SICE

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

🔓 OPEN ACCESS | Check for updates |

3D shape reconstruction of Japanese traditional puppet head from CT images by graph cut and machine learning methods

Hinata Ikeda[a], Hiroyuki Ukida[a], Kouki Yamazoe[a], Masahide Tominaga[b], Tomoyo Sasao[c] and Kenji Terada[a]

[a]Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Tokushima, Japan; [b]Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; [c]Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

ABSTRACT

In this study, we discuss the digital archiving of Japanese traditional puppets. We propose two methods for extracting the puppet head shape from computed tomography (CT) images. The first is the graph cut method, and the second is a machine learning method based on U-Net. According to the experimental results of the extraction of puppet heads from CT images, the U-Net-based method can extract puppet heads more accurately than the graph cut method. Moreover, the U-Net-based method can extract puppet heads with multiple materials. However, the extraction of metal parts is inaccurate because of metal artefacts in the X-ray CT images and insufficient learning data.

1. Introduction

In recent years, the number of people engaged in traditional Japanese crafts has been declining. According to a report [1], approximately 115,000 people were engaged in Japanese crafts in 1998, however by 2017, the figure had dropped to approximately 58,000. Therefore, an urgent issue is how to preserve the techniques of traditional crafts and pass them on to the future. As a solution to this problem, digital archiving has been attracting attention.

"Digital archiving" [2] is a technique for preserving data of object, such as three-dimensional (3D) shapes, colour, and gloss, semi-permanently. This technique has been studied and applied to national treasures, important cultural properties, old documents, etc. In the paper [2], the digital archiving of large monuments such as Angkor Wat is introduced. In papers [3,4], the digital archiving of rare old books is discussed.

In this study, we discuss a digital archiving method of puppets used in a traditional Japanese puppet theatre known as "Awa Ningyo Joruri," which has been played in Tokushima Prefecture of Japan. Figure 1 shows a scene of the puppet theatre [5]. Two or three people called "Ningyo Tsukai" operate one puppet, and this puppet show is facilitated by a narrator called "Tayu" and the music of "Shamisen" (a three-stringed Japanese musical instrument).

To promote Ningyo Joruri, we have discussed some methods for measuring the 3D shape data of puppets with a 3D scanner and manufacturing them with a 3D printer [6]. However, a 3D scanner can only measure the external shape of puppet. On the other hand, many Ningyo Joruri puppets have mechanisms in their heads that enable them to move their eyes and open/close their mouths. Therefore, to archive digital puppet information more precisely, we must reconstruct the inner parts of the puppet head.

In this study, we discuss a method for reconstructing the internal and external 3D shapes of the puppet head. X-ray computed tomography (CT) systems are primarily used for medical purposes to take X-ray images of the inside of the human body and construct tomographic images. However, the main material of the puppet head is dry wood. Thus, it is not displayed clearly in CT images because of low moisture. Human organ detection methods cannot be used directly for puppets.

To solve this problem, we examine two types of methods for reconstructing the puppet head shape. One is the graph cut method, which performs region segmentation based on the feature values of pixels in CT images [7–9]. The other is a method based on U-Net, a machine learning (ML) framework [10]. In the paper [10], only the wood part was extracted, however, in this study, the results of wood, paint, and metal part extraction are presented. We compare the results of the graph cut method and the ML method and discuss their advantages and disadvantages.

The structure of this paper is as follows: in Section 2, we introduce related studies on about the measurement and restoration methods of puppets, and the related studies on the head shape measurement from CT images are also shown. In Section 3, we describe CT

CONTACT Hiroyuki Ukida ✉ ukida@tokushima-u.ac.jp 🏢 Graduate School of Technology, Industrial and Social Sciences, Tokushima University, 2-1 Minamijosanjima-cho, Tokushima 770-8506, Japan

Figure 1. Scene of Japanese traditional puppet theatre (Awa Ningyo Joruri).

images of puppet heads. In Section 4, we show the shape reconstruction method of puppet heads using graph cuts. In Section 5, we show the shape reconstruction method based on ML (U-Net). In Section 6, we present and discuss the results of shape reconstruction using two proposed methods. Finally, Section 7 provides a summary of this study.

2. Related work

2.1. 3D shape measurement and reconstruction of puppets

Figure 2 shows examples of puppet heads. Puppet heads are carved out of wood by hand. Hence, the production of heads takes a lot of time, and they tend to be expensive. To popularize Ningyo Joruri (puppet theatre), puppets must be quick to produce and inexpensive. Recently, with the widespread use of 3D-computer-aided design (CAD) software and 3D printers, the design of puppet heads using 3D-CAD and their creation using 3D printers have been attempted. A website [11] sells puppet heads created in such a way.

We are also promoting a project to create puppets using a 3D printer by measuring their external shapes

with a 3D scanner to popularize Ningyo Joruri in general [12]. Figure 3(a) shows a puppet head fixed on a turn table, Figure 3(b) shows a scene of the measurement of the puppet head shape by a 3D scanner, and Figure 3(c) shows measured 3D head shape data interpolated by 3D-CAD for areas that could not be measured. Figure 3(d) shows the final created puppet head. The shape of this head is output by a 3D printer, and the eyes and mouth are painted with colours. The hair is covered with a wig. Furthermore, Figure 3(e) shows a completed puppet, with the head attached to a body clothed in a kimono dress. Both hands are also designed with 3D-CAD and output with a 3D printer. Based on this technology, we offer workshops on puppet making. We have also analysed the facial features of puppets and proposed a method for creating puppet heads that reflect human facial features [13].

Generally, puppet heads are often made with movable mechanisms for the eyes, mouth, neck, and other parts. Figure 4 shows the inside of a puppet head during production. The puppet head will be able to move its eyes and neck. Puppet heads created by traditional methods cannot be disassembled after completion, and it is impossible to measure the internal shape of the head with a 3D scanner. However, to realize the digital archiving of Ningyo Joruri, it is necessary to accurately measure the internal shape of the head. In this study, a shape reconstruction method using CT images of the head is discussed to measure the inside of the head without disassembling it.

Furthermore, in Ref. [14], an example of measuring and producing the puppet head shape from CT images is presented. This study aims to observe the inside of the head with CT images and to create a puppet head with a 3D printer using the external shape of the head. On the other hand, our research aims to identify the various materials (wood, paint, etc.) that make up the puppet head in CT images and to restore their shapes to realize accurate digital archiving.
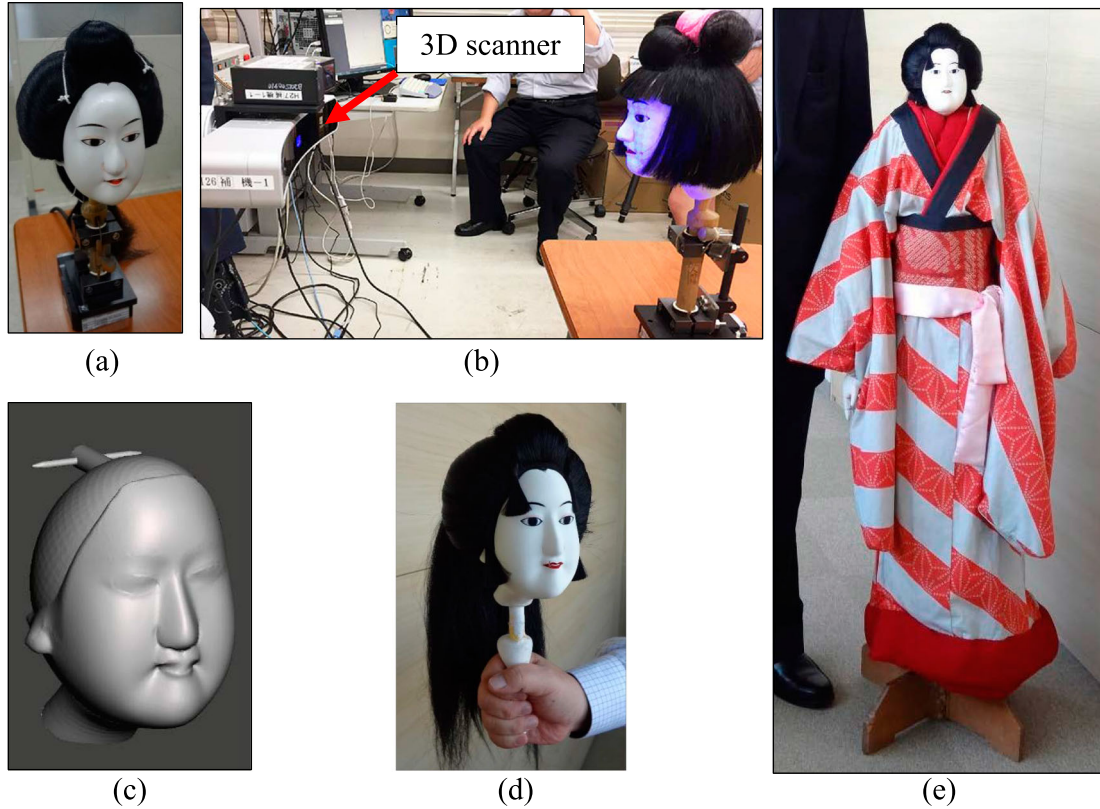

Figure 2. Examples of puppet heads.

Figure 3. Reproduction of puppet. (a) Puppet head. (b) Shape measurement by 3D scanner. (c) Measured 3D data and interpolated by 3D CAD. (d) Created puppet head. (e) Created puppet.
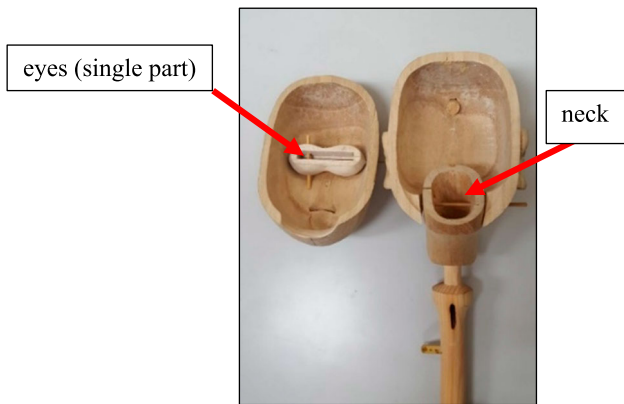

Figure 4. Inside of puppet head in production.

2.2. 3D shape reconstruction from CT images

X-ray CT is a device that reconstructs a tomographic image of the interior of an object by irradiating X-rays from various directions and measuring the absorptance (transmittance) of X-rays within the object. Different materials absorb X-rays at different rates, resulting in different pixel values for each material in the tomographic image. Therefore, we can identify the internal structure of an object by X-ray CT. (Hereafter, tomographic images are referred to as CT images, and pixel values in CT images are referred to as CT values.) In the human body, bones, muscles, and organs have relatively different CT values. They are easy to distinguish.

However, organs such as the stomach, and small and large intestines have similar CT values. They are difficult to distinguish.

Many methods have been explored for identifying multiple organs in CT images of the human body. Recently, an ML-based method has been proposed. A paper [15] proposes a method based on Ada Boosting. In this method, first, multiple discriminators are prepared that use features in CT images for identification. Then, from the results of each discriminator, the final identification result is obtained by a voting process. A paper [16] discusses a method for organ identification using a convolutional neural network (CNN), an ML framework.

On the other hand, this study aims to reconstruct the shape of a puppet head from CT images. The puppet head is made of wood (details will be described in Section 3). In addition, paint, hair, and metals are used. In particular, because wood is dry, it can be difficult to distinguish it from other materials.

Therefore, this study examines two types of material identification methods with reference to the human body. One is to identify materials using the graph cut method [17] based on CT values in CT images. This method is feature-based and different from ML. The other method uses U-Net [18], an ML framework. This method combines a feature extraction method by convolution processing and a feature restoration method

by inverse convolution processing. In this study, we show the results of using each of these methods to identify and extract the wood, paint, and metal parts that primarily comprise the head, and reconstruct the 3D shape. Furthermore, based on these results, we discuss a suitable method for obtaining the shape of the puppet head from CT images.

3. CT images of puppet heads

Figure 5 depicts the puppet head used in this study. The length of this head is approximately 170 mm long from the apex of the head to the neck. The eyes of puppet rotate up and down, and the head moves back and forth. Figure 6 shows a scene of a medical CT scan of the head. In this study, we use a CT system at the Faculty of Medicine, Tokushima University.

Figure 7 shows one of the CT images used in this study. 341 CT images are taken in the axial direction from the apex to the neck. The interval between images is 0.5 mm. The size of a CT image is 512×512 pixels, and the size of a pixel is 0.3×0.3 mm. Each pixel has 16 bits of data. Pixel values are defined as 0 for water and

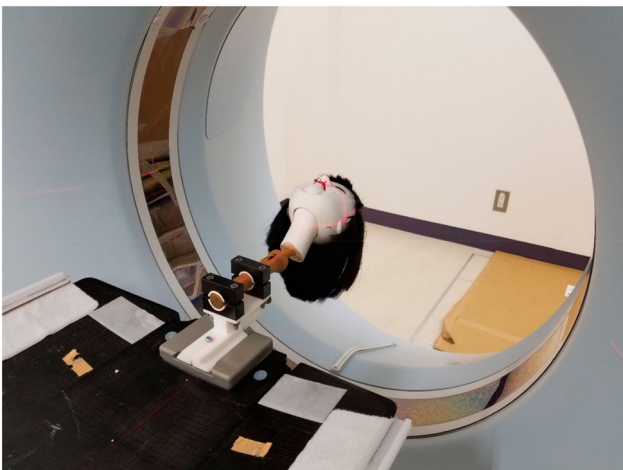**Figure 5.** Puppet head used in this study.

