IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 January 2023, accepted 25 January 2023, date of publication 27 January 2023, date of current version 1 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3240313

RESEARCH ARTICLE

Noise-Against Skeleton Extraction Framework and Application on Hand Gesture Recognition

JUN MA[1,2], XUNHUAN REN[2], HAO LI[2], WENZU LI[1,3], VIKTAR YUREVICH TSVIATKOU[2], AND ANATOLIY ANTONOVICH BORISKEVICH[2]

[1]CETC China Electronic Technology LLC, 220043 Minsk, Belarus
[2]Department of Infocommunication Technologies, Belarusian State University of Informatics and Radioelectronics, 220013 Minsk, Belarus
[3]Department of Intelligent Information Technologies, Belarusian State University of Informatics and Radioelectronics, 220013 Minsk, Belarus

Corresponding author: Xunhuan Ren (rxh1549417024@gmail.com)

This work was supported in part by the China Scholarship Council.

ABSTRACT Extracting stable skeletons from noisy images is a challenging problem since the skeletonization method is prone to be affected by inner and border noise. Although many methods have been proposed in the past for increasing the antinoise ability of skeletonization methods, most of them either only overcome border noise or, at the cost of lost topology, degrade the effects of two noises. In this paper, we propose a skeleton extraction framework to enhance the robustness of the existing skeletonization method against both inner and border noise. In our approach, we first use the different scales of Gaussian filters to smooth the input image and obtain multiple representations. Then, binarization and skeletonization were performed to produce a series of binary images and a series of skeletal images. Next, we use our measure on these binary and skeletal images to find the most suitable skeleton. Since our measure considers both the skeleton image changes and binary image changes caused by using a filter, the selected skeleton is sufficiently robust and has all the necessary skeletal branches. The inner noise experiment and border noise experiment are conducted for comparison. From the perspective of the measure of the rate of variation in the skeleton, the proposed framework can reduce the inner noise by approximately 92% and the border noise by approximately 40%. In addition, the experiment on static hand gesture recognition has demonstrated that the introduction of our framework can increase up to 11% mean recognition accuracy.

INDEX TERMS Gaussian filter, noise against, skeleton extraction framework, static hand gesture recognition.

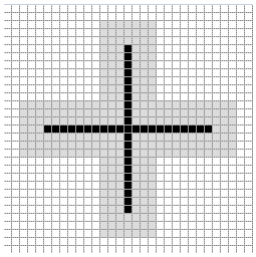
I. INTRODUCTION

Skeletons are popular descriptors since they can preserve the original topology and connectivity of an object [1] in an image. They are widely applied in many fields, such as hand gesture recognition [2], human action recognition [3], image matching [4], hepatic vascular analysis and vessel segmentation [5], sketch-based modeling [6], human character animation [7] and quantitative structure imaging [8].

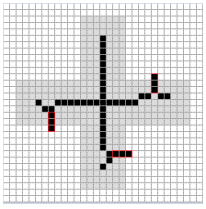
Skeletal images can be either obtained directly from the depth image captured by a depth camera such as Kinect or extracted by using the traditional skeletonization method from an image captured by a regular camera. For approaches based on a depth camera, the generated skeleton may not be affected by the lighting, shade, and color; therefore, recognition based on it tends to have a better result. However, the cost, size, and availability of the depth camera limits its use [9]. In contrast, the traditional method has a broader range of applications since it requires only a regular camera. For the traditional method, it is necessary to convert RGB images into grayscale images, followed by binarization to extract the region of interest (ROI) as a foreground object. Then, skeletons can be extracted by applying the skeletonization method on binary images. One of the challenges of the traditional method is that the produced skeleton is not stable due to the existence of noise.

The associate editor coordinating the review of this manuscript and approving it for publication was Hossein Rahmani.

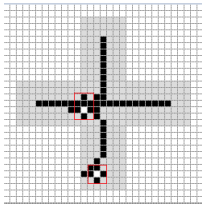
This noise can be observed in the binary image and may significantly influence the resulting skeleton. From the perspective of binary images, noise can be divided into two types: border noise and inner noise. Border noises are caused by tiny changes in the edge of the foreground, and this type of noise may produce many unwanted branches in the skeletal image. Inner noises appear on the inside of the foreground, and this type of noise may create many unnecessary skeletal rings. Both of these noises and their influences on the skeleton are shown in Fig. 1. In Fig. 1, the gray region is the foreground in the binary image, and the black region is the skeleton extracted by skeletonization. For visualization, the effects of the noise on the skeleton are marked with red rectangles.



(a) Skeleton extracted from a clean binary image



(b) Skeleton extracted from a binary image with border noise



(c) Skeleton extracted from a binary image with inner noise

FIGURE 1. Example of border noise and inner noise in a binary image and their effects on the skeleton.

Many different algorithms have been proposed in recent decades to obtain a more stable skeleton when using the traditional method. There are three different approaches to addressing the problem caused by noise: the skeletonization-based approach, the pruning-based approach, and the scale-space-based approach. The skeletonization-based approach concentrates on the improved robustness of the skeletonization method with respect to noise, such as [10], [11], [12], [13], [14], and [15]. The merit of this kind of method is that it can reduce the effects of border noise and does not require extra operations, which has a lower computational cost. The drawback is that they fail to alleviate the effects caused by inner noise. Pruning-based approaches, such as [16], [17], [18], [19], [20], [21], [22], and [23], tend to introduce the need for postprocessing after skeletonization to remove insignificant or unwanted branches. Almost all pruning-based algorithms are based on a salience measurement of the skeleton branches or their corresponding contour. Then, they remove the skeleton branches whose salience value measures less than a given threshold, which usually requires manual tuning. The merit of this kind of method is that it can significantly offset the problem caused by border noise, even when the extent of the border noise is significant. However, similar to the skeleton-based method, the pruning-based method is still unable to degrade the influence of inner noise. Scale-space-based approaches [24], [25], [26] adopt filters to smooth the image and remove noise.

The advantage of this approach is that it can deal with both inner and border noise. The defect of these methods is that they require an adequately set filter parameter. The reason is that when the filter parameter is small, the ability to filter noise is limited. When the filter parameter is large, the original geometrical and topotypical features may not be preserved.

In this paper, we propose a noise-against-skeletonization extraction framework, which can make the resulting skeleton more stable by degrading the influence caused by noise while retaining necessary skeleton branches. Our method first generates multiple skeletons for images according to different sets of filter parameters. Then, our framework automatically selects the most suitable filter parameter and its corresponding skeleton by using our proposed measure.

The main contributions of this paper are summarized as follows:

- We proposed a skeletonization extraction framework in which a novel measure based on both skeleton information and region information is used to select a suitable representation of the skeletons to strengthen the robustness of the existing skeletonization method.
- The robustness of the proposed framework against inner and border noise was proven during the artificial noise experiment. In the experiment, we also noticed that the proposed framework does not suffer from the distortion problem.
- We applied the proposed framework in the task of static gesture hand recognition and proved that the use of the proposed framework can help to improve recognition accuracy.

The remainder of this paper is organized as follows. Section II presents a review of several well-known denoising methods. Section III describes our framework. Inner and border noise experiments are detailed in Section IV. Section V presents a comparison of the results for static hand gesture recognition when using the skeleton extracted by our method and those of other methods. Section VI concludes this paper. Finally, limitations and future work are presented in Section VII.

II. RELATED WORKS

As we mentioned in the former section, there are three different denoising techniques, and in this section, we review several popular or recent methods for each technique.

A. SKELETONIZATION-BASED APPROACH

Skeletonization methods can be further divided into three basic types [1]: Voronoi-based methods, continuous curve propagation approaches, and digital approaches. Skeletonization-based denoising methods are skeletonization

algorithms with the ability to suppress border noise. Most of them belong to Voronoi-based [12], [14] and digital approaches [9], [13].

Durix et al. [13] recently proposed one-step compact skeletonization, an antinoise skeletonization Voronoi-based method. This method directly computes a simple skeleton with a few branches by propagating selected Voronoi circles within the shape while discarding propagation directions that designate negligible information. They demonstrated that their method could produce a clean skeleton and avoid creating branches caused by noise.

Among digital denoising skeletonization approaches, thinning algorithms have attracted much attention since these methods tend to use devised templates or criteria to extract skeletons, which are easy to modify to improve denoising. For example, in 1993, Shih and Wong [14] proposed an efficient thinning algorithm against border noise by applying 69 group templates. In recent years, Ma et al. [10] proposed a fully parallel skeletonization algorithm against border noise, which requires 13 group templates.

