the coincidence rate was 0.82 with the standard deviation of 0.09 for 66 CT cases.

**Keywords:** Segmentation, mammary gland, torso CT images, probabilistic atlas.

## 1. INTRODUCTION

Breast tumor detection is one of the most important tasks in computer-aided diagnosis. Currently, digital mammography is the most cost effective technique in breast cancer screening due to its high spatial resolution (about 50  $\mu\text{m}$ ). However, digital mammograms are 2-D projection of objects, and cannot show 3D structures in breast tissues that sometimes required in diagnosis. On the other hand, the remarkable progress in multi-slice X-ray CT tomography enables the acquisition of the whole human torso in as little as 20-30 seconds. The volumetric torso CT scan contains details of all anatomical structures in the torso region. This enables the detection of lesions in different organs and tissue regions in the human torso. In principle, these CT images may also be helpful in detecting breast tumor in mammary gland regions.

The identification of the mammary gland in torso CT images is a necessary primary processing step prior to the other processes such as tumor detection. Because the mammary gland regions do not have a unique CT number distribution in CT images and the location of mammary gland regions is not stable comparing with the other organs, the automated segmentation of mammary gland from the torso CT images is very difficult and has not been solved till now.

This research proposes an automated scheme to identify the mammary gland regions from the volumetric 3-D torso CT images. We describe the basic consideration of the method for mammary gland regions segmentation and show the outline of the scheme firstly, and then details of the method including probabilistic atlas construction and mammary gland segmentation in Section 2. Experimental results are shown in Section 3 and discussions on mammary gland segmentation are given in Section 4. Finally, conclusions are presented in Section 5.

\*zxr@fjt.info.gifu-u.ac.jp; phone 81-58-230-6510; fax 81-58-230-6514; Gifu University

## 2. METHODS

### 2.1 Mammary gland regions in CT images

The breasts consist of mammary gland with the associated skin and connected tissues. The mammary glands are modified sweat glands in the superficial fascia anterior to the pectoral muscles and anterior thoracic wall in human anatomy. The mammary glands consist of a series of ducts and associated secretory lobules. These converge to form 15 to 20 lactiferous, which open independently onto the nipple. The nipple is surrounded by a circular pigmented area of skin termed the areola<sup>1</sup>. On 3-D CT images, mammary gland regions appear as mass regions with an irregular shape and a density (CT number) distribution similar to the pectoral muscles. The volume and density of the mammary gland regions change greatly depending on the individual variation and the age of the patients as shown in Fig. 1, so that, the traditional segmentation methods based on CT numbers (such as gray-level thresholding, region growing) or shape-based method (such as snakes, active shape model) is not effective to solve the mammary gland regions segmentation problem completely.

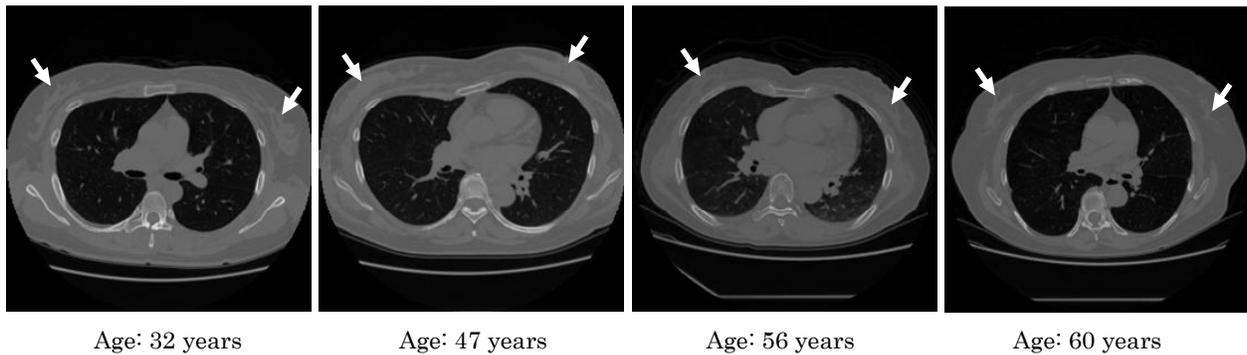


Fig. 1 Four slices of CT images with the mammary gland regions in four female patient cases.  
(White arrow: mammary gland regions)

### 2.2 Outline of mammary gland regions segmentation

The approach of mammary gland region segmentation can be considered as (1) finding the scale of the mammary gland regions in CT images approximately based on anatomical knowledge, and then, (2) deciding the mammary gland regions based on CT number precisely. The outline of this approach is shown in Fig. 2.