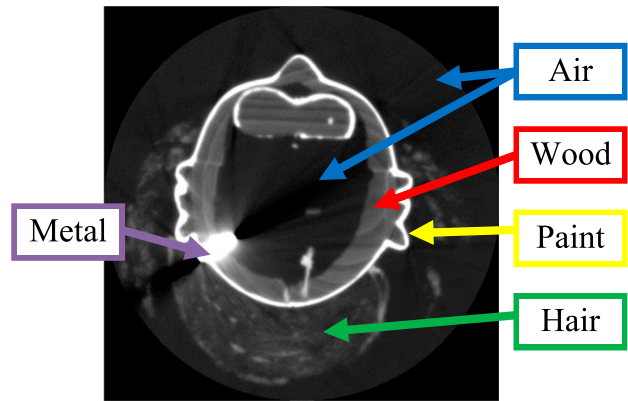**Figure 6.** Scene of a medical CT scan of puppet head.

Figure 7. CT image of a puppet head.

−1000 for air. The distribution of pixel values differs depending on the material.

As depicted in Figure 7, the head of the puppet is composed mainly of "wood," "paint," "hair," and "metal" materials. The regions of these four materials and an "air" region are contained in CT images. The characteristics of the pixel values for each material are listed below.

- Both "wood" and "hair" have relatively low values because of their low moisture content and overlapping pixel value distributions.
- The "paint" is made from shell powder and contains lime, which makes its pixel value relatively high.
- The "metal" is the nail used to attach the hair, which has a very high value and affects the surrounding pixel values.

In this study, we examine a method for extracting the wood, paint, and metal regions, which mainly constitute the shape of the head, among the four types of materials.

4. Shape reconstruction method of puppet head by graph cut method

In this section, we show material extraction methods for the 3D reconstruction of the puppet head. This method consists of two methods as follows:

(1) Rough region segmentation using thresholds obtained from histograms of intensities in a manually directed area in a CT image.
(2) Precise region extraction using a graph cut method from the results of 1.

We explain these methods in the next subsections.

Note that, in this method, we consider that the metal region is the same as the paint region because the CT values of the metal and paint are significantly higher than those of the other materials and the area of the metal region is smaller than that of the other regions.

4.1. Extracting materials based on histograms

This subsection shows a material extraction method based on histogram.

Step 1-1. For four regions (air, hair, wood, and paint), we obtain histograms from areas manually directed in some CT images. Figure 8 shows an example of manually directed areas in a CT image.

Step 1-1. 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 represents a type of region, x represents the intensity, and μ_k and σ_k denote the mean and standard deviation of region k, respectively.

Step 1-1. Estimate thresholds as cross points of normal distributions that are neighbouring two regions.

Using the estimated thresholds, we can roughly group of the four regions in CT images.

4.2. Precise extraction by graph cut method

The graph cut method efficiently estimates the combination of a label of pixel (object or background) in an image in a cost minimization criterion under the condition that some parts of the image are assigned labels beforehand [17].

In this study, we reconstruct the 3D shape of the wood and paint regions only. Because the shape of the hair deforms according to the attitude of the puppet head, it is difficult to model the 3D hair shape. Hence, the wood and paint regions are considered object labels, and the hair and air regions are the background.
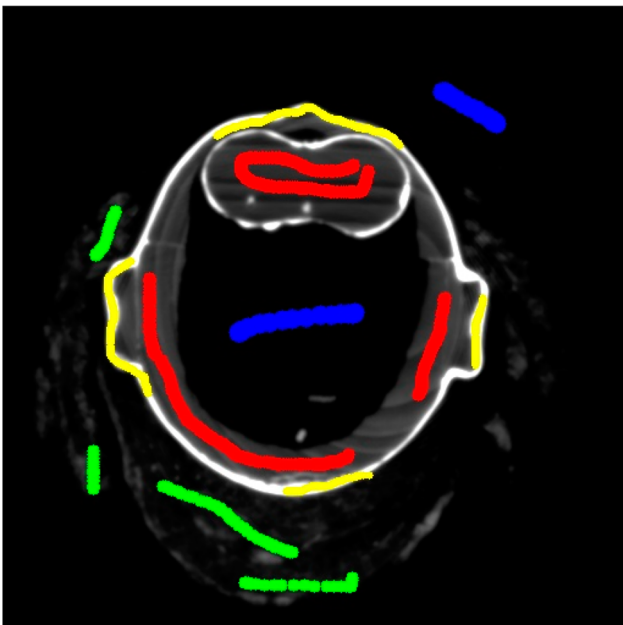
The cost function $E(L)$ used in the graph cut method is shown in Equation (2) as the linear combination of the region term $R(L)$ and the boundary term $B(L)$:

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

where L represents a label assigned to pixels ($L \in \{obj, bkg\}$) and λ represents 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 expressed as follows:

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

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

where f_u denotes the likelihood of a region and $g_{u,v}$ denotes the likelihood of a boundary between neighbouring pixels. U represents a set of pixels and u represents a pixel. N represents a set of two neighbouring pixels and u, v denotes a tuple of pixels. f_u and $g_{u,v}$ are expressed as follows:

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

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

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

Figure 9 shows the process of the graph cut method. In this figure, node u corresponds to a pixel, and it is connected to a neighbouring pixel (node v). We call this linkage "n-link," and we assume that this link has a cost
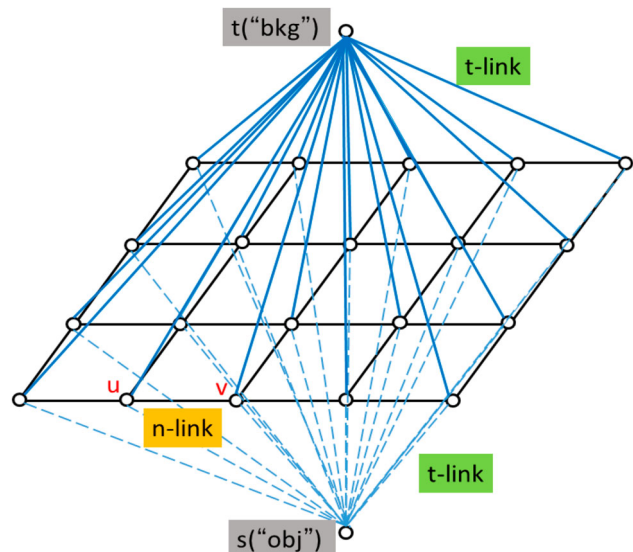

Figure 8. Manually directed areas in a CT image.

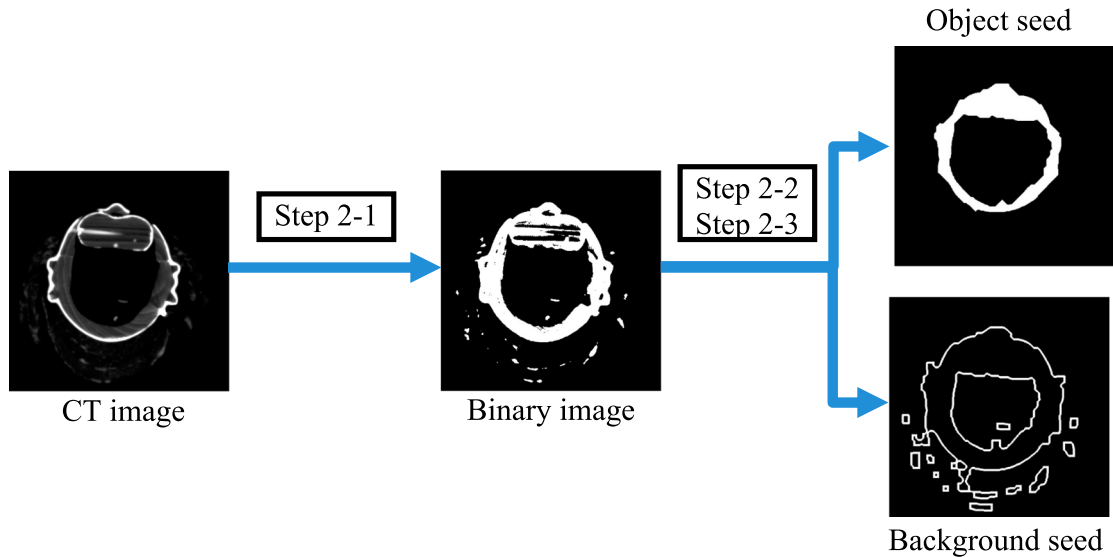**Figure 9.** Example of graph structure.

Figure 10. Seed region estimation.

estimated by Equation (6). Node u is also connected to the node s-labelled object "obj" and the node t-labelled 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 labelling "obj" or "bkg" to u. Therefore, when the sum of the 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 [19] to the cost minimization method.

During region extraction using the graph cut method, we must assign object or background labels to some parts of regions as "seeds." Generally, this "seed" is typically assigned by a user manually in the graph cut method. However, in this study, we attempt to assign a "seed" without user input through the following steps (Figure 10):

Step 2-1. By applying the method in Section 4.1 to a CT image, the wood and paint regions are extracted from the image.

Step 2-1. By using morphological methods (dilation and erosion), we fill holes and eliminate small regions (approximately 10×10 pixels) in the image in Step 2-1.

Step 2-1. We estimate a distance transformation image from the result of Step 2-2 and extract two regions as follows:

- One region consists of pixels with distances ranging from the maximum distance ($D_{\max}$) to $D_{\max} - \Delta_1$.
- The other region consists of some pixels with distances ranging from the minimum distance ($D_{\min}$) to $D_{\min} + \Delta_2$.

Here, Δ_1 and Δ_2 denote distance thresholds, and they are assigned some values in advance. As a result, we can obtain two seeds for the graph cut method. Using these seed regions, we extract the object (wood and paint) and background (hair and air) regions in a CT image.

Moreover, a puppet head has many CT images, which are aligned perpendicular to the image plane, thus, to assign seeds to all CT images, we use the following steps (Figure 11):

Step 3-1. In a neighbouring (upper or lower) CT image of a labelled CT image, first, region extraction based on histograms is applied. After that, we obtain an overlapped region as a "seed" region, comprising the wood and paint regions in this image and the object region in a labelled CT image. Using this seed region, we apply the graph cut method.

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

5. Shape reconstruction method of puppet head by ML

5.1. U-Net

In this section, we demonstrate how to extract material from CT images by ML. In Section 4, we consider that the paint and metal regions are the same regions. However, in this section, we consider that these are different regions for a precise puppet head restoration. On the other hand, since the deformation of hair occurs depending on the orientation of the puppet head in Figure 5, and the boundary between hair and air is not clear, hair is difficult to annotate as correct training data. Therefore, the hair and air regions are considered the background region as demonstrated in Section 4. Thus, we propose a method for distinguishing the wood, paint, metal, and background regions in CT images.

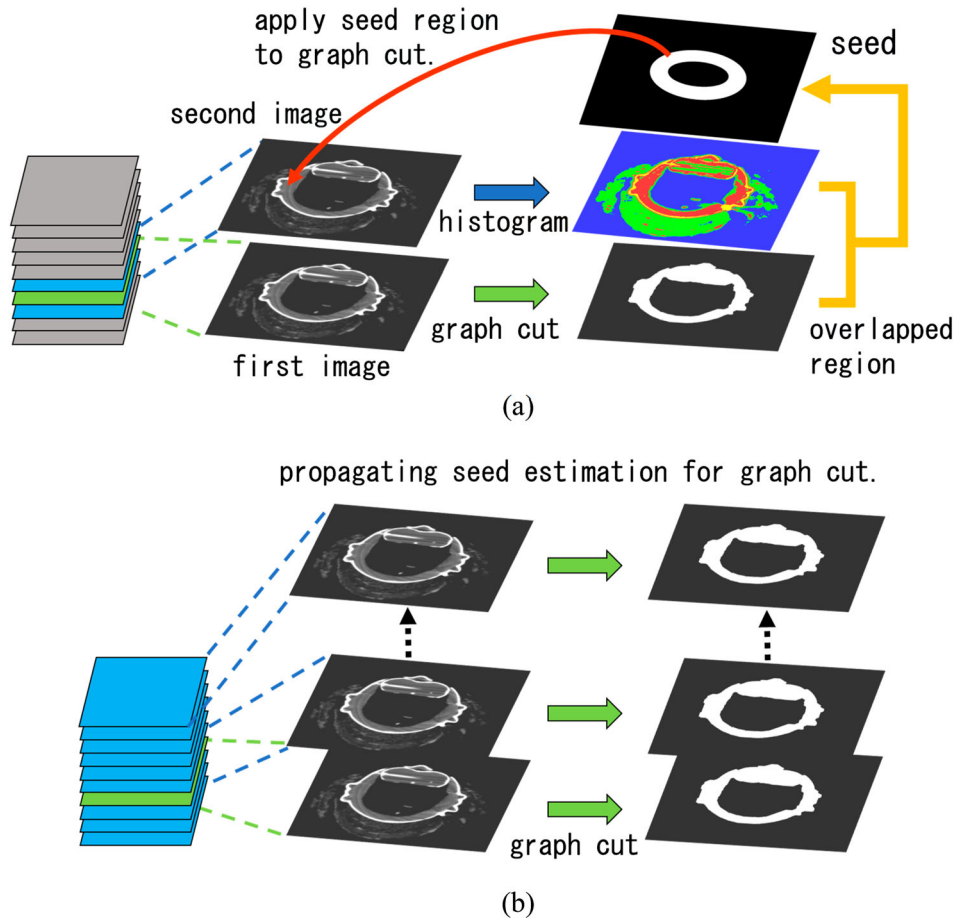U-Net is a network based on CNN. Figure 12 shows the configuration of the U-Net model used in this study.

