Curve-Structure Segmentation from Depth Maps: A CNN-Based Approach and Its Application to Exploring Cultural Heritage Objects

Yuhang Lu,[2] **Jun Zhou,**[2] **Jing Wang,**[2] **Jun Chen,**[2] **Karen Smith,**[3] **Colin Wilder,**[4] **Song Wang**[1, 2, *]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC
[3] South Carolina Institute of Archaeology and Anthropology, University of South Carolina, Columbia, SC
[4] Center for Digital Humanities, University of South Carolina, Columbia, SC

Abstract

Motivated by the important archaeological application of exploring cultural heritage objects, in this paper we study the challenging problem of automatically segmenting curve structures that are very weakly stamped or carved on an object surface in the form of a highly noisy depth map. Different from most classical low-level image segmentation methods that are known to be very sensitive to the noise and occlusions, we propose a new supervised learning algorithm based on Convolutional Neural Network (CNN) to implicitly learn and utilize more curve geometry and pattern information for addressing this challenging problem. More specifically, we first propose a Fully Convolutional Network (FCN) to estimate the skeleton of curve structures and at each skeleton pixel, a scale value is estimated to reflect the local curve width. Then we propose a dense prediction network to refine the estimated curve skeletons. Based on the estimated scale values, we finally develop an adaptive thresholding algorithm to achieve the final segmentation of curve structures. In the experiment, we validate the performance of the proposed method on a dataset of depth images scanned from unearthed pottery sherds dating to the Woodland period of Southeastern North America.

Introduction

Embellished designs on the surface of cultural heritage objects, such as pottery, shell, stone and wood contain important information for archaeologists (Zhou et al. 2017). These designs, if successfully identified and correlated, can be used to build chronologies and track trade networks of a region thousands of years ago. In archeology, most of these designs are found to be curve patterns stamped or carved by their makers. Therefore, it is of great interest to archaeologists to accurately segment the curve structures on the surface of unearthed fragments of cultural heritage objects and identify their underlying designs (Kampel and Sablatnig 2007; Halir 1999). Figure 1 shows several unearthed pottery sherds dating to the Woodland period of Southeastern North America. The curve structures on their surfaces reflect a portion of the curve pattern carved into wooden paddles and applied onto hand-built clay vessels designed by southeastern Native Americans around 2000 years ago. There are hundreds of
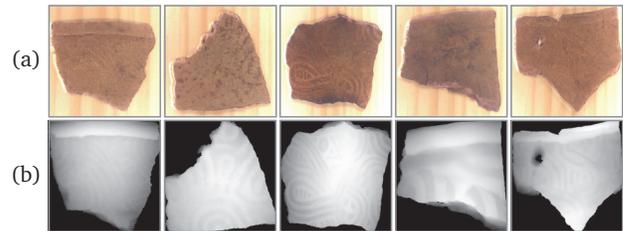
Figure 1: Five unearthed pottery sherds dating to the Woodland period of Southeastern North America. (a) RGB images. (b) Depth images where intensity indicates the depth.

thousands of such fragmented culture heritage objects stored in museums, which calls for more intelligent and automatic tools to explore them.

Clearly, accurately segmenting the curve structures stamped on the surface is the first step to explore these cultural heritage objects. In most cases, these curve structures do not bear distinctive colors and it is very difficult, if not impossible, to segment them from an RGB image of the sherd, e.g., Figure 1(a), taken by traditional cameras. In archeology, 3D scanners are usually utilized to produce a depth image of the object surface – with paddle stamping, the locations of curves exhibit a larger depth than the non-curve portion of surface, as shown in Figure 1(b).

However, three complexities may lead to very weak curve structures on the obtained depth map and make the curve-structure segmentation a very challenging problem. First, the carved paddles used for stamping are usually flat while the object surfaces are usually not. As a result, the paddle typically does not well fit the object surface, which leads to shallow curves at many locations. Second, purposeful smoothing of the stamped surface during vessel manufacture or weathering and erosion after vessel discard can lead to subtle depth differences between the curve and the non-curve portions of the surface. Third, erosion and weathering make the object surface highly rough, which is equivalent to adding random noise to the depth map of the initial object surface. With these three complexities, it is difficult to use a low-level image segmentation algorithm to accurately segment these depth images for curve structures, as shown by an example in Figure 2.
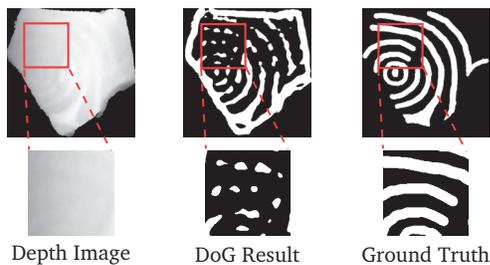
Depth Image   DoG Result   Ground Truth

Figure 2: An illustration of using low-level methods for curve-structure segmentation. Serious erosion in the red square leads to very low contrast in the depth image, and low-level method, such as DoG (Difference of Gaussian), may produce very poor segmentation.

