




Modality Conversion from Pathological Image to Ultrasonic Image Using Convolutional Neural Network

Takashi Ohnishi¹ (✉) , Shu Kashio², Takuya Ogawa², Kazuyo Ito³, Stanislav S. Makhanov⁴, Tadashi Yamaguchi¹, Yasuo Iwadata⁵, and Hideaki Haneishi¹

¹ Center for Frontier Medical Engineering, Chiba University,
Chiba 263-8522, Japan
t-ohnishi@chiba-u.jp

² Graduate School of Science and Engineering,
Chiba University, Chiba 263-8522, Japan

³ Graduate School of Engineering, Chiba University, Chiba 263-8522, Japan

⁴ Sirindhorn International Institute of Technology, Thammasat University,
Khlong Luang 12120, Pathum Thani, Thailand

⁵ Graduate School of Medicine, Chiba University, Chiba 260-8670, Japan

Abstract. Relation analysis between physical properties and microstructure of the human tissue has been widely conducted. In particular, the relationships between acoustic parameters and the microstructure of the human brain fall within the scope of our research. In order to analyze the relationship between physical properties and microstructure of the human tissue, accurate image registration is required. To observe the microstructure of the tissue, pathological (PT) image, which is an optical image capturing a thinly sliced specimen has been generally used. However, spatial resolutions and image features of PT image are markedly different from those of other image modalities. This study proposes a modality conversion method from PT image to ultrasonic (US) image including downscale process using convolutional neural network (CNN). Namely, constructed conversion model estimates the US from patch image of PT image. The proposed method was applied to the PT images and we confirmed that the converted PT images were similar to the US images from visual assessment. Image registration was then performed with converted PT and US images measuring the consecutive pathological specimens. Successful registration results were obtained in every pair of the images.

Keywords: Modality conversion · Pathological image · Ultrasonic image
Convolutional neural network

1 Introduction

In recent years, the physical properties of human tissue such as mechanical, optical and acoustic properties have been widely measured. In addition, these properties have been compared with microstructure such as the distribution of the cell nuclei and the running

direction of nerve fiber [1–3]. The microstructure of tissue can be acquired as pathological (PT) images which are optical images of thinly sliced specimens. Methodologies of multi-modal analysis using such PT images and other modal images have been widely developed. We also have been analyzing a relationship between acoustic parameters and microstructures of the human brain using PT images and microscopic ultrasonic (US) images.

To compare the physical properties and the microstructure at the same location using multi-modal images, accurate image registration is required. Previous studies employed landmark-based or semi-automatic methods [4–7]. However, a correction of local differences was too difficult because tissue characteristics in the PT image are not taken into consideration in these methods, which makes detection of the corresponding landmarks difficult. In this case, an intensity-based registration may be more promising.

When the PT image is used in the intensity-based registration, a spatial resolution of the PT image can be a hurdle because it is extremely higher than that of other image modalities such as computed tomography, magnetic resonance imaging and US image. For example, the spatial resolution of the PT image is approximately $230 \times 230 \text{ nm}^2$ while the spatial resolution of the US image measured by an US microscopic system is approximately $8 \times 8 \text{ }\mu\text{m}^2$ at most. Therefore, when a pixel is selected on the US image during the registration process, the corresponding pixel value is calculated from 35×35 pixels on the PT image. In this situation, the spatial resolution of the PT image is generally adjusted to almost the same as those of another image using an averaging and downsampling technique before the image registration. However, such a simple downscaling processing eliminates microscopic pattern that each organ inherently possesses and leads to a decline in registration accuracy.

To enhance each structural component in the PT image and achieve the highly accurate image registration, we introduced an image feature conversion method combining with the downscale process. This study tuned up a conversion method assuming the image registration between PT and US images. As a preliminary experiment, simple affine registration was conducted.