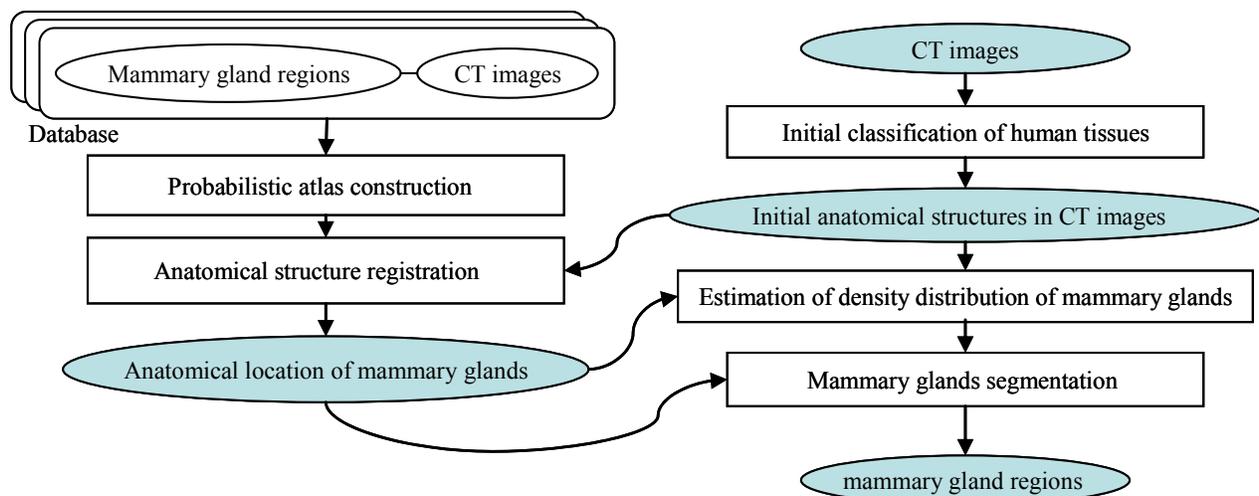


Fig. 2 Outline of the mammary gland regions segmentation

In order to estimate the anatomical location of the mammary gland regions in a patient case of CT images, our research proposes a method to construct a probabilistic atlas for mammary gland regions and skeletal muscles from a learning sample sets (each sample set includes a patient case of CT images and mammary gland regions that extracted manually). By deforming the probabilistic atlas and matching it to the inputted patient case, we generate a probabilistic image to show the mammary gland locations in CT images under a specific patient anatomical structure. Using the estimated probabilistic images, we can reduce scale of the target region for mammary gland segmentation and estimate the density (CT number) distribution of mammary glands using a histogram analysis. Combining the density distribution and anatomical location of mammary gland regions of the inputted CT case, we segment the mammary gland regions simply using a thresholding process and refine the results using a binary connected-component processing. The detail of each processing step is described in the following sections.

### 2.3 Initial classification of human tissues

The whole region in torso CT images is divided into 4 parts (Air, Fat, Muscle&Organ, and Bone) which have the unique density distributions that can be distinguished from the density (CT number) histogram of CT images<sup>2</sup>. A gray-level thresholding process is used for the region division. The optimized threshold values for segmenting each target region are estimated dynamically from the inputted images. A connected component processing including a combination of small region deletion and a binary morphological operator is used to refine the extracted regions and classify the air regions into two parts, one is air regions inside of human body, and another is the air region outside of human body. The body surface is identified by a 3-D contour tracing method using contour between the extracted human tissue and air regions outside of human body.

In order to normalize the spatial locations of mammary glands in CT images; the surface of thoracic cavity is also needed. The thoracic cavity is surrounded by the bone frame and closed by the diaphragm. The thoracic cavity is extracted using a ball-kernel based region growing process limited by the spatial position of bone frame within the chest region identified in step1. The parameter of the ball kernel is calculated and decided automatically from the size of circumscribed rectangle of bone frame for each patient case respectively<sup>3</sup>.

