Research on the Method of Displaying the Contour Features of Image to the Visually Impaired on the Touch Screen

Dapeng Chen, *Member, IEEE*, Jia Liu, *Member, IEEE*, Lei Tian, Xuhui Hu, and Aiguo Song, *Senior Member, IEEE*

***Abstract*— Conveying image information to the blind or visually impaired (BVI) is an important means to improve their quality of life. The touch screen devices used daily are the potential carriers for BVI to perceive image information through touch. However, touch screen devices also have the disadvantages of limited computing power and lack of rich tactile experience. In order to help BVI to access images conveniently through the touch screen, we built an image contour display system based on vibrotactile feedback. In this paper, an image smoothing algorithm based on convolutional neural network that can run quickly on the touch screen device is first used to preprocess the image to improve the effect of contour extraction. Then, based on the haptic physiological characteristics of human beings, this paper proposes a method of using the improved MH-Pen to guide the BVI to perceive image contour on the touch screen. This paper introduces the extraction and expression methods of image contours in detail, and compares and analyzes the effects of the subjects' perception of image contours in two haptic display modes through two types of user experiments. The experimental results show that the image smoothing algorithm is useful and necessary to help obtain the main contour of the image and to ensure the real-time display of the contour, and the contour expression method based on the motion direction guidance helps the subjects recognize the contour of the image more effectively.**

***Index Terms*— Contour display, image smoothing, touch screen interaction, motion direction guidance, BVI.**

Manuscript received April 25, 2021; revised September 24, 2021; accepted October 24, 2021. Date of publication October 27, 2021; date of current version November 4, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62003169, Grant 61773219, and Grant 61902179; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20200823; in part by the Jiangsu Innovation and Entrepreneurship Talent Program Project under Grant JSSCBS20200576; in part by the Natural Science Research Project of Jiangsu Higher Education Institutions under Grant 20KJB520029; in part by the Open Foundation of Jiangsu Key Laboratory of Remote Measurement and Control under Grant 2020YCCK-1; and in part by the Startup Foundation for Introducing Talent of NUIST under Grant 2020r010. *(Corresponding author: Dapeng Chen.)*

Dapeng Chen and Jia Liu are with the School of Automation, B-DAT, CICAEET, C-MEIC, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: dpchen@nuist.edu.cn).

Lei Tian is with the School of Automation, Nanjing Institute of Technology, Nanjing 211167, China.

Xuhui Hu and Aiguo Song are with the School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China.

Digital Object Identifier 10.1109/TNSRE.2021.3123394

I. Introduction

FOR a long time, people have been trying to convey image information to blind or visually impaired (BVI) to promote their development in object recognition and spatial cognition. Although BVI cannot directly acquire the information in the visual images, the lack of vision does not prevent them from establishing a "mental imagery" of the object through other compensatory mechanisms [1], [2]. Especially in supporting spatial behavior, the spatial information encoded by vision and touch is considered to be functionally equivalent in promoting the development of modal expression in the brain [3]. Therefore, the sense of touch can replace vision as an ideal physiological perception channel for BVI to obtain spatial information of objects.

The spatial structure features of objects in images, such as contours and shapes, are the basis for people to recognize objects through touch [4]. The traditional assistive technology to convey image spatial information to BVI is mainly realized by using touchable devices or materials [5]. These technologies generally need to first convert a visual image into a tactile image, and a person's finger or other area receives skin irritation corresponding to the image information during the process of touching the tactile image. In the early days, BVI mainly used hand-made materials such as embossed paper to perceive graphic information [6]. To improve the reusable rate of tactile images, a large number of studies have employed probes [7], vibrating elements [8], microelectrodes [9] and other elements as *taxels*, and designed refreshable tactile devices that display image information to people through stimulation methods such as tapping, vibration, and current [10]. However, these technologies generally have problems such as poor portability, high cost, and single function, which are difficult to be widely promoted and used in daily life.

In recent years, with the popularity of mobile smart devices, touch screen has become an important medium for people to interact with virtual environments. The commercialized touch screen-based mobile smart devices (referred to as touch screen devices) have the advantages of portability, low cost and data refreshability, which well solve the problems encountered by traditional assistive technology and provide an important interactive carrier for BVI to perceive image information [11].