In addition to the two types of antiborder-noise skeletonization approaches mentioned above, Yang et al. [12] proposed a new kind of method based on skeleton grafting. Their approach, inspired by tree grafting, generates a skeleton in a coarse-to-fine fashion.

Compared with normal skeletonization methods, noise-against-skeletonization methods are more robust but can deal only with border noise. They are still sensitive to inside noise.

B. PRUNING-BASED APPROACH

Pruning-based approaches are postprocessing approaches that are applied after the skeleton is obtained. These methods have an outstanding ability to remove unwanted branches.

One of the most cited pruning algorithms was proposed by Bai et al. [16]. That method is based on contour partitioning with discrete curve evolution. They linked every skeleton point to a boundary point that is tangential to its maximal circle and then deleted all skeleton points whose corresponding boundary point lies on the same contour segment.

Shen proposed a pruning method with a bending potential ratio [17] in which the pruning of the skeletal branches depends on the context of the boundary segment that corresponds to the branch.

A pruning method based on information fusion was proposed by Liu et al. [18]. They considered redundant branch length, region reconstruction, and local salience degree to determine the pruning of a branch.

Many other pruning methods have been proposed. Among them, Andres's method [23], Stelios's method [19], and Siyu's method [21] are also fascinating pruning methods. In addition, some pruning methods have been proposed along with skeletonization for some specified tasks in recent years, such as methods described in [27] and in [28].

The advantage of the pruning-based approach is that it can completely remove unwanted branches caused by border noise, even when the amount of border noise is significant. However, similar to the skeletonization-based method, they also fail to suppress the inner noise.

C. SCALE-SPACE-BASED APPROACH

The scale-space-based methods adopt Gaussian filters to smooth the image and to remove noise. They are promising methods since they can suppress inner and border noise.

A scale-space method was proposed by Hoffman and Wong [24] for thinning binary and grayscale images. First, they extracted the union of the topographical features, which consists of the peak, the ridge, and the saddle point, by applying the Gaussian filter, whose scale keeps increasing, to the original image. Then, these topographical features formed the skeleton.

Cai proposed a method [25] that can decrease the effects of noise by using oriented Gaussian filters, which help determine principal directions and help classify ridges, valleys, and edges. Their method is robust to interference from different types of noise. Their strategy is applied to handwriting and fingerprint image enhancement.

Houssem proposed an adaptive framework [26] that uses scale-space filtering to make thinning algorithms robust against noise. In their framework, multiple skeletons are first generated within various filtering scales, followed by using their proposed measure to select a suitable skeleton. Experiments have demonstrated that their method can yield good results in sketch images. However, their approach may break the original topology and connectivity since their evaluation measure considered only the skeleton information.

III. PROPOSED FRAMEWORK

The proposed framework is an improved version of Houssem's framework, which also first uses scale-space filtering to generate multiple representations of input images within a considerable scale of filtering. Then, our modified measure is used to select the most suitable skeleton. In our modified measure, we consider the relative variation between the skeleton generated from the filtered image and that generated from the original image; moreover, the relative change between the binary image produced from the filtered image and that produced from the original image is considered. In addition, the inside characteristics of the skeleton generated from the filtered image are also considered. Therefore, the selected skeleton tends to have few unwanted branches and rings, and all necessary skeletal branches are retained.

Before formally presenting the details of our framework, to simplify, there is only one object (connected component) in the foreground image. In addition, we assume that the input image is in grayscale.

The Gaussian kernel is used in our framework. The value of the element in the Gaussian kernel whose coordinate is

(x, y) can be denoted as $G_n(x, y, \sigma)$, which can be computed according to the following formulas:

$$G_n(x, y, \sigma) = \frac{G(x, y, \sigma)}{\sum_{x=1}^{k} \sum_{y=1}^{k} G(x, y, \sigma)} \tag{1}$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-k-1)^2+(y-k-1)^2}{2\sigma^2}} \tag{2}$$

$$k = 2 \times \lfloor 2\sigma \rfloor + 1 \tag{3}$$

where σ is the smoothing parameter of the Gaussian kernel that controls the scale and k is the kernel size, whose value depends on σ. $\lfloor . \rfloor$ is the ceiling function that computes the smallest integer that is greater than or equal to the input number.

By changing the value of σ, from the initial σ_{init} increasing to σ_{max}, whose value is a multiple of σ_{init}, with the step of σ_{init}, a series of different Gaussian kernels can be obtained. Then, these kernels are applied to the original input grayscale image to generate a series of filtered images. After binarizing those filtered grayscale images, we can obtain different binary images. Next, multiple skeletons can be extracted from these binary images by using the skeletonization algorithm.

A suitable σ can result in a proper binary image and a proper skeleton, in which the number of unwanted branches from the noise is as small as possible and the number of necessary branches is as large as possible. To select this suitable σ, we develop a novel measurement in which an improper σ may generate an overly large value, and a proper σ may produce a low value.

For a given σ, a Gaussian kernel can be determined. Then, the binary image and skeletal image derived from the grayscale image that is blurred by this Gaussian kernel can be denoted by B_σ and S_σ, respectively. Similarly, the binary and skeletal images derived from the original grayscale images can be denoted by B_o and S_o, respectively. The value of the proposed measure M for the given σ can be computed by using B_σ, S_σ, B_o and S_o, whose formula is shown in the following.

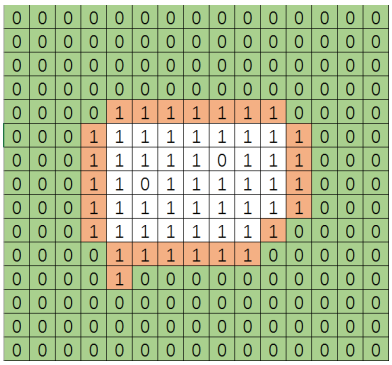
$$M(\sigma) = F_1(S_\sigma, B_o) + F_2(S_\sigma, S_o) + F_3(B_\sigma, B_o) \tag{4}$$

From Eq. 4, it is clear that the proposed measurement is composed of three functions: the F_1 function, F_2 function, and F_3 function. First, F_1 is defined as follows.

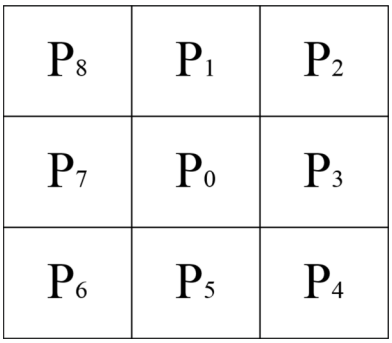
$$F_1(S_\sigma, B_o) = \sum_{i=1}^{H} \sum_{j=1}^{W} Sen(S_\sigma, B_o, i, j) \tag{5}$$

$$Sen(S_\sigma, B_o, i, j) = \begin{cases} 1 & T_b(i, j) > 2 \text{ or } B_o(i, j) = 0 \\ 5 & B_of(i, j) = 1 \\ 0 & otherwise \end{cases} \tag{6}$$

where H and W are the height and width of S_σ, respectively. B_{of} is defined as a particular binary image whose foreground pixels are the background pixels outside the foreground image in the B_o image. Fig. 2a shows a specific B_o image. Only those background pixels that are highlighted in green can be considered foreground pixels in the B_{of} image. T_b calculates the number of transitions from foreground pixel to background pixel in the 8-neighborhood (the definition is shown in Fig. 2b) of foreground pixel (i, j) in the S_σ image. $T_b(i, j)$ is defined in Eq. 7.



(a) Specific example of B_o in which the background pixels with green color form the foreground region in B_{of}



(b) Definition of the 8-neighborhood window

FIGURE 2. Illustration of the B_{of} and definition of 8-neighborhood.

$$T_b(i, j) = \sum_{k=1}^{8} transition(P_k) \tag{7}$$

$$transition(P_k) = \begin{cases} 1 & P_k = 1 \text{ and } P_{(k+1)mod8} = 0 \\ 0 & otherwise. \end{cases} \tag{8}$$

Function F_1 generates a large value in two cases. One of them is when there are many cross points in the skeleton, which denotes that it may still be a noisy skeleton with many unwanted branches or circles. Another case is when the skeleton image derived from the filtered image produces the distortion problem so that the skeletal pixel is located in the background region of the original pattern. We differentiate the penalization for the outer background (foreground pixels in B_{of}) and inner background because in principle, the skeleton should lie on the inside of the original object.