2 Methods

The proposed method consists of two steps as shown in Fig. 1. In the model construction step, the landmark-based registration with PT and US images was conducted. US image was moved to the coordinate system of PT image in this registration process. Because the original PT image was too large, a region of interest was set to the PT image. Rescaled PT image was generated using the simple average method and then binarized with discriminant analysis method. The landmarks were detected by AKAZE feature detector [8] from the binarized PT image. Outliers of the landmarks were removed by random sampling consensus [9]. This registration results must be visually confirmed by the operator. The conversion model was constructed with the original PT image and the registered US image using convolutional neural network (CNN) [10]. Figure 2a shows the flow of the conversion model construction. Some patch images were extracted from the original PT image. The conversion model estimates an US signal from each small patch image. Estimated US signals p_i were compared with

actual US signals l_i . CNN was optimized until the cost function becomes minimum. The cost function which is also called a loss function was defined as:

$$\text{Cost}(\mathbf{P}, \mathbf{L}) = \sum_{k=1}^N |p_k - l_k|. \quad (1)$$

Here, k and N describe the index of patch image and the total number of patch image input into the CNN, respectively. These processes were repeated until iteration number reached to a predefined limit. The framework of the CNN is shown in Fig. 2b. There were two convolution layers and two pooling layers followed by dropout and fully connected layers. CNN construction has to be conducted once before the actual registration.

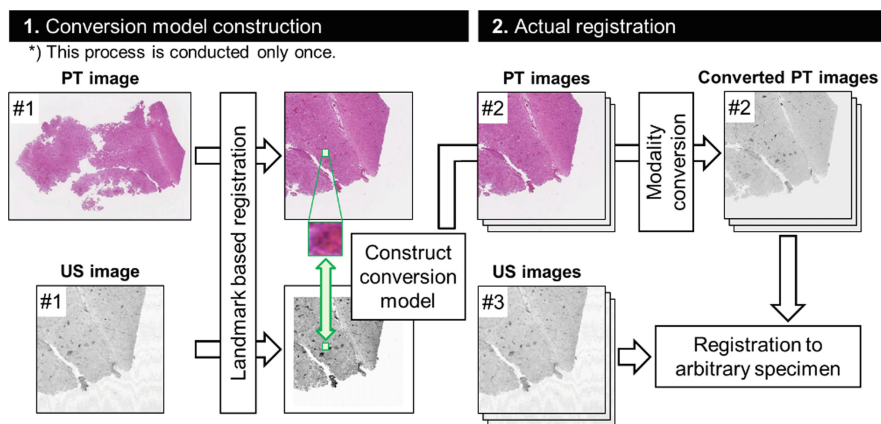


Fig. 1. Outline of the proposed method.

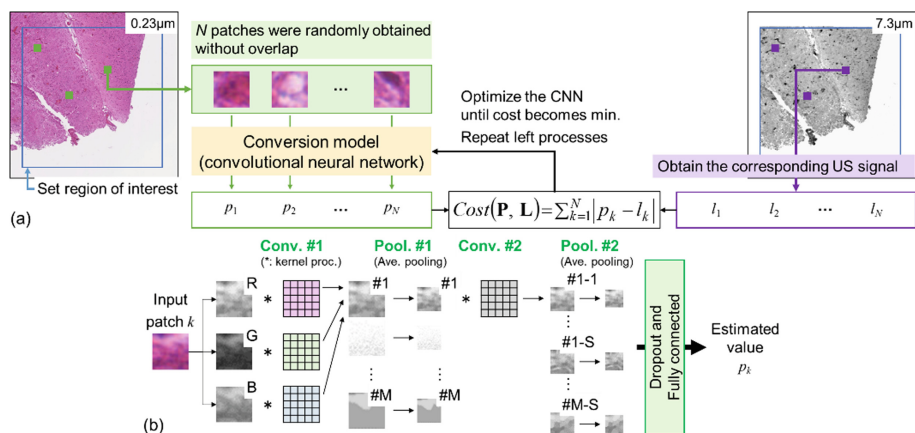


Fig. 2. (a) Flow of conversion model construction, (b) Framework of convolutional neural network.

In the second step, the obtained PT images for the image registration were converted by the constructed model. Affine registration including shift, rotation and scaling operations were then conducted. Normalized cross correlation (NCC) and Powell-Brent method were utilized as a similarity measure between converted PT and US images and an optimization method, respectively.

3 Experiments

3.1 Data

Brain tumors have been resected from four patients as a normal clinical procedure. After the surgery, the resected tumors were further resected into some pieces. These obtained pieces were named macro-specimens S1 to S4. This study has been approved by the Ethical Review Board of our University and we obtained informed consent from all the four patients participated in the study. Each resected macro-specimen was then undergone formalin fixation and paraffin embedding. Pathological specimens with 8- μ m thickness were then obtained by using a microtome. These specimens were then deparaffinized with xylene and cleaned with ethanol. For US measurement, the images of specimens in this status were captured. These specimens were further stained with hematoxylin-eosin (HE) and the PT images of the specimens were then captured.