Nevertheless, BVI still has great challenges in accessing various types of visual information such as maps, graphics, videos and ordinary images displayed on the touch screen. This is because, on the one hand, these information cannot be directly obtained by BVI through other senses except vision, and it is also difficult to effectively describe to them through sound or language [10], [12]. On the other hand, as a flat, featureless surface, the touch screen cannot provide people with meaningful skin stimulation like touching real objects [13]. Therefore, providing BVI with haptic stimulation corresponding to image information during touch screen interaction will help them perceive the image.

Current touch screen devices generally have feedback functions such as sound and vibration. Using these feedbacks, some studies have introduced methods based on vibration-sound prompts [12], varying sound intensity [14], and vibration-kinesthetic feedback [15] to help people access images or virtual environments non-visually. However, the vibration prompts provided by daily-used touch screen devices are too simple to describe the information of complex images, and sound is considered not as effective as touch in presenting the spatial information of the image to the BVI [16]. Therefore, the feedback function of the touch screen device can only be used to display image with simple 2D geometric shapes.

In order to enhance the haptic feedback ability of touch screen, some studies have modified the touch screen. For example, ultrasonic vibration [17], electrovibration [18] and electroadhesion [19] are utilized to change the friction or lateral force of a person's finger sliding on the touch screen. However, these surface tactile technologies provide global tactile stimulation. Although there are some studies that provide local haptic feedback for touch screen interaction [20], these surface haptic techniques still fail to accurately guide the movement direction of fingers on the touch screen.

When haptically exploring the characteristics of objects, people generally need to adopt appropriate exploratory procedures [21], [22]. For example, the continuous following of the contour line with the finger helps people establish a mental image of the layout of object. Thus, when the touch screen is used as the carrier for BVI to perceive image contour features, it is necessary to provide tactile stimulation with motion direction guidance function for the movement of hands or fingers on the touch screen. However, due to the frequent movement of touch screen devices, the touchscreen interaction often lacks a stable external reference frame, which cannot provide BVI with good positioning and guidance functions. These defects easily lead to BVI losing the tracking of the contour and shape of the object in the image during the process of touching the screen.

Based on this consideration, this study adopts a pen-type haptic device to provide motion direction guidance for BVI's movement on the touch screen. Pen-type haptic device is generally a self-contained system, which can be used as input or output, and has the advantages of portability and ease of use. By integrating multiple types of actuators internally, the pen-type device provides rich haptic stimulation for touch screen interaction, avoiding the modification of touch screen devices. Some pen-type devices used to improve the accuracy of touch screen interaction have been put into commercial applications, such as the Apple Pencil, the Microsoft Surface Pen, and the Samsung S-Pen. To achieve more complex touch screen applications, some researches have integrated actuators such as vibration tactile array [23], electromagnetic coil [24], DC motor [25], and magnetorheological fluid actuator [26] in pen-type device to realize the display of texture, friction, shape, hardness and other features of image. However, the existing pen-type haptic devices are still not well used to display the contour features of image on the touch screen.

In this paper, we developed a system that effectively displays image contour features to BVI on the touch screen through vibrotactile feedback. The main contributions of this study include: (1) In response to the limited computing power of touch screen devices and the real-time interaction requirements of haptic display, an image smoothing algorithm based on convolutional neural network (CNN) that can be quickly run on touch screen devices was designed to improve the effect of image contour extraction. (2) According to the exploration strategy used by BVI to perceive object contour and the characteristics of touch screen interaction, an image contour display method using vibrotactile feedback to guide the movement direction was proposed. (3) User experiments for simple letter images and natural images were performed separately to test the effectiveness of the image contour extraction algorithm and the haptic expression method. The experimental results show that the image smoothing algorithm based on CNN helps to extract the main contour features of the image faster and more effectively on the touch screen device. Meanwhile, the proposed contour display method based on the motion direction guidance significantly improves the accuracy of subjects' recognition of image contour on the touch screen, and reduces the time spent in the perception process.