As the second part of the measure, function F_2 is defined as follows:

$$F_2(S_\sigma, S_o) = \begin{cases} 0 & \alpha < \dfrac{Area(S_\sigma)}{Area(S_o)} \\ (1 - \dfrac{Area(S_\sigma)}{Area(S_o)} \cdot \dfrac{1}{\alpha}) max(H, W) & \alpha \geq \dfrac{Area(S_\sigma)}{Area(S_o)} \end{cases} \tag{9}$$

$$Area(S) = \sum_{i=1}^{H} \sum_{j=1}^{W} S(i, j) \tag{10}$$

where α is a threshold that controls the penalty. If the ratio of the number of foreground pixels of the skeleton extracted from the filtered image to that of the original skeleton is higher than α, then there is no penalization. Otherwise, penalization is introduced; in our framework, we set it to 0.42. $Area()$ is used to count the number of foreground pixels of the input binary or skeletal image. Overall, the F_2 function is used to penalize the case when there is a significant difference between the filtered skeleton and the original skeleton.

The last function F_3 is defined as follows:

$$F_3(B_\sigma, B_o) = F_{31}(B_\sigma, B_o) + F_{32}(B_\sigma, B_o) \tag{11}$$

$$F_{31}(B_\sigma, B_o) = \begin{cases} max(H, W) & \dfrac{|Area(B_\sigma) - Area(B_o)|}{Area(B_o)} < \beta \\ 0 & otherwise \end{cases} \tag{12}$$

$$F_{32}(B_\sigma, B_o) = \begin{cases} max(H, W) & N_{region}(B_\sigma) \neq N_{region}(B_o) \\ 0 & otherwise \end{cases} \tag{13}$$

In Eq. 12, β is a threshold for controlling the penalty. The penalty is introduced only when the relative difference in the number of foreground pixels is above this β. Here, we set it to 0.1. Function F_{31} is used to ensure that the changes in foreground pixels of the binary image caused by the introduction of the filter are within a reasonable range. In Eq. 13, $N_{region}(image)$ function is used to compute the total number of connected components in the input image. Therefore, we know that F_{32} is used to ensure that the number of connected components of binary image B_σ remains the same as that of the original binary image B_o. In B_o, there is only one connected component according to our assumption at the beginning of this section. Overall, function F_3 is used to detect the distortion caused by using an improper Gaussian kernel from the perspective of the binary image and to introduce enough penalty.

After obtaining the value of our measure for various σ, each of whose value is the sum of the value of three subfunctions, the best σ_{best} can be selected. A σ can be deemed a σ_{best} only when its derived skeleton and binary image produce the minimum output under our measure. Then, the derived skeleton from σ_{best} is considered the best skeleton, which becomes the output of our framework. The entire procedure of the proposed framework is summarized in Algorithm 1.

Algorithm 1 Procedure of the Proposed Framework

Input: Original grayscale image I_o, Initial Sigma σ_{init}, Maximum Sigma σ_{max}
Output: Skeleton S_{output}

```
M = []
S = []
B_o = Binarize(I_o)
S_o = Skeletonization(B_o)
for (σ = σ_init; σ <= σ_max; σ = σ + σ_init) do
    I_σ = Gaussian(I_o, σ)
    B_σ = Binarize(I_σ)
    S_σ = Skeletonization(B_σ)
    m = F_1(S_σ, B_o) + F_2(S_σ, S_o) + F_3(B_σ, B_o)
    M.append(m)
    S.append(S_σ)
end for
Index = argmin(M)
S_output = S[Index]
return S_output
```

IV. INNER AND BORDER NOISE EXPERIMENT

The proposed framework (PF) and Houssem's ATF framework are implemented in MATLAB on a core i7 Intel CPU. Since both of these frameworks require calling the skeletonization method to extract the skeleton, we also implemented the FPSA algorithm and SPSM algorithm, which are both skeletonization methods described in [10] and in [11].

To objectively evaluate the skeleton extracted by the PF, several performance measures are defined as follows:

- Number of endpoints (NEP): This measure counts the total number of endpoints in all foreground pixels in skeleton S, which can be computed by:

$$NEP(S) = \sum_{i=0}^{H} \sum_{j=0}^{W} EP(S(i, j)) \tag{14}$$

$$EP(S(i, j)) = \begin{cases} 1 & T_b(i, j) == 1 \\ 0 & otherwise \end{cases} \tag{15}$$

- Number of crosspoints (NCP): This measure counts the total number of cross points in all foreground pixels in skeleton S, which can be computed by:

$$NCP(S) = \sum_{i=0}^{H} \sum_{j=0}^{W} CP(S(i, j)) \tag{16}$$

$$CP(S(i, j)) = \begin{cases} 1 & T_b(i, j) > 2 \\ 0 & otherwise \end{cases} \tag{17}$$

- Rate of variation in the skeleton (RVS): This measure calculates the relative deviation of two skeletons, which are extracted from a noisy image and a clean image. The value of

RVS can be calculated by using the following formula:

$$RVS(S_n, S_c) = \frac{Area(S_n \bigtriangledown S_c) + Area(S_c \bigtriangledown S_n)}{2Area(S_c)} \quad (18)$$

where S_n denotes the skeleton extracted from a noisy image, S_c denotes the skeleton extracted from a clean image, and $S_c \bigtriangledown S_n$ denotes those foreground pixels that belong to set S_c but do not belong to set S_n. $Area()$, please refer to Eq. 10.

In later noisy experiments, the antinoise ability of six methods of skeletonization extraction is studied. Four methods are combinations of the frameworks and skeletonization methods. They are PF+FPSA, PF+SPSM, ATF+FPSA, and ATF+SPSM. The other two are the SPSM skeletonization method and the FPSA skeletonization method. For all extraction frameworks, the initial σ_{init} is set as one, and σ_{max} is set as 12.

To better conduct the comparison, a simple and clean human-shape image (named dude8), which comes from the well-known benchmark Kimia-99, is selected as the base image. Dude8 was used in both noisy experiments.

The extracted skeletons from dude8 for all six methods are presented in Fig. 3, and they are considered clean skeletons, which are later compared when computing the RVS with the skeletons extracted from noise.

From Fig. 3, there are many noticeable things. First, for the visualization, we use gray to denote the original pattern and black to indicate the extracted skeleton. All the experimental images presented later share the same style as this image. Next, according to the perception of human vision for this human shape, it is clear that a good skeleton should retain five skeletal branches, representing one head, two arms, and two legs. However, the skeletons extracted by the ATF+SPSM method and ATF+FPSA have only three and four skeletal branches, respectively, which demonstrates that they have suffered some distortion.

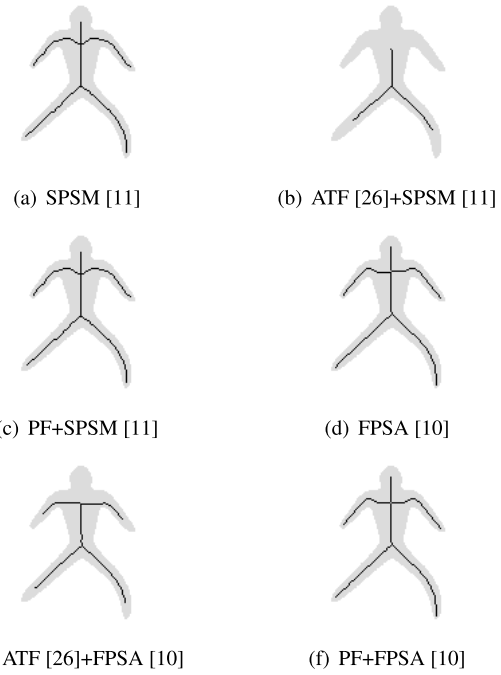


FIGURE 3. Skeletons extracted by six methods from a clean image.

