

Received October 20, 2021, accepted November 8, 2021, date of publication November 10, 2021, date of current version November 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3127229

# Deep Learning-Based Engraving Segmentation of 3-D Inscriptions Extracted From the Rough Surface of Ancient Stelae

YE-CHAN CHOI<sup>1</sup>, SHERIFF MURTALA<sup>1</sup>, BEOM-CHAE JEONG,  
AND KANG-SUN CHOI<sup>1</sup>, (Member, IEEE)

Future Convergence Engineering, I.P.C.E., KOREATECH, Cheonan, Chungnam 31253, Republic of Korea

Corresponding author: Kang-Sun Choi (ks.choi@koreatech.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2019R111A3A01060600, and in part by the BK-21 FOUR Program through National Research Foundation of Korea (NRF) under Ministry of Education.

**ABSTRACT** Ancient stelae are considered important historical sources. However, it is a challenge to recognize the inscriptions carved on stelae that have rough surfaces due to prolonged weathering. In this paper, we propose a deep learning-based method to extract engraved regions from the 3D scanned mesh of a stela. First, the uneven distribution of vertices in the mesh is transformed using a mesh subdivision method such that the vertices in the mesh are uniformly distributed. Then, surface features (depth, concave features, and local surface features) are extracted from the subdivided mesh. The depth represents the basic shape of the mesh and is obtained from the aligned mesh. The concave features effectively represent concave regions by using a Frangi filter, and the local surface features have the spin image technique applied to describe the fine shapes of neighboring vertices relative to a vertex. The mesh and the surface features are rasterized into feature images, and engraved regions are segmented from the feature images by using a FC-DenseNet. Our experiments confirm that the proposed method effectively extracts engraved regions of the inscriptions from the rough surface of a stela and it shows robustness to noisy and extremely abraded characters. The proposed method outperformed the second-best method, obtaining an F1 score, IoU, and SIRI of approximately 2.95%, 3.65%, and 7.53%, respectively.

**INDEX TERMS** Cultural heritage, relief extraction, mesh processing, engraving segmentation, image segmentation.

## I. INTRODUCTION

Archeological stelae recording events of the past have important value when it comes to studying political and cultural history. However, stelae have usually been exposed to prolonged physical and chemical weathering from the environments where they are located, making it difficult to recognize the characters engraved on them. Weathering causes damage of various sizes and shapes, such as scratches, cracks, dents, scaling, and spalling. The damaged regions are morphologically similar to engraved regions, and are difficult to distinguish from the strokes of the characters. Moreover, the colors of the surface, such as stains, moss marks, and stone patterns, make it more difficult to recognize character strokes.

The associate editor coordinating the review of this manuscript and approving it for publication was Charalambos Poulis<sup>1</sup>.

The direction of light can be used to detect thin character strokes through shadows, but it often makes them harder to find. To recognize characters, it is necessary to separate external factors such as colors and light from the surface and to effectively remove noise so as to extract only the engraved regions. Note that concave regions consist of not only engraved regions but also noise due to damage.

Rubbing is a traditional method applied directly to the surface of stelae to extract engraved regions [1]. An image is obtained by rubbing ink onto paper placed on the surface of a relic. The areas where the ink and paper touch the surface become black; elsewhere, the color of the paper is maintained, showing results similar to a binary image. Although external factors are excluded, and concave regions are extracted, the results can vary depending on the materials used, the environment, and the proficiency of the person using the

rubbing method. In particular, rubbing causes direct physical and chemical damage to the surface of the relic from ink contamination.

Recently, 3D scanning-based methods have been introduced to solve the problems associated with the rubbing method [2], [3]. The 3D scanning process does not require physical contact with the relic, and as such, the surface of the relic is not damaged, and the features of the surface are preserved, excluding the external factors. The transformed digital data can be easily reproduced and distributed.

The 3D scanning-based methods can be divided into those that focus on enhancing visualization and those for automatic extraction of engraved regions. The methods relating to visualization are shading-, curvature-, and morphology-based. The shading-based methods emphasize thin strokes by dynamically adjusting the light position or by adapting lighting effects to reveal surface details of 3D objects [4]–[6]. The curvature-based methods represent in color the convexity or concavity of the surface [7]–[9]. The modified curvature-based method estimates valleys and ridges of the surface using improved curvatures, and emphasizes the surface through adaptive filtering and colorization [10]. The morphology-based method highlights concave regions by subtracting smoothed mesh with a Laplacian filter [11]. Several methods combine multiple visualizations to improve the detection of engravings [12]–[14]. However, the visualization results capture both engraved regions and noise, so archaeological knowledge is required to analyze and interpret the results.

The methods automatically extracting engraved regions without subjective evaluation are as follows. The valley/ridges-based methods directly extract valleys and ridges from the surface [15]–[17]. The depth estimation-based relief extraction (DRE) method [18] estimates the relative depth of each vertex by using normal vectors of the virtual base surface, and then separates engraved regions with a depth threshold determined from using an expression maximization (EM) algorithm [19]. The curvature-based methods [20], [21] estimate engraved regions by detecting canal-like shapes with a Frangi filter [22], which was introduced to identify blood vessels using curvature from a medical image. However, since rough surfaces do not follow a Gaussian distribution, and since these methods lack the ability to distinguish between engraved regions and noise, a lot of noise is also extracted in the results.

Machine learning-based methods were proposed to distinguish engraved regions from noise. Texture-based methods are widely used [23]–[25]. The 2D texture-based method segments an input image containing multiple textures given a patch of a reference texture [24]. The 3D texture-based method [25] extracts texture features on local 3D surfaces, and classifies engraved regions using a support vector machine (SVM) [26]. In the deviation map-based method, a 3D mesh is projected into a 2D image to reduce the computational complexity and generate an enhanced deviation map [27]. Then, pixel-wise segmentation is performed

using a random forest model trained using the local surface patches [28]. The segment-based method involves the initial extraction of concave segments from the 3D mesh [29]. Features including local extrema, cross-sectional, and appearance information are extracted from the segments. The engravings are identified from the classification results obtained with an SVM model. The machine learning-based methods showed higher performance, compared to rule-based methods, but have limitations in that only local surface information is mainly used. In other words, they cannot utilize a global context, mainly using the local context for classification of engraved regions and noise.

Despite the previous work, it is still a challenge to separate only engraved regions from rough surfaces, except for noise. It is difficult to extract engraved regions from rough surfaces of stelae because the reference surface is not flat and contains a lot of noise. Although engraved regions can be separated from noise by considering the local surface characteristics, noise has local characteristics similar to the engraved regions. Noise appears on the surface of the stela as a result of prolonged weathering over a long period of time, just like the engraved regions. In contrast to the noise, the regions of engraved characters were originally carved on the surface of the stela. The strokes of the characters are short, shallow, and thin, and easily confused with, or interpreted as, noise. Thus, it is necessary to extract not only local shape features of engraved regions but also the global context for classification.

In this paper, we propose a method to segment the engraved regions of 3D inscriptions using deep learning. First, 3D scanned data are preprocessed by mesh subdivision to evenly distribute the position of vertices in the mesh. Then, surface features consisting of depth, concave features, and local surface features are extracted. The mesh and the surface features are rasterized into feature images, and the engraved regions are segmented using a FC-DenseNet. We select surface features that effectively represent the surface and utilize both local and global contexts via the FC-DenseNet to classify pixel-wise engraved regions.

The contributions of this paper are as follows.

- 1) We propose a method of segmenting the characters on a very rough 3D mesh with high-resolution by applying CNN-based 2D segmentation while preserving the fine features.
- 2) The proposed method shows the highest performance, both objectively and subjectively, when compared to conventional methods.
- 3) Multiple surface features are combined instead of using a single surface feature showing the better performance.
- 4) The proposed method utilizes both local and global contexts, unlike the approaches of conventional methods, to classify engraved regions achieving the higher performance.
- 5) The proposed method not only shows the best performance, but also performed well on extremely abraded

characters. The experiments show the proposed method is more robust to noise than conventional methods.

This paper is organized as follows. We review related studies in Section 2. Details of the proposed method are described in Section 3. We present evaluations of our results in Section 4, and conclude the paper in Section 5.

## II. RELATED WORKS

In this section, we briefly introduce the studies relevant to the proposed method. For clarity, these previous studies are divided into subsections relating to the Frangi filter [22], the spin image [30], and FC-DenseNet [31].

