3D Shape Reconstruction of Japanese Traditional Puppet Head from CT images

Hiroyuki Ukida[1], Kouki Yamazoe[1], Masahide Tominaga[2],
Tomoyo Sasao[3] and Kenji Terada[1]

[1] Graduate School of Technology, Industrial and Social Sciences, Tokushima University,
Tokushima, Japan
[2] Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan
[3] Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
ukida@tokushima-u.ac.jp

Abstract. In this paper, we discuss a 3D shape reconstruction method of Japanese traditional puppet heads for a digital archiving. Especially, to reconstruct an inner shape of head, we use CT images. First, we divide four regions (wood, hair, paint, and air) by thresholds based on manual directed regions. After that, we divide these regions by a graph cut method. And we also present a method to estimate 3D shape of parts in puppet head. This method is also based on a graph cut method. Moreover, we also discuss a method to distinguish material of puppet head by machine learning. Here we use "U-Net" to extract wood parts of puppet head from its CT images. And we show experimental results by these methods.

Keywords: X-ray CT images, Puppets, Shape reconstruction, Graph cut method, Machine learning.

1 Introduction

"Awa Ningyo Joruri" is traditional Japanese puppet theater, especially, it has been played in Tokushima prefecture. In recent years, however, there has been a shortage of successors to the Ningyo Joruri. To preserve this culture and pass it on to future generations, this research challenges a digital archive of puppets.

"Digital archiving" [1] is a technique to preserve an object's data such as, 3D shapes, color, gloss, and so on semi permanently. This technique has been studied and applied to the national treasures, the important cultural properties, old documents, etc. We have also discussed methods to measure 3D outer shape of puppet heads precisely using a 3D scanner, a turn table, and an arm robot [2-3].

Fig.1 shows examples of puppet heads. As a characteristic of "Awa Ningyo Joruri" puppets, many of them have some mechanisms in their heads such as moving eyes or opening/closing mouth and so on. Therefore, to archive puppet information more precisely, we must reconstruct inner parts of puppet head. Fig.2 shows an inside of puppet head in production. We can see parts of eyes and neck in it. But it is difficult to see the

inside of completed head because we cannot disassemble it. To observe the inner construction of puppet head, we use X-ray Computer Tomography (CT) imaging system. And we discuss a method to reconstruct a shape of puppet head from CT images.

Main material of puppet head is the dry wood. However, it is not displayed clearly in CT images because of a low moisture. Moreover, some other materials are used in a puppet head which are hair wig, whitewash paint on the surface and metal nail to fix wig on a puppet head. The ranges of intensities of these materials in CT images are partly overlapped that of dry wood. Therefore, to reconstruct the shape of puppet head, we must discuss a method to distinguish materials in CT images. In this paper, we propose a method to distinguish and extract materials in CT images using thresholding and graph cut method. And we also present a method to estimate 3D shape of parts in puppet head. This method is also based on a graph cut method.

Moreover, we also discuss a method to distinguish material of puppet head by machine learning. Here we use "U-Net" to extract wood parts of puppet head from its CT images. And we show experimental results by these methods.
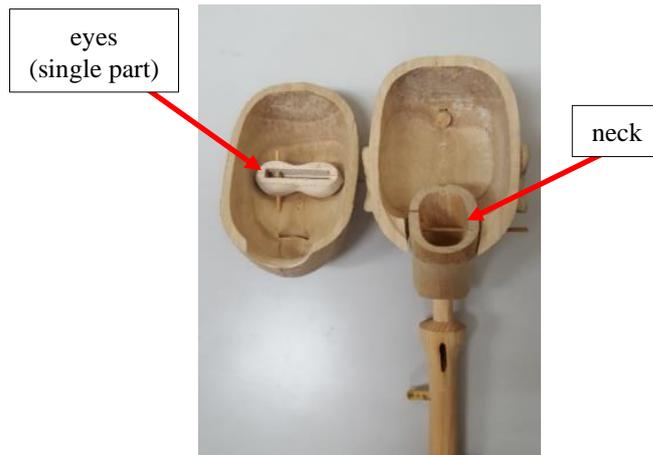

Fig. 1. Examples of puppet head.

Fig. 2. Inside of puppet head.