Figure 11. Object and background region estimation for all CT images. (a) Step 3-1. (b) Step 3-2.

U-Net consists of encoder and decoder parts. The U-Net model in this study has eight layers in both the encoder and decoder parts (the third to sixth layers are omitted in Figure 12). The layers are joined in a 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 using the reverse convolution layer.

However, the feature extraction in the encoder part discards the positional information in the image; thus, 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. Using this method, we can restore the image including the positional information. Details of each layer are shown below.

- Convolution layer

In the convolution layer, a convolution operation using a kernel (filter) is performed (Figure 13). The resulting matrix is smaller than the original matrix. The

kernel corresponds to the weight w in the network and is responsible for extracting the information to be conveyed to the next layer. The weight w is optimized to minimize the error through the learning process.

- Pooling layer

The pooling layer generates a new matrix by extracting only the highest-valued elements in a square region (max pooling) or by calculating the average value in a square region (average pooling). In the example in Figure 14, the max pooling method generates a 2×2 matrix. This operation provides robustness against object misalignment in the input image. In other words, if the objects are the same, they will be recognized as the same object even if their positions are slightly different.

- Reverse convolution layer

In the reverse convolution layer, the input image is enlarged before the convolution operation is performed, which makes the size of the output feature map larger than that of the input feature (Figure 15).
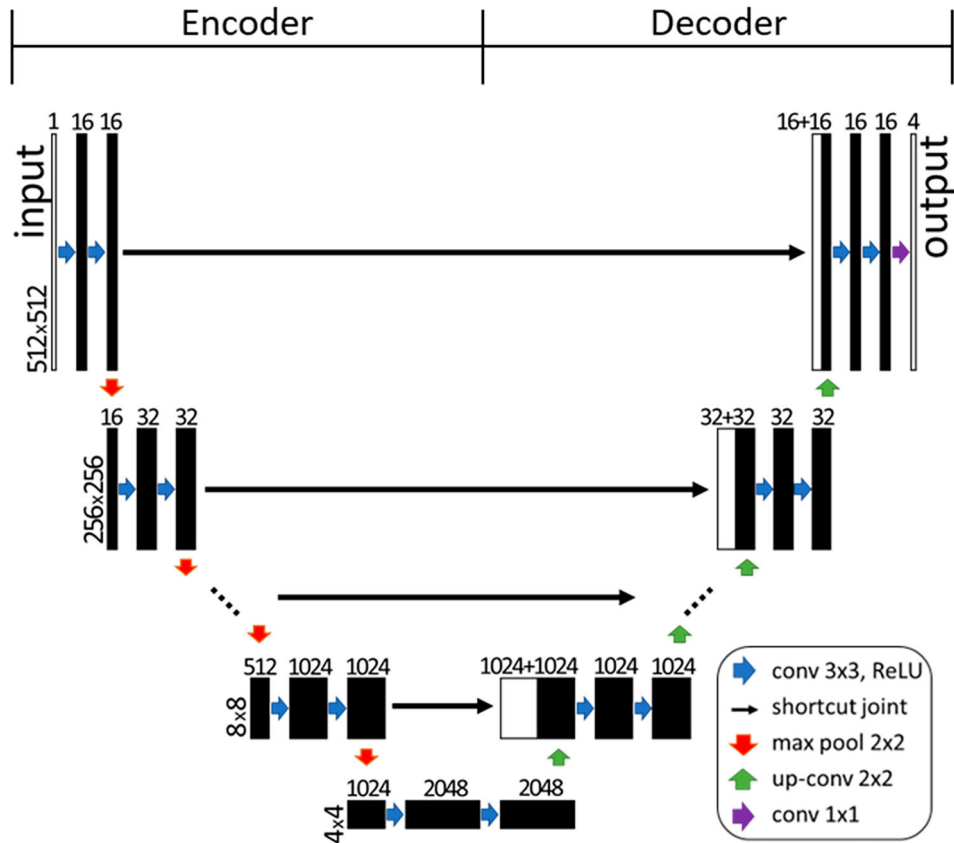
- Shortcut joint
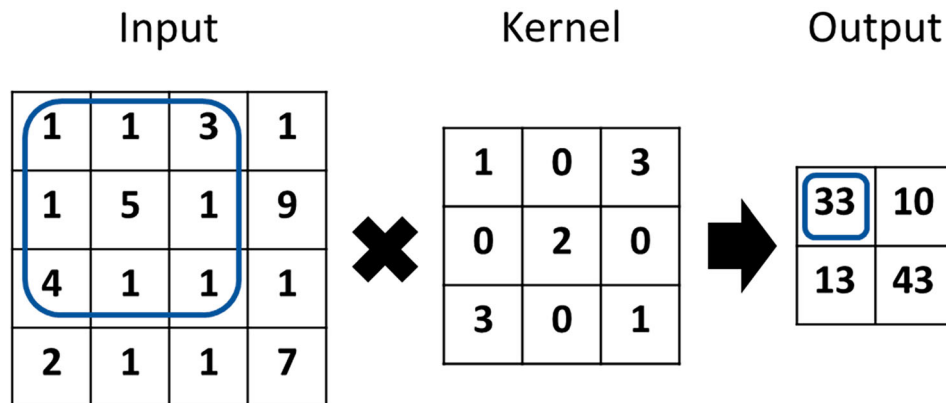
Figure 12. Configuration of U-Net.

Figure 13. Convolution layer.

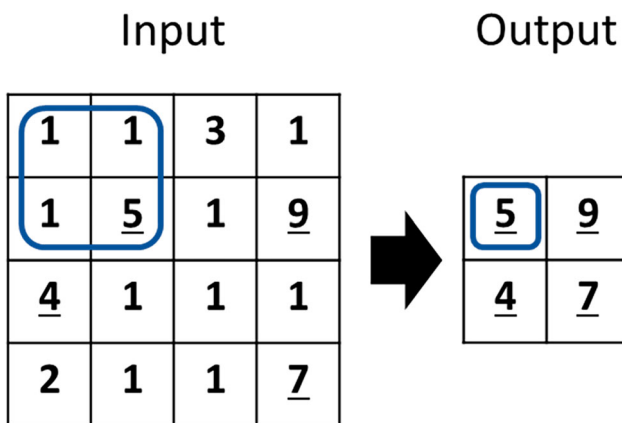**Figure 14.** Pooling layer.

Shortcut joints join two 3D arrays in the depth direction to create a new 3D array. In the example in Figure 16, a 4 × 4 × 3 array is concatenated with a 4 × 4 × 3 array to produce a 4 × 4 × 6 array.

The output data of U-Net (Figure 12) are four-channel data like images. These channels correspond to labels of materials (wood, paint, metal, and background), and the pixel value at the coordinate (x, y) is the probability of being a material corresponding to each channel. Hence, we obtain a label (material) that has maximum probability at each co-ordinate (x, y). This is the final result of the material extraction from a CT image.
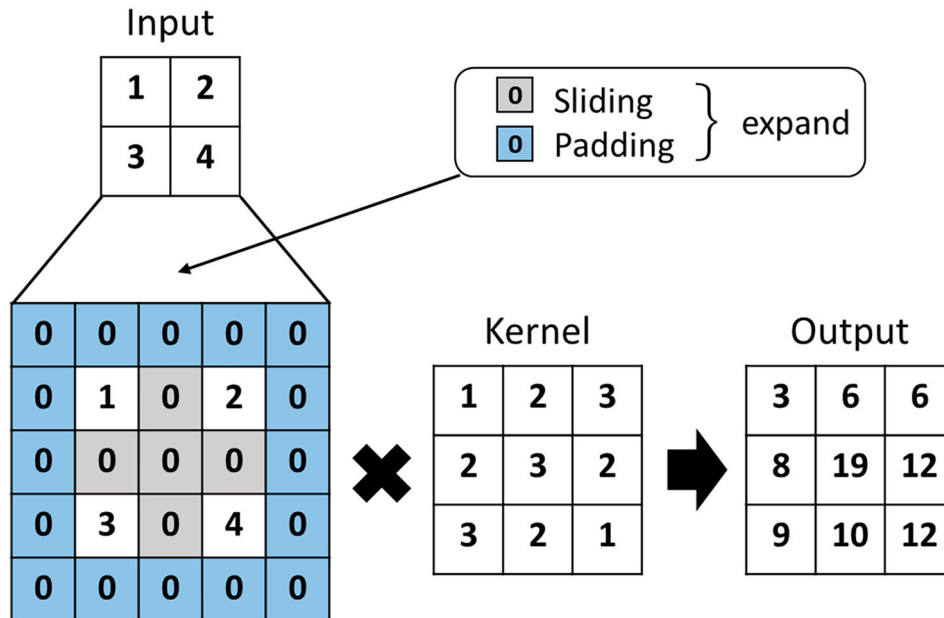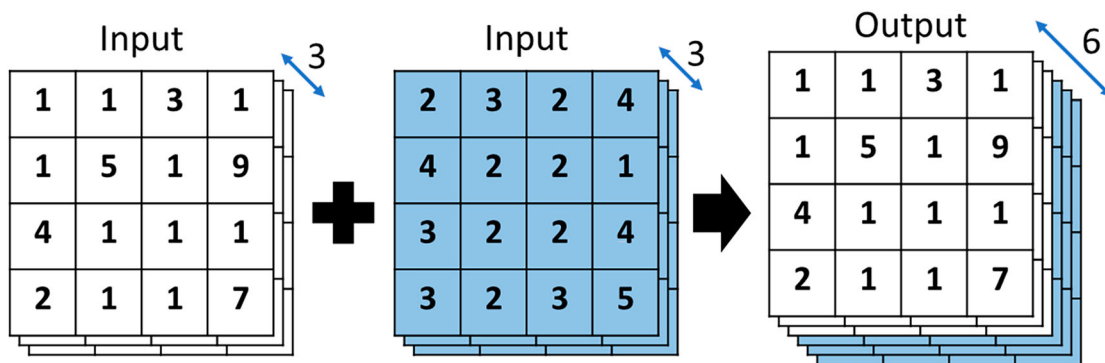
Figure 15. Reverse convolution layer.

Figure 16. Shortcut joint.

5.2. Learning data

As described in Section 3, 341 CT images of the puppet head are taken. In this study, we use n% CT images as training image data among N (all) CT images. Here, m CT images are selected at equal intervals (every p images) from the top of the puppet head (Figure 17). For example, when we use approximately 5% of 341 CT images, the training data are selected for every 20 CT images. In this case, we use 18 CT images as the training data. The number of CT images suitable for the training data is discussed in the experiments in Section 6.2. The image data for training also need to represent the correct material regions corresponding to the input CT image (we call it "correct label image"). Here, correct label images are obtained by manually checking and correcting regions after the CT image is divided into regions with threshold values based on the pixel values of each material. The specific procedure is as follows.