### 2.4 Extraction of nipple locations

The nipples are important landmarks to show the centers of mammary gland regions. We identify the positions of nipples using a shape analysis on body surface using the following three processing steps: (1) we calculate the 3-D curvatures of the body surface extracted in Section 2.3 and find out the convex points as the candidates of the nipples; (2) we deform the 3-D body surface (Fig. 3(a)) to a 2-D thin-plate (Fig.3 (b)) by a body stretching process and generate a range map (Fig. 3(c)) to show the statistical nipple locations based a CT database. Within range map (ellipse line in Fig. 3(b)), we reduce the false positive of the nipple candidates identified using the curvatures of body surfaces in the step 1.

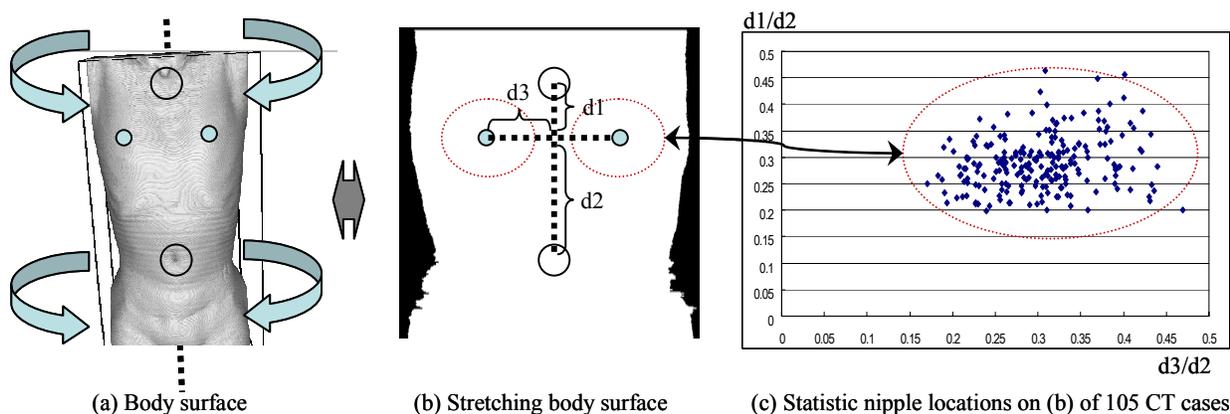


Fig.3 Estimation of nipple locations based on body surface stretching.

(3) we decide the nipple position on right (or left) side of human body respectively within the range map generated in step 2 by searching the nipple candidates  $p$  that has the maximum ratio  $A(p)$  defined in the following equation.

$$A(p) = S1/S2 \tag{1}$$

$S1$ : Volume of the human tissues inside a ball with the center  $p$  and radio  $d=10$  (Voxel).

$S2$ : Volume of the muscles inside a ball with the center  $p$  and radio  $d=10$  (Voxel).

The position of left (right) nipple is used in identification of the mammary gland region in the next processing step.

### 2.5 Probabilistic atlas construction

The locations of the mammary gland regions are investigated using a database (learning sample dataset) employed in this study. The mammary gland regions in all cases in the database are pre-segmented manually by a radiologist. A standard coordinate space is defined on a model human torso and the positions of all torso regions in the model are known. An individual case in the database is transformed into this standard coordinate space by normalizing and deforming the anatomical structures of the human torso. A point distribution model is used to express the anatomical structures surrounding the mammary gland region. 400 points are selected from the nipple; body surface and thoracic cavity surface (Fig. 4(a)). The points on the body surface are selected from the crossing points of a grid with the equal interval on the body surface stretching image. The length of each interval is decided by the distance between left and right nipple and jugular notch to navel (Fig. 4(b)). The points on the surface of the thoracic cavity are selected directly in 3-D based on extracted thoracic cavity and bone structures (Fig. 4(c)). The correspondence between the anatomical structures in different patient case is decided by the correspondence between the point distribution models in different patient case that is solved easily during the point selections.