2   3D Shape Reconstruction from CT Images

2.1   Extracted regions in CT images

Fig.3 shows an example of a CT image of a puppet head. In this study, we assume four regions in CT images of a puppet head.

1. Air region inside and outside of the head.
2. Hair wig region put on the head.
3. Dry wood region of the head.
4. Paint region on the surface of head.

In an actual puppet head, other materials (ex. metal nails) are also included. But the sizes of these materials are not large. Therefore, we distinguish above four regions from CT images.

Material extraction methods we propose consist of two methods as following:

1. Rough region segmentation using thresholds obtained from histogram of intensities in manual directed area in a CT image.
2. Precise region extraction using a graph cut method from the result of 1.
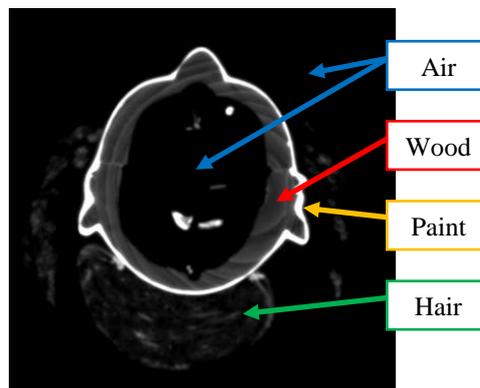
We explain these methods in next subsections.

Fig. 3. A CT image of a puppet head.

2.2   Extracting materials based on histogram

This subsection shows a material extracting method based on histogram.

- Step 1-1.   About four regions (air, hair, wood, and paint), we obtain histograms from areas directed by manual in some CT images.
- Step 1-2.   Each histogram is approximated to the normal distribution by equation (1):

$$f_k(x) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) \qquad (1)$$

where k is a kind of region, x is intensity, μ_k and σ_k are mean and standard deviation of region k.

- Step 1-3. Estimate thresholds as cross points of normal distributions which are neighboring two regions.

By using estimated thresholds, we can divide four regions in CT images roughly.

2.3　Precise extraction by graph cut method

The graph cut method estimates the combination of pixel's label (object or background) in images in a criterion of a cost minimization efficiently under the condition in which some parts in image are assigned labels beforehand [4].

In this study, we reconstruct 3D shape of dry wood and paint parts only. Because the shape of hair wig deforms depend to the attitude of puppet head, it is difficult to model the 3D hair shape. Hence, the dry wood and paint regions are considered as object label and hair and air regions are background.

The cost function $E(L)$ used in the graph cut method is shown in equation (2) as the linear combination of the region term $R(L)$ and the boundary term $B(L)$:

$$E(L) = R(L) + \lambda \cdot B(L) \qquad (2)$$

where L is label assigned to pixels ($L \in \{\text{obj, bkg}\}$), λ is a weight between the region term and the boundary term (non-negative value).

The region term $R(L)$ and the boundary term $B(L)$ are described as:

$$R(L) = \sum_{u \in U} f_u(L_u) \qquad (3)$$

$$B(L) = \sum_{\{u,v\} \in N} g_{u,v}(L_u, L_v) \qquad (4)$$

where f_u is a likelihood of a region, $g_{u,v}$ is a likelihood of a boundary between neighboring pixels. U is a set of pixels and u shows a pixel. N is a set of neighboring two pixels and $\{u, v\}$ is a tuple of pixels. f_u and $g_{u,v}$ are described as follows:

$$f_u(L_u) = -\ln \Pr(I_u|L_u) \qquad (5)$$

$$g_{u,v}(L_u, L_v) = \begin{cases} \frac{\exp\{-\beta(I_u - I_v)^2\}}{\text{dist}(u,v)} & (L_u \neq L_v) \\ 0 & (L_u = L_v) \end{cases} \qquad (6)$$

$\Pr(I_u|L_u)$ is a likelihood of pixel I_u in each region. This is approximated to a normal distribution. β is constant and $\text{dist}(u,v)$ is a distance between neighboring pixels.