For macro-specimen S1, the sectioning by the microtome was performed repeatedly and 19 consecutive pathological specimens were obtained. US measurement and PT image acquisition were performed on the only first pathological specimen. A pair of PT and US images acquired in this process was utilized to construct the conversion model. For the other pathological specimens, US and PT images were acquired from odd and even numbered pathological specimens, respectively. As for macro-specimens S2–4, one pathological specimen was obtained from each macro-specimen and a pair of US and PT images was acquired in each macro-specimen just as the pair of PT and US image of S1.

US images were obtained as two-dimensional echo amplitude map in depth direction at each scan point of pathological specimens. We used two ultrasonic microscopic systems. One is a modified version of a commercial product (AMS-50SI, Honda Electronics Co., Ltd, Japan) and was used for S1. The other is an in-house developed system and was used for S2–4. In both systems, a ZnO wave transducer (Fraunhofer IMBT, St. Ingbert, Germany) with 250 MHz center frequency was commonly used. This transducer was attached to the X-Y stage and scanned with 8- μ m pitch in each direction. Echo amplitude was calculated from acquired RF echo signal at each scan point and used as pixel value of US image. Image size and pixel size were 300×300 to 800×800 pixels and $8.0 \times 8.0 \mu\text{m}^2/\text{pixel}$. Detailed calculation way of the echo amplitude was described in [11].

For PT image acquisition, HE-stained pathological specimens were digitalized with a virtual slide scanner (NanoZoomer S60, Hamamatsu Photonics K.K., Japan). Image size and pixel size were approximately 12000×12000 pixels and $228 \times 228 \text{ nm}^2$, respectively.

3.2 Results and Discussions

In this study, two kinds of experiments were conducted. In the first experiment, applicability of the conversion model was evaluated with PT and US images obtained from the same macro-specimen S1. A conversion model was constructed with a pair of PT and US images and applied to other nine PT images. Image registration was then undergone. The first US image was used as reference image and the other images were registered into the first US image. To evaluate the versatility of the conversion model, another experiment was conducted with the images of S2–4. A conversion model was constructed with the images of S2 and applied to the PT images of S3 and S4. The patch size for the conversion model was set to 32×32 pixels. Namely, the pixel size after conversion were $7.30 \times 7.30 \mu\text{m}^2$. The number of iteration, the number of patch image (batch size), a learning rate and a dropout rate for CNN were set to 20,000, 100, 0.001 and 0.5, respectively.

Experiment 1: Study on the Consecutive Specimens Resected From a Patient

Figure 3 shows a result of conversion model construction. As yellow arrows in Fig. 3c indicate, black spots were clearly enhanced after conversion. From visual assessment, the converted PT image was similar to the US image compared with the simply downscaled PT image. The constructed conversion model was applied to other PT images. Conversion results are shown in Fig. 4. Because US images corresponding to PT images were not acquired, we could not evaluate the conversion effect quantitatively. However, we confirmed that image features of all converted images were similar to those of US image shown in Fig. 3d.

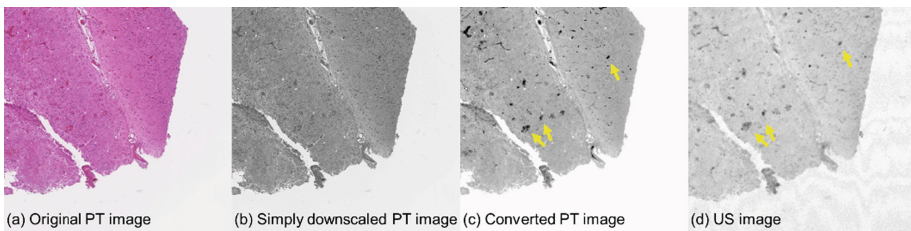


Fig. 3. Result of conversion model construction. (a) Original and (b) Simply downsampled PT image, (c) Converted PT image, (d) US image. Yellow arrows indicate enhanced regions by the conversion. It should be noted that image size of the original PT image was larger than that of other three images in practice. (Color figure online)

Registration results are shown in Fig. 5. All images including both US and PT images were successfully registered. Because image features of converted PT and US images were similar, NCC provided acceptable results.