In this paper, we propose a new supervised learning approach to segment such curve structures that were weakly stamped on object surface. The basic idea is that, in most applications, such as exploring cultural heritage objects in archeology, the underlying designs of the curve structures bear certain geometries and patterns. For example, most of the curve structures consist of smooth curve segments. Furthermore, many curves in the structures show good parallelism against each other. These characteristics give the material a visually distinctive style (Smith and Knight 2012). Consideration of these high-level geometry and pattern information may help improve the accuracy and reliability of curve-structure segmentation. While it is difficult to handcraft the features of all relevant curve geometry and pattern in an application, we expect the proposed approach can automatically learn these features from a set of training data with labeled ground truth.

In practice, the curve structures of interest have width, which may vary along the curve and need to be inferred in segmentation. However, it is well known that the curve geometry and pattern are independent of the curve width. Mixing all of them may substantially increase the difficulty of feature learning for segmentation. In this paper, we handle them separately by developing a three-step curve-structure segmentation algorithm. In the first step, a Fully Convolutional Network (FCN) is employed to extract the skeleton of curve structures, and estimate a scale value at each skeleton pixel. This scale value reflects the curve width at the corresponding skeleton pixel. In the second step, we propose a dense prediction network to refine the curve skeletons. In the third step, we develop an adaptive thresholding algorithm to achieve the final segmentation of curve structures with width by considering the estimated scale values.

For the experiments, we collected the depth image of a set of pottery sherds excavated from archaeological sites associated with the Swift Creek paddle-stamped tradition of southeastern North America. Ground-truth curve structure segmentation are manually constructed. We evaluate the proposed method on the collected depth images and compare its performance against several other existing algorithms. We also evaluate the segmentation results in the task of design matching in archeology.

Related Work

General-purpose image segmentation has been studied for many decades, resulting in many image segmentation algorithms. For example, by considering only low-level pixel intensities, many edge detection (Wang, Kubota, and Siskind 2004; Arbelaez et al. 2011), region growing/splitting (Tremeau and Borel 1997), pixel clustering (Li and Chen 2015), and graph-based algorithms (Shi and Malik 2000; Wang and Siskind 2001; 2003) have been developed for segmenting an image into multiple regions. By further considering mid-level cues like boundary smoothness, many active-contour and level-set algorithms have been developed to segment foreground objects from background (Chan and Vese 2001; Vese and Chan 2002). In principle, these general-purpose image segmentation algorithms can be easily adapted to handle our problem of segmenting curve structures from depth images, by treating depth value as intensity. However, their segmentation performances are usually poor when the depth image is noisy and the desired curve structures are weak. In the experiments, we include several general-purpose segmentation algorithms, such as DoG, LevelSet (Vese and Chan 2002), and GrabCut (Rother, Kolmogorov, and Blake 2004), as comparison methods.

Deep-learning based algorithms, particularly the CNN-based algorithms, have been recently used for image segmentation, by learning high-level features of the desired segments in a supervised way (Badrinarayanan, Kendall, and Cipolla 2015; Zheng et al. 2015). The most influential one is the Fully Convolutional Network proposed by Long, Shelhamer, and Darrell (2015). It transforms traditional fully connected layers to convolution layers, thus enabling to train and predict a whole image at a time. To improve the localization of object boundaries, Chen et al. (2016) proposed a framework to combine Conditional Random Field (CRF) with FCNs. However, if we directly apply these deep-learning based segmentation algorithms to our problem of segmenting curve structures, it may produce non-curve segments because the CNNs are trained directly on the color/intensity images. In this paper, we will train CNNs on the curve-skeleton images to better learn the curve-geometry and curve-pattern features. More related to our work is Deep Skeleton (Shen et al. 2016), which also uses CNNs for skeleton extraction. However, Deep Skeleton is not specifically developed for curve structures and may produce many false positive skeletons. In the experiments, we include Deep Skeleton as a comparison method.

Curve-structure segmentation from RGB or gray-scale images have been studied in many specific applications. For example, Lorigo et al. (2001) utilized an energy criterion based on intensity and local boundary smoothness to extract blood vessels in medical images. Tao, Prince, and Davatzikos (2002) constructed a statistical shape model to extract sulcal curves on the outer cortex of human brain. Zou et al. (2012) proposed a tree-based algorithm to detect curve-like cracks from pavement images. However, these methods all rely on specific assumptions in respective applications and it is not easy to extend the segmentation algorithm developed for one application to another application.

Using computer vision and machine learning techniques

to explore cultural heritage objects has been attracting more interest in recent years. However, most of them are focused on the classification and matching of object fragments. For example, in (Smith et al. 2010; Makridis and Daras 2012; Rasheed and Nordin 2015), various archaeological fragments are classified based on color and texture features. In (Zhou et al. 2017), an extended Chamfer matching algorithms is developed to identify the design of a pottery sherd by matching the curve structures on the sherd to all the known designs, where the curve structures on the sherds are segmented with manual assistance. In this paper, we focus on accurate segmentation of curve structures on the surface of sherds, which is a fundamental step before the classification and matching.