Fig.4 shows a process of graph cut method. In this figure, the node u is corresponding to a pixel, and it connected to a neighbor pixel (node) v. We call this linkage "n-link" and we assume that this link has a cost estimated by equation (6). The node u is also connected to the node s labeled object ("obj") and the node t labeled background

("bkg"). These linkages are called "t-link" and they have costs estimated by equation (5). Cutting off a link from u to t or s can be regarded as labeling "obj" or "bkg" to u. Therefore, when the sum of costs of these linkages is minimum, the label assigned u minimize equation (2). In this study, we apply the minimum cut / maximum flow algorithm by Boykov [5] to the cost minimization method.
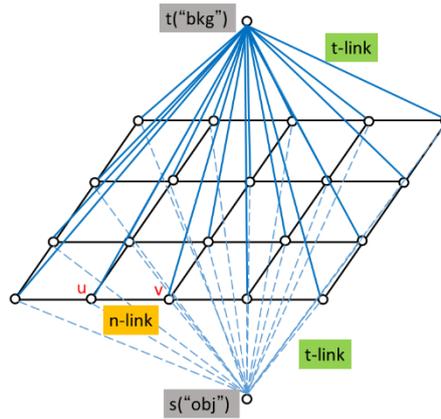

Fig. 4. Example of graph structure.

In the method of region extraction by graph cut, we must assign the object or background labels to some parts of regions as "seed" certainly. In general, this "seed" is usually assigned by user manually in graph cut method. But, in this study, we try to assign "seed" without user input by using following steps.

- Step 2-1.   By applying a method in subsection 2.2 to a CT image, dry wood and paint regions are extracted in this image.
- Step 2-2.   By using morphologic methods (dilation and erosion), we fill holes and eliminate small regions (about 10 by 10 pixels) in the image of Step 2-1.
- Step 2-3.   We estimate a distance transformation image from the result of Step 2-2, and extract two regions which are:
  - One region consists of some pixels which distance is from maximum distance (D_{max}) to $D_{max} - \Delta_1$,
  - Other region consists of some pixels which distance is from minimum distance (D_{min}) to $D_{min} + \Delta_2$,

where Δ_1 and Δ_2 are distance thresholds and they are assigned some values previously. As a result, we can get two seeds for graph cut method.

Moreover, CT images of a puppet has many images, and they are aligned perpendicular to the image plane, so to assign seeds to all CT images, we use following steps:

- Step 3-1.   In a neighbor (upper of lower) CT image of a labeled CT image, firstly, the region extraction based on histogram is applied. After that, we obtain an overlapped region as "seed" in which the dry wood region in this image and the object

region in a labeled CT image. By using this seed region, we apply the graph cut method.

- Step 3-2. By propagating results of Step 3-1 to the upper and lower CT images, we extract object regions in all CT images.

2.4   Extraction of inner parts

In subsection 2.3, we explained the object (wood and paint) and background (hair and air) extraction method in CT images. In this section, we present a parts extraction method in puppet head. Herein, we show an eye part extraction method as an example. In puppet head, its eye is one part because two eyes are connected as shown in Fig.2.

In a preliminary experiment, we confirmed that the sagittal CT images (from left to right) could obtain the 3D shape of the eye part more accurately than the normal axial CT image (from bottom to top). This is because that the 2D shape of eye part is simple in sagittal plane. On the other hand, CT images can be used in three directions: axial, sagittal, and coronal (from front to back). For parts other than the eyes, other orientations may be appropriate. Therefore, the part shape is obtained for each of the three directions, and the final part shape is reconstructed by integrating them.

In the following, the method using the case of extracting an eye from a sagittal CT image as an example is illustrated. The range of an eye part is specified manually in advance as shown as red lines in Fig.5 (a).