### A. FRANGI FILTERS

The curvature-based relief extraction (CRE) method applies a Frangi filter to extract engraved regions from relics [20]. CRE exploits the advantages of the Frangi filter by using the principal curvature to extract canal-like concave regions from a 3D mesh. The concavity,  $C$ , of vertex  $v_i$  is given as

$$C_i = \begin{cases} \exp(-\frac{\Gamma^2}{2\alpha^2})(1 - \exp(-\frac{\Pi^2}{2\beta^2})), & k_1 \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where  $k_1$  and  $k_2$  ( $|k_1| \geq |k_2|$ ) are the maximum and minimum principal curvatures at a vertex. The vertex index,  $i$ , is omitted for clarity. The parameters  $\Gamma$  and  $\Pi$  represent the ratio of the principal curvature  $|k_2/k_1|$  and the magnitude  $\sqrt{k_1^2 + k_2^2}$ , respectively. When  $k_1 \geq 0$  (concave), the first term indicates how similar the local surface is to a canal by using a ratio of the principal curvatures. The second term represents the depth of the local surface using a magnitude of the principal curvature. The principal curvatures calculation requires the computation of the second derivatives of the surfaces. In the proposed methods, the principal curvatures are obtained using Rusinkiewicz's method [9] that utilizes tensor representation.

However, CRE is not as effective at extracting engravings with coarse surfaces as it is in extracting engravings of flat murals. In particular, the principal curvature is obtained in quadratic differential form, which is weak against noise, resulting in very messy boundaries of the extracted regions.

The modified curvature-based relief extraction (MCRE) method applies Gaussian smoothing to reduce the noise of the principal curvature and increase the performance of the Frangi filter [21]. MCRE extracts engraved regions by applying a threshold value according to obtained concavity  $C$ . Subsequently, the dual curvature-based relief extraction (DCRE) method extracts only deeply engraved regions by using two Frangi filters with different parameters [21]. MCRE and DCRE can effectively extract concave regions. Nevertheless, the methods are less accurate in the detection and removal of noise, and there are difficulties in adjusting the parameters according to the degree of weathering on each characters.

The segment-based relief extraction (SRE) method is employed with MCRE to obtain segments [29]. Features consisting of appearance-based, cross section-based, and local

extrema-based characteristics are extracted for each segment. Then, SRE classifies each segment into either an engraved region or noise by using an SVM, and it does not need to adjust the parameters depending on the degree of weathering [26]. The result obtained by SRE is more accurate, compared to other methods.

### B. SPIN IMAGES

A spin image is a 3D shape descriptor that is invariant to rotation and translation [30]. The descriptor was mainly introduced for object recognition, surface matching, and facial feature point detection [32], [33]. We apply the spin image as a 2D histogram to represent the distribution of peripheral vertices at each vertex. The projection function,  $S_i(j)$ , that has position  $(\rho, \gamma)$  with respect to the neighboring vertex  $v_j$  from a vertex  $v_i$ , is as follows:

$$S_i(j) \rightarrow (\rho, \gamma), \quad (2)$$

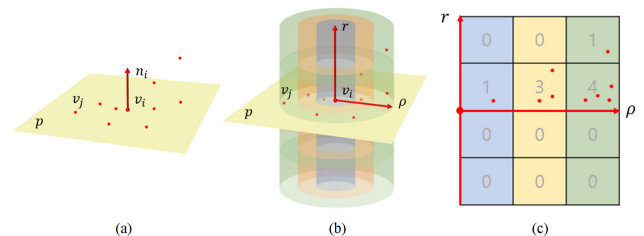
$$\rho = \sqrt{\|\mathbf{v}_j - \mathbf{v}_i\|^2 - (\mathbf{n}_i \cdot (\mathbf{v}_j - \mathbf{v}_i))^2}, \quad (3)$$

$$\gamma = \mathbf{n}_i \cdot (\mathbf{v}_j - \mathbf{v}_i), \quad (4)$$

where  $\mathbf{n}_i$  is the normal vector of  $v_i$ . The neighborhood vertices are mapped relative to reference vertex  $v_i$ .

Then, spin image  $S_i$  is obtained as a 2D histogram of  $S_i(j)$  with a cylindrical subspace having the number of rows,  $n_\gamma$ , and columns,  $n_\rho$ , in the  $\gamma$  and  $\rho$  directions, respectively. The bin sizes of the 2D histograms are denoted as  $\Delta\gamma$  and  $\Delta\rho$ .

Fig. 1 shows a representation of a spin image. A space with a reference vertex and neighboring vertices is illustrated in Fig. 1 (a). The geometric structure of the spin image is shown in Fig. 1 (b) where  $n_\gamma = 4$ ,  $n_\rho = 3$ , and the bin sizes,  $\Delta\gamma$  and  $\Delta\rho$ , are arbitrary. The generated spin image is a  $4 \times 3$  (h  $\times$  w) histogram, as shown in Fig. 1 (c). Thus, the spin image captures the distribution of vertices located within a cylindrical subspace with radius of  $n_\rho$  and a height of  $n_\gamma$ .



**FIGURE 1.** A representation of the spin image: (a) a space with reference vertex  $v_i$  and neighboring vertices  $v_j$ , (b) the geometric structures of cylindrical subspaces, and (c) the generated spin image as a 2D histogram.

### C. FC-DenseNet

Image segmentation is considered an essential element in scene understanding, and is a topic being addressed significantly in computer vision and machine learning [34], [35]. Image segmentation means classifying images on a pixel-wise basis, and it is being researched and utilized in various

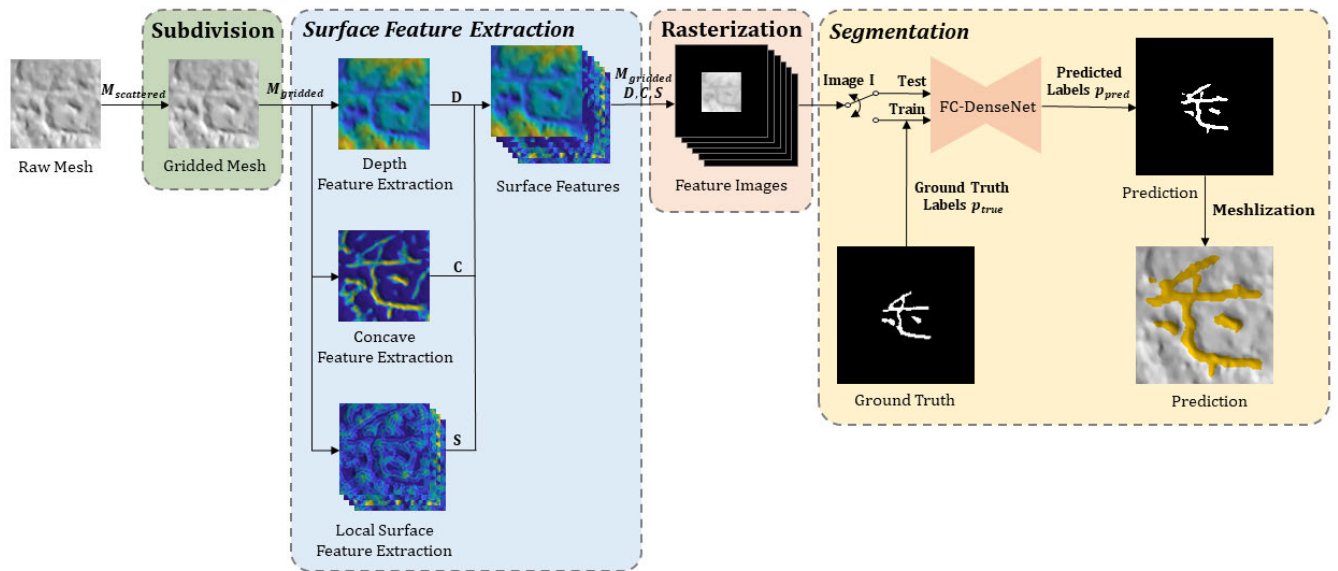


FIGURE 2. An overview of the proposed method.

environments, such as face recognition, medical imaging, and autonomous driving. In pixel-wise classification, the global context is very important because the local context alone is not sufficient.

Recently, the advances in deep learning-based convolutional neural networks (CNNs) have achieved high performance in semantic segmentation applications. After the development of the fully convolutional network [36], many models based on encoder-decoder structures were proposed, such as DeconvNet [37], SegNet [38], and U-Net [39]. These models extract features using the encoder structure, and restore them to high-resolution images through the decoder structure.

In particular, U-Net was proposed for cell image segmentation in medical images, and achieved high performance despite being trained on a small data size [39]. Bezmaternykh *et al.* applied U-Net to segment characters from historical documents [40]. Their research has characteristics similar to the subject of this paper because the characters were extracted from ancient materials that contain a lot of noise due to weathering. Although the data available for the study were not sufficient owing to a limited number of ancient artifacts, U-Net had acceptable results from character segmentation.

When the number of layers in a CNN is increased, performance can be increased. However, information from previous layers can be lost due to vanishing gradients, which occurs when networks are deep. To solve this problem, DenseNet was presented [41]. DenseNet includes a dense block architecture in which feature maps of previous layers are concatenated with feature maps in the next layers. Through this structure, a dense block preserves information with a strong gradient flow. Since the features of each layer are connected and preserved, DenseNet has fewer channels, reducing the

number of parameters. And the dense block has the effect of regularization, and it reduces overfitting.