Proposed Method

The proposed method consists of three steps. First, we train an FCN to detect the skeletons of the curve structures in the depth image. This FCN network also estimates a scale value at each detected skeleton pixel to reflect the curve width at this skeleton pixel. Second, we train a dense prediction convolutional network to identify and prune false positive skeleton pixels. Finally, we develop a scale-adaptive thresholding algorithm to recover the curve width and achieve the final segmentation of curve structures.

Step I: Detecting Curve Skeletons using FCN

In this paper skeletons are the center lines of the curve structures and they are of one-pixel width. By ignoring the curve width, the skeletons reflect the geometry and pattern of the curve structures. Therefore, in the first step, we train a FCN to detect the skeletons of the curve structures from an input depth image. Just like image segmentation, skeleton detection can be formulated as a pixel-labeling problem: skeleton pixel has a label 1 and non-skeleton pixel has a label 0.

We design an FCN, as illustrated in Figure 3, to label skeleton pixels. It follows the encoder-decoder architecture developed in (Long, Shelhamer, and Darrell 2015). Encoders 1 and 2 are small convnets made up of two 3×3 convolutional layers, two ReLu layers and one 2×2 max-pooling layer. Encoder 3 is a small convnet made up of three 3×3 convolutional layers, three ReLu layers and one 2×2 max-pooling layer. After an encoder, the image size will be reduced to 1/4. Therefore, the receptive field sizes of feature maps generated by the three encoders are 2×2, 4×4, and 8×8, respectively. After each encoder, a fully connected layer is employed to match the number of feature maps with the number of labels. In order to generate pixelwise prediction result, the fully connected layers are implemented by 1×1 convolutional layers. These results are denoted as S_1, S_2 and S_3, respectively, as shown in Figure 3. Note that the size of S_1, S_2 and S_3 are successively downsampled by factors of 2, 4, and 8 from the original image size. The decoders are three deconvolution layers with a kernel size of 4×4 and a stride of 2. The kernels are fixed to perform bilinear interpolation (Xie and Tu 2015).

The use of multiple encoders/decoders can extract image features in different levels of details. To make full use of all
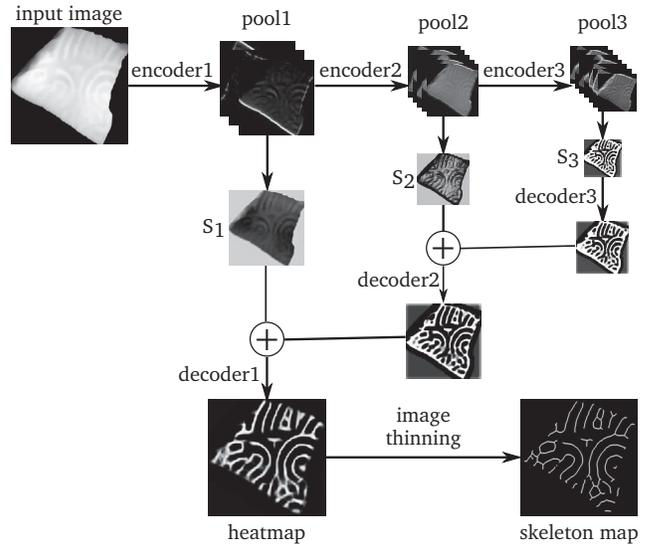

Figure 3: FCN used for skeleton detection.

the extracted features, the decoders are organized in a way of stepwise accumulation when fusing them together. The output skeleton heat map S can be computed by

$$S = softmax(\Psi^{(2)}(S_1 + \Psi^{(2)}(S_2 + \Psi^{(2)}(S_3)))) \quad (1)$$

where Ψ indicates the upsampling operation performed by the decoders and its associated superscript is the upsampling factor, e.g., $\Psi^{(2)}$ indicates an upsampling of map by a factor of 2. With the skeleton heat map S, we apply a common image thinning algorithm (Lam, Lee, and Suen 1992) to generate the single-pixel width skeleton map.

Inspired by (Shen et al. 2016), we can compare the three score maps S_1, S_2 and S_3 to estimate the scale at each detected skeleton pixel. The scale value at a skeleton pixel reflects the local curve width at this pixel. More specifically, since different encoders correspond to different receptive field sizes, at each pixel the receptive field size of the encoder with the largest score reflects the scale at this pixel. Before we compare the score of different maps, we need to first upsample them to the original image size. This way, the scale $s(x, y)$ at the skeleton pixel (x, y) can be computed by

$$s(x, y) = \arg \min_{k \in \{1,2,3\}} \hat{S}_k(x, y) \quad (2)$$

where $\hat{S}_k = \Psi^{(2^k)}(S_k)$ is the upsampled score map of S_k. Later we will use the estimated scale values to help recover the curve width.