- Step 4-1. Select a CT image including an eye part and specify seed regions of object (eye part) and background (otherwise) manually. In Fig.5 (b), white line is object seed and black line is background seed.
- Step 4-2.   By using these seed regions and the graph cut method, we extract eye part area in selected CT image (Fig.5 (c)).
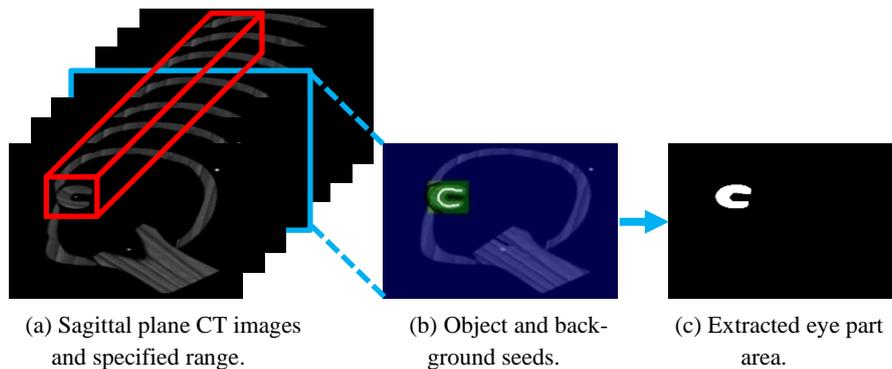

| (a) Sagittal plane CT images and specified range. | (b) Object and background seeds. | (c) Extracted eye part area. |

Fig. 5. Extraction of eye part.

By using results of Step 4-2 and following method, we estimate object and background seeds automatically in CT images adjacent to the left and right sides of the selected CT image and extract the eye part.

- Step 5-1. Estimate binarized images of object and background from eye part extracted image in Step 4-2.
- Step 5-2. Estimate a distance transform image from an object image in Step 5-1, and extract pixels whose values are from D_{min} to D_{max}.
- Step 5-3. In an adjacent CT image, we estimate object seeds in which the object region extracted in subsection 2.3 and extracted pixels in Step 5-2 overlap.
- Step 5-4. Similarly, we estimate a distance transform image from a background image, and extract pixels whose values are from D_{min} to D_{max}.
- Step 5-5. In the adjacent CT image, we estimate background seeds in which the background region extracted in subsection 2.3 and extracted pixels in Step 5-4 overlap.
- Step 5-6. By using the object and background seeds, we estimate eye part area in the adjacent CT image.

By applying the method from Step 5-1 to Step 5-6 in specified CT images, we can extract the 3D shape of eye part. In addition, the above method is applied to CT images of the axial and coronal direction to obtain the shape of the eye in each case. For the three eye shapes obtained, the common part is used as the final eye shape. The same process is applied to other parts to obtain their shapes.

2.5   Extraction by machine Learning

In this section, we explain the method to distinguish materials from CT images by machine learning. Here, we extract only wood parts by dividing CT images into wood and other regions. As a method of two-class segmentation, we use "U-Net" [6] in this study.

U-Net is a network based on convolutional neural networks (CNNs). Fig.6 shows the configuration of the U-Net used in this study. U-Net consists of an encoder part and a decoder part. It has eight layers in both the encoder and decoder parts. The layers are joined in U-shape. In the encoder part, features are extracted from the input image by convolutional and pooling layers. In the decoder part, the extracted features are used to restore the image by the inverse convolution layer.

However, in general, the feature extraction in the encoder part discards the positional information in the image, so even if the decoder part restores the image, it will not be able to recover the same image as the original. Therefore, U-Net introduces shortcut joints. This is a method that concatenates the output from one reverse convolution layer of the decoder part with the features from the encoder part at the same level to perform the next reverse convolution process. By using this method, we can restore the image including the positional information.
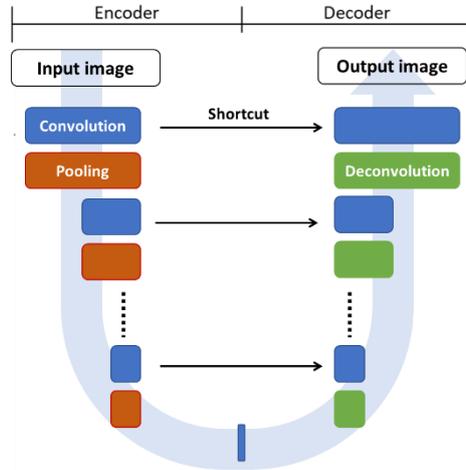
Fig. 6. Configuration of U-Net.