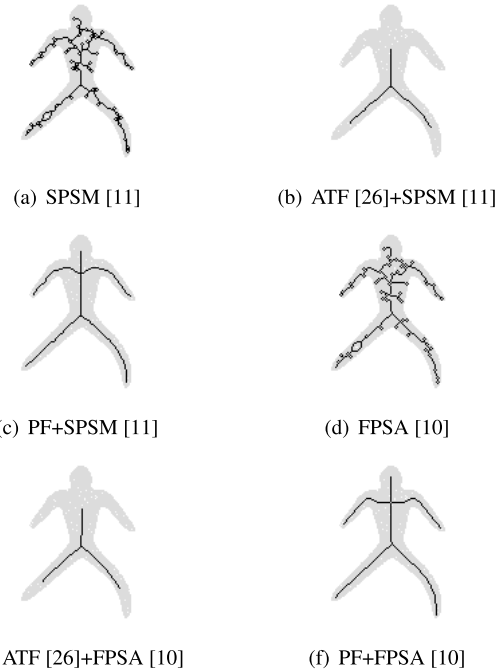
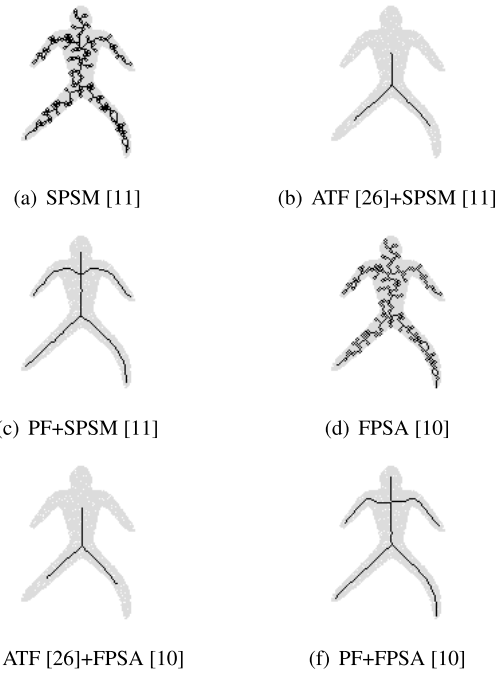


FIGURE 4. Skeletons extracted by six methods from images under 2% inner noise.

A. INNER NOISE EXPERIMENT

The inner noise experiment is divided into five subtests according to the distinct noise levels, which start at 2% and gradually rise to 10%. In each subtest, inner noise is randomly added to the inner region of the original object. Then, six methods are used to extract the skeletons. Figs. 4, 5, and 6 present the skeletons extracted from the noisy image under noise levels of 2%, 6%, and 10% by using six methods. In addition, the RVS, NEP, and NCP values of the skeleton extracted by each distinct method are measured and recorded in each subtest. The whole inner noise experiment was independently conducted 100 times, and the statistical descriptions of the parameters of RVS NEP and NCP are presented in Table 1, Table 2, and Table 3.

Figs. 4 to 6 present an intuitive sense of the robustness of the inner noise for the various methods. In Fig 4, it is clear that the two pure skeletonization methods, SPSM and FPSA, are prone to effects by the inner noise and produce many meaningless rings, even when the noise level is very small. In contrast, the four framework-based methods can create a relatively stable skeleton under 2% inner noise. Their skeletons have only nuances with that used in Fig. 3. From the perspective of skeleton completeness, in Fig. 4, the two ATF-based methods have both suffered from the problem of skeleton distortion since both ATF+SPSM and ATF+FPSA have only three skeletal branches. In contrast, the two

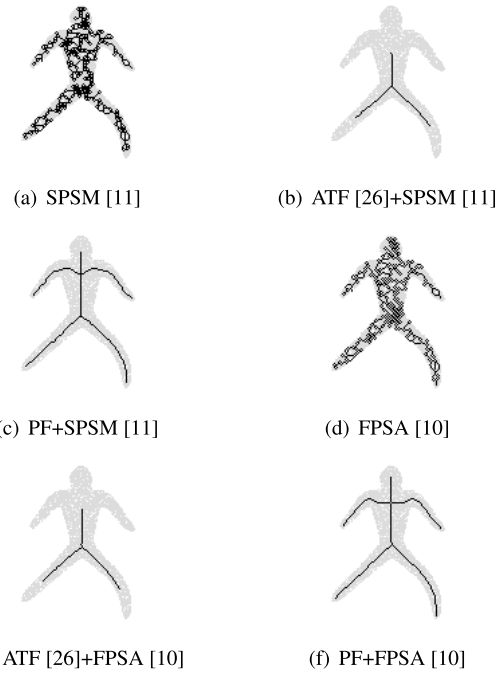


(a) SPSM [11] (b) ATF [26]+SPSM [11]

(c) PF+SPSM [11] (d) FPSA [10]

(e) ATF [26]+FPSA [10] (f) PF+FPSA [10]

FIGURE 5. Skeletons extracted by six methods from images under 6% inner noise.



(a) SPSM [11] (b) ATF [26]+SPSM [11]

(c) PF+SPSM [11] (d) FPSA [10]

(e) ATF [26]+FPSA [10] (f) PF+FPSA [10]

FIGURE 6. Skeletons extracted by six methods from images under 10% inner noise.

methods based on our framework can generate a complete skeleton with all five skeletal branches.

In Fig. 5 and Fig. 6, as the noise level of the inner noise increases, increasingly more rings appear in the results of the SPSM and FPSA methods. In contrast, all the framework-based methods produce skeletons similar to the one shown in Fig. 4 so that it is believable that they are robust to the inner noise. Among the framework-based methods, one difference between the methods based on the PF and those based on the ATF framework is that the methods based on the PF can appropriately preserve all the necessary skeletal branches, while the ATF-based method cannot.

TABLE 1. Mean value of RVS of six methods under different levels of inner noise, each level noise experiment was repeated 100 times.

Noise Level	2%	4%	6%	8%	10%
SPSM [11]	0.909	1.506	1.900	2.262	2.545
ATF [26]+SPSM [11]	0.314	0.338	0.354	0.339	0.340
PF+SPSM [11]	0.098	0.133	0.086	0.092	0.242
FPSA [10]	0.799	1.311	1.638	1.922	2.139
ATF [26]+FPSA [10]	0.533	0.522	0.556	0.568	0.559
PF+FPSA [10]	0.071	0.097	0.100	0.117	0.204

TABLE 2. Mean value of NEP of six methods under different levels of inner noise, each level noise experiment was repeated 100 times.

Noise Level	2%	4%	6%	8%	10%
SPSM [11]	2.92	2.08	1.93	1.84	1.93
ATF [26]+SPSM [11]	3.07	3.07	3.07	3.04	3.05
PF+SPSM [11]	5.00	5.00	5.00	5.00	5.00
FPSA [10]	3.14	2.16	2.23	1.94	2.18
ATF [26]+FPSA [10]	3.16	3.17	3.04	3.03	3.02
PF+FPSA [10]	5.00	5.00	5.00	5.00	5.00

In Table 1, the mean value of RVS of the SPSM method and that of the FPSA method are much higher than that of the other four framework-based methods in all subtests, and they keep increasing with the increment of the noise level. This is consistent with our visual perception in Fig. 4 to Fig. 6. Among these methods, two methods based on the PF have the lowest two values in terms of RVS for each subtest, which demonstrates that they are the two most robust methods to inner noise among these methods.

In addition, in Table 1, by comparing each data listed in the third row with each data recorded in the first row, it is noticed that using the PF can reduce noise by approximately 92% on average. Similar results can be obtained when comparing each data point listed in the sixth row with each data point recorded in the fourth row.

From Table 2, we can also learn that the average NEP for both PF-based methods is five, which means that the proposed method has the potential to maintain skeleton completeness. From Table 3, the proposed framework-based methods can also maintain the value of NCP sufficiently stably with an increasing noise level.

B. BORDER NOISE EXPERIMENT

Similar to the inner noise experiment, the border noise experiment also consists of five subtests according to the noise level. The initial noise level is 30%, which gradually escalates to 50% in steps of 5%. In each subtest, border noise is randomly added to the boundary of the original image, and then the six methods are used to extract the skeleton from them. Each subtest is conducted independently 100 times.

TABLE 3. Mean value of NCP of six methods under different levels of inner noise, each level noise experiment was repeated 100 times.

Noise Level	2%	4%	6%	8%	10%
SPSM [11]	52.78	75.12	85.58	96.67	104.05
ATF [26]+SPSM [11]	1.96	2.00	1.94	1.96	1.96
PF+SPSM [11]	2.01	2.00	2.00	2.02	2.08
FPSA [10]	85.28	144.00	181.32	216.05	241.93
ATF [26]+FPSA [10]	2.00	2.00	2.00	2.00	2.00
PF+FPSA [10]	4.72	4.52	4.52	4.28	4.23

(a) SPSM [11] (b) ATF [26]+SPSM [11]

(c) PF+SPSM [11] (d) FPSA [10]

(e) ATF [26]+FPSA [10] (f) PF+FPSA [10]

FIGURE 7. Skeletons extracted by six methods from an image under 30% border noise.