FC-DenseNet combines U-Net with DenseNet, providing a deep network for image segmentation [31]. FC-DenseNet has very deep networks but a small number of parameters. The performance of FC-DenseNet is promising in the extraction of local and global contexts and has the advantages of both U-Net and DenseNet.

### III. THE PROPOSED METHOD

In this section, we describe the proposed method for segmentation of engravings by using deep learning. An overview of the proposed method is shown in Fig. 2. The process is divided into mesh subdivision, surface feature extraction, rasterization, and image segmentation. First, we perform mesh subdivision that transforms the vertices in the mesh from nonuniform to uniform distribution. We then extract features consisting of depth, and concave and local features to characterize the surface. Thereafter, the rasterization method is applied to convert the mesh and features to feature images. Finally, the engraved regions are segmented and identified using FC-DenseNet.

The 3D scanned data are expressed as a polygon mesh,  $M = \{V, E, F\}$ . Set  $V$  contains the list of vertices representing the position of the 3D surface, and it is expressed as  $V = \{v_i : 1 \leq i \leq n_V\}$ , where  $v_i = (x_i, y_i, z_i) \in \mathbb{R}^3$ .  $V$  is determined by sampling the surface of a real 3D target through a 3D scan. Set  $E$  consists of a list of edges connecting two vertices.  $E$  is expressed as  $E = \{e_k : 1 \leq k \leq n_E\}$ , where  $e_k = (i_1^k, i_2^k)$ . The set containing all triangulated faces connecting three edges is expressed as  $F = \{f_k : 1 \leq k \leq n_F\}$ , where  $f_k = (i_1^k, i_2^k, i_3^k)$ . The notations,  $n_V$ ,  $n_E$ , and  $n_F$  are the numbers of vertices, edges, and faces, respectively, in the mesh.

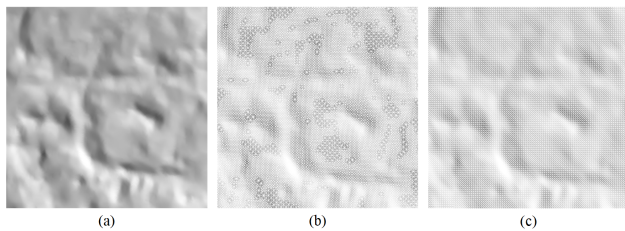


The region of interests (ROIs) where the characters can exist are manually selected. The process of the proposed method works based on the vertices in the ROIs.

**A. MESH SUBDIVISION**

The 3D scanned mesh,  $M_{scattered}$ , has uneven distributions of vertices. During the 3D scanning process, the vertices of well-scanned areas are dense, and the vertices of poorly scanned areas are sparse. The unequal distribution of vertices is not suitable for feature extraction. In particular, for a histogram, a spin image is highly dependent on the number of vertices. Thus, mesh subdivision is essential in order to equalize the distribution of the vertices so they have even distribution at any vertex. To transform the uneven distribution of vertices in  $M_{scattered}$  into a mesh with evenly distributed vertices ( $M_{gridded}$ ), linear triangular interpolation [42] is applied. After the faces are regenerated,  $M_{gridded}$  is grid-shaped, and the vertices are uniformly distributed. This preprocessing is done before the extraction of features.

Fig. 3 illustrates an example of mesh subdivision. The surface of a raw mesh and the corresponding mesh with the positions of vertices are shown in Fig. 3 (a) and Fig. 3 (b), respectively. Observe the uneven distribution of vertices in the poorly scanned areas of the mesh in Fig. 3 (b). The mesh in Fig. 3 (c) is the output of the preprocessed mesh after mesh subdivision. The sparse areas in the mesh are removed, and the vertices are uniformly distributed, thereby resulting in the same mesh resolution across the surface of the stela.



**FIGURE 3.** An example of mesh subdivision: (a) the raw mesh, (b) visualization of edges with the white color from (a), and (c) the gridded mesh after mesh subdivision of (b).

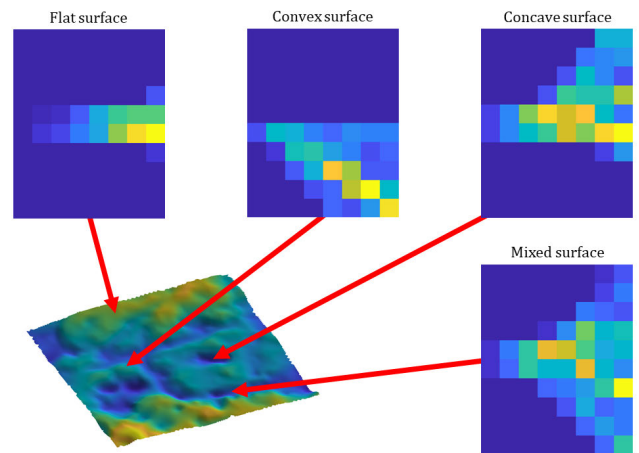
**B. SURFACE FEATURE EXTRACTION**

1) DEPTH

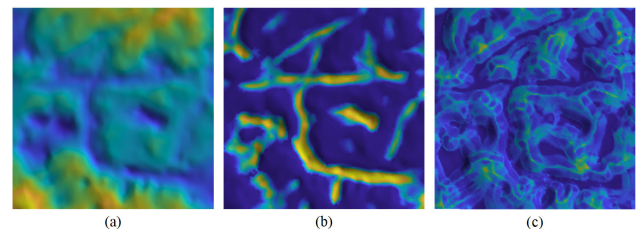
The depth of the mesh represents the basic surface feature of the stela. We use principal component analysis for the vertices in the ROI of a character to remove the slope of the z-axis of mesh  $M_{gridded}$ , and we approximate the surface into a plane. The z-value of each vertex in the aligned mesh is used as the depth, which is normalized as

$$D_i = s \cdot (z_i - z_{min}), \tag{5}$$

where  $s$  preserves the scale for the depth, and normalizes the depth into the range  $[0, D_{max}]$ . Since depth deviation is different for each character, scaling is adapted.



**FIGURE 4.** Characteristics of a spin image according to surface shape.



**FIGURE 5.** An example of feature extraction: (a) the depth  $D$ , (b) concave feature  $C$ , and (c) the local surface feature  $S$  for a channel.

However, certain characters are located on the outer boundary of the stela, and the depth decreases rapidly along the outside boundary, making the value of  $D_{max}$  large. Therefore, the following expression is applied to limit the depth:

$$D_i = \max(0, D_i + (1 - D_{max})). \tag{6}$$

Equation (6) holds the maximum value to 1, and values below 0 are clipped to adjust the distribution into the range  $[0, 1]$ .

2) CONCAVE FEATURES

We extract all concave regions of the mesh using MCRE [21]. The Frangi filter effectively extracts concave regions using the principal curvature. However, since the principal curvature is obtained in quadratic differential form, it is vulnerable to noise, and the process requires denoising. Gaussian smoothing is applied to the mesh to address this problem, but the proper parameters must be applied because a high value for standard deviation from Gaussian smoothing removes engraved regions. After Gaussian smoothing, the principal curvatures of the mesh are obtained using Rusinkiewicz’s method [9]. Then, concave feature  $C_i$  of vertex  $v_i$  is obtained using (1) and has a value in the range  $[0, 1]$ .

3) LOCAL SURFACE FEATURES

To extract a local surface feature, we apply the spin image because of its effectiveness in describing the surface features

of 3D objects [30]. The local surface feature denoted as  $S_i$  is obtained for every vertex  $\mathbf{v}_i$ . Spin images generated on the surface of a mesh are illustrated in Fig. 4. Depending on the shape of the local surface, spin images are classified into four main forms. For a vertex located on a flat surface, a straight spin image is obtained, while for a vertex located on a convex surface, a downward spin image is generated, and for a vertex in concave areas, an upward spin image is acquired. The last form is for a vertex located in mixed surface areas, so a mixed-form spin image is obtained.

However, the number of vertices in the spin image shell increases as the area covered by the spin image expands. Therefore, if the variables have different sizes, a process for fitting the units is necessary. The spin image is normalized for the outside shell volume of the cylinder using the following expression:

$$S_i(x, y) = \frac{S_i(x, y)}{2\pi n_\gamma \Delta\gamma (2x - 1)(\Delta\rho)^2}, \quad (7)$$

where  $x \in \{1, 2, 3, \dots, n_\rho\}$  and  $y \in \{1, 2, 3, \dots, n_\gamma\}$ .