The rest of this paper is organized as follows. Section II introduces the implementation process of CNN-based image smoothing algorithm in detail, and compares it with four commonly used algorithms in terms of smoothing effect and running time. In Section III, we modified the previously designed MH-Pen to have the function of guiding the movement direction on the touch screen. In Section IV, we applied the pen-type haptic device to display the contour features of the image, and verified the effectiveness of the proposed image contour display system through two user experiments. Finally, we conclude this paper in Section V.

II. Extract the Contour Features of the Image

A. Image Smoothing Filter Based on CNN

The image usually contains rich feature information such as object structure (contour and shape) and surface details (texture and small contour) presented in the form of aliasing. Too much detailed information not only does not help the overall perception of the object, but also brings confusion to the integration of spatial information. Therefore, the image for BVI perception needs to be smoothed on the surface of the object while maintaining the structural features, so as to generate a tactile image that meets the low-bandwidth characteristics of the touch [5], [22]. This requirement coincides with the goal of image smoothing.

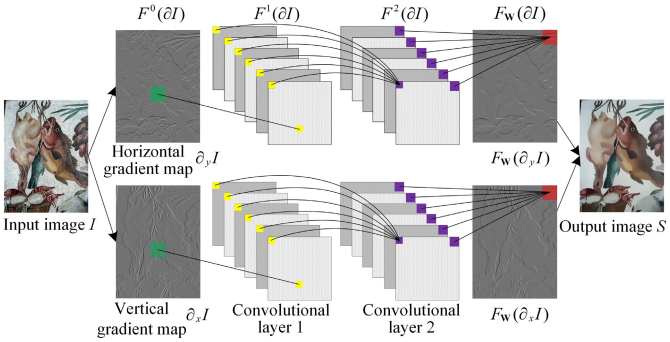


Fig. 1. The structure of the CNN filter.

Traditional image smoothing methods mainly include local filtering-based methods and global optimization-based methods [27]. Among them, the optimization-based methods obtain the desired image edge recognition and filtering effect by solving the optimal solution of the objective function composed of the data fidelity term and the regularization term. For example, according to the different properties shown by the structure and texture, the relative total variation (RTV) [28] uses different penalty functions to process the two features in an optimization framework, which effectively obtains image with clear structure and smooth surface. However, since the objective function requires multiple iterative calculations to obtain the optimal solution, the global optimization-based methods generally have problems such as high computational overhead and time-consuming. These shortcomings bring great challenges to real-time interactive operations. Especially for the application of this project that needs to display image contour on the touch screen device through haptic feedback, the image smoothing algorithm based on global optimization will be more difficult to implement quickly and effectively.

The main factors that limit the application of image smoothing operations on touch screen devices are the time and calculations consumed by the smoothing process. In order to ensure the real-time performance of touch screen interaction, this research established a simplified image smoothing model based on CNN under the condition of supervised learning [29].

For a color input image I, a smooth image $L(I)$ is obtained under the filtering of the RTV filter, where $L()$ represents the non-linear filtering process of the RTV. Our goal is that any input image I can approximate the smoothing effect of $L(I)$ through a CNN-based feedforward deep network $F_{\mathbf{W}}(I)$, where F represents the structure of the CNN, and $\mathbf{W}$ represents the network parameters that control the feedforward process. Fig. 1 shows the CNN structure that simulates the smoothing effect of the RTV filter, which is called CNN filter.

In this structure, the input layer is the horizontal and vertical gradient maps $\partial_y I$ and $\partial_x I$ of the input image I, where ∂I represents the gradient map. The convolutional layer 1 contains 256 feature maps, which are obtained by convolving the gradient map ∂I by a convolution kernel with a size of 16×16, and then calculating the tanh function. The function of the convolutional layer 1 is to map each local color block in the gradient map ∂I into a 256-dimensional pixel vector. Then, a convolution kernel with a size of 1×1 is convolved with all the feature maps in the convolutional layer 1, and the convolutional layer 2 is obtained by calculating the tanh function. Therefore, the convolutional layer 2 also contains 256 feature maps. The function of this process is to perform a weighted average of the processed pixel vectors in the convolutional layer 1 to perform a "smoothing" operation. Finally, a convolution kernel with a size of 8×8 is used to perform a convolution operation on the convolutional layer 2 to restore sharp edges, that is, "edge recognition" processing, so as to obtain the final smoothed gradient map.