TABLE 4. Mean value of RVS of six methods under different levels of border noise, each level noise experiment was repeated 100 times.

Noise Level	30%	35%	40%	45%	50%
SPSM [11]	0.348	0.398	0.447	0.506	0.556
ATF [26]+SPSM [11]	0.335	0.312	0.300	0.298	0.299
PF+SPSM [11]	0.225	0.247	0.265	0.287	0.301
FPSA [10]	0.350	0.393	0.454	0.496	0.557
ATF [26]+FPSA [10]	0.378	0.374	0.402	0.392	0.370
PF+FPSA [10]	0.226	0.239	0.269	0.288	0.311

Figs. 7, 8 and 9 present the extracted skeleton under noise levels of 30%, 40% and 50%, respectively.

It is clear that in Fig. 8, the SPSM method and FPSA method produced seven and six skeletal branches, respectively. These are two and one more than the number of skeleton branches they extracted in a clean image, respectively. The ATF+SPSM and ATF+FPSA methods still suffered the problem of excessive erosion. PF+SPSM and PF+FPSA can generate a satisfactory skeleton, although there are some tiny position changes in the skeleton. In Fig. 9 and Fig. 10, we can see that as the border noise level increases, an increasing number of unwanted branches appear in the results from the

(a) SPSM [11] (b) ATF [26]+SPSM [11]

(c) PF+SPSM [11] (d) FPSA [10]

(e) ATF [26]+FPSA [10] (f) PF+FPSA [10]

FIGURE 8. Skeletons extracted by six methods from an image under 40% border noise.

(a) SPSM [11] (b) ATF [26]+SPSM [11]

(c) PF+SPSM [11] (d) FPSA [10]

(e) ATF [26]+FPSA [10] (f) PF+FPSA [10]

FIGURE 9. Skeletons extracted by six methods from an image under 50% border noise.

SPSM and FPSA methods. ATF+SPSM and ATF+FPSA are still hindered by a distortion defect. Only two proposed framework-based methods can suppress border noise and maintain the original structure.

TABLE 5. Mean value of NEP of six methods under different levels of border noise, each level noise experiment was repeated 100 times.

Noise Level	30%	35%	40%	45%	50%
SPSM [11]	7.10	7.92	9.20	10.60	11.59
ATF [26]+SPSM [11]	3.06	3.03	3.00	3.02	3.01
PF+SPSM [11]	5.00	5.00	5.00	5.00	5.00
FPSA [10]	7.24	8.11	9.26	10.94	11.84
ATF [26]+FPSA [10]	3.65	3.73	3.60	3.68	3.73
PF+FPSA [10]	5.00	5.00	5.00	5.00	5.00

TABLE 6. Mean value of NCP of six methods under different levels of border noise, each level noise experiment was repeated 100 times.

Noise Level	30%	35%	40%	45%	50%
SPSM [11]	4.24	5.18	7.55	9.17	11.48
ATF [26]+SPSM [11]	1.90	1.80	1.94	1.92	1.84
PF+SPSM [11]	2.07	2.07	2.16	2.12	2.17
FPSA [10]	6.58	7.62	9.68	11.93	14.3
ATF [26]+FPSA [10]	2.00	2.02	2.00	2.01	2.00
PF+FPSA [10]	4.00	4.02	4.00	3.84	3.72

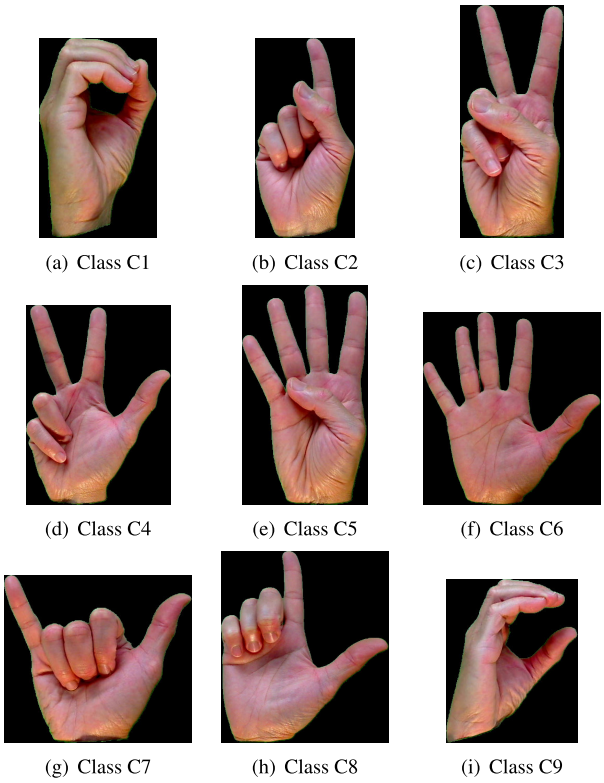
Table 4, Table 5, and Table 6 present the mean values of RVS, NEP, and NCP, respectively, from which it can be seen that with an increasing level of border noise, the values of all three parameters of SPSM and FPSA rapidly increase. In contrast, this trend in the framework-based methods is not apparent, especially when considering NEP and NCP. In addition, it is noted that under various noise levels, those methods based on our frameworks still have a strong ability against border noise, whose RVS is lowest when compared with the other techniques that use the same skeletonization method. In addition, the NEP of the method based on our framework is always 5, which demonstrates that using our framework may not introduce distortion.

In addition, from Table 4, by comparing each data listed in the third row with each data recorded in the first row, it is noticed that using the PF can reduce noise by approximately 40% on average. Similar results can be obtained when comparing each data point listed in the sixth row with each data point recorded in the fourth row.

From the inner noise experiment and border noise experiment, it is confirmed that the PF can enhance the robustness of the skeletonization algorithm, and the introduction of the PF will not cause skeleton distortion. Therefore, the PF is an excellent option for improving the stability of existing skeletonization algorithms.

V. STATIC HAND GESTURE RECOGNITION EXPERIMENT

To further explore the performance of SPSM, ATF+SPSM, PF+SPSM, FPSA, ATF+FPSA, and PF+FPSA, static hand gesture recognition experiments are conducted. All the static hand gesture images used in this experiment are part of a well-known public dataset named MU_ HandImages_ASL [29]. Nine different hand gesture classes are considered in our experiment, and in each category, there are 70 RGB images. As a result, there are a total of 630 RGB images. Examples of nine classes of hand gestures are shown in Fig. 10.



(a) Class C1 (b) Class C2 (c) Class C3

(d) Class C4 (e) Class C5 (f) Class C6

(g) Class C7 (h) Class C8 (i) Class C9

FIGURE 10. Example of nine classes of static hand gestures.

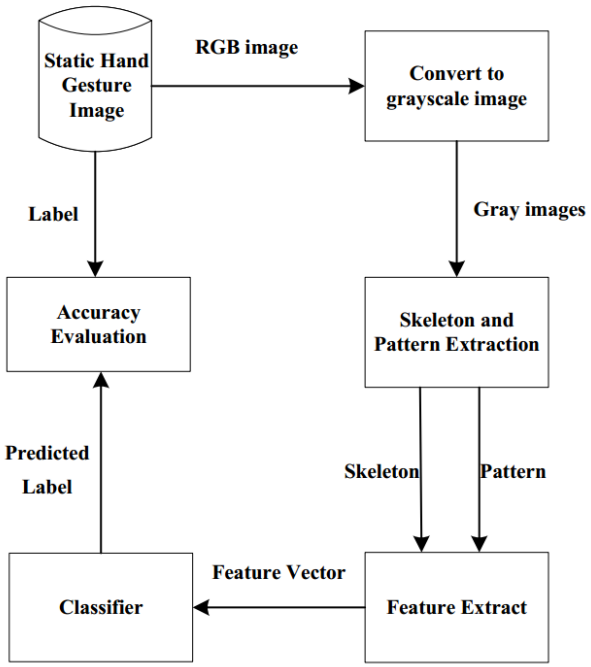


FIGURE 11. Block diagram of the entire procedure from feature extraction to model evaluation.

A. OVERVIEW

In our experiment, for static hand gesture recognition, we adopted a standard machine learning pipeline, which includes feature extraction, model training, and evaluation.