(1) The range of pixel values for each material in a CT image is distributed as depicted in Figure 18(a). In particular, there are some overlaps in the pixel
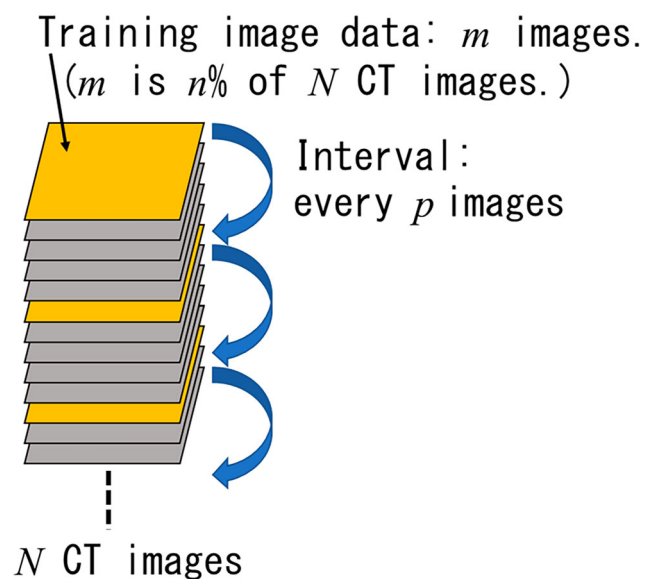

Figure 17. Training data selection from CT images.

value range between hair and wood. From this distribution, a threshold value representing the range of each material is set as depicted in Figure 18(b). Note that Figure 18 is based on the experimental
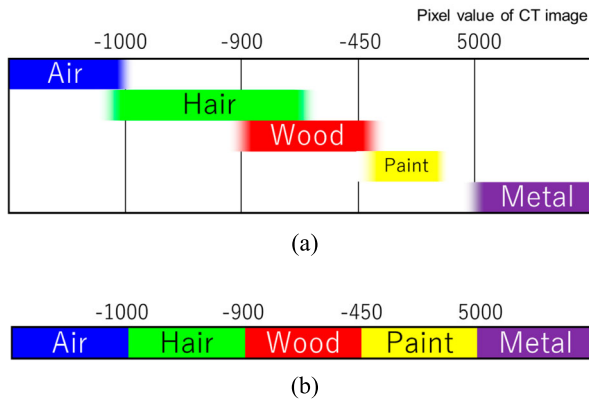
Figure 18. Material extraction by thresholds in CT image. (a) Range of pixel values for each material. (b) Threshold values.

results in Section 6.1 which is an experiment of the method described in Section 4.1.

(2) Using the thresholds in Figure 18(b), we obtain an image divided into regions for each material (Figure 19(a)) from CT images. The subsequent process manually corrects pixels whose pixel values are incorrectly extracted because of overlap.

(3) To distinguish the material of each pixel during the manual correction, which is easier for humans, a colour image (Figure 19(b)) is generated by super-imposing the original CT image (grey-scale image) and the image of Figure 19(a).

(4) For the image shown in Figure 19(c), manually check and correct material areas. Thus, four binary

images corresponding to four material regions are obtained.

(5) Finally, the isolated noises are removed to obtain the correct label images of the four material regions. Figure 19(d) shows obtained correct label image and Figure 19(e) shows the wood label image as examples.

In addition, to evaluate the experimental results, correct label images were generated not only for the 18 training images but also for all CT images (341 images).

6. Experimental results

In this section, we show the experimental results of the puppet shape extraction from CT images. Sections 6.1 and 6.2 show the material extraction and 3D shape reconstruction results by each method. Section 6.3 shows the discrimination rate of the material extraction. Here, we compare the results of both methods. Section 6.4 shows the material extraction results using two puppet heads by ML.

6.1. Experimental results by graph cut method

As shown in Section 4.1, we obtain histograms from manually directed areas in some CT images about four regions. Figure 20 shows directed areas on 10 CT images. These CT images are mainly eyes and nose
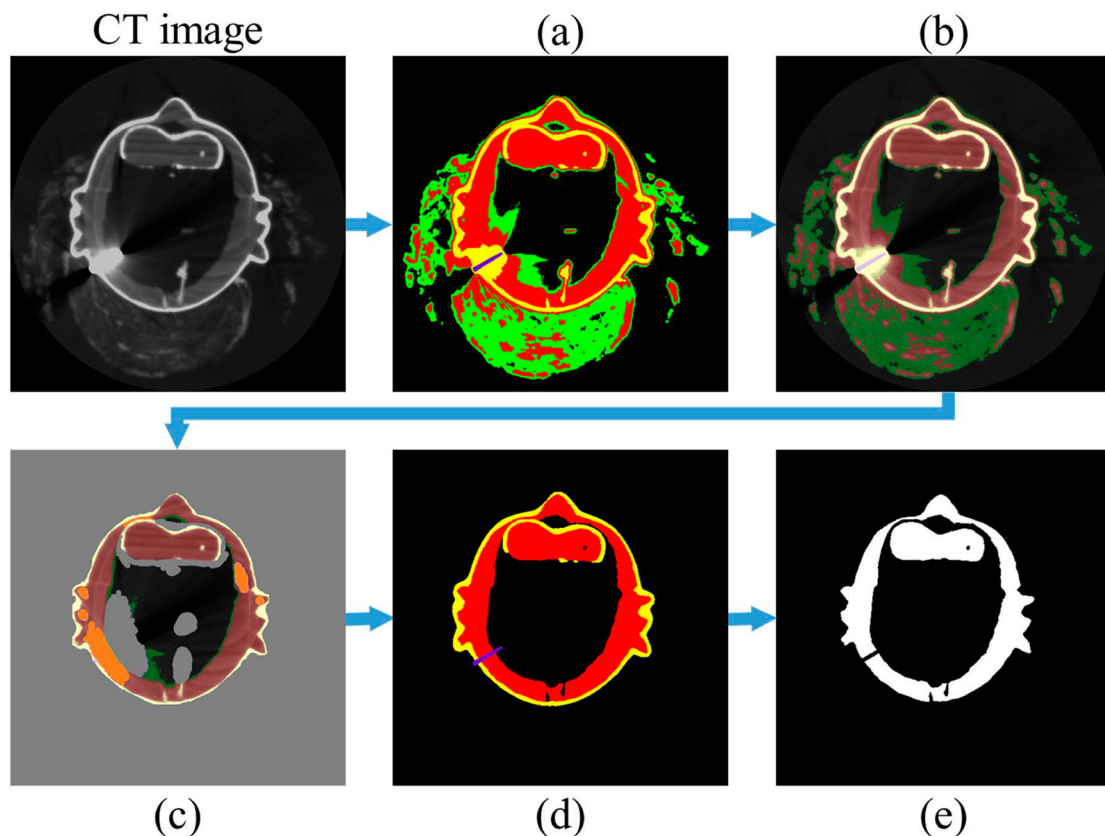

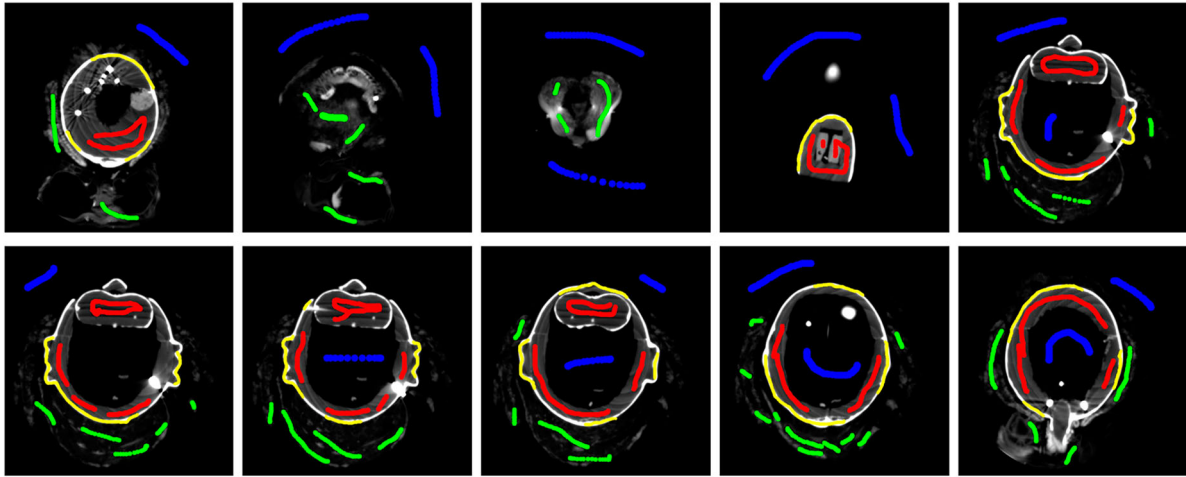**Figure 19.** Manual estimation of correct label image.

Figure 20. Manually directed areas (blue:air, green:hair, red:wood, yellow:paint).
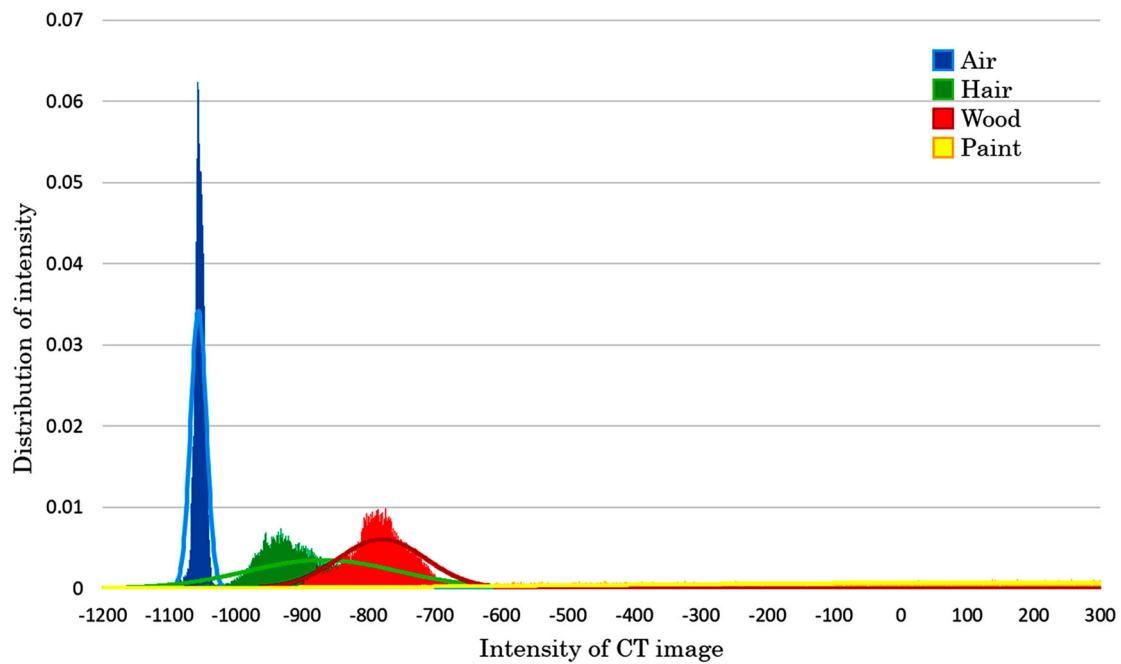

Figure 21. Probability distribution and normal distribution curve of CT value in each region.

Table 1. Estimated means, variances, and thresholds.

	Air	Hair	Wood	Paint
Mean	−1056.3	−869.2	−780.0	150.7
Variance	135.9	13,042.2	4348.2	321,814.0
Threshold		−1026	−850	−616

parts, and they also include the upper and lower parts of puppet head. (Figure 20 includes Figure 8.)

Figure 21 shows histograms and normal distributions of the four regions (air, hair wig, wood, and paint) denoted in Section 4.1. Table 1 shows the means and variances of normal distributions and thresholds for dividing these regions.

Figure 22 shows the results of the region segmentation process. Figure 22(a) shows the input CT image. Figure 22(b) shows the results of segmentation by histogram. Some parts of wood (red) are estimated as hair

(green), and some parts of hair are also estimated as wood.

Figure 22(c) shows the separated wood and paint regions from Figure 22(b). Figure 22(d) shows the results of the dilation and erosion method. Figure 22(e) shows the result of distance transformation. White pixels in Figure 22(f) represent regions with the label "obj," whereas, those in Figure 22(g) represent regions with the label "bkg."

Figure 22(h) shows the segmentation results of the graph cut method. The object part (wood and paint) is white, and the background part (air and hair) is black. In these images, most of the hair regions can be extracted as the background, whereas some parts of the hair regions are extracted as the object.