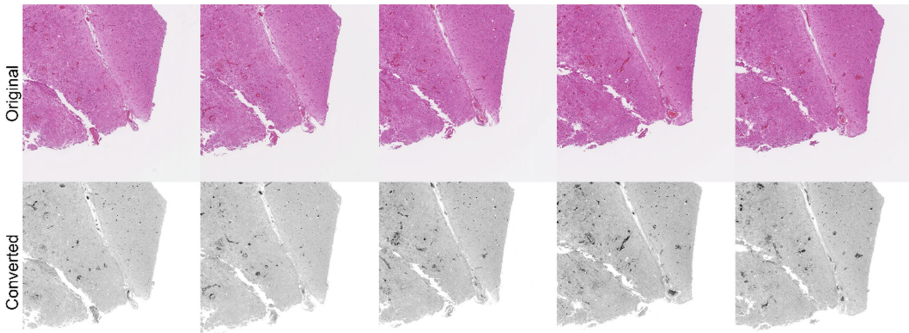


Fig. 4. Conversion results of S1. Top: Original PT images. Bottom: Converted PT images using the same model in Fig. 3.

Experiment 2: Study on the Specimens Resected From Different Patients

A conversion model was constructed with the image dataset of S2 and applied to the image datasets of S3 and S4. Resultant images are shown in Fig. 6. NCC between the US image and each downsampled PT image was calculated at the best match position. The best match positions for each image were manually decided by the operator. Calculated NCCs are shown in Table 1.

Some structures in the converted PT image of S2 were slightly enhanced as shown in enlarged region of Fig. 6. For image dataset of S3, tendency of conversion result was similar to S2. We confirmed the effect of the modality conversion, however it was less than that in the previous experiment. For the image dataset of S4, pixel values of the converted PT image were almost the same and image contrast became low. Specific region could not be enhanced after modality conversion. The cause of this result was that there were many necrosis regions on S4. On the other hand, learning image dataset did not include such region. Namely, the variety of pathological structures to be learnt was not sufficient to provide a versatile conversion model. Learning image dataset must be generated from patch images whose pathological structure are distinctive and diverse.

Every NCC with modality converted PT image was higher than that with simply downsampled PT image. Even though simple downscale method led to high NCC, the proposed method further improved it because the conversion model was optimized with S2. NCCs of the other two datasets were also improved and achieved more than 0.9 using the conversion model. Although we could not visually confirm the effectiveness of the proposed method, histogram or spatial distribution of pixel value was similar to that of the US image. From these results, we can expect that the proposed method will be able to produce the better registration than the simple downscale method.

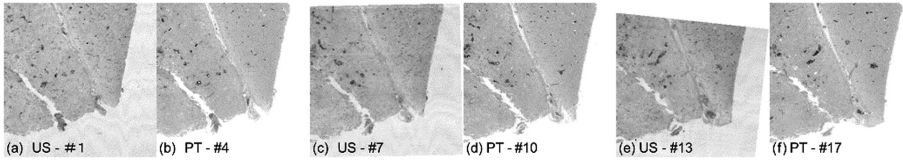


Fig. 5. Example of registration results. (a) Reference US image, (c)(e) Registered US images, (b)(d)(f) Registered PT images. The numbers describe pathological specimen number.