The structure of the above CNN can be expressed as:

$$\begin{cases} F^0(\partial I) = \partial I \\ F^n(\partial I) = \tanh(\mathbf{W}^n * F^{n-1}(\partial I) + b^n), \ \ n = 1, 2 \\ F_{\mathbf{W}}(\partial I) = \mathbf{W}^n * F^{n-1}(\partial I) + b^n, \ \ n = 3 \end{cases} \tag{1}$$

where n represents the index of the number of layers. $F^0(\partial I)$, $F^1(\partial I)$, $F^2(\partial I)$, and $F_{\mathbf{W}}(\partial I)$ represent the input gradient map, convolutional layer 1, convolutional layer 2, and output gradient map, respectively. $\mathbf{W}^n$ represents the convolution kernel matrix, b^n is the bias parameter, and tanh() is the hyperbolic tangent function. In order to simplify the network, the horizontal and vertical gradients share weights, that is, $\partial_y I$ and $\partial_x I$ use the same $\mathbf{W}^n$. In addition, no pooling layer is used in this CNN. This is because the pooling layer may weaken the location characteristics of the image, which is unfavorable for maintaining the edge of the image.

As a simulation of the smoothing effect of the RTV filter, the CNN model in Fig. 1 selects $\partial L(I)$ as the ground-truth of $F_{\mathbf{W}}(\partial I)$. Therefore, for D training image pairs $(I_0, L(I_0)), (I_1, L(I_1)), \ldots, (I_{D-1}, L(I_{D-1}))$ with ideal smoothing effects, the following energy function is used as the constraint for training CNN:

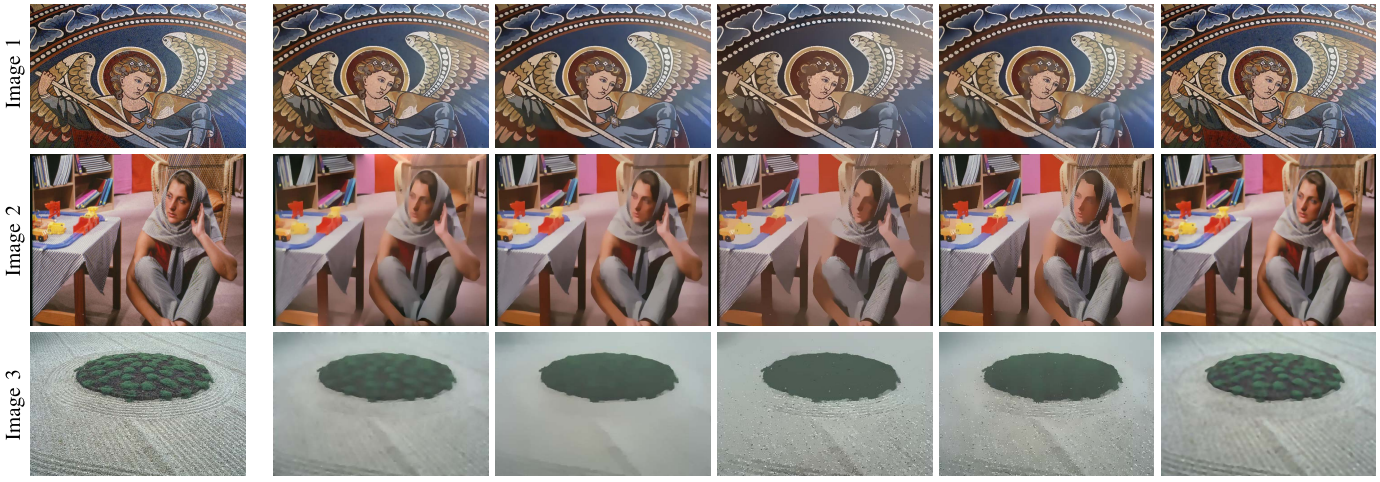
$$\frac{1}{D}\sum_{i=0}^{D}\left\{\frac{1}{2}||F_{\mathbf{W}}(\partial I_i) - \partial L(I_i)||^2 + \lambda\,\Phi(F_{\mathbf{W}}(\partial I_i))\right\} \tag{2}$$