As discussed previously, the spin image is a 2D histogram, so we apply flattening to represent it with feature vectors for each vertex. We apply min-max normalization for all  $S_i$  to adjust values into the range [0, 1]:

$$S_i^{n_\gamma \times n_\rho} \rightarrow S_i^{1 \times n_\gamma n_\rho}, \quad (8)$$

$$S_i = \frac{S_i - S_{min}}{S_{max} - S_{min}}. \quad (9)$$

The spin image bins of the mesh are uniquely divided into  $n_\gamma n_\rho$  channels such that each channel of  $S_i$  has different types of local surface image representation for the mesh.

Representations of the three surface features extracted from the raw mesh in Fig. 3 (a) are illustrated in Fig. 5 (a) to Fig. 5 (c). It is important to note that a single channel representing the image of the local surface feature is shown in Fig. 5 (c).

### C. RASTERIZATION

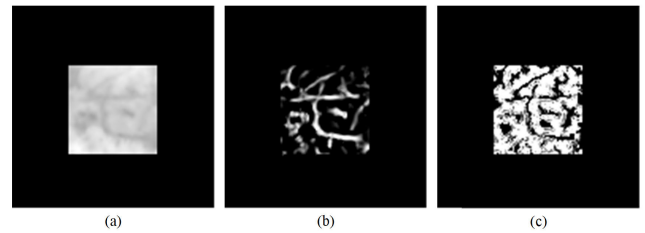
The mesh and the three corresponding features extracted from the mesh are transformed into feature images  $I$  such that the engraved regions are extracted using FC-DenseNet. The position of vertex  $\mathbf{v}_i$  of the mesh is mapped to the pixel position. Concurrently, the surface features are also mapped to the pixel intensity:

$$I_{x,y} \leftarrow (x_i, y_i), \quad (10)$$

$$I_{intensity} \leftarrow D_i, C_i, S_i. \quad (11)$$

When the resolution of the mesh is maintained, the computation cost increases due to the large number of pixels. However, if mesh resolution is reduced, fine features cannot be extracted. Therefore, feature images are generated by applying subsampling to reduce the size of the input images for FC-DenseNet while preserving the resolution of the features.

In addition, the characters in an ancient stela have different sizes, allowing multiple characters to be located in the kernel



**FIGURE 6.** Generated feature images with zero padding (which forces only one character per image): (a) depth feature image, (b) concave feature image, and (c) local surface feature image for a channel.

of the CNN. This leads to learning imbalances, because small characters can be learned in large numbers. Another problem that needs to be addressed is characters that are located on the boundary of the kernel that truncates the character regions. Consideration of the global context is significant in removing the noise that shares features similar to the engraved regions. The training of truncated characters degrades the performance of FC-DenseNet because the global context is difficult to extract. We experimentally confirmed that segmentation results are better than when a single character exists in the CNN’s kernel without truncation. Therefore, we apply zero-padding to the outer region of the ROI so that only one character exists in the kernel. Fig. 6 shows the result of the rasterization process on a sample mesh.

### D. IMAGE SEGMENTATION

Engraved regions are segmented by using FC-DenseNet in the image domain. The engraving segmentation of the mesh domain demands large amounts of memory and incurs computational costs due to the large number of vertices in the mesh. It is difficult to extract engravings directly in the mesh domain due to the large number of vertices in the mesh. A mesh patch of a character contains about 70k vertices, while the majority of existing 3D segmentation methods work on small point clouds such as 4,096 point clouds [43]. Furthermore, sufficient training datasets are not available because the number of ancient artifacts is limited. To address the computation complexity, we convert the mesh to image dimensions similar to [27], and use FC-DenseNet [31] for image segmentation. FC-DenseNet shows excellent performance with the extraction of local and global contexts, and has the advantages of both U-Net and DenseNet.

The structure of the modified version of FC-DenseNet that we used in this study is shown in Fig. 7. The FC-DenseNet consists of the encoder structure used for contracting paths, and the decoder structure for extending paths. Details of the FC-DenseNet structure are presented in Fig. 8. The input first passes through the  $3 \times 3$  convolution layer. Then, encoding is performed on the input data through the dense block and transition down. And decoding is applied via the dense block and transition up. In the contracting path, all feature maps of the dense block are linked to the feature maps of the next stages, but in the expanding path, only the output channels

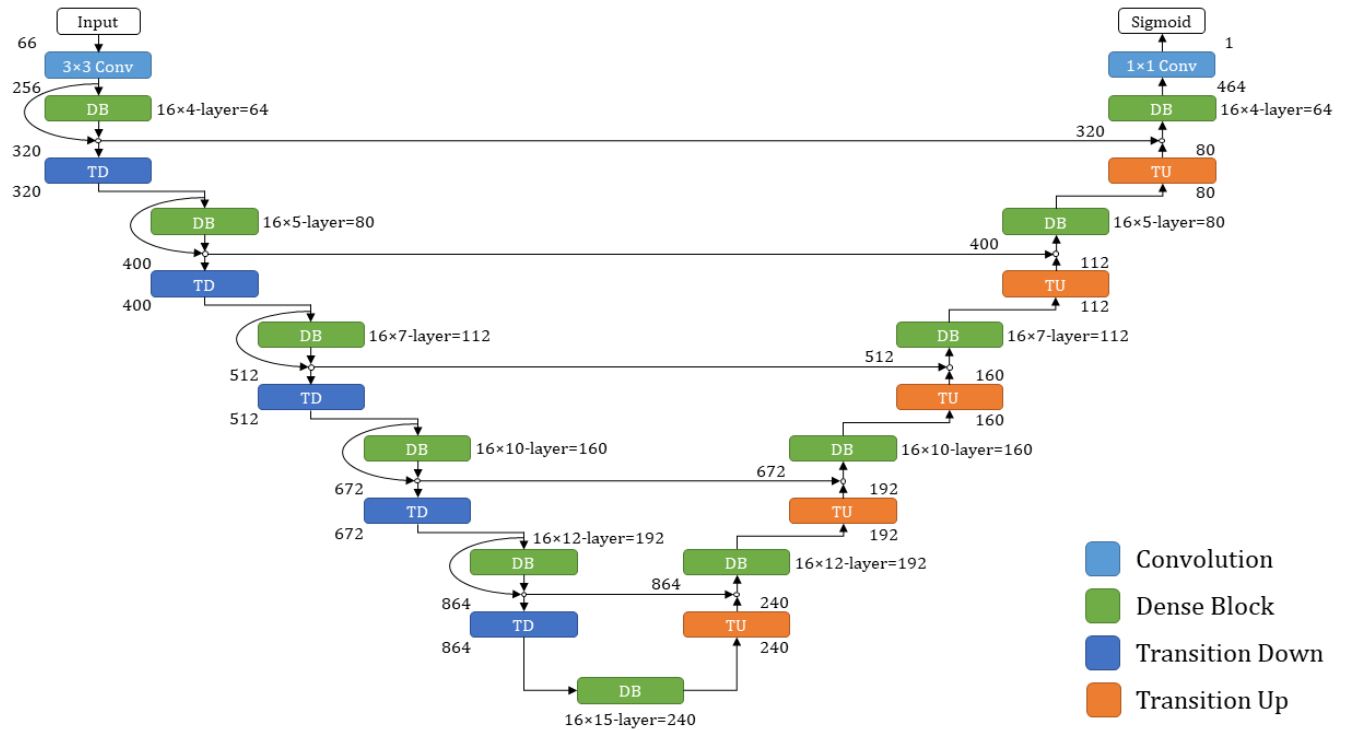


FIGURE 7. The architecture of the modified FC-DenseNet.

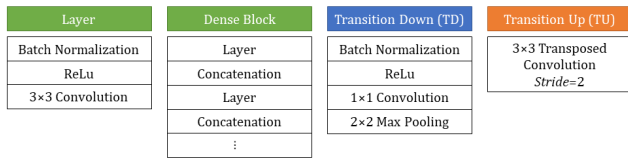


FIGURE 8. The details of the FC-DenseNet blocks.

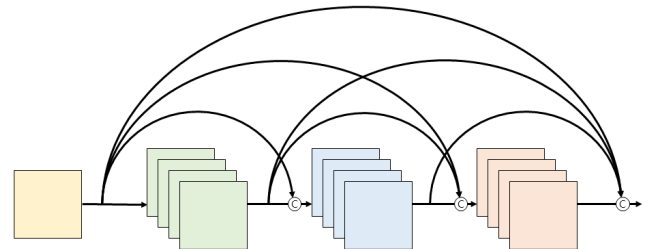


FIGURE 9. An example of a dense block.

of the dense block are linked to the feature maps of the next stages. After processing the last  $1 \times 1$  convolution layer, the sigmoid function is adapted to segment the engraved regions for pixel-wise classification.