3　　Experiments

3.1　　Puppet head and CT images

Fig.7 shows a puppet head we use in this experiment. There is a mechanism to rotate eye up and down in this head. Fig.7 shows two of CT images of this head. These are intersection of puppet head at red lines A and B in Fig.8. ((b) is as same as Fig.3.) The black parts are air, the gray parts are dry wood and hair wig, and the white parts are paint on the wood surface.
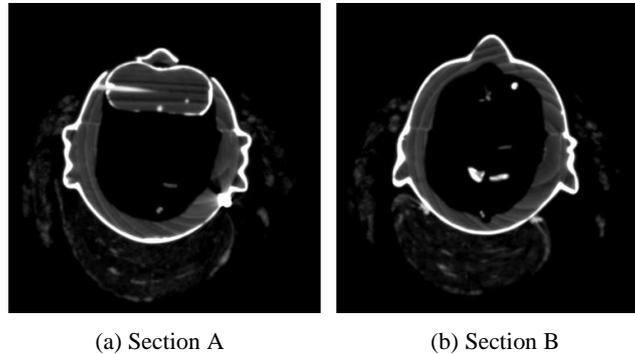

Fig. 7. Puppet head for experiments.

<table>
<tr><td>(a) Section A</td><td>(b) Section B</td></tr>
</table>

Fig. 8. CT images of puppet head.

3.2 Experimental results of puppet shape extraction

Fig.9 shows histograms and normal distributions of four regions (air, hair wig, dry wood, and paint) denoted in section 2.2. Table 1 shows means and variances of normal distributions and thresholds to divide these regions. Fig.10 shows results of region segmentation of Fig.8. As shown in Fig.9, some parts of dry wood (red) are estimated as hair wig (green), and some parts of hair wig are also estimated as dry wood.

Fig.11 shows separated wood and paint region from Fig.10 (a) denoted in subsection 2.3. Fig.12 shows result of dilation and erosion method. Fig.13 shows result of distance transformation. In Fig.14 (a), white pixels are a region which assigns label "obj". In Fig.14 (b), white pixels are regions which assigns label "bkg".

Fig.15 shows the segmentation results by the graph cut method. The object part (dry wood and paint) is white, and the background part (air and hair wig) is black. In these images, most of hair wig regions can be extracted as the background, but some parts of hair wig regions are extracted as the object.

Fig.16 shows reconstructed 3D shape of puppet head. (a) shows extracted dry wood, paint, and hair wig from histogram thresholds. (b) shows extracted objects (dry wood and paint) by the graph cut method. In the result of (b), at the top of puppet head, the region of hair wig is remained. This is because that the hair wig is dense at this part to fix hair wig to wood parts, the intensity of hair wig is almost as same as the dry wood. Hence, we must discuss another method to segment these parts.

3.3 Experimental results of inner parts extraction

Fig.17 shows a result of eye part extraction by a method in subsection 2.4. (a) shows the shape obtained from CT images in the axial direction. Similarly, (b) is the shape obtained from the CT images in the sagittal direction and (c) is the shape obtained from the CT images in the coronal direction. (d) is the shape extracted only from the common part of (a), (b), and (c). In Fig.17 (a), (b) and (c), some wood parts of face are added. Because of the existence of wood parts around these parts, the proposed graph cut method cannot distinguish pixels which are parts or not. However, as shown in (d), by

obtaining only the common part, the unnecessary part is removed, and the eye shape is obtained correctly.

Fig.18 shows a result of neck part extraction by same method. (a), (b), and (c) show the results obtained from CT images in the axial, sagittal, and coronal directions. (d) is the shape extracted only from the common part of (a), (b), and (c). In the case of the neck parts, contrary to the case of the eyes, a part of the neck shape is missing in all the results (a), (b), and (c). This case is also that the proposed graph cut method cannot distinguish pixels which are parts or not. Therefore, it is necessary to discuss a method how to extract the wood part more accurately.
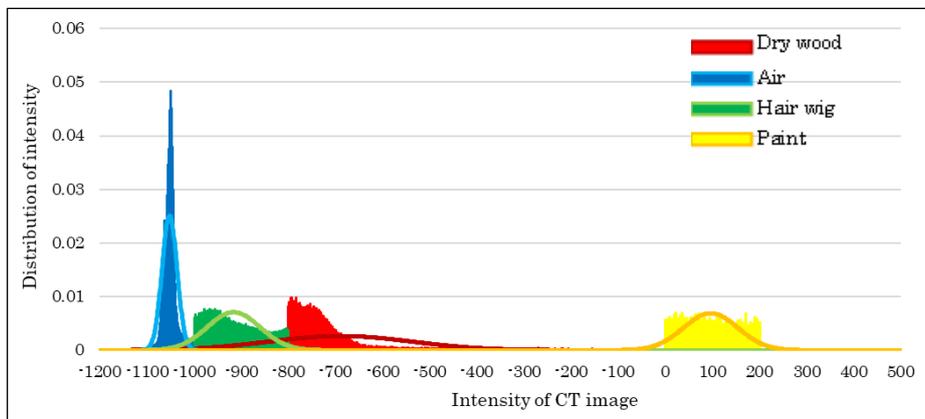
Fig. 9. Probability distribution and normal distribution curve of CT value in each of region.