where $(\partial I_i, \ \partial L(I_i))$ is a pair of training examples in the gradient domain, and λ is the regularization weight. $\Phi(F_{\mathbf{W}}(\partial I_i))$ is a regularization term used to sparse the gradient. This item suppresses the color change while edge-preserving smoothing, and helps to strengthen the edge and generate decent weight initialization in the neural network. $\Phi()$ is the Charbonnier penalty function, and $\Phi(x) = \sqrt{(x^2 + \varepsilon^2)}$. The stochastic gradient descent method is used to minimize equation (2).

We use the BSD500 [30] image library as the training sample, and randomly collect one million 64×64 image patches from it. For a given image patch I_i, $L(I_i)$ is first obtained by the RTV filter. Then apply the gradient operator to get ∂I_i and $\partial L(I_i)$. In each training step for one sample, the weight update process of CNN is:

$$\mathrm{W}_{t+1} = \mathrm{W}_t - \eta\left\{(F_{\mathbf{W}}(\partial I_i) - \partial L(I_i))^T + \lambda\frac{(F_{\mathbf{W}}(\partial I_i))^T}{\sqrt{(F_{\mathbf{W}}(\partial I_i))^2 + \varepsilon^2}}\right\}\frac{\partial F_{\mathbf{W}}(\partial I_i)}{\partial \mathrm{W}} \tag{3}$$

where η is the learning rate, which is set to decay from 0.001 during the training process. The gradient is further back propagated through $\partial F_{\mathbf{W}}(\partial I_i)/\partial \mathrm{W}$.



(a) Input image (b) CNN filter (c) RTV filter [28] (d) L_0 filter [31] (e) fast WLS [32] (f) RGF [33]

Fig. 2. Comparison of smoothing effects of three images under five algorithms.

The CNN is trained according to the above process, and the smooth gradient map $F_{\mathbf{W}}(\partial I)$ after training needs to be reconstructed in the color domain. We use S to represent the final output color smooth image. Image reconstruction needs to consider the structural information in the input image to guide the smoothing in the gradient domain. Therefore, these two items are put together:

$$||S - I||^2 + \beta \left\{ ||\partial_x S - F_{\mathbf{W}}(\partial_x I)||^2 + ||\partial_y S - F_{\mathbf{W}}(\partial_y I)||^2 \right\} \tag{4}$$

where $||S - I||^2$ is the color confidence that uses the input image to guide the smoothed image for reconstruction. $||\partial_x S - F_{\mathbf{W}}(\partial_x I)||^2$ and $||\partial_y S - F_{\mathbf{W}}(\partial_y I)||^2$ represent the sum of squared gradient errors of the input and output images in the vertical and horizontal directions, respectively. β is a parameter that balances the two loss functions. For a given test image I, it is first converted to the gradient domain, and ∂I is input to CNN to obtain a smooth gradient map $F_{\mathbf{W}}(\partial I)$. Then use the best β value to solve the image reconstruction formula (4), and the final image S can be obtained.

B. The Effect and Speed of CNN Filter for Image Smoothing

In order to show the smoothing performance of the CNN filter, we compared it with four traditional algorithms in terms of smoothing effect and running time. The four traditional algorithms selected are RTV filter [28], L_0 filter [31], fast WLS [32], and RGF [33]. Each algorithm selects its optimal parameters to acquire the best image smoothing effect. We selected three different types of images and smoothed them under five algorithms. The results are shown in Fig. 2.