During the feature extraction procedure, the original RGB static hand gesture image is first converted into a grayscale image, and the skeleton and its corresponding binary image (pattern image) are extracted. For the implemented framework-based method, it is easy to extract both the skeleton and pattern image from the grayscale image (the skeleton should be derived from this pattern) because the operation of image binarization is embedded in the framework. However, since skeletonization methods can process only binary images, SPSM and FPSA must introduce an extra binarization operation. The result of the binarization is considered a pattern and saved. After obtaining the skeleton and pattern image, the next step is to convert them into a 7-dimensional feature vector, which is presented in the following subsection.

Four well-known classification models are adopted in the current experiment: the decision tree model (DT), the bagging tree model (BT), the support vector machine model (SVM), and the k-nearest neighbor model (KNN). All these models are created in MATLAB, and their parameters are set as default values. For better training and evaluation of these models, the original dataset is randomly divided into two subsets, the training-validation set and the testing set. There are 500 images in the training-validation set, which is used in model training, and 130 images in the testing set, which is used in model evaluation. During the training procedure, a 10-fold cross-validation strategy is adopted. Since there is a balanced multiclass classification task in the present experiment, accuracy is used to evaluate the models' performance. The formula for the accuracy is shown as follows:

$$Acc(f, D) = \frac{1}{m}\sum_{i=1}^{m} I(f(x_i) = y_i) \tag{19}$$

$$I(expression) = \begin{cases} 1 & expression\ is\ True \\ 0 & expression\ is\ False \end{cases} \tag{20}$$

where D is a set consisting of all the feature vectors and corresponding labels, and m is the number of the pair of feature vectors x and its corresponding actual label y. $f(x)$ is the predicted output of a classifier when the input feature vector is x.

In Fig. 11, a block diagram of the entire procedure from feature extraction to model evaluation is presented.

B. FEATURE VECTOR EXTRACTION FROM SKELETON AND PATTERN IMAGES

After a skeletal image and its pattern image are obtained from an input image, it is necessary to transform the skeletal images along with its pattern image to a 7-dimensional feature vector that is used in later classification. This 7-dimensional vector includes the number of endpoints (NEP), number of crosspoints (NCP), whether the inner hole exists or not (EIH), the rate of deviation of the thickness of the endpoints (RDTE), the average distance between the thickest point in the pattern image and the endpoints in the skeletal image (ADTPE), the distance between the pattern thickest point and the skeletal thickest point (DPSP) and the average angle between the endpoint and the main axis (AAEP).

The values of NEP and NCP can be calculated by using the formula in Eq. 14 and Eq. 16, the EIH can be determined by using an all-ones matrix to subtract the skeleton matrix and count the number of regions. EIH is one when the number of regions is above one; otherwise, it is zero.

Before presenting the definition of RDTE, ADTPE, and DPSP, the concept of thickness is first introduced. The thickness of a pixel is defined by the distance between this pixel and its closest pixel located on the boundary in the pattern image. Boundary pixels are composed of the foreground pixel, whose 4 neighbors have at least one background pixel. In Fig 2 (a), the region marked by brown is the boundary.

For a given skeleton that has n endpoints, all endpoints can form a set S_{EP}, in which the i-th endpoint is denoted as S_{EP_i}. The thickness of S_{EP_i} can be denoted as T_{EP_i}. The set formed by all T_{EP_i} is denoted as $T_{S_{EP}}$. Then, the RDTE for this skeleton can be computed by using the following formula:

$$RDTE = \begin{cases} 0 & n \le 1 \\ \sum_{i=1}^{n} \dfrac{\sqrt{(T_{EP_i} - \frac{1}{n}\sum_{i=1}^{n} T_{EP_i})^2}}{max(T_{S_{EP}}) - min(T_{S_{EP}})} & n > 1 \end{cases} \tag{21}$$

We assume the coordinates of the thickest pixel in the pattern image are P_x and P_y, and its thickness is T_p. We suppose that in a skeletal image, there are n endpoints. The coordinates of the i-th endpoint are denoted as EP_{i_x} and EP_{i_y}. Then, the ADTPE can be calculated by using the following formula:

$$ADTPE = \begin{cases} 0 & n = 0 \\ \dfrac{\sum_{i=1}^{n} \sqrt{(P_x - EP_{i_x})^2 + (P_y - EP_{i_y})^2}}{nT_p} & n > 0 \end{cases} \tag{22}$$

Assuming the coordinate of the thickest pixel in the pattern image is P_x and P_y, and the coordinate of the thickest pixel in the skeletal image is S_x and S_y, the DPSP can be calculated according to the following formula:

$$DPSP = \sqrt{(P_x - S_x)^2 + (P_y - S_y)^2} \tag{23}$$

Before obtaining the value of the AAEP, the main axis is defined by the thickest point in the pattern image and the farthest endpoint in the skeletal image from that point. Based on that, it is easy to calculate the relative angle of the remaining endpoint to these axes, and the AAEP is the mean of these angles. If the number of endpoints is less than 2, the AAEP is set as 0.

C. RESULTS OF THE EXPERIMENT

Since the classifier accuracy may drift when using different test sets and validation sets, thus influencing our core task of studying how different skeletonization methods influence

final recognition accuracy, we conducted the training and evaluation on different classifiers 80 times independently, in which images in the test set and validation set were randomly selected.

Table 7 presents the maximum validation accuracy when training different classifiers on feature vectors derived from skeletons extracted by different skeletonization methods. In the table, we can see that under different classifiers, the maximum validation accuracy of PF+SPSM outperforms SPSM and ATF+SPSM, and the maximum validation accuracy of PF+FPSA exceeds that of FPSA and ATF+FPSA. For example, when using the BT classifier, the validation accuracy of PF+SPSM and PF+FPSA reaches 89.60%. The confusion matrix for maximum validation accuracy when using the BT classifier and various skeletonization methods can be seen in Fig. 12. Tables 8 and 9 present the mean validation accuracy and min validation accuracy when using different skeletonization methods and classifiers, from which we can observe that the validation accuracy of using the PF-based skeletonization method is higher than that of using other skeletonization methods.

In addition, by comparing each data recorded in the third row and each data recorded in the first row in Table 8, it is clear that the maximum increase in the mean accuracy on the validation set is approximately 11%, which denotes that the use of the PF can greatly improve recognition accuracy.

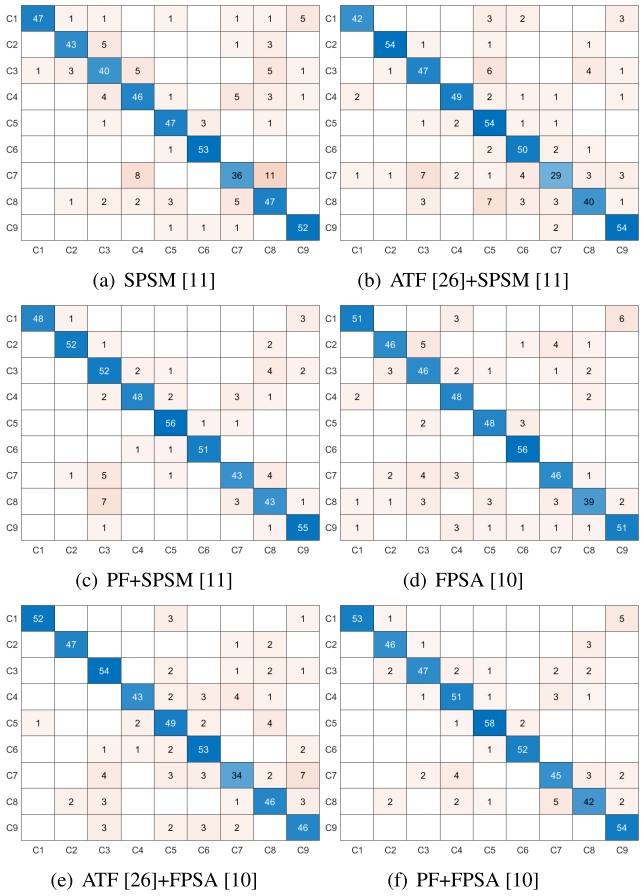
TABLE 7. Max accuracy on the validation set under the different skeleton extraction techniques and various classifiers among 80 individual experiments.

MAX accuracy	DT	BT	KNN	SVM
SPSM [11]	72.80%	82.20%	73.64%	65.60%
ATF [26]+SPSM [11]	73.60%	83.80%	79.71%	59.40%
PF+SPSM [11]	80.20%	89.60%	84.09%	70.80%
FPSA [10]	75.00%	86.20%	79.37%	66.40%
ATF [26]+FPSA [10]	74.80%	84.80%	84.36%	58.80%
PF+FPSA [10]	81.20%	89.60%	85.65%	66.40%

TABLE 8. Mean accuracy on the validation set under the different skeleton extraction techniques and different classifiers among 80 individual experiments.