Table 1. Estimated means, variances, and thresholds.

	air	hair wig	dry wood	paint
mean	-1051.9	-916.1	-689.4	96.3
variance	253.0	3166.9	22796.7	3347.3
threshold	-1014		-822	-131

(a) Section A   (b) Section B
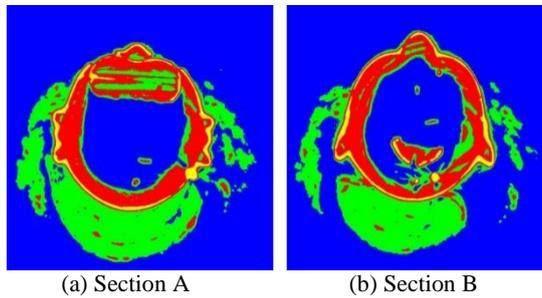
Fig. 10. Results of region segmentation by histogram.

Fig. 11. Extracted wood and paint regions from Fig.10(a).
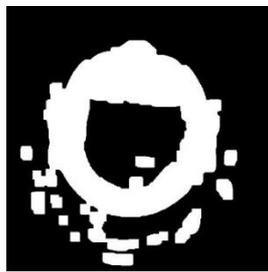
Fig. 12. Result of dilation and erosion method.
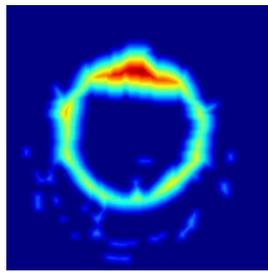
Fig. 13. Result of distance transformation image.
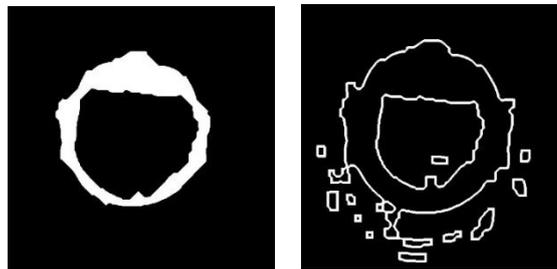
(a) Object label.   (b) Background label.

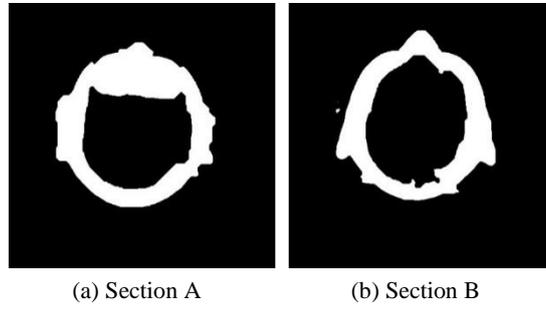**Fig. 14.** Seeds of labels for graph cut.

(a) Section A   (b) Section B

Fig. 15. Results of extracted regions by graph cut method.

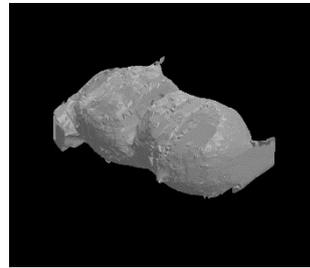(a) Before applying graph cut.