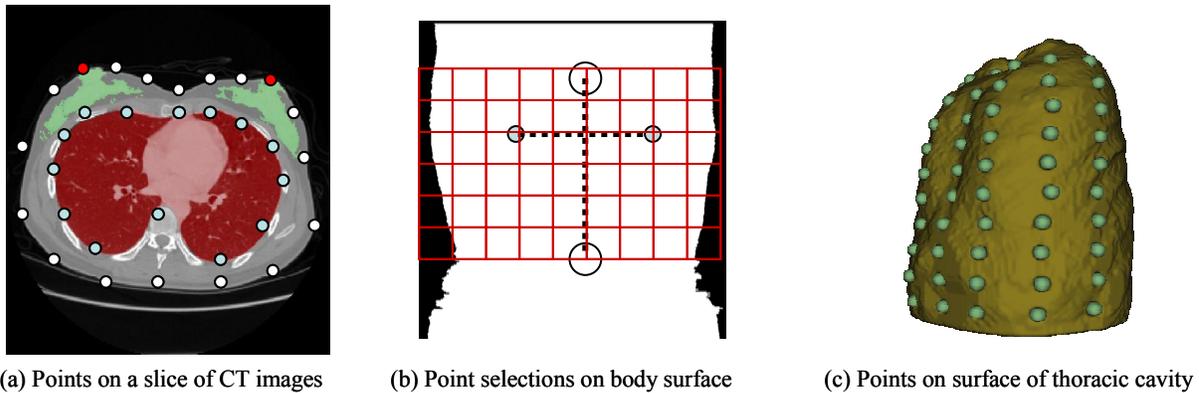


Fig. 4 A point distribution model that shows the anatomical structures surrounding the mammary gland regions.

We normalize the anatomical structures of a patient case by moving each point in the point distribution model to a standard position and deforming the whole CT images using Thin-plate splines (TPS) algorithm<sup>5</sup>. After the anatomical structure normalization for each CT image in learning sample dataset, we investigate appearances of the mammary glands region and skeletal muscles region in the standard coordinate space and generate the probability of mammary gland and muscles in each voxel position respectively. We call those probability images under the standard coordinate space as probabilistic atlas (Fig.5(a)).

### 2.6 Estimating the locations of mammary gland region in a specific CT case

The probability of mammary glands region in a specific CT case is generated from the probabilistic atlas. The point distribution model of the CT case is decided using the method described in Section 2.5 firstly. Using the spatial correspondence between the points in CT cases and the position in the standard coordinate space, we transfer and deform the probabilistic atlas onto the specific coordinate space of the CT case using the TPS algorithm. The deformed probabilistic atlas shows the mammary gland locations in the inputted CT images. The same process is also applied to generate a probability image to show the location of the chest muscle regions by deforming the probabilistic atlas of muscles from the standard coordinate space to the specific coordinate space of the inputted CT case (Fig.5(b)).

## 2.7 Estimation of the density distribution of the mammary gland region

The density (CT number) distribution of mammary glands region is approximated as a normal distribution  $N(\mu, \sigma)$ . The value  $\mu$  and  $\sigma$  are estimated by the histogram of CT images within the scale that the probability of mammary glands in each voxel is greater than 0. In fact, we select the maximum (peak value) of the histogram as the  $\mu$  and FWHM (Full Width at Half-Maximum) as the  $\sigma$ .

## 2.8 Segmentation of mammary glands region based on the probabilistic atlas

The identification of mammary gland regions is performed using the following steps: (1) Combing the probability images of mammary gland and chest muscle images estimated in Section 2.6, and Section 2.7 to generate a contour (Fig.5(c)) that separating the of mammary gland and chest muscle regions on the inputted CT images. (2) Using the contour line generated in step 1 mammary gland probability image, we decide the mask regions (Fig.5(d)) of the mammary glands regions on CT images. (3) Using the estimated density distribution in 2.7, we extract the mammary glands using a thresholding process within the mask region generated in step 2 (Fig. 5(e)).