Step II: Refining Skeletons using Dense Prediction Convnet

While we expect the FCN trained in Step I can learn curve geometry and pattern features in detecting skeletons, we find that it still detects many false positive skeletons, as shown in Figure 4. In this step, we further train a supervised classifier to identify and prune such false positives by learning more
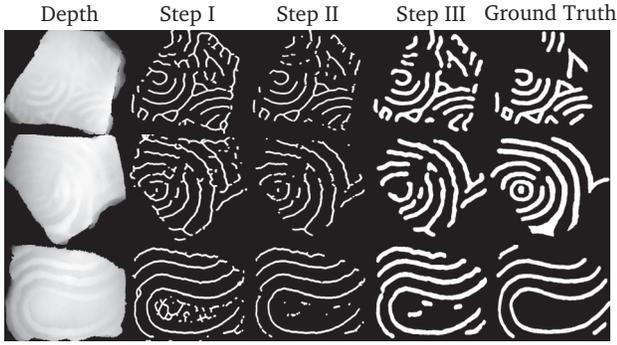
Depth   Step I   Step II   Step III   Ground Truth

Figure 4: Example results after each step of the proposed method.

curve features. Specifically, for each skeleton pixel (x, y) detected in Step I, we take a neighboring 45×45 window in the original depth image around the pixel (x, y) as the input and train a dense prediction convnet to determine whether (x, y) is a true skeleton pixel or a false positive.

On real images, detecting a skeleton with small dislocation to its real position is totally fine and unavoidable – even a manually labeled skeleton may not be perfectly aligned with the real center line of the curve structures. Therefore, our aim is not to directly train a hard classifier to distinguish skeleton pixels and non-skeleton pixels. Instead, we hope to train a soft classifier where a skeleton probability is outputted at each pixel. To achieve this goal, in the training we transform a binary skeleton map to a skeleton probability map by

$$D(x, y) = \frac{1}{1 + \min_{(x', y') \in P} \sqrt{(x - x')^2 + (y - y')^2}} \quad (3)$$

where P is the set of skeleton pixels in the binary skeleton map. Using D as output of the network, the binary classification problem is converted to a regression problem. Accordingly, we need to use a sigmoid function instead of softmax in the last layer of the proposed dense prediction convnet.

In this paper, we propose to use a convnet consisting of three convolutional layers, three max-pooling layers and two fully connected layers. Its specific configuration is summarized in Table 1. For a testing image, let the set of the skeleton pixels detected in Step I be $\hat{P}$ and the skeleton probability map generated by the prediction convnet in this step be D, we prune the low-probability (< 0.5) skeleton pixels in $\hat{P}$ to achieve a refined set of skeleton pixels as

$$P = \{(x, y) | (x, y) \in \hat{P}; \ D(x, y) \geq 0.5\} \quad (4)$$

Sample results of skeleton map after this step of refinement can be found in Figure 4.

Step III: Curve-Structure Segmentation by Recovering Curve Width

In this step, we recover the width of curve structures from the skeleton map derived in Step II, with the help of the scale values derived in Step I. Note that the width of the

Table 1: The configuration of network for Step II, where n, k, s, p stand for the number of outputs, kernel size, stride and padding size respectively.

Type	Configuration
Sigmoid	-
Fully Connected	n:2
Dropout	ratio:0.5
Fully Connected	n:512
MaxPooling	k:2 × 2, s:2
Convolution	n:128, k:3 × 3, s:1, p:1
Batch Normalization	-
MaxPooling	k:2 × 2, s:2
Convolution	n:64, k:3 × 3, s:1, p:1
Batch Normalization	-
MaxPooling	k:2 × 2, s:2
Convolution	n:32, k:3 × 3, s:1, p:1
Input	45 × 45 gray-scale image

curve structures is not a constant and it may vary along the skeleton. Denote the original depth image by I and let P be the set of refined skeleton pixels detected on I after Step II. For each skeleton pixel $(x, y) \in P$, we have a scale value $s(x, y) \in \{1, 2, 3\}$ derived in Step I. We construct the curve-structure segmentation, in the form of a binary map C of the same size as I, using the following algorithm 1.

Algorithm 1 Curve-Structure Segmentation by Recovering Curve Width

Input: Depth image I, Refined skeleton P, Scale values s
Output: Binary segmentation map C

1: Initialize all the elements in C to zero.
2: **for** each skeleton pixel $(x, y) \in P$ **do**
3:     Compute neighborhood:
    $\mathbb{N} = \left\{ (x', y') | \sqrt{(x - x')^2 + (y - y')^2} \leq 2^{s(x,y)} \right\}.$
4:     **for** each pixel $(x', y') \in \mathbb{N}$ **do**
5:         **if** $I(x', y') \geq \frac{I(x,y) + \min_{(x'', y'') \in \mathbb{N}} I(x'', y'')}{2}$ **then**
6:             $C(x', y') = 1$
7:         **end if**
8:     **end for**
9: **end for**

From the steps 3 and 5 of this algorithm, we can see that the curve width at each skeleton pixel is determined by both the scale value s at this pixel and the depth values I at and around this pixel. This algorithm does not require the detected skeleton to be exactly aligned with the center line of the curves – a small dislocation of the skeletons may not change the final segmentation if the dislocated skeletons are still located inside the underlying curves. Sample results after Step III are shown in Figure 4.