(b) After applying graph cut.
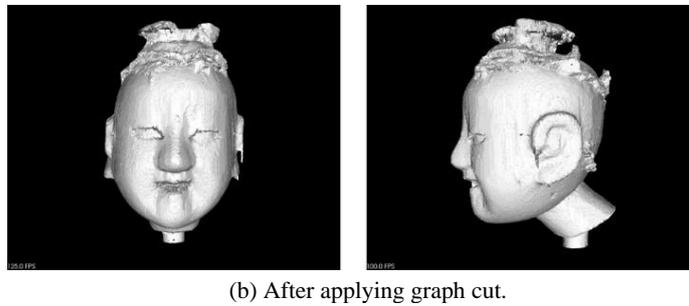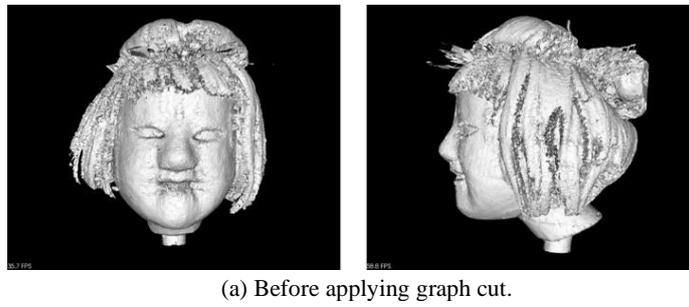
Fig. 16. Reconstructed 3D shape of puppet head.
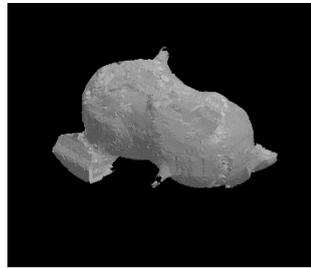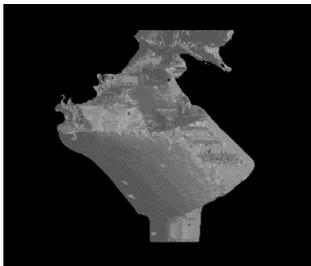
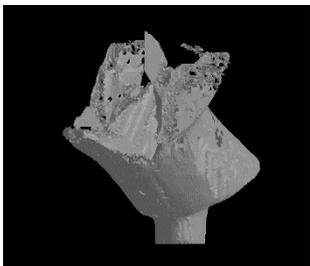(a) Axial direction.
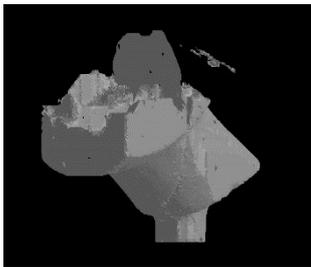
(b) Sagittal direction.
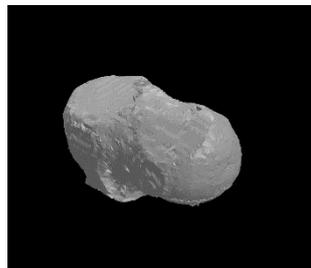
(c) Coronal direction.
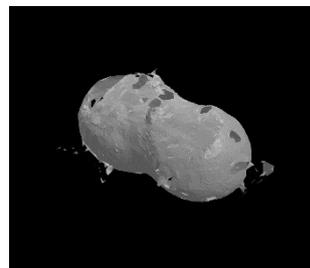
(d) Common part.

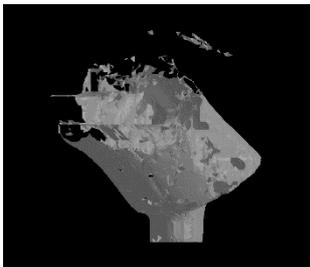**Fig. 17.** Reconstructed 3D shape of eye part.

(a) Axial direction.

(b) Sagittal direction.

(c) Coronal direction.

(d) Common part.

Fig. 18. Reconstructed 3D shape of neck part.

3.4 Experimental results of puppet shape extraction by U-Net

In this section, we show the extraction results of wood parts using machine learning (U-Net). First, 18 CT images are selected as training data from the 341 CT images used in Section 3.2. These images are selected from the entire CT image at equal intervals. For these CT images, only the wood part is extracted using the histogram thresholds shown in Section 2.2, and then the region of the wood part is manually modified to make the image data representing the ground truth in the training data. In practice, the ground truth is manually estimated for all CT images (341 images) to obtain the extraction accuracy.