In Fig. 2, the Image 1 is a pattern composed of multiple small tiles. The contour of the image has clear edges, but there are obvious gray changes around each tile. By comparing the overall and details of the smoothing result, it can be seen that the RTV filter removes fine texture features very well, and makes the contour of the entire image complete and clear, thus obtaining the best smoothing effect. The smoothing effect of the L_0 filter is not ideal. It has problems such as unclean texture removal, excessive smoothing of local contours, and significant color shift. The fast WLS algorithm also has the problems of unclean texture removal and excessive smoothing of local contours. The RGF algorithm maintains the integrity of the contour very well, but there are problems with unclean texture removal and blurred image. The CNN filter obtains an image smoothing effect similar to the RTV filter, which removes the texture relatively cleanly, while maintaining the integrity and clarity of the contour. Image 2 contains a texture with obvious directionality inside. It can be seen from the smoothing effect that the texture removal effect of the L_0 filter and the fast WLS algorithm is poor, and both have the problem of excessively smoothing the contours. Both the RTV filter and the RGF algorithm effectively remove the texture, but the image will have a certain blur effect. Image 3 is a natural landscape image, and the two areas in the image have strong contrast. Both the L_0 filter and the fast WLS algorithm can smooth the dark green areas in the image, but they still cannot remove the texture well. The RGF algorithm makes the original image more blurred, but it does not get a good smoothing effect. The RTV filter obviously filters out the texture very well and keeps the two regions producing a significant contrast difference, which is very beneficial for contour extraction.

Through the comparison, it can be seen that the RTV filter extracts the structural features of the image very well, and has a good filtering effect on fine textures. Although the L_0 filter and the fast WLS algorithm have a certain removal effect on randomly distributed textures, they both have the problem of excessively smoothing of the contours. In particular, the L_0 filter also causes the color shift of the image. The RGF algorithm as a whole is to blur the original image. The CNN filter obtains a smoothing effect similar to the RTV filter in these three images. Nevertheless, the CNN filter still has unique advantages for this research.

During the smoothing process of each image using each algorithm, we also recorded the time consumed for smoothing. The results of the time statistics are shown in Fig. 3.

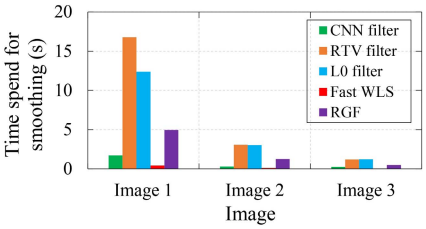


Fig. 3. The time spent to smooth the three images using five algorithms.

The configuration of the server used to run the five algorithms includes i7-6900K CPU and NVIDA TITAN Xp GPU. The sizes of Image 1, Image 2, and Image 3 are 1600×1067, 720×576, and 481×321, respectively.

As shown in Fig. 3, the fast WLS algorithm has the fastest image smoothing speed, while the RTV filter and L_0 filter take longer. The smoothing time of the CNN filter is only about one-tenth that of the RTV filter. Since the image can be smoothed quickly, and a good image smoothing effect can be obtained, the CNN filter has a significant advantage over the other four algorithms. Especially for touch screen devices, their computing power is much lower than the server currently used to run these five algorithms. Therefore, if the CNN filter is transplanted to a touch screen device to run, it can not only obtain a good image smoothing effect, but also meet the real-time requirements of haptic interaction.

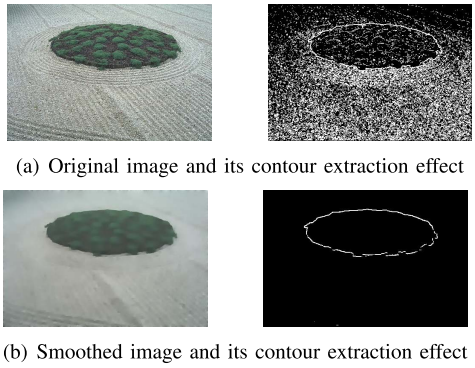
C. Extract Contour Features From Smoothed Image

After training and testing on an external server, we transplanted the proposed CNN model to a touch screen device based on the Android system, and employed the Sobel operator to extract the image contour. Sobel operator is a simple edge detection method with small amount of calculation, fast calculation speed and good edge positioning effect, which is very suitable for running on smart devices. The image processed by the Sobel operator also needs to be binarized in order to convert the grayscale image into a black and white image. The selection of the threshold in this study is automatically determined by the Otsu algorithm [34].

Fig. 4 shows the contour extraction effect of Sobel operator and Otsu algorithm on Image 3. It can be seen that the contours obtained by directly processing the original image are chaotic, while clean and continuous contour features can be extracted after smoothing by CNN filter.