The architecture of the dense block is presented in Fig. 9. The number of channels for each convolution in the dense block is referred to as *growth rate* parameter. Since the feature maps of a dense block are concatenated with each result of each convolution, the number of channels in the feature map increases in the form of an equivalent sequence through the *growth rate*. The *growth rate* of the dense block is 16.

Transition down reduces the resolution through max pooling, and transition up increases the resolution through transposed convolutions with a *stride* of 2 while maintaining the same number of channels.

We removed the dropout layers, and initiated a convolution of 256 channels owing to the large number of channels in our input images, unlike the original model.

DICE loss is applied during training with the proposed method [44]. The loss function is expressed as

$$DICE\ Loss = 1 - \frac{2\sum p_{true}p_{pred}}{\sum p_{true} + \sum p_{pred} + \epsilon}, \quad (12)$$

where  $p_{true}$  is ground-truth labels,  $p_{pred}$  is predicted labels, and  $\epsilon$  is a very small value to prevent division by zero. Generally, the character image has a small proportion of character regions, compared to background regions, so there are many true negatives (TNs). The loss is suitable for segmenting the characters because it does not consider TNs. The similarity between the regions for  $p_{true}$  and  $p_{pred}$  is well represented by the loss, and when the loss decreases, the result of the segmentation becomes closer to  $p_{true}$ .

After prediction, The  $p_{pred}$  are converted into meshes using the linear triangular interpolation for visualization [42].

## IV. EXPERIMENTS

### A. DATASET

In this paper, we automatically extract only the character regions (except for noise) from the stela called Musul-ojakbi, a Korean treasure (No. 516) that was created in the year of “Musul” (578 AD) during the Silla dynasty. The stela was built to commemorate construction of a reservoir, and records





FIGURE 10. The Musul-ojakbi stela and its rough surface.

information that includes the construction period, the location and size of the reservoir, the number of workers, the names and appointments of the construction managers, and the name of the inscripitor. The information was written in Chinese in the Idu script. Musul-ojakbi is regarded as a highly valuable source in various archeological fields, such as Korean history, facility history, and Idu script history during the Silla period. However, despite its historical value, it is difficult to read the characters due to prolonged weathering.

The reference surface of Musul-ojakbi was not polished. Therefore, the characters were carved on the unflattened surface of natural stone. As a result, some strokes were even carved upon noise. Musul-ojakbi and its very rough surface is shown in Fig. 10. Moreover, the strokes of the characters are short, shallow, and thin. In the 3D scanned data from Musul-ojakbi, the dimensions of the front area, excluding the sides and back of the stela, are  $982.4 \times 650.8 \times 24.3$  mm (H  $\times$  W  $\times$  D). The total number of characters is estimated at 169, in which the average character size is  $30.6 \times 29.1 \times 2.0$  mm. However, engraved strokes have an average width of 1.97 mm, and an average depth of 0.21 mm. It is noteworthy that the average depth of the strokes is 0.21 mm, while the depths of the characters and the entire mesh are 2.0 mm and 24.3 mm, respectively.

Musul-ojakbi contains a lot of noise due to weathering over a long period of time. The noise regions have features

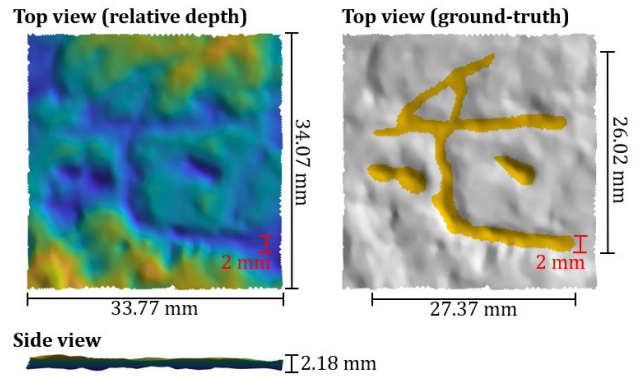


FIGURE 11. A 3D scan of a character from Musul-ojakbi. The left mesh shows the raw mesh from colorizing the depth. It shows many dents and cracks. The right image is ground truth. Noise crosses engraved regions, making boundaries of the engraved regions ambiguous.

TABLE 1. Categorization of the quality of the 169 characters on Musul-ojakbi based on human perception.

	High	Medium	Low	Bad	Totals
Used	32	54	28	20	134
Excepted	5	15	8	7	35
Totals	37	69	36	27	169

similar to the engraved regions due to the weathering, making it difficult to segment the engraved regions. The noise sometimes crosses engraved regions and covers strokes entirely. An example of a character in Musul-ojakbi is Fig. 11. It contains many dents and cracks, which makes it difficult to distinguish the engraved region. In addition, a lot of noise connects engraved regions, so the boundaries of inscriptions are often ambiguous. Therefore, a global context is required to separate the engraved regions from the noise.

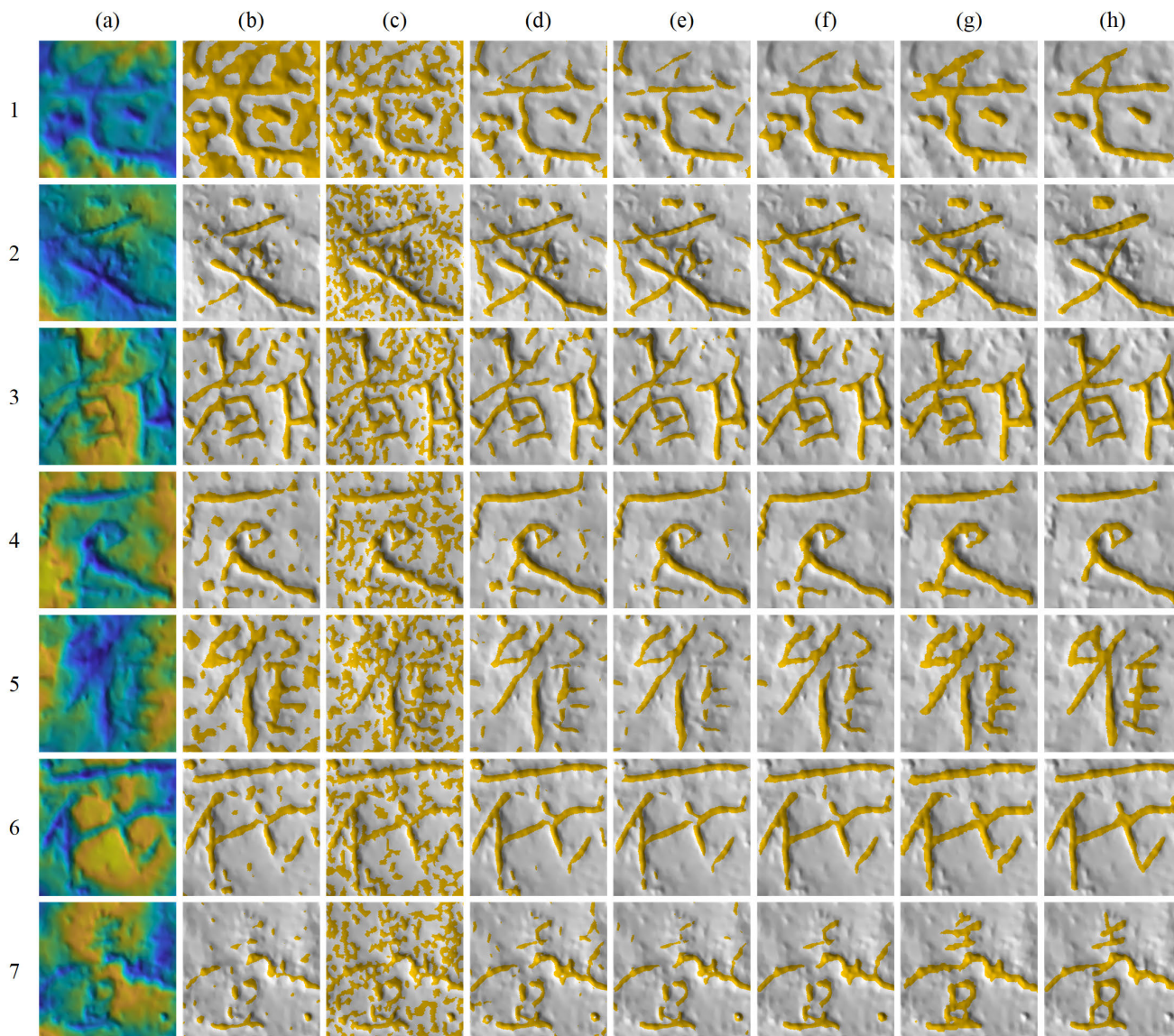
### B. ENVIRONMENTS

Consideration for the centerline of a stroke is more important for recognition of a character than the boundaries of the stroke, which is similar to a character font, provided the centerline is not broken. Furthermore, the importance of the outer boundaries is low, because the boundaries of engraved regions are ambiguous due to noise and abrasions.