Next, we train the U-Net model shown in Section 2.5 using the training data. Then, we identify the wood parts in 341 CT images including the training data.

Fig.19 (a) shows the 3D shape of wood parts extracted by U-Net and Fig.19 (b) shows wood parts of the ground truth. The discrimination rate for the ground truth is 99.33%. On the other hand, for the experimental results in section 3.2, the discrimination rate for the ground truth is 96.07%. This result shows that the machine learning method is more accurate to extract wood parts. However, as shown in Fig.19 (a), the hair at the top of the head is still present. The bundled hair is dense and the intensity in the CT image is similar to that of wood, which makes it difficult to identify. In addition, metal nails are used on the top of the head to hold the hair in place. It changes the pixel intensity around the nails, so we think that this influence makes it difficult to identify wood parts correctly.

In Fig.19 (a), the handle under the neck is not extracted. This part can be extracted by the graph cut method. The surface of the wood is often coated with paint, but the handle is not. Therefore, in machine learning, the part adjacent to the paint is identified as wood. But areas that are not adjacent to the paint are not identified as wood. From these points, it is also necessary to consider what kind of CT images are appropriate as training data.
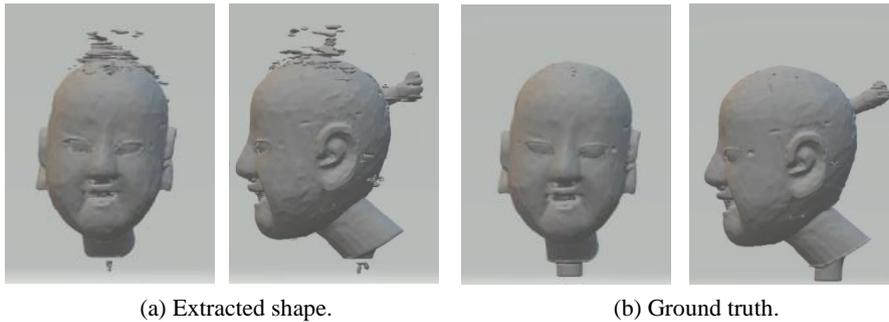

(a) Extracted shape.　　　　　　　　　(b) Ground truth.

Fig. 19. Reconstructed 3D shape by U-Net.

4 Conclusion

In order to reconstruct 3D shape of puppet head, we discuss a segmentation method for some materials of the puppet head in CT images. We propose two extraction methods to estimate the puppet head shape: one is a rough extraction based on histogram, and the other the graph cut method to extract precisely. We also propose the inner parts extraction method based on the graph cut method. Moreover, we discuss the wood part distinguish method by U-Net which is a framework of the machine learning.

From the results of experimental results using real puppet head and its CT images, 3D shape of the dry wood and paint region can be extracted but some parts of the hair wig are still remained in cases of both the graph cut and machine learning methods.

As the future works, we improve the method using machine learning. But we also use the graph cut method to assist in the manual extraction for training data. We study how to extract materials other than wood, and how to extract various puppets with high accuracy.

Acknowledgment

References

1. Katsushi Ikeuchi and Takeshi Oishi et al., 3D Digital Archiving, University of Tokyo Press (2010) (in Japanese).
2. Hiroyuki Ukida, "3D Shape Measurement System by Integration of Partial Shapes and Automatic Defects Detection," Journal of JSNDI, vol.67, no.7, pp.324-328 (2018) (in Japanese).
3. Hiroyuki Ukida, Tomoyo Sasao, Kenji Terada and Atsuya Yoshida, "Calibration Method of 3D Shape Measurement System Using 3D Scanner, Turn-Table and Arm-Robot," Proc. of the SICE Annual Conference 2019, pp.136-141 (2019).
4. Hiroshi Ishikawa, "Graph Cut," IPSJ SIG Technical Report, 2007-CVIM-158, no.26, pp. 193-204 (2007) (in Japanese).
5. Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.26, no.9, pp.1124–1137 (2004).
6. Hiroshi Fujita et al., Introduction to Medical Image Processing in Python, Ohmsha, Tokyo (2020) (in Japanese).