Mean accuracy	DT	BT	KNN	SVM
SPSM [11]	68.08%	79.65%	70.41%	61.40%
ATF [26]+SPSM [11]	67.70%	80.60%	76.84%	55.87%
PF+SPSM [11]	76.64%	87.61%	81.16%	66.42%
FPSA [10]	70.43%	83.71%	75.47%	62.88%
ATF [26]+FPSA [10]	70.22%	82.19%	81.06%	56.64%
PF+FPSA [10]	77.81%	87.28%	83.58%	62.44%

Similar to Table 7, Table 10 presents the maximum testing accuracy when testing different classifiers by using feature vectors derived from skeletons extracted by various skeletonization methods. The testing accuracy based on the PF+SPSM method is higher than that of the SPSM method and ATF+SPSM, and the accuracy of the PF+FPSA method is higher than that of FPSA and ATF+FPSA. The maximum testing accuracy based on PF+SPSM is 94.29% and that based on PF+FPSA is 93.45% when using the BT classifier.



(a) SPSM [11] (b) ATF [26]+SPSM [11] (c) PF+SPSM [11] (d) FPSA [10] (e) ATF [26]+FPSA [10] (f) PF+FPSA [10]

FIGURE 12. Confusion matrix for maximizing the validation accuracy when using a BT and different skeletonization methods, which corresponds to the third column in Table 7.

TABLE 9. Min accuracy on the validation set under the different skeleton extraction techniques and different classifiers among 80 individual experiments.

Min accuracy	DT	BT	KNN	SVM
SPSM [11]	64.20%	77.40%	67.35%	57.00%
ATF [26]+SPSM [11]	56.40%	78.00%	73.17%	51.60%
PF+SPSM [11]	71.60%	84.40%	78.35%	63.00%
FPSA [10]	65.80%	81.40%	72.13%	59.60%
ATF [26]+FPSA [10]	64.40%	79.60%	78.04%	53.60%
PF+FPSA [10]	73.40%	85.00%	80.67%	57.00%

TABLE 10. Max accuracy on the testing set under different skeleton extraction techniques and different classifiers among 80 individual experiments.

Max accuracy	DT	BT	KNN	SVM
SPSM [11]	76.92%	86.88%	81.48%	69.99%
ATF [26]+SPSM [11]	78.79%	87.67%	84.23%	66.11%
PF+SPSM [11]	84.44%	94.29%	92.36%	74.54%
FPSA [10]	85.34%	91.47%	84.57%	71.88%
ATF [26]+FPSA [10]	83.09%	92.19%	91.48%	66.65%
PF+FPSA [10]	89.96%	93.45%	91.55%	69.64%

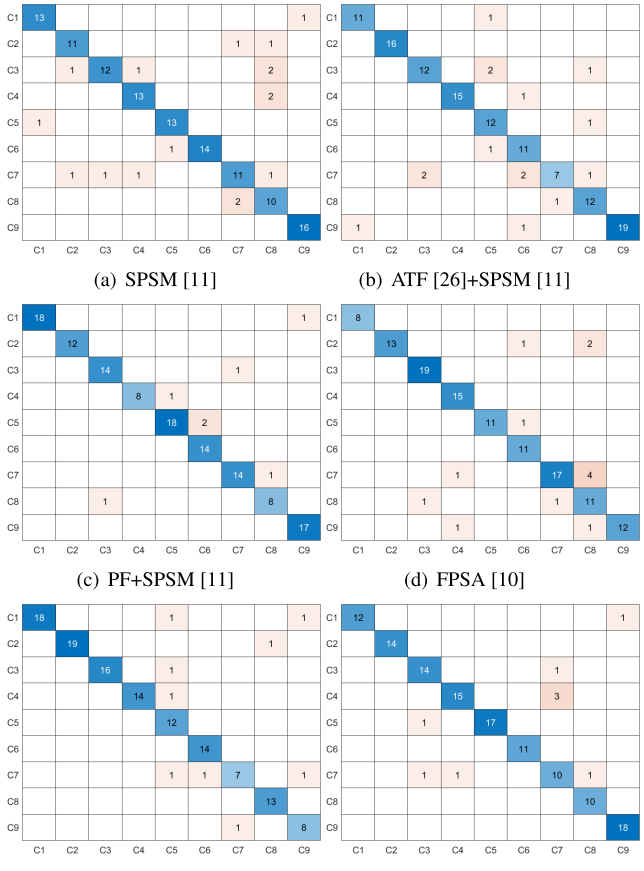
Similar to Fig. 12, Fig. 13 shows the confusion matrix for maximum testing accuracy when using a BT and different skeletonization methods. Table 11 and Table 12 present the

TABLE 11. Mean accuracy on testing set under different skeleton extraction techniques and different classifiers among 80 individual experiments.

Mean accuracy	DT	BT	KNN	SVM
SPSM [11]	69.31%	80.25%	71.11%	61.45%
ATF [26]+SPSM [11]	68.71%	80.62%	77.30%	55.58%
PF+SPSM [11]	77.45%	87.96%	82.09%	67.67%
FPSA [10]	70.88%	84.02%	76.49%	63.44%
ATF [26]+FPSA [10]	70.62%	82.49%	82.12%	57.13%
PF+FPSA [10]	79.21%	88.20%	84.97%	62.86%

TABLE 12. Min accuracy on testing set under different skeleton extraction techniques and different classifiers among 80 individual experiments.

Min accuracy	DT	BT	KNN	SVM
SPSM [11]	59.70%	71.67%	56.44%	50.19%
ATF [26]+SPSM [11]	49.93%	70.51%	65.86%	47.01%
PF+SPSM [11]	65.72%	79.45%	74.14%	59.61%
FPSA [10]	57.64%	76.21%	65.27%	51.26%
ATF [26]+FPSA [10]	61.97%	73.86%	73.48%	49.54%
PF+FPSA [10]	68.91%	81.22%	78.38%	51.61%



(a) SPSM [11] (b) ATF [26]+SPSM [11]

(c) PF+SPSM [11] (d) FPSA [10]

(e) ATF [26]+FPSA [10] (f) PF+FPSA [10]

FIGURE 13. Confusion matrix for maximum testing accuracy when using a BT and different skeletonization methods, which corresponds to the third column in Table 10.

mean and min testing accuracy when using different skeletonization methods and classifiers.

In addition, by comparing each data recorded in the third row and each data recorded in the first row in Table 11, it is clear that the maximum increase in the mean accuracy on the testing set is also approximately 11%.

Using the PF can significantly increase validation and testing accuracy for four classifiers in the static gesture recognition experiment. The reason is that the skeleton extracted by the PF is more robust. Additionally, it can preserve the necessary branches to avoid the appearance of distortion.

VI. CONCLUSION

In this work, we proposed a novel noise-against-skeleton extraction framework. Our framework can enhance the robustness of the existing skeletonization method for both inner and border noise. Two noise experiments demonstrated the robustness of the proposed framework to noise. In addition, the proposed framework can appropriately preserve the essential skeleton and avoid the distortion problem. These factors make it promising for applications along with existing skeletonization methods in the pattern recognition field. The results of experiments on static hand gesture recognition support the opinion that using the proposed framework can improve the validation accuracy and the testing accuracy of four well-known classifiers.

VII. LIMITATIONS AND FUTURE WORK

One of the limitations of this paper is that in static hand gesture recognition experiments, we only focus on the influence of introducing the proposed framework to the recognition accuracy and temporarily omit some important factors.

In the future, it is possible to further improve recognition accuracy by considering other important factors. Since feature selection may greatly affect the recognition accuracy, it is necessary to find more important features and from them to obtain the optimal features. Besides, the configuration of the classifiers may alter the recognition accuracy as well, therefore, it is necessary to put more effort into tuning the classifiers. It will also be interesting to consider other classifiers. In addition, in our paper, there are only nine different types of static hand gestures, and it will be exciting to explore the performance by including more types of static hand gestures.

Deep learning methods to extract skeletons have emerged in these years, such as [30] and [31]. Their robustness to noise is still far from satisfying. It is also attractive to combine our framework and these new methods in the future.