Next, we show a confusion matrix as the identification results. Since the graph cut method assigns two labels which are the object and background, this matrix
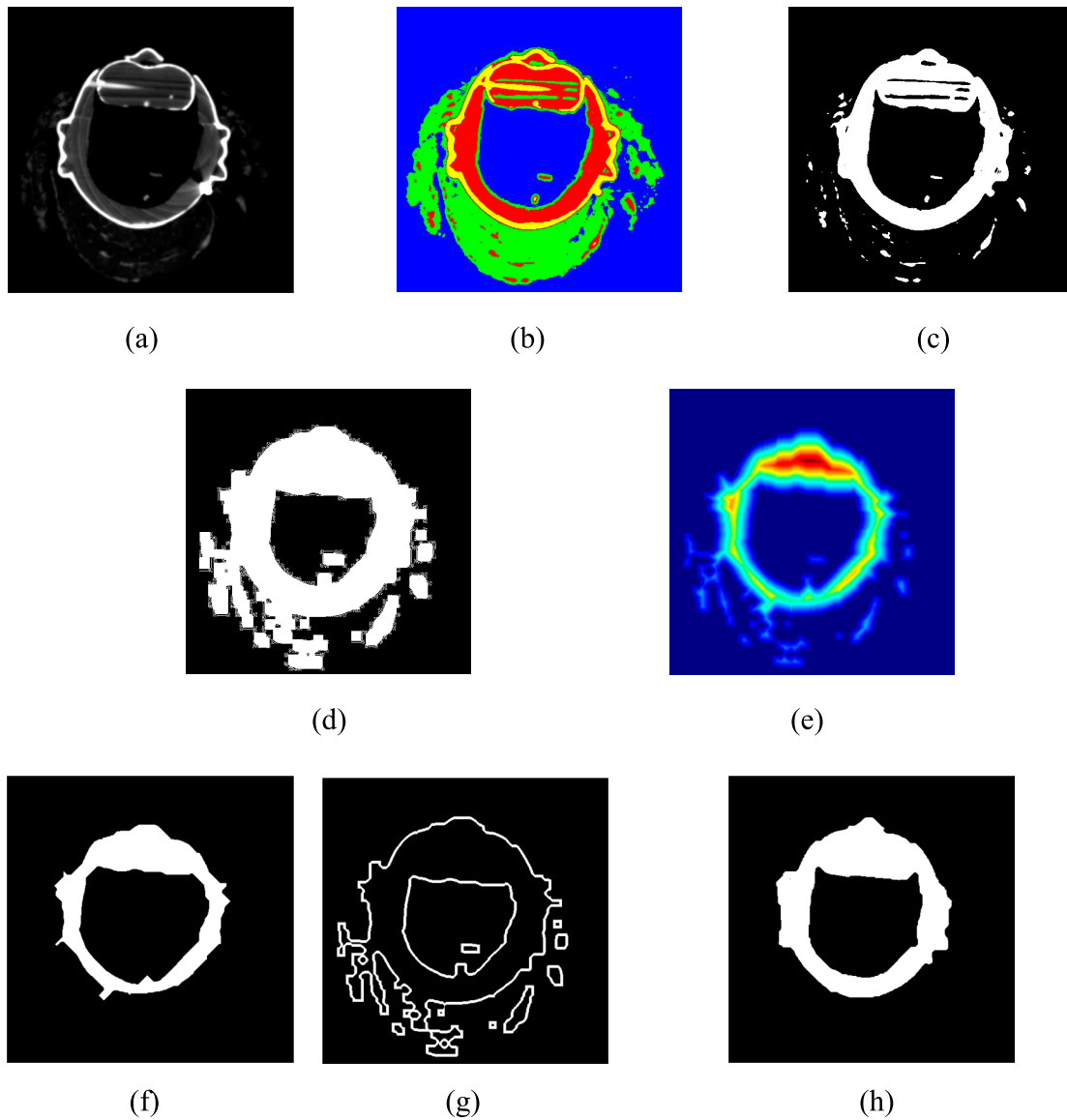
Figure 22. Experimental results of the proposed graph cut method. (a) Input CT image. (b) Segmentation by histogram. (c) Wood and paint regions. (d) Results of dilation and erosion method. (e) Distance transformation image. (f) Object seed. (g) Background seed. (h) Result of graph cut method.

has 2×2 elements as shown in Table 2. (The most right-side column is the total number of correct labelled pixels.) Here, we use the correct label images used in ML (U-Net) as the true value, and the metal in the correct label images is included in object (wood and paint). In Table 2, the numbers without brackets are the number of pixels identified in the CT images, and the bracketed percentages represent the ratio to the number of the correct label in each row. The numbers underlined in bold indicate the highest discrimination rate for each correct label. From these results, we can see that a lot of pixels are misidentified.

Figure 23 shows the reconstructed 3D shape of the puppet head. Figure 23(a) shows the extracted wood, paint, and hair from the histogram thresholds. Figure 23(b) shows the extracted objects (wood and paint) by the graph cut method. In the result in Figure 23(b), at the top of the puppet head, the hair region remains. This

Table 2. Confusion matrix by the graph cut method.

		Identified label		
		Object	Background	Total
Correct label	Object	**7,867,585**	417,207	8,284,792
	(%)	**(94.96)**	(5.04)	(100.0)
	Background	1,942,526	**79,163,786**	81,106,312
	(%)	(2.40)	**(97.60)**	(100.0)

is because hair is dense at this part to fix hair to wood parts, and the intensity of hair is almost the same as that of dry wood. As a result, these parts are extracted as the wood region.

6.2. Experimental results by ML

In this section, first, we show the identification results of the ML method about various numbers of training data. Here, we compare the results of three types of the

(a)

(b)
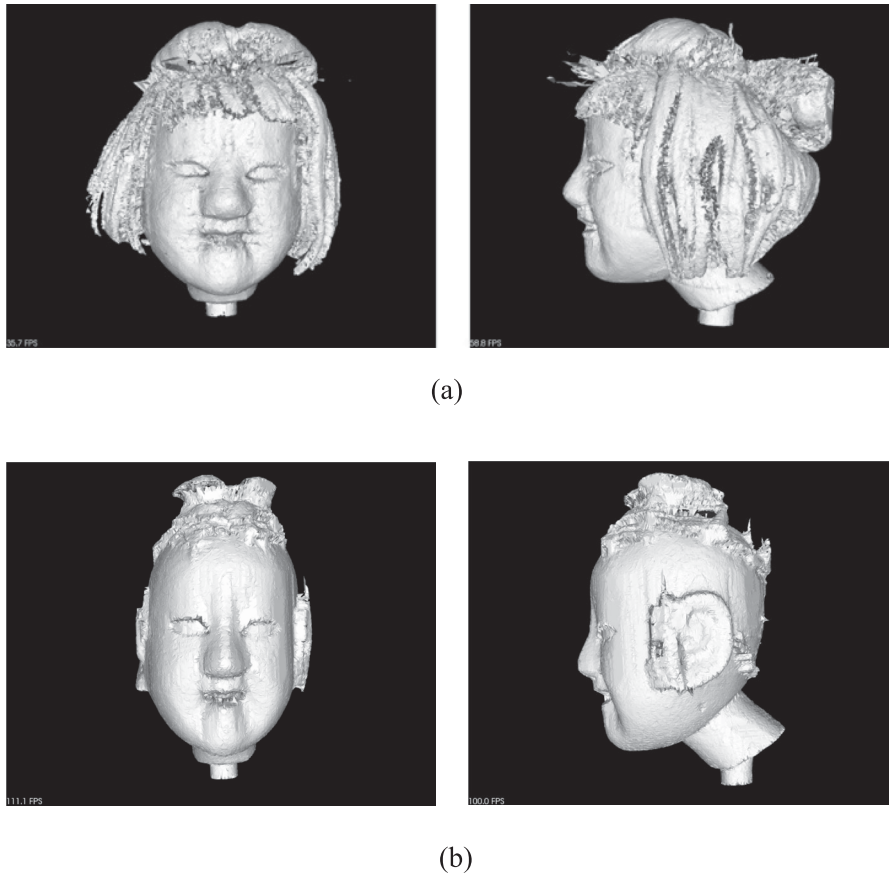
Figure 23. Reconstructed 3D shape of puppet head using the graph cut method. (a) Before applying graph cut. (b) After applying graph cut.

numbers of training data (m). (Refer Figure 17 about m, n, and p.)

- TD1: $m = 9 \left(n \fallingdotseq 2.5\%, p = 40\right)$
- TD2: $m = 18 \left(n \fallingdotseq 5\%, p = 20\right)$
- TD3: $m = 35 \left(n \fallingdotseq 10\%, p = 10\right)$

Here, "TD" means "Training Data." Figure 24 shows nine correct label images used in TD1.

As the identification results, we show confusion matrices. Since there are four materials (wood, paint, metal, and background (hair and air)), the confusion matrix in this experiment has 4×4 elements as shown in Table 3. N_{ij} is the number of pixels which is identified as the label j and this correct label is i in all CT images. S_i is all number of correct label i pixels. It can be obtained by Equation (7):

Figure 24. Correct label images (TD1: nine images). (There are no labels in No.1 and No.41.)

Table 3. Confusion matrix for identification result of four materials.

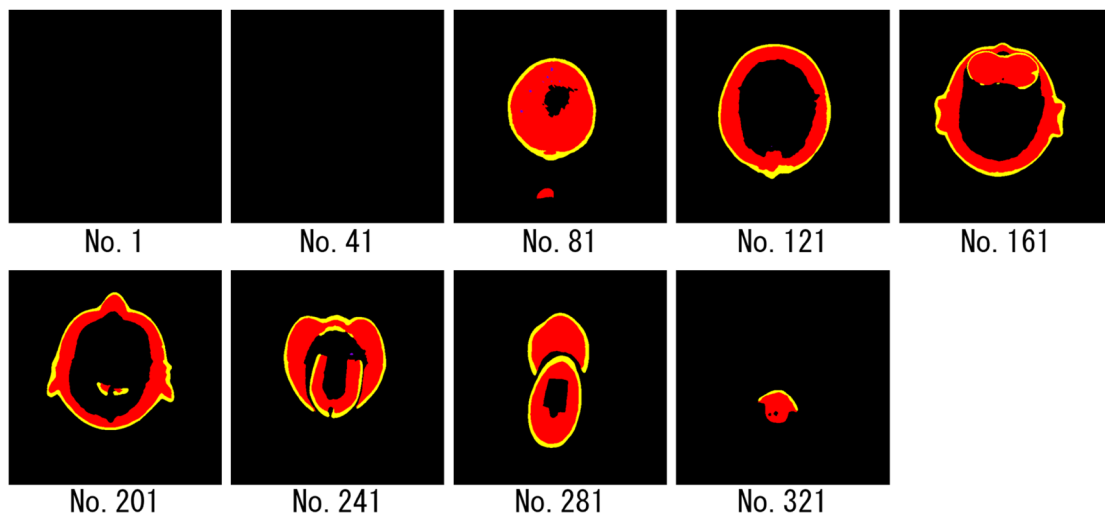|		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	N_{11}	N_{12}	N_{13}	N_{14}	S_1
	(%)	(r_{11})	(r_{12})	(r_{13})	(r_{14})	(100.0)
	Pain	N_{21}	N_{22}	N_{23}	N_{24}	S_2
	(%)	(r_{21})	(r_{22})	(r_{23})	(r_{24})	(100.0)
	Metal	N_{31}	N_{32}	N_{33}	N_{34}	S_3
	(%)	(r_{31})	(r_{32})	(r_{33})	(r_{34})	(100.0)
	Background	N_{41}	N_{42}	N_{43}	N_{44}	S_4
	(%)	(r_{41})	(r_{42})	(r_{43})	(r_{44})	(100.0)
IoU		IoU_1	IoU_2	IoU_3	IoU_4	
PA				PA		

$$S_i = \sum_{j=1}^{L} N_{ij}, \qquad (7)$$

where L is the number of labels ($L = 4$). And,

$$r_{ij} = (N_{ij}/S_i) \times 100. \qquad (8)$$

We also evaluate the identification rate for each material by the intersection over union (IoU). The IoU_j for the label (material) j is obtained by Equation (9):

$$IoU_j = \frac{N_{jj}}{\sum_{k=1}^{L} N_{kj} + \sum_{k=1}^{L} N_{jk} - N_{jj}}. \qquad (9)$$

Moreover, the pixel accuracy (PA) is introduced as an evaluation of the overall identification result. PA is obtained by Equation (10):

$$PA = \frac{\sum_{i=1}^{L} N_{ii}}{\sum_{i=1}^{L} \sum_{j=1}^{L} N_{ij}}. \qquad (10)$$

Tables 4–6 show the identification results for three types of the training data (TD1, TD2, and TD3). IoU_j and PA are added under the confusion matrix. The numbers underlined in bold indicate the highest discrimination rate for each correct label. From these results, the accuracy of TD1 (nine images) is lower than other results. But the accuracies of TD2 (18 images) and TD3 (35 images) are almost the same. Therefore, from the viewpoint of the efficiency of the learning process, TD2 (18 training data) is suitable for this experiment. In the following, we conduct experiments using 18 training data.

Figures 25 and 26 show the extraction results of four material regions using U-Net. The input CT images in these figures are not used in the training data. These are the areas around the eyes of puppet. Metal parts do not exist in Figure 25, but they exist in Figure 26.

In Figures 25 and 26, (a) is the input CT image, (b) is the result of material extraction, (c) is the correct label image, and (d) is the difference image between (b) and (c). The red area is the extracted wood area, the yellow is the paint region, the purple is the metal region, and the black is the background (air and hair). In (d), the white area represents an incorrectly identified area.