In the experiments, we used a 3D scanned mesh of Musul-ojakbi. The mesh consists of approximately 23 million vertices with an average resolution of 0.25 mm.

We analyzed and classified the possible characters originally engraved on the mesh based on weathering. The number of characters in each classification is presented in Table 1. The characters were classified based on human observation. We classified the characters into 4 classes according to the surface state with the guidance of an archeologist. The estimated number of characters on the stela is 169. These characters were put into four classifications: high, medium, low, and bad. In particular, bad characters are difficult for users with no experience in epigraphy to recognize without guidance from an expert. Since the distinction of engraved regions is





**FIGURE 12.** A subjective comparison of methods: (a) input meshes, (b) DRE [18], (c) CRE [20], (d) MCRE [21], (e) DCRE [21], (f) SRE [29], (g) the proposed method, and (h) ground truth.

ambiguous due to prolonged weathering, labeling of ground truth (GT) is subjective. To reduce bias in subjectivity while generating GT, one person produced GT for all 169 characters with the guidance of the archeologist.

For evaluation, 134 characters from the total number were selected. The remaining 35 characters were not included. The characters have overlapping ROIs, making it difficult to use for learning and evaluation. For the qualitative evaluation, we excluded the characters of bad quality from 134 characters, because the extremely damaged characters deteriorate the performance of the model. For the noise robustness evaluation, we used 134 characters including bad quality.

Since the data size is small, we applied fivefold validation for investigating the reliability of the proposed method. The dataset was randomly divided into five subsets, while the

characters of each quality were included in each subset at the same rate. The four subsets were used as the training set, and the other subset was used as the validation to tune the hyperparameters of the model. The process was repeated five times, leading to five results for each model. The scores of each model were averaged to identify the results through a cross validation.

The hyperparameters used in this experiment are as follows. Since the average resolution of the mesh is 0.25 mm, and the resolution of dense areas of the mesh is more than the average, a lower value is required in the mesh subdivision, so 0.2 mm was selected for the grid resolution. The scale  $s = 1/6$  was adapted to obtain depth. To obtain concave characteristics, the values  $\sigma = 1.0$ ,  $\alpha = 0.6$ , and  $\beta = 0.1$  were selected. For local surface features, we obtained

**TABLE 2.** The quantitative results from fivefold validation, compared to conventional methods.

	Accuracy	Precision	Recall	F1	IoU	SIRI [45]
DRE [18]	0.9538	0.5189	0.7287	0.5367	0.3779	0.3030
CRE [20]	0.9244	0.2820	0.8084	0.4142	0.2633	0.2048
MCRE [21]	0.9791	0.6619	0.7808	0.7090	0.5561	0.4670
DCRE [21]	0.9817	0.7381	0.7026	0.7112	0.5597	0.4681
SRE [29]	0.9862	<b>0.8164</b>	0.7762	0.7867	0.6586	0.5681
Ours: D+S+C	<b>0.9875</b>	0.8070	<b>0.8322</b>	<b>0.8162</b>	<b>0.6951</b>	<b>0.6434</b>
Ours: D (Depth)	0.9866	0.7956	0.8215	0.8032	0.6766	0.6286
Ours: S (Local surface features)	0.9871	0.8032	0.8232	0.8096	0.6856	0.6300
Ours: C (Concave features)	0.9874	0.8040	<b>0.8346</b>	0.8152	0.6931	0.6354

$10 \times 8$  spin images by using  $\Delta\gamma = 0.15$ ,  $n_\gamma = 10$ ,  $\Delta\rho = 0.35$ , and  $n_\rho = 8$ . The mesh and features were then converted to  $128 \times 128$  images via rasterization, which were then subsampled to 0.6 mm. The size of a depth feature image is  $128 \times 128 \times 1$ , the concave feature image is  $128 \times 128 \times 1$ , and the local surface feature image is  $128 \times 128 \times 80$ . However, certain channels in the local surface feature images are described as zero matrices, and they were removed. Consequently, the size of the local surface feature was reduced to  $128 \times 128 \times 64$ . The size of all the surface feature images is  $128 \times 128 \times 66$ . Since the proposed method is scale-dependent, we applied only data augmentation for rotation and translation. The weight initialization method in FC-DenseNet is HeUniform. The model was trained with  $LR = 0.0001$ , with a mini-batch size of 8 using the Adam optimizer.

The performance of the proposed method was compared with engravings obtained with DRE [18], CRE [20], MCRE [21], DCRE [21], and SRE [29]. However, these conventional methods operate in a mesh dimension, making it difficult to directly apply a quantitative evaluation owing to the different resolutions. We performed a quantitative comparison against the proposed method by transforming the estimated engraving results of each method into image dimensions. The boundary accuracy of a character has low importance when recognizing the character, so evaluation of the image dimension with a low resolution can be more accurate for comparison.

### C. THE QUALITATIVE EVALUATION

The segmentation results from engravings of seven characters selected from the Musul-ajakbi stela are shown in Fig. 12. Fig. 12 (a) visualizes the roughness of the surfaces by applying color maps to the depth differences. DRE shows unsatisfactory results in Fig. 12 (b). DRE estimates the relative depth of each vertex, and then separates the engraved regions using the EM algorithm, which assumes a Gaussian mixture model. However, due to weathering, the depth deviation is so severe that the depth distribution does not follow a Gaussian distribution. Therefore, DRE showed unacceptable estimated engraving from the EM algorithm for certain characters with a high or low threshold. CRE was significantly affected by noise in the principal curvatures. With CRE, the engravings were extracted along with

**TABLE 3.** The F1 score comparison from engraving segmentation based on the categories for surface weathering.

	High	Medium	Low	Bad
DRE [18]	0.5788	0.5434	0.4759	0.4424
CRE [20]	0.4264	0.4048	0.4182	0.3971
MCRE [21]	0.7700	0.6906	0.6749	0.6256
DCRE [21]	0.7754	0.6964	0.6667	0.6081
SRE [29]	0.8653	0.7718	0.7254	0.6490
Our method	<b>0.8744</b>	<b>0.8104</b>	<b>0.7611</b>	<b>0.7010</b>

noise, as shown in Fig. 12 (c). The engravings identified with MCRE, shown in Fig. 12 (d), were more accurate than those detected with CRE, and the improvement is associated with Gaussian smoothing in MCRE to effectively reduce the noise of the principal curvature. The results from DCRE, shown in Fig. 12 (e), provided only deeper concave regions by using the two Frangi filters. There is considerable reduction in the noise from DCRE compared to MCRE. However, the drawback in DCRE is reduction of the stroke thickness, which resulted in failure to detect shallow or thin strokes. SRE classifies engraved regions using SVM, and results are illustrated in Fig. 12 (f). SRE effectively eliminated more noise than MCRE. However, since it classifies engraved regions based on candidate segments, it showed unsatisfactory results when the candidate segments were not well obtained. In SRE, some noise was connected to the engraved regions, and there were misclassified regions because the local context was primarily used for classification. The results from the proposed method are presented in Fig. 12 (g), and are most similar to GT in Fig. 12 (h). Comparing the results of the proposed method with SRE, the incorrectly detected engraved regions of SRE were accurately classified. The proposed method is more reliable than MCRE, DCRE, CRE, and DRE. Therefore, the visualization results show that the proposed method using both local and global contexts performed better than SRE and other existing methods.

### D. THE QUANTITATIVE EVALUATION

The character regions of the rasterized feature images account for 3.4% of the pixels on average in the  $128 \times 128$  images. This occurs because the size of the characters varies greatly. This basically results in a very high accuracy that considers TNs. Though higher accuracy is better, it is not the appropriate metric for a proper comparison. A higher precision