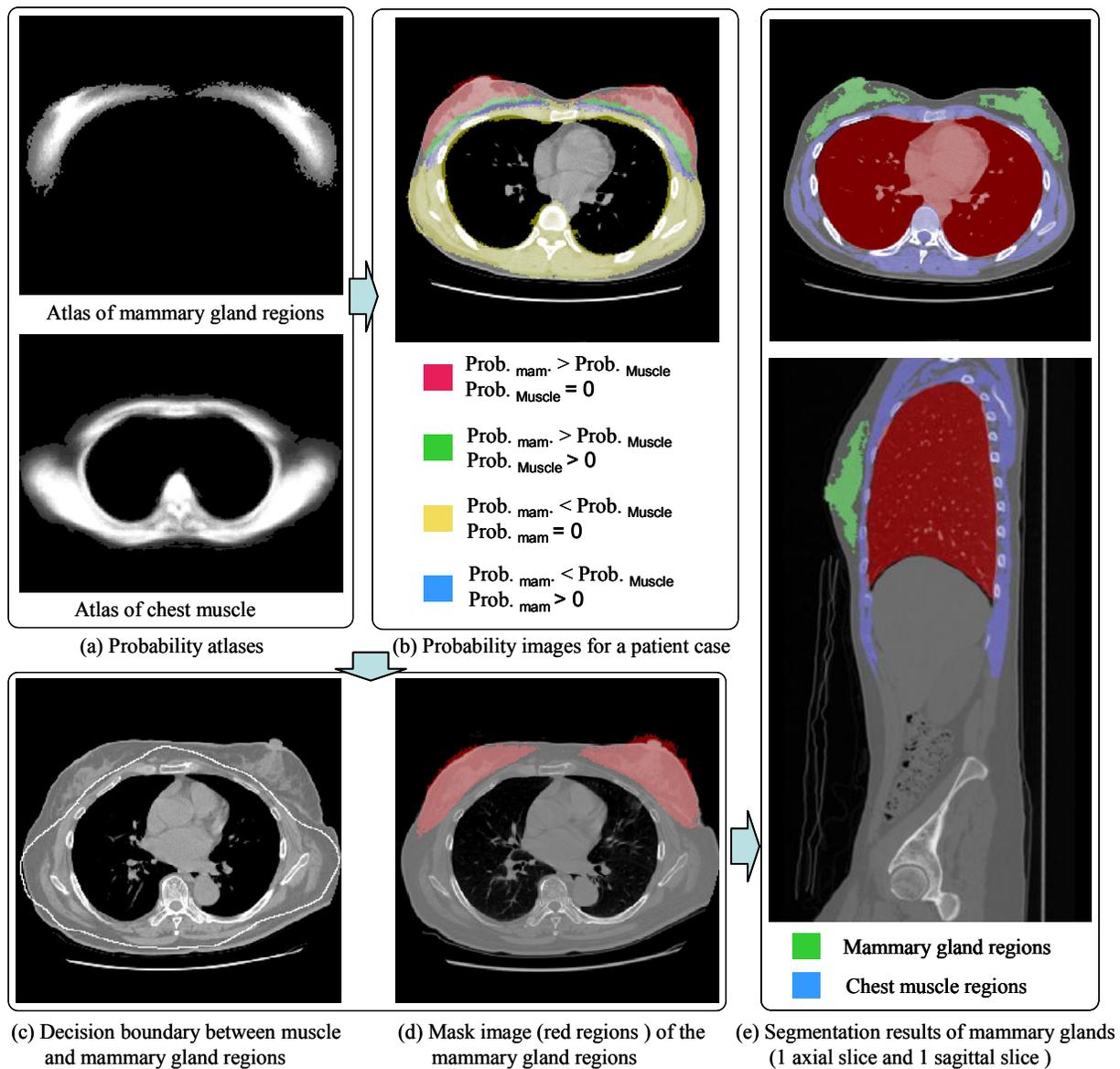


Fig. 5 Segmentation of mammary gland regions

### 3. EXPERIMENTS

Non-contrast CT scans of 66 females were used for mammary gland segmentation in this study. The age of the patients ranges from 40 to 80 years old. Each CT image has an isotropic spatial resolution of about 0.6 mm and density (CT number) resolution of 12 bits. The gold standard of mammary gland regions was identified manually by Author K.S. (a radiologist specialized in breast cancer). Accuracy of the mammary gland regions segmented by the proposed scheme was evaluated by using the coincidence ratio between the segmented result and the corresponding gold standard. The leave-one-out method (using 65 cases to construct an atlas and segment the mammary gland based on the atlas from the rest 1 case) was used in the experiment for performance evaluation. An example of the experimental result was shown in Fig 6 and Fig 7. By comparing gold standard of the mammary gland regions and segmentation result, we find that our scheme can provide satisfactory result on segmentation of false positive in CT images. The coincidence ratio between the segmentation result and gold standard for each CT case is graphically shown in Fig.8 for 66 cases. The mean value of the coincidence ratio was 0.82 with the standard deviation of 0.09.