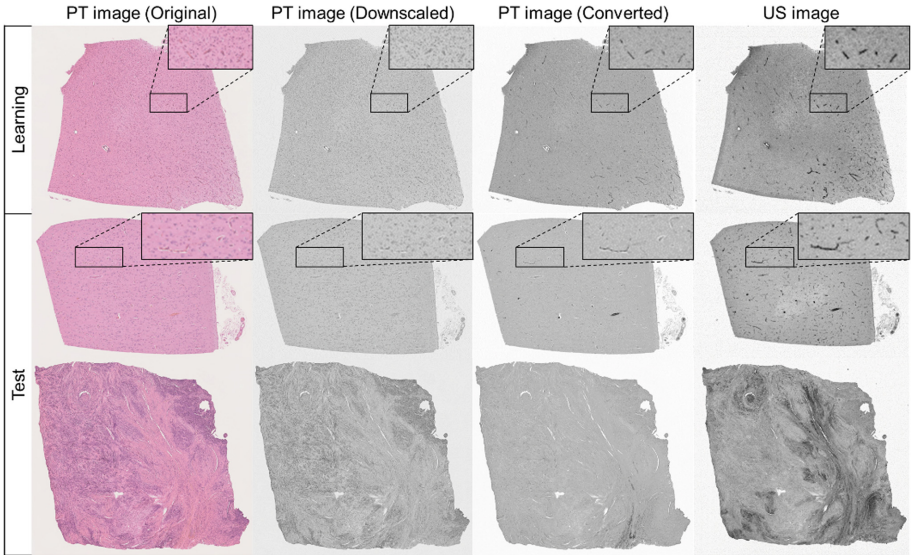


Fig. 6. Conversion results of S2-4. Top: Learning image dataset (S2), Middle and bottom: Test image dataset (S3 and S4).

Table 1. Normalized cross correlation between ultrasonic and each downscaled pathological image at best match position.

	S2	S3	S4
Simply downscaled	0.956	0.796	0.803
Modality converted	0.978	0.955	0.915

4 Conclusions

To conduct the image registration with ultrasonic image, we proposed a CNN-based modality conversion method for pathological image. From visual assessment, converted pathological images were similar to the ultrasonic image compared with the simply downscaled pathological image. Therefore, we will be able to obtain the highly accurate registration result without additional intelligent and/or complicated registration method.

We found that the registration results were successfully obtained using classical similarity measure. For future work, we would like to increase the various image datasets because the CNN can not estimate acceptable US signal from unseen image pattern. Quantitative assessment of the image registration and comparison with the landmark-based registration method will be also conducted. In addition, because pathological specimen might be locally deformed in the staining process, a non-rigid registration method would be introduced.

Compliance with Ethical Standards

Funding. This study was partly supported by MEXT KAKENHI, Grant-in-Aid for Scientific Research on Innovative Areas, Grant Number 17H05278.

Conflict of Interest. The authors declare that they have no conflict of interest.

Ethical Approval. All procedures performed in study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent. Informed consent was obtained from all individual participants in the study.

References

1. Ban, S., Min, E., Baek, S., Kwon, H.M., Popescu, G., Jung, W.: Optical properties of acute kidney injury measured by quantitative phase imaging. *Biomed. Opt. Express* **9**(3), 921–932 (2018)
2. Nandy, S., Mostafa, A., Kumavor, P.D., Sanders, M., Brewer, M., Zhu, Q.: Characterizing optical properties and spatial heterogeneity of human ovarian tissue using spatial frequency domain imaging. *Biomed. Opt.* **21**(10), 101402-1–101402-8 (2016)
3. Rohrbach, D., Jakob, A., Lloyd, H.O., Tretbar, S.H., Silberman, R.H., Mamou, J.: A novel quantitative 500-MHz acoustic microscopy system for ophthalmologic tissues. *IEEE Trans. Biomed. Eng.* **64**(3), 715–724 (2017)
4. Choe, A.S., Gao, Y., Li, X., Compton, K.B., Stepniewska, I., Anderson, A.W.: Accuracy of image registration between MRI and light microscopy in the ex vivo brain. *Magn. Reson. Imaging* **29**(5), 683–692 (2011)
5. Goubran, M., et al.: Registration of in-vivo to ex-vivo MRI of surgically resected specimens: a pipeline for histology to in-vivo registration. *J. Neurosci. Method* **241**, 53–65 (2015)
6. Elyas, E., et al.: Correlation of ultrasound shear wave elastography with pathological analysis in a xenografic tumour model. *Sci. Rep.* **7**(1), 165 (2017)
7. Schalk, S.G., et al.: 3D surface-based registration of ultrasound and histology in prostate cancer imaging. *Comput. Med. Imaging Graph.* **47**, 29–39 (2016)
8. Alcantarilla, P.F., Nuevo, J., Bartoli, A.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: *Proceedings of British Machine Vision Conference*, pp. 13.1–13.11 (2013)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)

10. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Netw.* **8**(1), 98–113 (1997)
11. Kobayashi, K., Yoshida, S., Saijo, Y., Hozumi, N.: Acoustic impedance microscopy for biological tissue characterization. *Ultrasonics* **54**(7), 1922–1928 (2014)