REFERENCES

[1] P. K. Saha, G. Borgefors, and G. S. D. Baja, "A survey on skeletonization algorithms and their applications," *Pattern Recognit. Lett.*, vol. 76, pp. 3–12, Jun. 2016.

[2] T. Zhang, H. Lin, Z. Ju, and C. Yang, "Hand gesture recognition in complex background based on convolutional pose machine and fuzzy Gaussian mixture models," *Int. J. Fuzzy Syst.*, vol. 22, no. 4, pp. 1330–1341, Jun. 2020.

[3] B. Su, H. Wu, M. Sheng, and C. Shen, "Accurate hierarchical human actions recognition from Kinect skeleton data," *IEEE Access*, vol. 7, pp. 52532–52541, 2019.

[4] C. Yang, O. Tiebe, K. Shirahama, and M. Grzegorzek, "Object matching with hierarchical skeletons," *Pattern Recognit.*, vol. 55, pp. 183–197, Jul. 2016.

[5] J. Ma, J. Wang, J. Li, and D. Zhang, "Real-time skeletonization for sketch-based modeling," *Comput. Graph.*, vol. 102, pp. 56–66, Feb. 2022.

[6] L. Mourot, L. Hoyet, F. Le Clerc, F. Schnitzler, and P. Hellier, "A survey on deep learning for skeleton-based human animation," in *Computer Graphics Forum*, vol. 41, no. 1. Hoboken, NJ, USA: Wiley, 2022, pp. 122–157.

[7] J. Zhang, F. Wu, W. Chang, and D. Kong, "Techniques and algorithms for hepatic vessel skeletonization in medical images: A survey," *Entropy*, vol. 24, no. 4, p. 465, Mar. 2022.

[8] P. K. Saha, "Skeletonization and its application to quantitative structural imaging," in *Proc. Int. Conf. Frontiers Comput. Syst.* Cham, Switzerland: Springer, 2023, pp. 233–243.

[9] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: A review of techniques," *J. Imag.*, vol. 6, no. 8, p. 73, Jul. 2020.

[10] J. Ma, X. Ren, V. Y. Tsviatkou, and V. K. Kanapelka, "A novel fully parallel skeletonization algorithm," *Pattern Anal. Appl.*, vol. 25, no. 1, pp. 169–188, Feb. 2022.

[11] J. Ma, X.-H. Ren, T. V. Yurevich, and V. K. Kanapelka, "A novel sub-iterative parallel skeletonization method," *J. Comput.*, vol. 32, no. 6, pp. 83–97, 2021.

[12] C. Yang, B. Indurkhya, J. See, and M. Grzegorzek, "Towards automatic skeleton extraction with skeleton grafting," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 12, pp. 4520–4532, Dec. 2021.

[13] B. Durix, G. Morin, S. Chambon, J.-L. Mari, and K. Leonard, "One-step compact skeletonization," in *Proc. 40th Annu. Conf. Eur. Assoc. Comput. Graph.-Eurograph.*, 2019, pp. 1–5.

[14] F. Y. Shih and W.-T. Wong, "Fully parallel thinning with tolerance to boundary noise," *Pattern Recognit.*, vol. 27, no. 12, pp. 1677–1695, Dec. 1994.

[15] B. Durix, S. Chambon, K. Leonard, J.-L. Mari, and G. Morin, "The propagated skeleton: A robust detail-preserving approach," in *Proc. Int. Conf. Discrete Geometry Comput. Imag.* Cham, Switzerland: Springer, 2019, pp. 343–354.

[16] X. Bai, L. J. Latecki, and W.-Y. Liu, "Skeleton pruning by contour partitioning with discrete curve evolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 449–462, Mar. 2007.

[17] W. Shen, X. Bai, R. Hu, H. Wang, and L. J. Latecki, "Skeleton growing and pruning with bending potential ratio," *Pattern Recognit.*, vol. 44, no. 2, pp. 196–209, Feb. 2011.

[18] H. Liu, Z.-H. Wu, X. Zhang, and D. F. Hsu, "A skeleton pruning algorithm based on information fusion," *Pattern Recognit. Lett.*, vol. 34, no. 10, pp. 1138–1145, Jul. 2013.

[19] S. Krinidis and M. Krinidis, "Empirical mode decomposition on skeletonization pruning," *Image Vis. Comput.*, vol. 31, no. 8, pp. 533–541, Aug. 2013.

[20] L. Serino and G. S. D. Baja, "A new strategy for skeleton pruning," *Pattern Recognit. Lett.*, vol. 76, pp. 41–48, Jun. 2016.

[21] G. Siyu, H. Pingping, L. Zhigang, W. He, and L. Min, "A skeleton pruning method based on saliency sorting," in *Proc. 14th IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI)*, Nov. 2019, pp. 593–599.

[22] W. Shen, X. Bai, X. Yang, and L. J. Latecki, "Skeleton pruning as trade-off between skeleton simplicity and reconstruction error," *Sci. China Inf. Sci.*, vol. 56, no. 4, pp. 1–14, 2013.

[23] A. Solís Montero and J. Lang, "Skeleton pruning by contour approximation and the integer medial axis transform," *Comput. Graph.*, vol. 36, no. 5, pp. 477–487, Aug. 2012.

[24] M. E. Hoffman and E. K. Wong, "Scale-space approach to image thinning using the most prominent ridge-line in the image pyramid data structure," in *Proc. SPIE*, vol. 3305, pp. 242–252, Apr. 1998.

[25] J. Cai, "Robust filtering-based thinning algorithm for pattern recognition," *Comput. J.*, vol. 55, no. 7, pp. 887–896, Jul. 2012.

[26] H. Chatbri and K. Kameyama, "Using scale space filtering to make thinning algorithms robust against noise in sketch images," *Pattern Recognit. Lett.*, vol. 42, pp. 1–10, Jun. 2014.

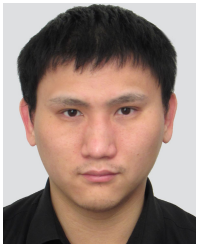
[27] C. Zhou, G. Yang, D. Liang, X. Yang, and B. Xu, "An integrated skeleton extraction and pruning method for spatial recognition of maize seedlings in MGV and UAV remote images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4618–4632, Aug. 2018.

[28] M. M. Kamani, F. Farhat, S. Wistar, and J. Z. Wang, "Skeleton matching with applications in severe weather detection," *Appl. Soft Comput.*, vol. 70, pp. 1154–1166, Sep. 2018.

[29] A. Barczak, N. Reyes, M. Abastillas, A. Piccio, and T. Susnjak, "A new 2D static hand gesture colour image dataset for ASL gestures," *Res. Lett. Inf. Math. Sci.*, vol. 15, pp. 12–20, Jan. 2011.

[30] S. Fang, K. Li, and Z. Li, "CAMION: Cascade multi-input multi-output network for skeleton extraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 2952–2957.

[31] Z. Huang, Y. Wang, Z. Chen, X. Gao, R. Feng, and X. Li, "Context attention network for skeleton extraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 2946–2951.



JUN MA received the B.S. degree from the Lanzhou University of Technology, in 2015, and the M.S. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, machine learning, and computer vision.

XUNHUAN REN received the B.S. degree from the Lanzhou University of Technology, in 2015, and the M.S. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018, where she is currently pursuing the Ph.D. degree. Her current research interests include image processing and information theory.

HAO LI received the B.S. and M.S. degrees in computer science from Henan University. He is currently pursuing the Ph.D. degree with the Belarusian State University of Informatics and Radioelectronics. His research interests include image processing, mobile crowd-sensing networks, and signal processing.

WENZU LI received the B.S. degree from the Lanzhou University of Technology, in 2016, and the M.S. degree from Francisk Skorina Gomel State University, in 2018. He is currently pursuing the Ph.D. degree with the Belarusian State University of Informatics and Radioelectronics. His research interests include machine learning, intelligent tutorial systems, and knowledge base and semantic web.

VIKTAR YUREVICH TSVIATKOU received the Ph.D. degree from the Belarusian State University of Informatics and Radioelectronics, in 1999. He works at the Belarusian State University of Informatics and Radioelectronics, where he is currently a Professor and the Dean of the Department of Infocommunication Technologies, Faculty of Information Security. His research interests include digital image processing, pattern recognition, signal processing, and information theory.

ANATOLIY ANTONOVICH BORISKEVICH is a Professor with the Belarusian State University of Informatics and Radioelectronics. His research interests include digital signal processing, machine learning, and deep learning.