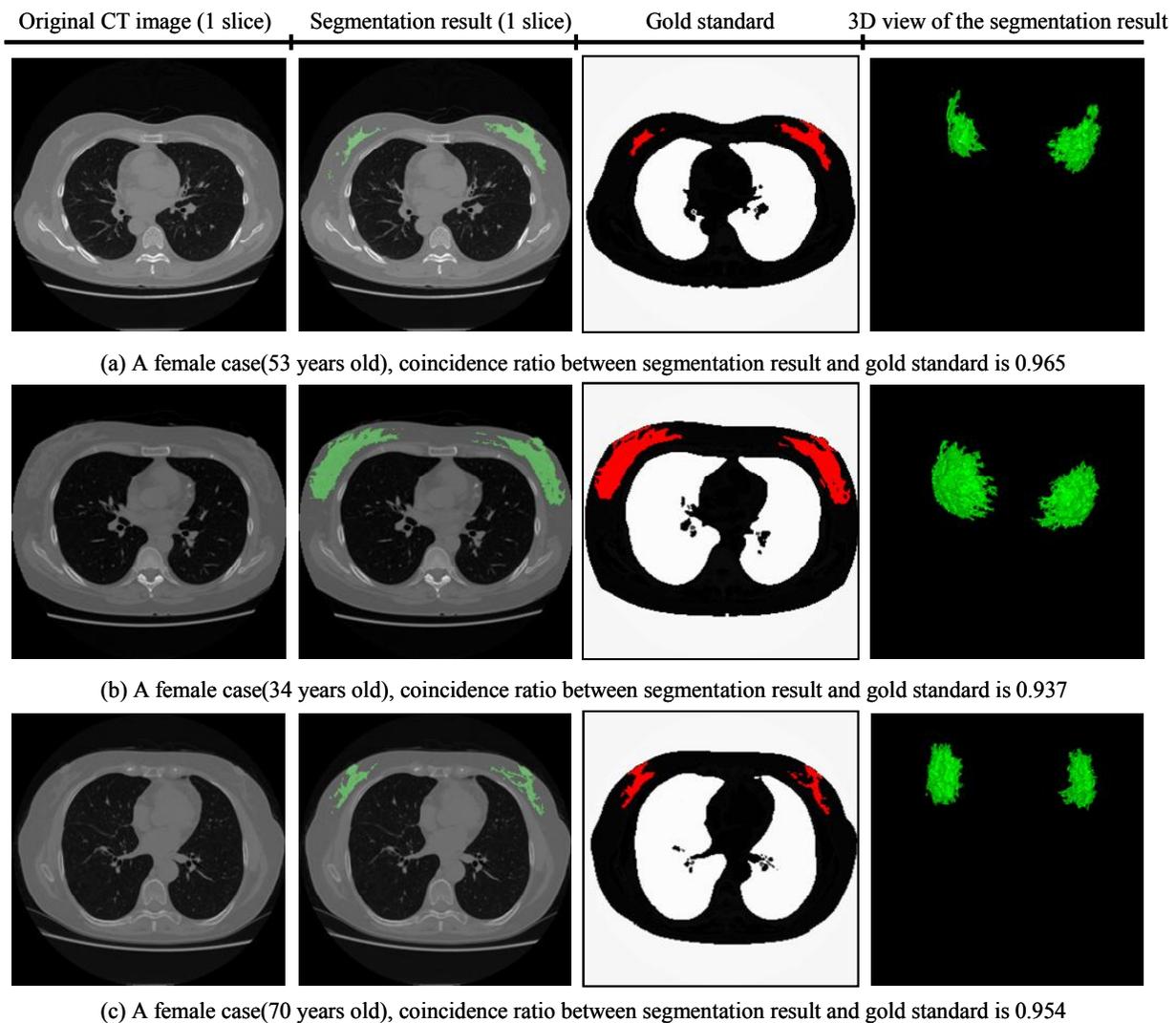


Fig. 6 Segmentation results (green) and gold standards (red) of mammary gland region.