Design Matching

One important application of the segmented curve structures in archeology is the task of design matching. In the later
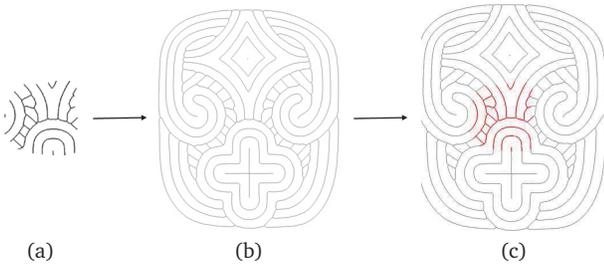
Figure 5: An illustration of design matching. (a) (Thinned) curve structures U segmented on the sherd. (b) (Thinned) full design V. (c) Partial matching between U to V with minimal Chamfer distance. Original design illustration copyrighted by Frankie Snow. Used with permission.

experiments, we will use this task to evaluate the performance of curve-structure segmentation. As shown in Figure 5(c), a design is a full curve pattern of the paddle that are used for stamping the object surface. In the past decades, archaeologists have restored a small number of full designs by manually examining thousands of sherds (Broyles 1968; Snow 1975). The goal of design matching is to identify whether the segmented curve structures are originated from a known design. This is a classical partial matching problem and the key component is the definition of a matching score or distance.

In this paper, we use the classical Chamfer matching (Barrow et al. 1977; Zhou et al. 2017) for this purpose. As shown in Figure 5, we first thin both the segmented curve structures and the considered design into one-pixel wide skeletons and denote them as U and V, respectively. We then transform U to match the design V and compute the Chamfer distance

$$d'_{CM}(U_{\mathbf{T}}, V) = \frac{1}{|U|} \sum_{\mathbf{u} \in U_{\mathbf{T}}} \min_{\mathbf{v} \in V} \|\mathbf{u} - \mathbf{v}\|_2 \qquad (5)$$

where $U_{\mathbf{T}}$ is the curve pattern U after the transform T, $\mathbf{u} \in U_{\mathbf{T}}$ indicates all the skeleton-pixel coordinates $\mathbf{u}$ in the transformed partial pattern $U_{\mathbf{T}}$, and $\mathbf{v} \in V$ indicates all the skeleton-pixel coordinates $\mathbf{v}$ in the curve pattern V. $|U|$ is the total number of skeleton pixels in the partial pattern U. Eq. (5) actually finds the nearest skeleton-pixel coordinate in V for each skeleton-pixel coordinate in $U_{\mathbf{T}}$, records its Euclidean distance $\|\mathbf{u} - \mathbf{v}\|_2$ and finally averages over all the skeleton-pixel coordinates in $U_{\mathbf{T}}$. The matching distance between U and V is then defined by

$$d(U, V) = \min_T d'_{CM}(U_{\mathbf{T}}, V) \qquad (6)$$

with T covers all possible translations and rotations. The scaling transforms is not considered here because both U and V have known actual sizes.

Experiment

In this section, we validate the effectiveness of the proposed method from three perspectives. First, we evaluate the proposed method in terms of the classical metrics of precision, recall and F-measure and compare it against other six comparison methods. Second, we conduct experiment to justify the usefulness of each step in our method. Third, we evaluate the curve-structure segmentation results in the task of design matching.

Dataset

For this study, we collected the depth images of 1,174 pieces of pottery sherds that are excavated in various archaeological sites located in southeastern North America. We used a linear array 3D laser scanner, NextEngine, to get the point cloud of sherd surfaces with the resolution of 100 points per mm^2. Then their depth images are sampled with the same resolution, i.e., each pixel in depth image covers $0.01mm^2$. The average size of the collected depth images is 446×421. We have 530 of these depth images with manually labeled ground-truth curve-structure segmentations. Among all 530 images, we randomly pick 250 for training and the remaining 280 for testing.

To train the FCN in Step I, we thin all the ground-truth curve structures to one-pixel width skeletons, using a standard image thinning algorithm (Lam, Lee, and Suen 1992). Data augmentation is employed here to generate sufficient training data. Specifically, we first split the whole image into small blocks with a size of 100×100. Then these blocks are rotated, scaled and flipped with the same scheme as in (Shen et al. 2016). Finally, 141,696 blocks are used in FCN training in Step I. As for the network training in Step II, we randomly take 44,906 window images with a size of 45×45 around the skeleton pixels identified in Step I for training.

Implementation Details

For the purpose of better training, the parameters of encoders in the skeleton extraction network are initialized with the pre-trained FCN-8s model (Long, Shelhamer, and Darrell 2015). The parameters of decoder are fixed to perform bilinear interpolation (Xie and Tu 2015). The maximum number of training iterations is set as 20,000, with a mini-batch size of 10. The base learning rate is 1×10^{-7} and decays to 1×10^{-8} after 10,000 iterations. Momentum and weight decay are set to 0.9 and 5×10^{-4} respectively.