**Table 4.** Confusion matrix for identification results using TD1.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**5,801,628**	99,075	32	614,032	6,514,767
	(%)	**(89.05)**	(1.52)	(0.00)	(9.43)	(100.0)
	Pain	67,682	**1,567,981**	4	117,274	1,752,941
	(%)	(3.86)	**(89.45)**	(0.00)	(6.69)	(100.0)
	Metal	355	**12,412**	393	3924	17,084
	(%)	(2.08)	**(72.65)**	(2.30)	(22.97)	(100.0)
	Background	502,290	78,549	46	**80,525,427**	81,106,312
	(%)	(0.62)	(0.10)	(0.00)	**(99.28)**	(100.0)
IoU		81.88	80.70	2.29	98.39	
PA				98.33		

Table 5. Confusion matrix for identification results using TD2.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**6,191,791**	92,567	1531	228,878	6,514,767
	(%)	**(95.04)**	(1.42)	(0.02)	(3.51)	(100.0)
	Pain	23,089	**1,705,733**	261	23,858	1,752,941
	(%)	(1.32)	**(97.31)**	(0.01)	(1.36)	(100.0)
	Metal	803	2607	**10,530**	3144	17,084
	(%)	(4.70)	(15.26)	**(61.64)**	(18.40)	(100.0)
	Background	243,187	98,699	628	**80,763,798**	81,106,312
	(%)	(0.30)	(0.12)	(0.00)	**(99.58)**	(100.0)
IoU		91.30	87.62	53.99	99.26	
PA				99.20		

Table 6. Confusion matrix for identification results using TD3.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**6,390,563**	34,366	29	89,809	6,514,767
	(%)	**(98.09)**	(0.53)	(0.00)	(1.38)	(100.0)
	Pain	35,532	**1,698,437**	17	18,955	1,752,941
	(%)	(2.03)	**(96.89)**	(0.00)	(1.08)	(100.0)
	Metal	4042	1250	742	**11,050**	17,084
	(%)	(23.66)	(7.32)	(4.34)	**(64.68)**	(100.0)
	Background	532,165	72,531	52	**80,501,564**	81,106,312
	(%)	(0.66)	(0.09)	(0.00)	**(99.25)**	(100.0)
IoU		90.18	91.26	4.32	99.11	
PA			99.11			
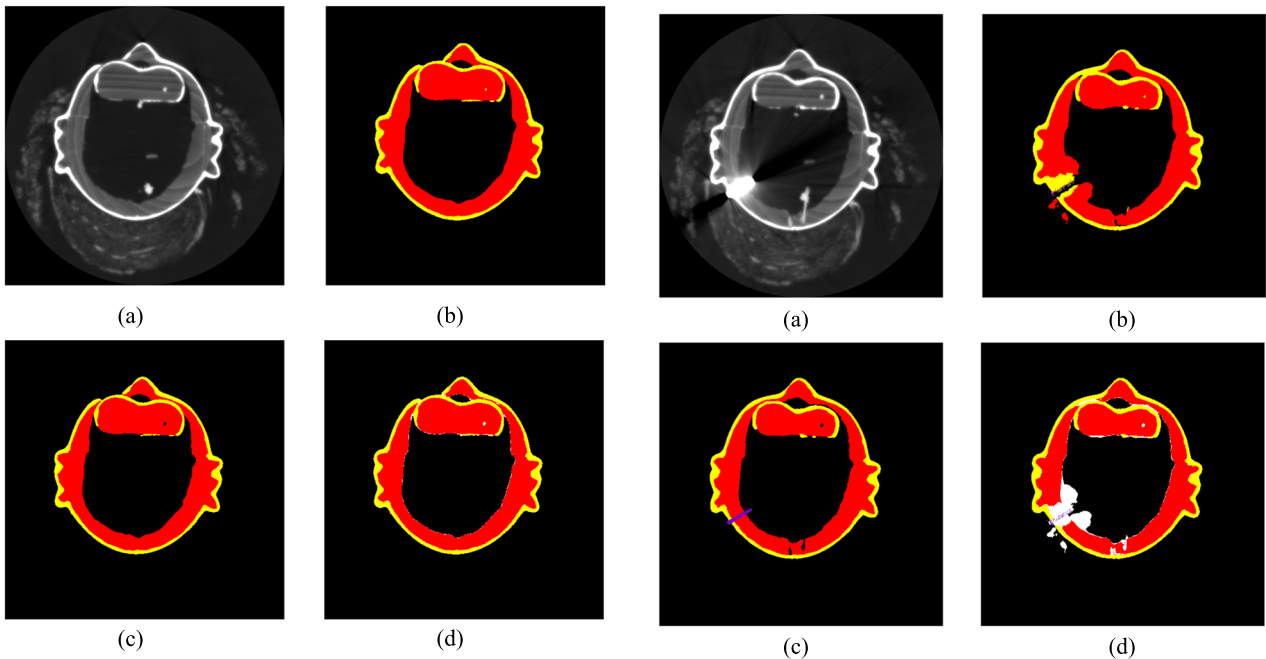
Figure 25. Results of wood and paint extraction without metal parts. (a) CT image. (b) Extracted wood and paint parts. (c) Correct label image. (d) Difference image.

Figure 26. Results of wood and paint extraction with metal parts. (a) CT image. (b) Extracted wood and paint parts. (c) Correct label image. (d) Difference image.

The results in Figures 25 and 26 show that hair around the head, which overlaps the range of pixel values with the wood area, is not almost detected. Therefore, we can say that ML (U-Net) can correctly discriminate between wood and hair. On the other hand, the results in Figure 26 show that some areas around the metal parts are not correctly discriminated as wood. This is the area where metal artefacts occur. Hence, we think that this effect is not sufficiently learned in this experiment.

Next, the 3D shapes of the wood extraction are depicted in Figure 27. Figure 27(a) shows the result of the extraction of all materials before the extraction process, Figure 27(b) shows the results of the extracted regions, and Figure 27(c) shows the correct label image (ground truth). The red area indicates the wood region, the yellow area indicates the paint region, the purple area indicates the metal region, and the black area indicates the background region. Figure 27(b) shows

that some areas on the top of the head are incorrectly extracted as wood regions. The reason for this is that, as shown in Figure 26, there are metal parts for fixing the hair on the top of the head as well, which are affected by metal artefacts.

In addition, the neck area at the bottom of the head has areas where the wood area is not correctly extracted. The following are possible reasons for this. The wood region of the head is surrounded by paint. In U-Net training, the positional relationship between the wood region and the surrounding paint region is considered a feature for wood region extraction.

Figure 28 shows 3D shape views of the inside of a head. The head contains small parts such as strings close to the threshold of wood as well as hair, but these parts have also been removed.

In the results of U-Net, it fails the correct extraction of the metal regions. To solve this problem, we should consider for finding the part of the nail in advance using
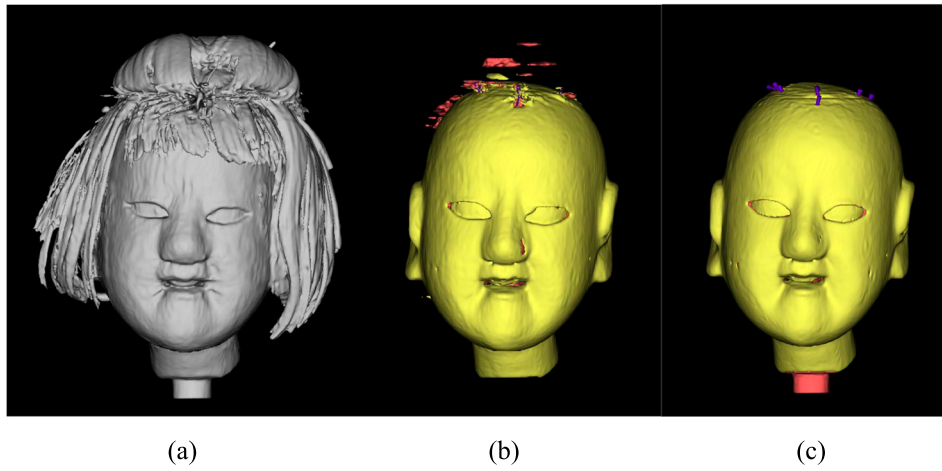
Figure 27. 3D shapes of extracted materials. (a) Before the extraction process. (b) Extracted regions. (c) Ground truth.
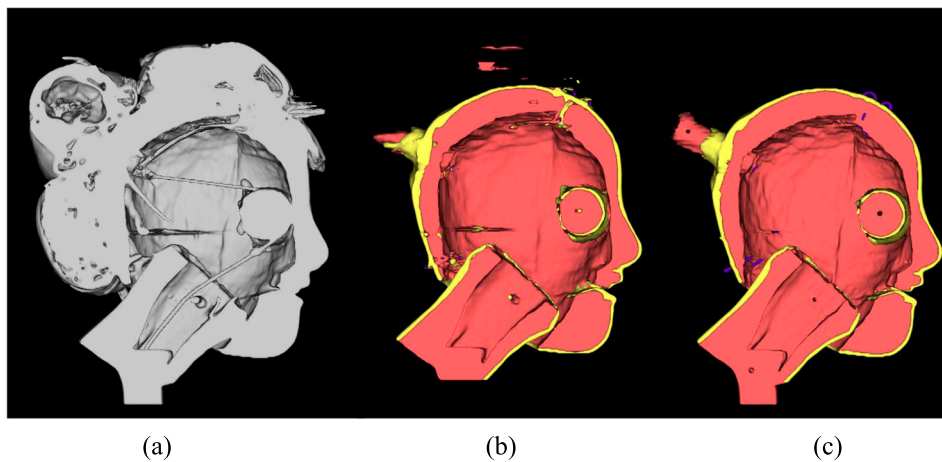

Figure 28. Inside shapes of extracted materials. (a) Before the extraction process. (b) Extracted regions. (c) Ground truth.

a method different from ML. For example, the use of the condition that the CT value of the metal is significantly high compared to the others and the shape of the nail is long and thin.

6.3. Comparison of proposed methods

In this section, we compare both methods, namely, the graph cut and ML (U-Net) methods for reconstructing a puppet head from CT images.

6.3.1. Comparison of head part extraction

Figures 29 and 30 show the extraction results of each method using the same CT image. Figure 29 shows the results without metal and Figure 30 shows the results with metal. In Figures 29 and 30, (a) are CT images (same as Figures 25 and 26). Both (b) show the results of the graph cut method, and both (c) show the results of U-Net. White indicates correct extraction results, and red indicates incorrect extraction results. In the graph cut method, wood and paint are extracted
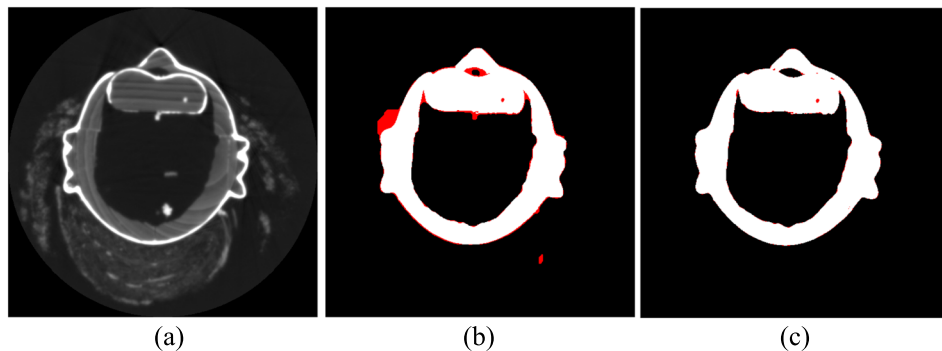

Figure 29. Extraction results by each method without metal regions. (a) CT image. (b) Graph cut method. (c) U-Net.
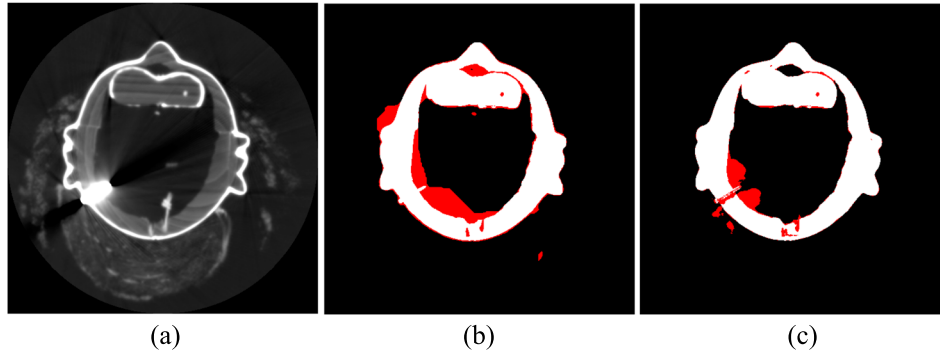
| (a) | (b) | (c) |

Figure 30. Extraction results by each method with metal regions. (a) CT image. (b) Graph cut method. (c) U-Net.

Table 7. Exchanged confusion matrix of two labels from the results of U-Net method (Table 5).

		Identified label		Total
		Object	Background	
Correct label	Object	**8,028,912**	255,880	8,284,792
	(%)	**(96.91)**	(3.09)	(100.0)
	Background	342,514	**80,763,798**	81,106,312
	(%)	(0.42)	**(99.58)**	(100.0)

as objects, and metal parts are also included in the paint. Therefore, the U-Net results are shown for wood, paint, and metal. Both (Figures 29 and 30) show that the U-Net method correctly identifies the object (puppet head). The same trend can be seen in the other CT images.

6.3.2. Comparison of accuracy

Next, the material identification rates of the two methods are compared. As mentioned above, the graph cut method extracts wood, paint, and metal as an object, and the U-Net method also evaluates wood, paint, and metal as an object.