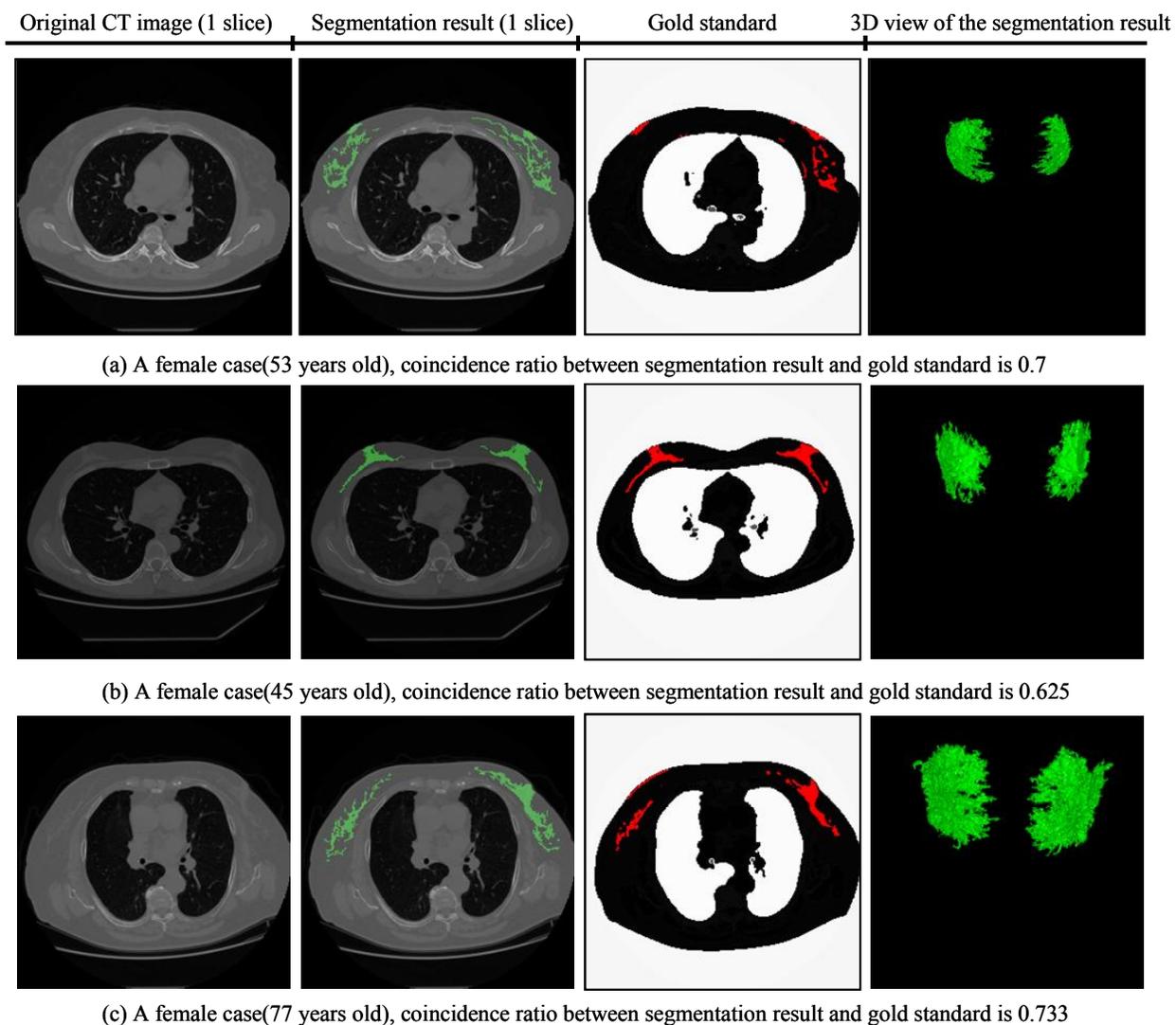


Fig. 7 Segmentation results (green) and gold standards (red) of mammary gland region.