Because the dense prediction convnet in Step II is relatively lightweight, we choose to train it from scratch. The maximum number of training iterations is set to 100,000, with a mini-batch size of 10. The base learning rate is 1×10^{-3}, and it decays in an inverse way with the parameter $\gamma = 10^{-3}$ and $power = 0.75$. Momentum and weight decay are set to be the same as the FCN in Step I.

F-measure based Segmentation Performance

To evaluate the effectiveness of our method of curve-structure segmentation, we select six widely-used segmentation methods for comparison – Difference of Gaussian (DoG), Level Set (Vese and Chan 2002), GrabCut (Rother, Kolmogorov, and Blake 2004), Fully Convolutional Network (FCN) (Long, Shelhamer, and Darrell 2015), Deep Skeleton (Shen et al. 2016) and DeepLab (Chen et al. 2016). The experiment is conducted on the 280 testing images as
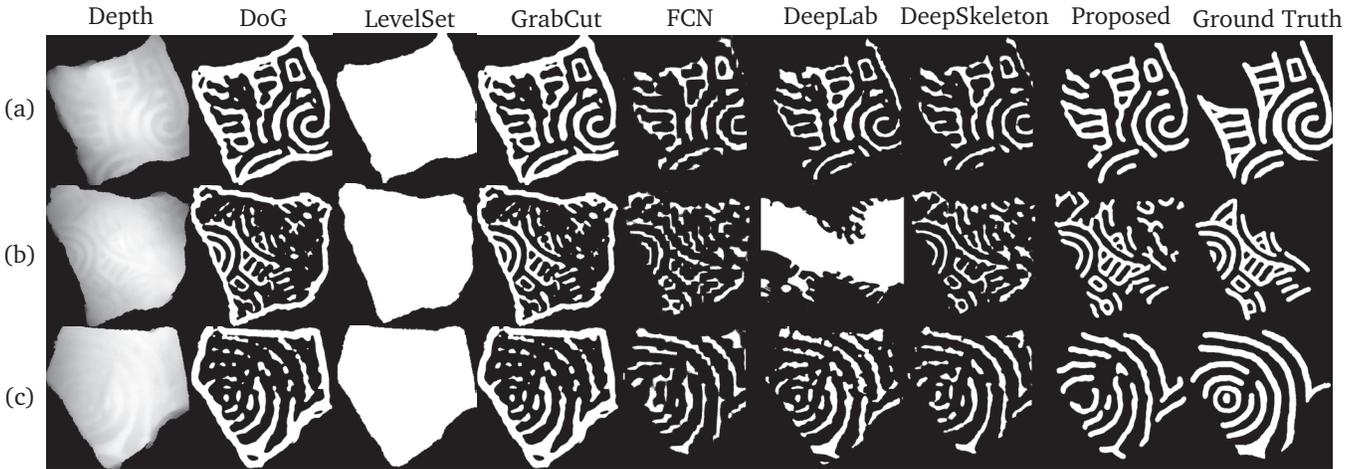
Figure 6: Examples of the curve-structure segmentation result from the proposed method and six comparison methods.

Table 2: Precision, recall and F-measure of the proposed method and six comparison methods, averaged over 280 test images.

Methods	Precision	Recall	F-measure
DoG	0.366	0.774	0.490
LevelSet	0.262	0.938	0.399
GrabCut	0.357	0.671	0.448
FCN	0.589	0.472	0.514
DeepLab	0.585	0.670	0.583
DeepSkeleton	0.634	0.690	0.654
Proposed	0.660	0.827	**0.731**

described above, and the evaluation criteria is the traditional F-measure of $\frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$.

For most of these comparison methods, we keep the default settings in their source codes. But there are several exceptions need to be clarified. Since there is no default setting in DoG, we determine its parameters by trial-and-error. The best performing setting we found is: $k_1 = k_2 = 45$, $\sigma_1 = 11$, $\sigma_2 = 5$, where k and σ are the kernel size and standard deviation of Gaussian filters. The filtered images are transformed to curve maps with the threshold 1. In Grab-Cut, an initialization of the foreground object is required, for which we simply use the DoG result. In Deep Skeleton, we calculated the ground-truth scale maps by applying distance transform on ground-truth segmentation maps. Performance of all methods, averaged over all 280 testing images, are summarized in Table 2.

We can see that the proposed method achieves the best F-measure, and outperforms the second best (Deep Skeleton) by 7.7%. Figure 6 shows the segmentation results on three sample images, using all seven methods. In these images, we can observe that DoG actually enhances the difference between adjacent pixels. As a purely low-level method, it may not capture deep and shallow curves simultaneously. GrabCut was initialized by DoG, but its performance be-

comes even worse. One major reason might be that the data and smoothness energy defined in GrabCut are not sufficiently discriminative to segment the curve structures and non-curve object surface in such a low-contrast image. This is probably the same reason that makes Level Set fail. As expected, the three CNN-based comparison methods, i.e., FCN, Deep Skeleton and DeepLab, normally achieve better performances than the low-level methods. However, their segmentation results usually contain many false positives and the boundaries of the segmented curve structures are quite rough. While the proposed method does not achieve the first place in either precision or recall, it achieves the best performance in final F-measure.