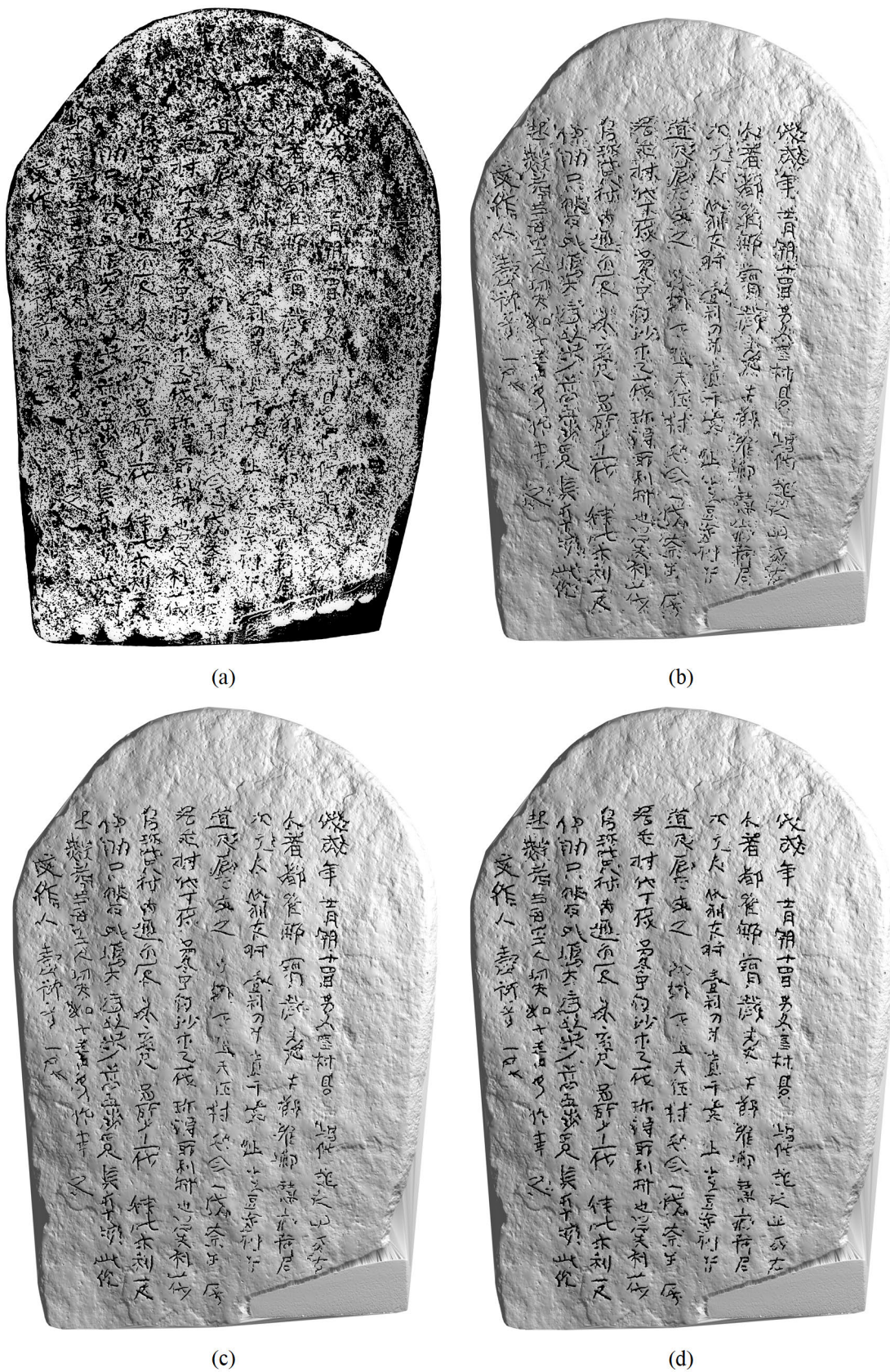


FIGURE 13. Comparison of the engraving segmentation results for the entire Musul-ogakbi stela: (a) rubbing with inverted colors, (b) MCRE [21], (c) SRE [29], and (d) the proposed method.

indicates that the detection of noise in the background was reduced, and a higher recall means that segmentation results of the engravings are acceptable. If the scores for precision and recall are simultaneously high without bias on either side of precision or recall, the backgrounds are less noisy, and the engravings are well extracted, making the character more recognizable. It is important that neither precision nor recall is biased in character recognition. The F1 score is a harmonic mean of precision and recall, as well as a good evaluation indicator. Intersection over union (IoU) represents the degree to which the regions overlap as a ratio of the intersection region and the union region of the predicted regions and the true regions.

We also considered the segmented inscription recognition index (SIRI) [45], a metric that quantifies the subjective recognition score from character recognition. Conventional SIRI divides engraved regions into inner and outer regions [46]. Noise is divided into regions close to and far from the engraved regions. Breaking the centerlines of strokes has a great impact on character recognition, but conventional SIRI simply divides the engraved regions into inner and outer regions, and assigns the same weight to the same region. To address this problem, generalized SIRI [45] applies the level set function to the engraved regions by assigning high weight values to the center of the stroke and decreasing the weight values as the centerlines extend outwards. The equations of generalized SIRI are as follows:

$$FN_w = \sum w_{FN} \log(|levelset(FN)|), \quad (13)$$

$$FP_w = \sum w_{FP} FP, \quad (14)$$

$$SIRI = \frac{2\sum TP}{2\sum TP + FN_w + FP_w}. \quad (15)$$

Thus, the generalized SIRI assigns stiff penalties for breaking the strokes and character boundaries. The parameters  $w_{FN}$  and  $w_{FP}$  were 18.4 and 2.7, respectively.

The quantitative results from fivefold validation are presented in Table 2. The proposed method outperformed other methods in accuracy, recall, F1 score, IoU, and SIRI [45] at 0.0013, 0.0238, 0.0295, 0.0365, and 0.0753 respectively. The highest accuracy of the proposed method proves that the engraving segmentation is excellent than the conventional methods. In the case of precision, the proposed method shows the similar score to SRE. However, since our data is a skewed data, precision does not change even if the strokes of the characters are thinned or disappear. The recall of the proposed method shows the highest score, indicating that the strokes are well extracted like the GTs. The best SIRI of the proposed method means that the engraving segmentation shows the highest subjective recognition. IoU indicates that the predicted regions and the GT regions are also highly overlapped, and SIRI indicates that the results show the highest subjective results. Therefore, although the engraving segmentation of the proposed method shows that the detection of noises in the backgrounds is similar to the second-best method, the strokes of the characters are well extracted like the GTs. The

proposed method shows objectively and subjectively the best performance compared to the conventional methods.

The scores of each feature were evaluated. Each single feature outperformed SRE in all metrics except precision. The result of the single depth feature shows that the proposed deep learning-based method using both local and global contexts performed better than SRE using the local context only. This indicates that the global context must also be used to extract engraved regions of inscriptions with coarse surfaces.

### E. ROBUSTNESS TO NOISE

The F1 score comparisons from engraving segmentation according to the degree of abrasion are presented in Table 3. The proposed method showed the highest performance on all states of abrasion. Although the proposed method and SRE produced similar scores for high-quality characters, the performance of the proposed method was significantly higher for the other states, regardless of the degree of abrasion. In particular, for engraving segmentation in the bad state, we used a model of the fivefold validation that was trained using only characters in the high, medium, and low states. The bad characters included characters that are difficult for an ordinary person to recognize. Nevertheless, the performance of the proposed method with characters in a bad state was acceptable. Therefore, the proposed method quantitatively and qualitatively outperformed the existing methods. The satisfactory performance of the proposed method is linked to its robustness to noise.

### F. THE PREDICTED ENGRAVED REGIONS FOR THE ENTIRE STELA

The engraving segmentation for all 169 characters on Musul-ojakbi is illustrated in Fig. 13. Compared to the other methods, engraving prediction from the proposed method showed the best performance. The proposed method achieved very good performance compared to rubbing (the most common investigation method for archeology). The proposed method extracted the engraved regions accurately, with less influence from noise, compared to MCRE. And the proposed method showed the best results by classifying the engraved regions misclassified by SRE into engraved regions.

### V. CONCLUSION

In this paper, we proposed a deep learning-based approach to the extraction of engravings from a 3D scanned mesh of a severely weathered stela. The deep learning model was trained using a basic depth feature, a concave feature, and a local surface feature. Frangi filters and the spin image were employed in the extraction of concave features and local surface features, respectively. The local and global contexts of the features were considered in FC-DenseNet for the segmentations. Through experiments, we confirmed that each subfeature of the proposed features is effective in describing the surface of the stela. The proposed method outperformed



conventional methods objectively and subjectively. F1 score, IoU, and SIRI values from the proposed method were 2.95%, 3.65%, and 7.53%, respectively. Furthermore, the proposed method showed robustness to noise and achieved an acceptable F1 score of approximately 5.2%—higher than the second-best method with extremely abraded characters.