III. Improvement of the Pen-Type Haptic Device

When perceiving the contour of an object through touch, people need to combine kinesthetic feedback to form a "mental imagery" as the fingers follow the contour. For haptically perception of the image contour displayed on the touch screen, it is necessary to add tactile feedback to the finger or tool-mediated contour following process. However, the lower spatial resolution of the finger reduces the bandwidth of the displayable image information [35]. One way to enhance the ability of tactile perception is to use multiple fingers to touch at the same time [36]. Therefore, we improved the previously



(a) Original image and its contour extraction effect

(b) Smoothed image and its contour extraction effect

Fig. 4. Contour of Image 3 extracted by Sobel operator and Otsu algorithm.

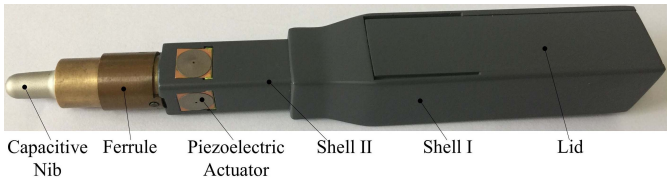


Fig. 5. Prototype of the MH-Pen2.

designed MH-Pen [26] (called MH-Pen2 in this paper, and the prototype is shown in Fig. 5). MH-Pen2 embeds a disc-type piezoelectric actuator on each of the four faces of the Shell II, which provides vibrotactile feedback to the four fingers when a user holds the pen with one hand. This layout enlarges the image information in the small area where the pen tip contacts the touch screen, and provides directional guidance for the user's movement on the touch screen. The size of the piezoelectric actuator (PowerHap 2.5G, TDK Corp., Japan) is $9 \times 9 \times 1.25\ mm^3$, and the maximum vibration acceleration is about 1.68 g ($g = 9.81\ m/s^2$).

In order to guide the user to correctly follow the contour of the virtual object, the interactive system also needs to detect the posture of the MH-Pen2 to drive the piezoelectric actuator in the corresponding position to vibrate. Therefore, we added a posture detection circuit in MH-Pen2. In recent years, low-power, high-sensitivity, and small-sized micro-electromechanical systems (MEMS) have been widely used in 3D motion tracking and attitude capture [37]. MEMS consists of a three-axis accelerometer, gyroscope, and magnetometer. It measures the user's posture by sensing gravity, angular velocity, and geomagnetic field. In this study, a commercial inertial measurement unit (IMU) called MPU6050 (InvenSense, USA) is integrated in the MH-Pen2 to track the pen's posture.

IV. Image Recognition Experiments

We performed two types of user experiments to test the effectiveness of the proposed contour display system.

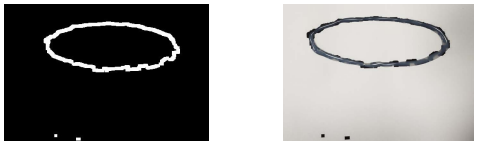
A. Participants and Apparatus

Twenty healthy right-handed subjects (six visually impaired people, seven females, Age: 22.55 ± 2.68, mean ± SD)

participated in the experiment. All subjects gave informed consent and were paid for their participation. Six visually impaired subjects were completely blind, two of them were congenitally blind, and the other four were adventitiously blind after the age of 10. This study invited 14 subjects with normal vision to participate in the experiment to expand the number of samples. Some studies have shown that there is no statistically significant difference between blindfolded-sighted subjects and blind people in learning, integrating, and representing graphical information [6], [38]. Therefore, it is generally accepted in the field to use blindfolded-sighted subjects in the preliminary efficacy test of assistive technology. All subjects had at least received high school education, no subjects reported any defects in their tactile perception ability, and they did not know the purpose of the study. All subjects need to wear eye masks in both experiments, requiring them to perceive the contours of objects only through vibrotactile feedback and kinesthetic feedback. This experimental procedure was approved by the ethics committee of our university.