Usefulness of Each Step

Intuitively, the three steps of our method can be replaced by other alternatives or simply ignored. To justify the usefulness of each step, we design three additional experiments, in each of which, we modify or remove one step of the proposed method, and then check its influence to the final segmentation performance.

Modifying Step I: Step I of the proposed method is skeleton extraction. Actually, the FCN we used in this step can be trained to produce curve-structure segmentation directly. However, we choose to extract skeletons first, and then take additional steps to recover the curve width. In this experiment, we make several adjustments in the FCN in Step I to let it output curve structures with width directly. For this purpose, we just use the ground-truth segmentation as the output for training and remove extra upsampling layers in FCN. All the implementation parameters keep unchanged. Sample results of this modified method are shown in Figure 7(b) . We can see that these results contain more false positives and rougher segmentation boundaries. Quantitatively, F-measure of the proposed method decreases from 0.731 to 0.665 if we make this modification to Step I.

Removing Step II: Step II of the proposed method employs a dense prediction convnet as a pixel-wise classifier to refine skeletons extracted by FCN in Step I. To justify its
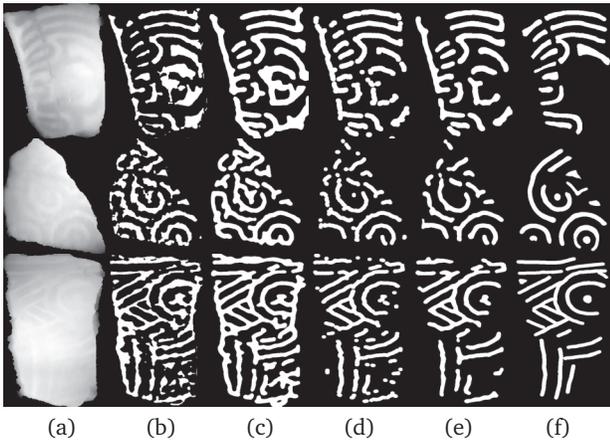
Figure 7: Sample segmentation results of the proposed method with modifications to each step. (a) Input depth image. (b) Segmentation result after modifying Step I. (c) Segmentation result after removing Step II. (d) Segmentation result after modifying Step III. (e) Segmentation result of the proposed method without any modification. (f) Ground-truth segmentation.

usefulness, we remove this step and recover curve width directly from the skeletons generated in Step I. Sample results are shown in Figure 7(c). We can see that the removal of Step II leads to more false positives. Quantitatively, F-measure of the proposed method decreases from 0.731 to 0.662 if we remove Step II.

Modifying Step III: Simple morphological dilation seems to be a very intuitive approach to recover curve width in Step III. In this experiment, we modify Step III by replacing it with a dilation operation with a radius of 15 pixels, which is the best parameter after we try and test all different values. Sample results are shown in Figure 7(d). While the dilation produces very smooth curve structures, they do not align well with the ground truth. Quantitatively, F-measure of the proposed method decreases by 3.5% if we make this modification to Step III.

Design-Matching Performance

In this experiment, we evaluate curve segmentation results in the task of design matching. We take the depth images of 292 sherds with known full designs and in total they come from 29 different designs. The matching distance is the minimal Chamfer distance as defined above.

We use the Cumulative Matching Characteristics (CMC) ranking metric to evaluate the design-matching performance. For each sherd curve-pattern U, we match it against all 29 designs by Chamfer matching. We then sort these 29 designs in terms of the matching distance and pick the top L matching designs with the smallest matching distances. If the ground-truth design of this sherd is among the identified top L designs, we count it as a correct matching under rank L. We repeat this for all 292 sherds and calculate the accuracy, i.e., the percentage of the correctly matched sherds, under each rank L, $L = 1, 2, \cdots, 29$. This way, we can draw a
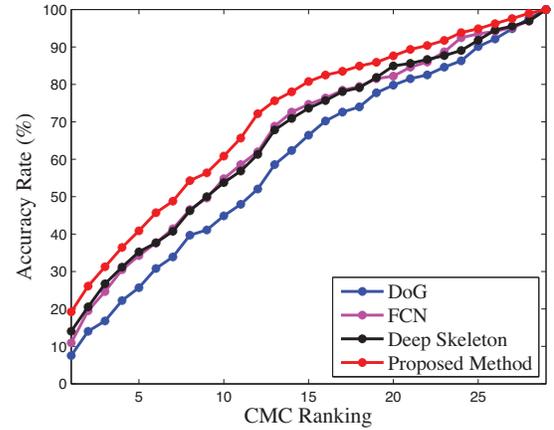

Figure 8: CMC curves of the proposed method and three comparison methods.

CMC curve in terms of rank L as shown in Figure 8, which reflects the performance of curve-structure segmentation – The higher the CMC curve, the better the segmentation performance.

Besides the proposed method, we select three other representative comparison segmentation methods for performance evaluation in this experiment. These three comparison methods are DoG, FCN and Deep Skeleton. Figure 8 shows the CMC curves of the proposed method and these three comparison methods in the task of design matching. The proposed method achieves a CMC rank-1 rate of 20% and a CMC rank-15 rate of 78%, which are much better than the other three comparison methods.