#### 4. DISCUSSION

We confirmed that the majority of mammary gland regions in each CT case were extracted correctly. However, the over-extraction (false positive) was observed in many CT cases (Fig. 7). The over-extractions were occurred as the misclassification between chest muscle and mammary gland regions. Because the volume of mammary gland is small comparing the chest muscle regions, the over-extraction of the chest muscle regions could cause a large decrease of the coincidence ratio. We found that the accuracy of the result with the coincidence ratio larger than 0.8 (the cases shown in Fig. 6) was acceptable to show the mammary glands region on CT images.

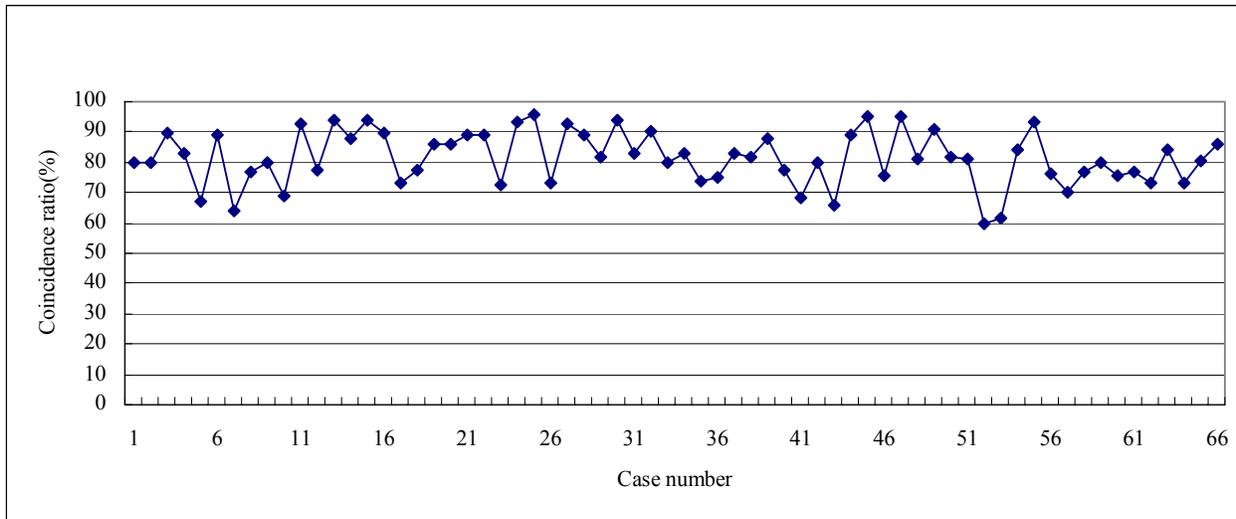


Fig. 8 Evaluation results based on gold standard of mammary gland regions for 66 CT cases.

## 5. CONCLUSION

We developed a fully automated scheme for segmenting the mammary gland regions on non-contrast torso CT images. The proposed method, using the probabilistic atlas and density (CT number) estimation, was efficient and robust for the mammary gland segmentation. We confirmed that our scheme could extract the mammary gland regions in 66 torso CT scans successfully and provide satisfactory results of the volume measurement for mammary glands that maybe useful at an initial stage for the breast tumor detection in CAD system.

## ACKNOWLEDGEMENTS

Authors thank to members of the Fujita Laboratory. This research was supported in part by research grants of Grant-in-Aid for Scientific Research on Priority Areas, in part by Ministry of Health, Labour, and Welfare under a Grant-In-Aid for Cancer Research, and in part by the Knowledge Cluster Initiative of the MEXT, Japanese Government.

## REFERENCES

1. R. L. Drake, W. Vogl, A. W.M. Mitchell, Gray's anatomy for students, *Elsevier*, 2005.
2. X.Zhou, T.Hayashi, T.Hara, H.Fujita, R.Yokoyama, T.Kiryu, and H.Hoshi, Automatic segmentation and recognition of anatomical lung structures from high-resolution chest CT images, *Computerized Medical Imaging and Graphics*, 30 (5), pp. 299-313, 2006.
3. X. Zhou, T. Hara, H. Fujita, R. Yokoyama, T. Kiryu, M. Kanematsu, and H. Hoshi, Preliminary study for automated recognition of anatomical structure from torso CT images, *Proc. of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, paper#340, 2005.
4. F. L. Bookstein, "Principal Warps:Thin-Plate Splines and the Decomposition of Deformations", *IEEE Trans. PAMI*, vol.11, (6), pp. 567-585, 1989.