The results by U-Net shown in Table 5 are changed to the confusion matrix in the case of two labels (the object and background). This result is shown in Table 7. The *PA* of Tables 2 and 7 are shown below:

- Graph cut method: 97.36%
- U-Net: 99.33%

These results show that the U-Net method is more accurate for the entire CT image.

But the graph cut method estimates based on manually directed regions in 10 CT images as shown in Figure 20. In contrast, U-Net uses the correct label images manually annotated on all pixels in 18 CT images. Therefore, the above results show that the graph cut method compares unfavourably with the U-Net method. As a result of the comparison under the same conditions, we show the results of the graph cut method using the correct label images for the U-Net as the manually directed regions.

However, in the graph cut method shown in Section 4, the first process to calculate the threshold value to divide the region is to obtain histograms for the four materials: air, hair, wood, and paint. On the other hand, the U-Net extracts background (air and hair), wood, paint, and metal materials. This is due to the difficulty of accurately annotating the hair in creating the correct label image, and we prioritize the extraction of wood and paint regions for the reconstruction of a puppet head by the 3D printer. But the graph cut method can be applied when the background (air and hair) and the object (wood, paint, and metal) can be divided. Hence, the histogram is calculated by regarding the air and hair as one area (background), and the threshold with wood area can be obtained.

Figure 31 depicts histograms and normal distributions of three materials obtained from 18 correct label images. Note that the histogram of the metal region is not obtained because the graph cut method does not consider the metal region. Table 8 shows means and variances of normal distributions and thresholds to divide these materials. The threshold which divides wood and hair is −850 in Table 1, but it changes to −888 in Table 8. This is because the air and hair are regarded as one region.

Table 9 shows the confusion matrix of the identified object and background by the graph cut method based on the threshold in Table 8. The *PA* of this result is shown below:

- Graph cut method (Table 9): 97.17%

Figure 32 shows reconstructed 3D shapes of extracted object parts. The *PA* value is 0.19% lower than the results of manually directing areas. But more hair parts are remained than Figure 23(b). To investigate this reason, the same CT image around the eyes of the puppet head is binarized at the thresholds of −850 and −888, and the extracted images are shown in Figure 33. Threshold −888 is closer to the hair range. Large parts of hair which is not the object are extracted. Because these parts are remained as the object seeds in the graph cut method, a large number of hair areas are remained.
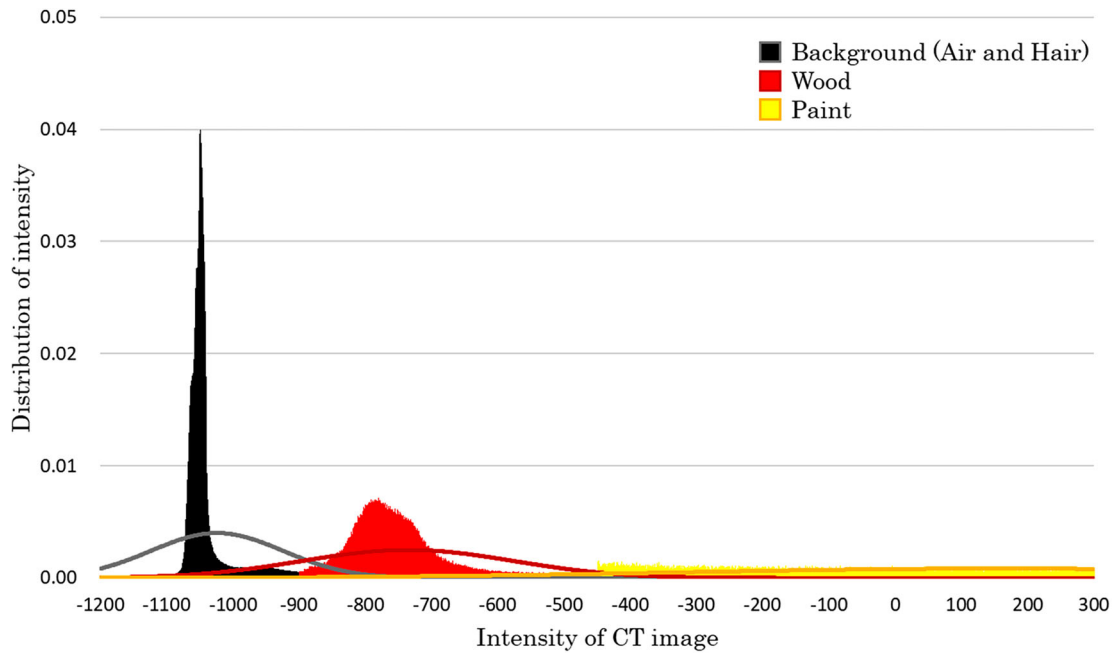
Figure 31. Probability distributions and normal distributions using 18 correct label images.

Table 8. Estimated means, variances, and thresholds using 18 correct label images.

	Background	Wood	Paint
Mean	−1023.9	−736.0	160.1
Variance	9910.6	25,440.5	226,150.0
Threshold		−888	−429

Table 9. Confusion matrix by the graph cut method using 18 correct label images.

		Identified label		
		Object	Background	Total
Correct label	Object	**8,169,424**	115,368	8,284,792
	(%)	**(98.61)**	(1.39)	(100.0)
	Background	2,412,353	**78,693,959**	81,106,312
	(%)	(2.97)	**(97.03)**	(100.0)

The graph cut method requires the specification of regions that can be used as seeds reliably, therefore seed regions are generally annotated manually. The correct label images used in this experiment are different from manual directed areas in the way of the region specification. Hence, the thresholds change, uncertain seeds increase, and accuracy becomes low. Changes in threshold values are small relative to the range of pixel values in a CT image. However, because of its large influence, the method using the threshold and the graph cut is more unstable than the ML (U-Net) method.

6.4. Two heads extraction by the ML method

In this section, we show the extraction results of another puppet head shown in Figure 34 using U-Net. From here, the puppet head in Figure 5 is called "No.1" and the head in Figure 34 is called "No.2." The size of a CT image of No.2 is 512 × 512 pixels, it is the same as that
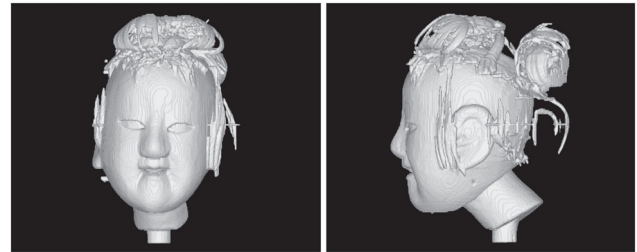

Figure 32. Reconstructed 3D shape of puppet head using the graph cut method by 18 correct label images.
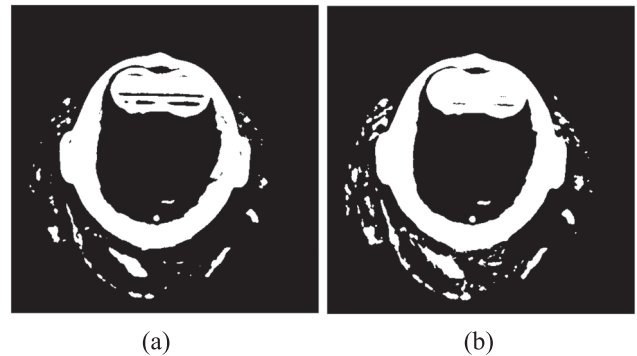

(a)   (b)

Figure 33. Binarized results of CT images. (a) Threshold: −850. (b) Threshold: −888.

of No.1, but the size of head of No.2 is larger than that of No.1, the number of CT images is 361, 20 more than that of No.1. A part of the hair extends to the outside area of the CT image.

For head No.2, 19 images, which is approximately 5% of all CT images, are selected as training data in the same way as in Figure 17. The following three types of training data are prepared from the combination with the training data of No.1, and the identification results

Figure 34. Another puppet head (No.2).

of head No.1 and No.2 are obtained for each of them. Note that the results of EX A-1 correspond to Table 5 shown in Section 6.2.

- Training data A: 18 images of Head No.1.
  - o EX A-1. Identification of head No.1. (Experimental results in Section 6.2, Table 5.)
  - o EX A-2. Identification of head No.2.
- Training data B: 19 images of Head No.2.
  - o EX B-1. Identification of head No.1.
  - o EX B-2. Identification of head No.2.
- Training data C: 18 images of head No.1 and 19 images of head No.2 (total of 37 images).
  - o EX C-1. Identification of head No.1.
  - o EX C-2. Identification of head No.2.

Tables 10–14 show the confusion matrix, IoU_j, and PA as the identification results from EX A-2 to EX C-2. Including Table 5, when the correspondence between

Table 10. Confusion matrix for identification results of EX A-2.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**7,097,128**	192,466	2249	863,723	8,155,566
	(%)	**(87.02)**	(2.36)	(0.03)	(10.59)	(100.0)
	Pain	31,949	**1,907,761**	239	63,861	2,003,810
	(%)	(1.59)	**(95.21)**	(0.01)	(3.19)	(100.0)
	Metal	208	3679	**11,708**	4620	20,215
	(%)	(1.03)	(18.20)	**(57.92)**	(22.85)	(100.0)
	Background	1,102,733	395,195	737	**82,955,728**	84,454,393
	(%)	(1.31)	(0.47)	(0.00)	**(98.23)**	(100.0)
IoU		76.39	73.51	49.95	97.15	
PA				97.19		

Table 11. Confusion matrix for identification results of EX B-1.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**6,278,286**	99,740	1189	135,552	6,514,767
	(%)	**(96.37)**	(1.53)	(0.02)	(2.08)	(100.0)
	Pain	42,912	**1,673,707**	294	36,028	1,752,941
	(%)	(2.45)	**(95.48)**	(0.02)	(2.06)	(100.0)
	Metal	776	2402	**11,581**	2325	17,084
	(%)	(4.54)	(14.06)	**(67.79)**	(13.61)	(100.0)
	Background	1,465,178	145,566	1281	**79,494,287**	81,106,312
	(%)	(1.81)	(0.18)	(0.00)	**(98.01)**	(100.0)
IoU		78.25	83.66	58.35	97.80	
PA				97.84		

Table 12. Confusion matrix for identification results of EX B-2.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**7,912,431**	115,860	2614	124,661	8,155,566
	(%)	**(97.02)**	(1.42)	(0.03)	(1.53)	(100.0)
	Pain	52,712	**1,916,172**	427	34,499	2,003,810
	(%)	(2.63)	**(95.63)**	(0.02)	(1.72)	(100.0)
	Metal	608	3580	**13,436**	2591	20,215
	(%)	(3.01)	(17.71)	**(66.47)**	(12.82)	(100.0)
	Background	345,404	157,175	1406	**83,950,408**	84,454,393
	(%)	(0.41)	(0.19)	(0.00)	**(99.40)**	(100.0)
IoU		92.50	84.03	54.48	99.21	
PA				99.11		

Table 13. Confusion matrix for identification results of EX C-1.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**6,338,401**	96,839	817	78,710	6,514,767
	(%)	**(97.29)**	(1.49)	(0.01)	(1.21)	(100.0)
	Pain	20,022	**1,705,074**	280	27,565	1,752,941
	(%)	(1.14)	**(97.27)**	(0.02)	(1.57)	(100.0)
	Metal	2246	**8485**	5725	628	17,084
	(%)	(13.15)	**(49.67)**	(33.51)	(3.68)	(100.0)
	Background	512,990	125,184	894	**80,467,244**	81,106,312
	(%)	(0.63)	(0.15)	(0.00)	**(99.21)**	(100.0)
IoU		89.91	85.97	30.01	99.08	
PA			99.02			

Table 14. Confusion matrix for identification results of EX C-2.

		Identified label				
		Wood	Paint	Metal	Background	Total
Correct label	Wood	**7,979,343**	112,416	703	63,104	8,155,566
	(%)	**(97.84)**	(1.38)	(0.01)	(0.77)	(100.0)
	Pain	42,188	**1,935,928**	133	25,561	2,003,810
	(%)	(2.11)	**(96.61)**	(0.01)	(1.28)	(100.0)
	Metal	2791	**9818**	6316	1290	20,215
	(%)	(13.81)	**(48.57)**	(31.24)	(6.38)	(100.0)
	Background	284,514	174,498	729	**83,994,652**	84,454,393
	(%)	(0.34)	(0.21)	(0.00)	**(99.46)**	(100.0)
IoU		94.04	84.15	29.00	99.35	
PA			99.24			

the training data and the head is different, the identification rate become low slightly. But, in case C, where both training data are used, both heads are identified with high accuracy. However, in all results, the identification rate of metals is low.

Figures 35 and 36 show the identification results of head No.2 in EX C-2. These are corresponding to Figures 25 and 26. Figure 35 is the case without metal parts and Figure 36 is the case with metal parts. From these results, we can see that the identification accuracy of the metal parts is low.

Figures 37 and 38 show a 3D view of the extraction results. In the CT images, the face of this puppet is slightly tilted upward, hence these are rotated so that they are in the frontal plane. A part of the hair is not shown in the result before the extraction process, because the hair is spread outside the area of the CT image. Although a part of the hair remains, the extraction result is similar to that of No.1.