Conclusion

In this paper, we put forward a novel and challenging image segmentation problem: weak curve-structure segmentation from noisy depth images, which has important applications in archeology for exploring large collections of fragmented cultural heritage objects. We developed a new three-step supervised-learning based method to address this problem, by first extracting and refining the skeletons of underlying curve structures and then producing the final segmentation by recovering the curve width at each skeleton pixel. In the experiment, we tested the proposed method on a dataset of depth images scanned from unearthed pottery sherds from southeastern North America. We found that the proposed method performs better than several widely used low-level and deep-learning based image segmentation methods in terms of F-measure.

References

Arbelaez, P.; Maire, M.; Fowlkes, C.; and Malik, J. 2011. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(5):898–916.

Badrinarayanan, V.; Kendall, A.; and Cipolla, R. 2015. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*.

Barrow, H. G.; Tenenbaum, J. M.; Bolles, R. C.; and Wolf, H. C. 1977. Parametric correspondence and chamfer matching: Two new techniques for image matching. Technical report, SRI International Menlo Park Ca Artificial Intelligence Center.

Broyles, B. J. 1968. Reconstructed designs from swift creek complicated stamped sherds. Southeastern Archaeological Conference Bulletin.

Chan, T. F., and Vese, L. A. 2001. Active contours without edges. *IEEE Transactions on Image Processing* 10(2):266–277.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2016. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv preprint arXiv:1606.00915*.

Halir, R. 1999. An automatic estimation of the axis of rotation of fragments of archaeological pottery: A multi-step model-based approach. In *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*.

Kampel, M., and Sablatnig, R. 2007. Rule based system for archaeological pottery classification. *Pattern Recognition Letters* 28(6):740–747.

Lam, L.; Lee, S.-W.; and Suen, C. Y. 1992. Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(9):869–885.

Li, Z., and Chen, J. 2015. Superpixel segmentation using linear spectral clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1356–1363.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Lorigo, L. M.; Faugeras, O. D.; Grimson, W. E. L.; Keriven, R.; Kikinis, R.; Nabavi, A.; and Westin, C.-F. 2001. Curves: Curve evolution for vessel segmentation. *Medical Image Analysis* 5(3):195–206.

Makridis, M., and Daras, P. 2012. Automatic classification of archaeological pottery sherds. *Journal on Computing and Cultural Heritage* 5(4):15.

Rasheed, N. A., and Nordin, M. J. 2015. Archaeological fragments classification based on rgb color and texture features. *Journal of Theoretical & Applied Information Technology* 76(3).

Rother, C.; Kolmogorov, V.; and Blake, A. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics* 23(3):309–314.

Shen, W.; Zhao, K.; Jiang, Y.; Wang, Y.; Zhang, Z.; and Bai, X. 2016. Object skeleton extraction in natural images by fusing scale-associated deep side outputs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 222–230.

Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8):888–905.

Smith, K. Y., and Knight, V. J. 2012. Style in swift creek paddle art. *Southeastern Archaeology* 31(2):143–156.

Smith, P.; Bespalov, D.; Shokoufandeh, A.; and Jeppson, P. 2010. Classification of archaeological ceramic fragments using texture and color descriptors. In *Proceedings of the Computer Vision and Pattern Recognition Workshops*, 49–54.

Snow, F. H. 1975. Swift creek designs and distributions: A south georgia study. *Early Georgia* 3(2):38–59.

Tao, X.; Prince, J. L.; and Davatzikos, C. 2002. Using a statistical shape model to extract sulcal curves on the outer cortex of the human brain. *IEEE Transactions on Medical Imaging* 21(5):513–524.

Tremeau, A., and Borel, N. 1997. A region growing and merging algorithm to color segmentation. *Pattern Recognition* 30(7):1191–1203.

Vese, L. A., and Chan, T. F. 2002. A multiphase level set framework for image segmentation using the mumford and shah model. *International Journal of Computer Vision* 50(3):271–293.

Wang, S., and Siskind, J. M. 2001. Image segmentation with minimum mean cut. In *Proceedings of the IEEE International Conference on Computer Vision*, 517–524.

Wang, S., and Siskind, J. M. 2003. Image segmentation with ratio cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(6):675–690.

Wang, S.; Kubota, T.; and Siskind, J. M. 2004. Salient boundary detection using ratio contour. In *Advances in Neural Information Processing Systems*, 1571–1578.

Xie, S., and Tu, Z. 2015. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 1395–1403.

Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; and Torr, P. H. 2015. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 1529–1537.

Zhou, J.; Yu, H.; Smith, K.; Wilder, C.; Yu, H.; and Wang, S. 2017. Identifying designs from incomplete, fragmented cultural heritage objects by curve-pattern matching. *Journal of Electronic Imaging* 26(1):011022–011022.

Zou, Q.; Cao, Y.; Li, Q.; Mao, Q.; and Wang, S. 2012. Cracktree: Automatic crack detection from pavement images. *Pattern Recognition Letters* 33(3):227–238.