## REFERENCES

- [1] R. Pan, Z. Tang, and W. Da, "Digital stone rubbing from 3D models," *J. Cultural Heritage*, vol. 37, pp. 192–198, May 2019.
- [2] K. Liu and J. Ma, "3D-scanning and computer reverse engineering technology to preserve inscriptions at Beihai park," *Int. J. Simul., Syst.*, vol. 17, no. 26, pp. 38.1–38.7, Jan. 2016.
- [3] A. Abdelhafiz, "Integrating digital photogrammetry and terrestrial laser scanning," Ph.D. dissertation, Inst. Geodesy Photogramm., Tech. Univ. Braunschweig, Braunschweig, Germany, Feb. 2009.
- [4] S. Rusinkiewicz, M. Burns, and D. DeCarlo, "Exaggerated shading for depicting shape and detail," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1199–1205, Jul. 2006.
- [5] R. Vergne, R. Pacanowski, P. Barla, X. Granier, and C. Schlick, "Radiance scaling for versatile surface enhancement," in *Proc. Symp. Interact. 3D Graph.*, Feb. 2010, pp. 143–150.
- [6] L. Ammann, P. Barla, G. Guennebaud, X. Granier, and P. Reuter, "Surface relief analysis for illustrative shading," *Comput. Graph. Forum*, vol. 31, no. 4, pp. 1481–1490, Jun. 2012.
- [7] H. T. Ho and D. Gibbins, "Curvature-based approach for multi-scale feature extraction from 3D meshes and unstructured point clouds," *IET Comput. Vis.*, vol. 3, no. 4, pp. 201–212, Jan. 2009.
- [8] N. Mellado, G. Guennebaud, P. Barla, P. Reuter, and C. Schlick, "Growing least squares for the analysis of manifolds in scale-space," *Comput. Graph. Forum*, vol. 31, no. 5, pp. 1691–1701, Aug. 2012.
- [9] S. Rusinkiewicz, "Estimating curvatures and their derivatives on triangle meshes," in *Proc. Symp. 3D Data Process., Vis., Transmiss.*, Sep. 2004, pp. 486–493.
- [10] M. Kolomenkin, I. Shimshoni, and A. Tal, "Prominent field for shape processing and analysis of archaeological artifacts," *Int. J. Comput. Vis.*, vol. 94, no. 1, pp. 89–100, Aug. 2011.
- [11] H. Pires, L. Gonçalves-Seco, J. Fonte, M. J. C. Santos, and O. Sousa, "Morphological residual model: A tool for enhancing epigraphic readings of highly eroded surfaces," in *Information Technologies for Epigraphy and Cultural Heritage*. Paris, France: Sapienza Univ. Editrice, Jan. 2014, pp. 133–144.
- [12] A. Gilboa, A. Tal, I. Shimshoni, and M. Kolomenkin, "Computer-based, automatic recording and illustration of complex archaeological artifacts," *J. Archaeol. Sci.*, vol. 40, no. 2, pp. 1329–1339, Feb. 2013.
- [13] B. Ridel, P. Reuter, J. Laviolle, N. Mellado, N. Couture, and X. Granier, "The revealing flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts," *J. Comput. Cultural Heritage*, vol. 7, no. 2, pp. 6.1–6.18, Jul. 2014.
- [14] A. Tal, "3D shape analysis for archaeology," in *3D Research Challenges in Cultural Heritage*. Berlin, Germany: Springer, Jan. 2014, pp. 50–63.
- [15] M. Kolomenkin, I. Shimshoni, and A. Tal, "Demarcating curves for shape illustration," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 157.1–157.9, Dec. 2008.
- [16] M. Kolomenkin, I. Shimshoni, and A. Tal, "On edge detection on surfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2767–2774.
- [17] Y. Ohtake, A. Belyaev, and H.-P. Seidel, "Ridge-valley lines on meshes via implicit surface fitting," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 609–612, Aug. 2004.
- [18] R. Zatzarinni, A. Tal, and A. Shamir, "Relief analysis and extraction," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 136.1–136.9, Dec. 2009.
- [19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
- [20] K. Lawonn, E. Trostmann, B. Preim, and K. Hildebrandt, "Visualization and extraction of carvings for heritage conservation," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 801–810, Jan. 2017.
- [21] E.-J. Heo, J.-M. Sa, and K.-S. Choi, "Relief extraction from stone monuments," *IEIE Trans. Smart Process. Comput.*, vol. 7, no. 4, pp. 321–324, Aug. 2018.
- [22] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98*, vol. 1496. Berlin, Germany: Springer, Oct. 1998, pp. 130–137.
- [23] L. Armi and S. F. Ershad, "Texture image analysis and texture classification methods—A review," Apr. 2019, *arXiv:1904.06554*.
- [24] I. Ustyuzhaninov, C. Michaelis, W. Brendel, and M. Bethge, "One-shot texture segmentation," Jul. 2018, *arXiv:1807.02654*.
- [25] S. Liu, R. R. Martin, F. C. Langbein, and P. L. Rosin, "Segmenting geometric reliefs from textured background surfaces," *Comput.-Aided Des. Appl.*, vol. 4, no. 5, pp. 565–583, Jan. 2007.
- [26] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [27] M. Zeppezauer, M. S. Georg Poier, C. Reinbacher, S. Schuster, C. Breiteneder, and H. Bischof, "Interactive 3D segmentation of rock-art by enhanced depth maps and gradient preserving regularization," *J. Comput. Cultural Heritage*, vol. 9, no. 4, pp. 19.1–19.30, Sep. 2016.
- [28] L. Breiman, "Random forest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Jan. 2001.
- [29] Y.-C. Choi, S. Murtala, B.-C. Jeong, and K.-S. Choi, "Relief extraction from a rough stele surface using SVM-based relief segment selection," *IEEE Access*, vol. 9, pp. 4973–4982, 2021.
- [30] A. E. Johnson, "Spin-images: A representation for 3D surface matching," Ph.D. dissertation, Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Aug. 1997.
- [31] S. Jégou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 1175–1183.
- [32] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.
- [33] K.-S. Choi and D.-H. Kim, "Angular-partitioned spin image descriptor for robust 3D facial landmark detection," *Electron. Lett.*, vol. 49, no. 23, pp. 1454–1455, Nov. 2013.
- [34] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," Nov. 2020, *arXiv:2001.05566*.
- [35] I. Ulku and E. Akagunduz, "A survey on deep learning-based architectures for semantic segmentation on 2D images," May 2021, *arXiv:1912.10230*.
- [36] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
- [37] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.
- [38] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, vol. 9351. Cham, Switzerland: Springer, Oct. 2014, pp. 234–241.
- [40] P. V. Bezmaternyykh, D. A. Ilin, and D. P. Nikolaev, "U-Net-Bin: Hacking the document image binarization contest," *Comput. Opt.*, vol. 43, no. 5, pp. 825–832, Oct. 2019.
- [41] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.
- [42] I. Amidror, "Scattered data interpolation methods for electronic imaging systems: A survey," *Proc. SPIE*, vol. 11, no. 2, p. 157–176, Apr. 2002.
- [43] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3D point clouds: A survey," Jun. 2020, *arXiv:1912.12033*.
- [44] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, Sep. 2016, pp. 179–187.
- [45] B.-C. Jeong, Y.-C. Choi, S. Murtala, and K.-S. Choi, "A study on the generalization method of segmented inscription recognition index to improve 3-D inscription recognition quality," *J. Inst. Electron. Inf. Eng.*, vol. 58, no. 6, pp. 79–86, Jun. 2021.
- [46] B.-C. Jeong, Y.-C. Choi, S. Murtala, and K.-S. Choi, "Segmented inscription recognition index for measuring the inscription quality extracted from 3-D data," *IEIE Trans. Smart Process. Comput.*, vol. 9, no. 6, pp. 445–451, Dec. 2020.



**YE-CHAN CHOI** received the B.S. degree in electrical engineering from KOREATECH, South Korea, where he is currently pursuing the M.S. degree in Interdisciplinary Program in creative engineering. His research interests include geometry processing, image segmentation, text segmentation, and machine learning.



**BEOM-CHAE JEONG** received the B.S. degree in electrical engineering from Kongju National University, South Korea, in 2019. He is currently pursuing the M.S. degree in the Interdisciplinary Program in creative engineering with KOREATECH. His research interests include image segmentation, text extraction, and 3D data processing.



computer vision, machine learning, MIMO systems, spatial modulation, and social network analysis.

**SHERIFF MURTALA** received the B.Eng. degree in electrical engineering from the University of Ilorin, Nigeria, in 2010, and the M.Eng. degree in communication engineering from the Federal University of Technology, Minna, Nigeria, in 2017. He is currently pursuing the Ph.D. degree in information and communication engineering with KOREATECH, South Korea. He worked as a Technical Officer at Salbodi Group Ltd., Nigeria, from 2012 to 2014. His research interests include



**KANG-SUN CHOI** (Member, IEEE) received the B.S. degree in electronic engineering from Korea University, in 1997, the M.S. degree, in 1999, and the Ph.D. degree in nonlinear filter design, in 2003. In 2011, he joined the School of Electrical, Electronics and Communication Engineering, Korea University of Technology and Education, where he is currently an Associate Professor. From 2008 to 2010, he was a Research Professor with the Department of Electronic Engineering, Korea University. From 2005 to 2008, he worked at Samsung Electronics, South Korea, as a Senior Software Engineer. From 2003 to 2005, he was a Visiting Scholar with the University of Southern California. His research interests include multimedia compression, video processing, and computational photography. He was a recipient of an IEEE International Conference on Consumer Electronics Special Merit Award, in 2012.

...