Furthermore, to improve the accuracy of ML methods, it is necessary to increase the training data. In the future, we will need to increase the training data when extracting and evaluating other head shapes. However, we need to prepare the training data (correct label images) of puppet heads by ourselves. In this study, as shown in Section 5.2, we manually correct the threshold images to obtain the correct label images. However, because the graph cut method is more accurate than the thresholding method, it will be effective to correct the graph cut method results manually to obtain correct label images. Then, by using such training data,
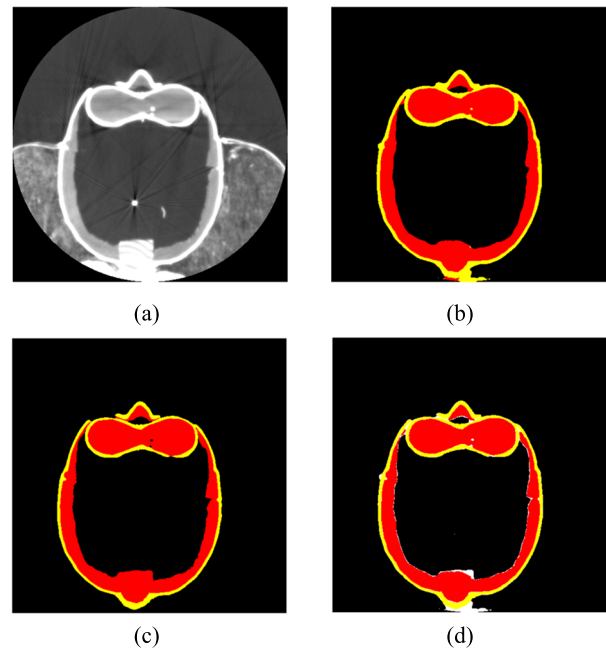

(a)　　　　　　　　　(b)

(c)　　　　　　　　　(d)

Figure 35. Results of wood and paint extraction without metal parts (puppet head No.2). (a) CT image. (b) Extracted wood and paint parts. (c) Correct label image. (d) Difference image.

the head is extracted using the U-Net method. It is necessary to study such a system in the future.

7. Conclusion

In this study, for the digital archiving of Japanese traditional puppet theatre (Awa Ningyo Joruri), we propose
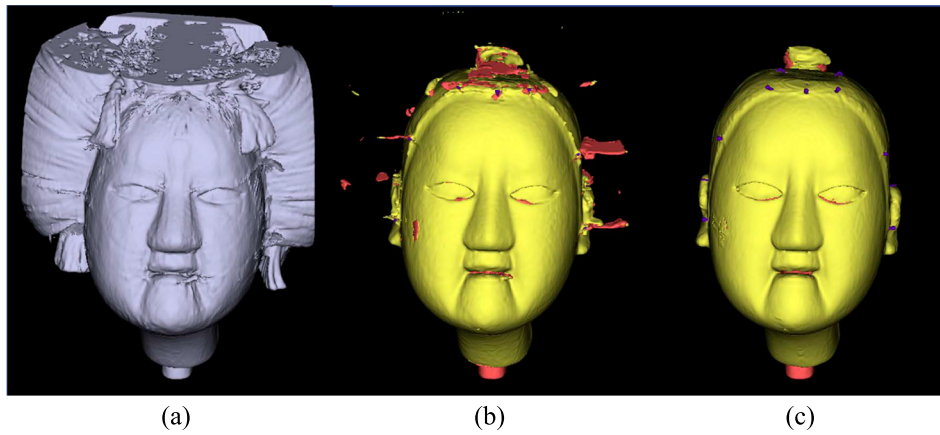
Figure 37. 3D shapes of extracted materials (puppet head No.2). (a) Before the extraction process. (b) Extracted regions. (c) Ground truth.
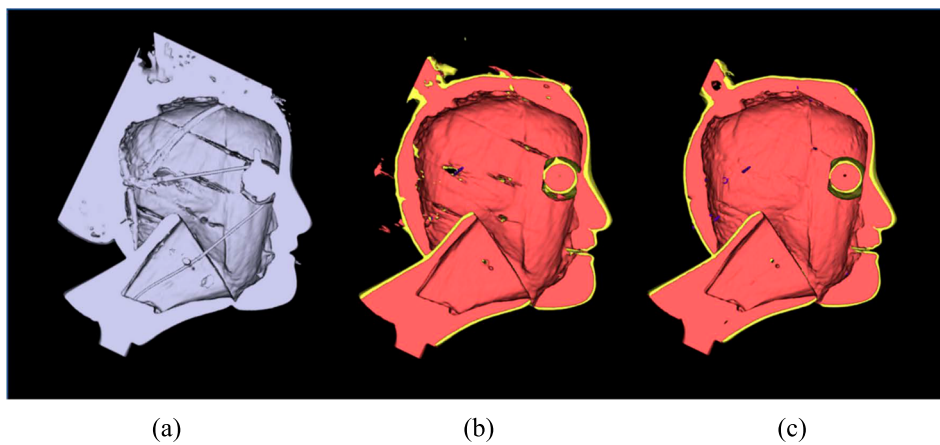

Figure 38. Inside shapes of extracted materials (puppet head No.2). (a) Before the extraction process. (b) Extracted regions. (c) Ground truth.
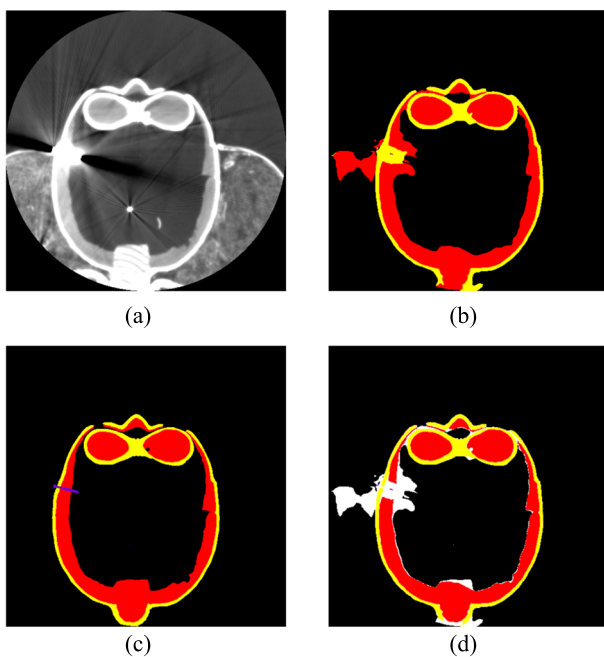

Figure 36. Results of wood and paint extraction with metal parts (puppet head No.2). (a) CT image. (b) Extracted wood and paint parts. (c) Correct label image. (d) Difference image.

two methods for extracting the puppet head shape from CT images. One is the graph cut method, and the other is the U-Net method, an ML approach.

From the experimental results of the extraction of the puppet head from CT images, the U-Net method can extract the puppet head more accurately than the graph cut method. Moreover, we can show that the U-Net method can extract a puppet head with multiple materials. However, the extraction of metal parts is inaccurate because of the metal artefacts in the X-ray CT images and insufficient learning data.

In future studies, because hair is not subjected for identifying region in this study, hence, it is necessary to discuss the extraction of air and hair separately to reconstruct the perfect puppet head as digital archiving. we will discuss how to improve the extraction accuracy of each material using the U-Net method. In the proposed method, each CT image is extracted independently, and we will also investigate an extraction method using 3D information consisting of adjacent CT images. Furthermore, we plan to improve the accuracy of the learning model using CT images of other

puppet heads and to realize the digital archiving of puppet heads by extracting many heads. Hence, we will discuss the improvement of the extraction accuracy of each material using not only the U-Net but also other ML frameworks, such as PSPNet, U2Net, U-Net++, SegNet, and so on.
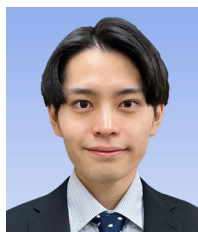
Acknowledgment

Disclosure statement

Notes on contributors

Hinata Ikeda He received his B.E. from Tokushima University in 2022. He is currently a Master's student at the Graduate School of Advanced Technology and Science, Tokushima University, Japan. His main research interests include image processing and machine learning.

Hiroyuki Ukida He received his B.E. and M.E. from Okayama University in 1992 and 1994, respectively, and a Ph.D. degree in informatics from Kyoto University in 2003. He is currently an Associate Professor at Tokushima University. His main research interests include image processing, robot vision, and machine learning. He is a member of IEEE.

Kouki Yamazoe He received his B.E. from Tokushima University in 2021. He is currently a Master's student at the Graduate School of Advanced Technology and Science, Tokushima University, Japan. His main research interests include shape restoration and machine learning.

Masahide Tominaga He received his Ph.D. degree from Tokushima University in 2011. He is currently a Lecture at Tokushima University. His main research interests include medical image analysis and radiation therapy. He is a member of the JSRT, JSMP, and JASTRO.
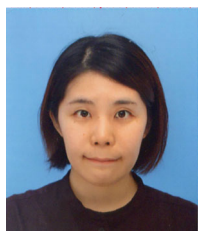
Tomoyo Sasao She received her B.E. from Tokyo Metropolitan University in 2011 and her M.E. and Ph.D. degrees in environmental studies from the University of Tokyo in 2013 and 2016, respectively. She is currently a Project Assistant Professor at the University of Tokyo. Her main interests include participatory design, urban computing, and human–computer interaction.

Kenji Terada He received a Ph.D. degree from Keio University in 1995. He is currently a Professor at Tokushima University. His main research interests include computer vision and image processing. He is a member of the SICE, IEEJ, IEICE, JSPE, and ISCIE.

Funding

References

[1] Report on the Results of a Survey on the Utilization of Traditional Crafts as a Regional Resource. Administrative Evaluation Bureau, Ministry of Internal Affairs and Communications, Japan; 2022. p. 39–53. Japanese.

[2] Ikeuchi K, Oishi T. 3D digital archiving. Tokyo: University of Tokyo Press; 2010.

[3] Aoki M, Kashimura M, Ozawa S. Preservation and display system for rare book research. J Inst Image Electron Eng Jpn. 2004;33(5):776–782.

[4] Kashimura M, Takamiya T. Digital archiving of rare books – preset and prospect. J Print Sci Technol. 2004;41(3):17–26.

[5] Awa Jurobe Yashiki [Internet]. [cited 2022 Oct 17]. Available from: https://joruri.info/jurobe/.

[6] Ukida H, Sasao T, Terada K, et al. Calibration method of 3D shape measurement system using 3D scanner, turn table and arm robot. Proceedings of the SICE Annual Conference; 2019 Sept 10–13; Hiroshima, Japan; 2019. p. 136–141.

[7] Hatakenaka Y, Ukida H, Tominaga M, et al. 3D shape reconstruction of puppet head from CT images. Proceedings of the SICE Annual Conference; 2020 Sept 23–26; Chiang Mai, Thailand; 2020. p. 701–705.

[8] Ukida H, Hatakenaka Y, Tominaga M, et al. Inner and outer 3D shape reconstruction of puppet head from CT images. Proceedings of SPIE 11794, 15th International Conference on Quality Control by Artificial Vision; 2021 July 16; Tokushima, Japan. DOI: 10.1117/12.2588748

[9] Ukida H, Yamazoe K, Tominaga M, et al. 3D shape reconstruction of Japanese traditional puppet head from CT images. Frontiers of Computer Vision (CCIS). 2022;1578:49–63. doi:10.1007/978-3-031-06381-7_4.

[10] Ikeda H, Ukida H, Yamazoe K, et al. 3D shape reconstruction of puppet head from CT images by machine learning. Proceedings of the SICE Annual Conference; 2022 Sept 10–13; Kumamoto, Japan; 2022. p. 592–597.

[11] DEKO [Internet]. [cited 2022 Oct 17]. Available from: https://3ddeko.shop/ Japanese.

[12] TOKUSHIMA UNIVERSITY Digital Fabrication PROJECT [Internet]. [cited 2022 Oct 17]. Available from: https://dentomake.com/ Japanese.

[13] Sasao T, Ukida H, Takana T, et al. Development and evaluation of a method for generating 3D models of joruri puppets reflecting the facial image features of human. J Jpn Acad Fac Stud. 2022;21(2):5–13. Japanese.

[14] Production of Ningyo Joruri by X-ray CT and 3D Printer [Internet]. [cited 2022 Oct 17]. Available from: http://advancedvislab.com/?page_id = 1074 Japanese.

[15] Zhou X, Wang S, Chen H, et al. Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning.

Comput Med Imaging Graph. 2012;36:304–313. DOI: 10.1016/j.compmedimag.2011.12.004

[16] Zhou X, Takayama R, Wang S, et al. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. Med Phys. 2017;44(10):5221–5233.

[17] Ishikawa H. Graph cut. IPSJ SIG technical report. CVIM. 2007;158(26):193–204. Japanese.

[18] Fujita H. Introduction to medical image processing in Python, Ohmsha, Tokyo; 2020. Japanese.

[19] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell 2004;26(9):1124–1137.