Experimental apparatus includes the MH-Pen2 and a Samsung Tablet PC. The touch screen has a diagonal length of 10.1 inches, an optical resolution of 2560×1600 pixels, and a screen pixel density of 299 pixels/inch. To reduce the loss of contour tracking when the subject's finger slides on the touch screen, current studies generally set the display width of the detection line on the screen to 0.35 inches (about 9 mm) [13], [14]. However, due to the limited space of the touch screen, a line width of 9 mm will significantly reduce the effective resolution of the touch screen and the complexity of the displayable image. To this end, the literature [39] specifically studied the influence of parameters such as line width and parallel line interval on the BVI perception of vibrating tactile lines on touch screen. They found that for a single simple line detection task, a detection line with a width of 1 mm achieves a resolution similar to a wider detection line. For two detection lines displayed in parallel, a line width of 2 mm and a line spacing of at least 4 mm should be maintained. In addition, literature [40] found that since the diameter of the stylus tip is smaller than the fingertip, when a user tries to touch a small target on the screen, the pen gets higher perceptual accuracy than the fingertip. These research conclusions show that for different types of interaction methods and tools, the setting of the line width is variable, and the use of a pen with a smaller tip diameter to interact with the touch screen improves the perception accuracy. Therefore, for the MH-Pen2 with a pen tip diameter of about 4 mm, we set the width of the detection line to 2 mm.

In the process of continuous or piecemeal touch, BVI combines the kinesthetic information prompts to gradually integrate the extracted information in the brain, thereby establishing a mental image of the object layout. Discontinuous contours or missing traced contours will affect the recognition and understanding of objects. However, in Fig. 4(b), the width of the contour line extracted by the Sobel operator is very narrow, which does not meet the requirements. For this reason, we thicken the contour line to about 2 mm through dilation operation, as shown in Fig. 6(a).



(a) (b)

Fig. 6. Follow-up processing of Fig. 4(b). (a) Thickened contour after dilation operation. (b) A raised line map made by hand with thermoplastic material.

In addition, when the pen tip of MH-Pen2 slides along the contour of the object, it will form a contact circle with the contact point as the center and a diameter of 4 mm in the touch screen. When working, the touch screen detects the sliding direction of the pen tip in real time. According to the direction of the contour line falling inside the contact circle and the posture of the MH-Pen2 detected by the IMU, the system drives one or more piezoelectric actuators at the corresponding positions to generate vibration prompts that are consistent with the direction of the contour line and the direction of the user's movement. The subject adjusted the sliding direction according to the vibration prompts during the contour following process.

Before the formal experiment, each subject was familiar with the use of the entire interactive system through training examples. The training examples utilized images similar to but different from the formal experiment for subjects to perceive. For sighted subjects, they can first see the shape and contour of the image displayed on the touch screen. Then, with visual participation, sighted subjects employed the MH-Pen [26] with no direction prompt function and the MH-Pen2 shown in Fig. 5 to slide along the contour of the object, respectively. The subjects were able to feel the difference in vibrotactile feedback between the two pen-type devices, and learned to combine kinesthetic feedback to gain an understanding of the contours of objects. Finally, as in the formal experiment, sighted subjects were blindfolded and the training example was repeated. For visually impaired subjects, we printed the contour of the image used for training on white paper, and hand-made the raised line map with thermoplastic material, as shown in Fig. 6(b). At the same time, we presented the image contour corresponding to the raised line map on the touch screen. During training, a visually impaired subject can repeatedly touch the raised line map with both hands, or slide along the contour line on the touch screen with MH-Pen and MH-Pen2 respectively, and can ask the experimenter for help at any time. After all subjects confirmed that they were familiar with the use of the two devices, the characteristics of vibrotactile feedback, and the image contour perception process, the training process ended.

B. Experiment 1: Letter Recognition Task

The purpose of this experiment was to test the effect of direction guidance function on contour perception. Two vibrotactile feedback modes were exploited to display the contours of objects. The Mode 1 uses only the linear resonant actuator integrated in the MH-Pen to express the contact state

TABLE I
STATISTICS OF THE CORRECT RATE OF LETTER RECOGNITION (%)

	Mode 1								Mode 2							
Letter	A	B	D	E	F	H	P	R	A	B	D	E	F	H	P	R
A	**60**	0	0	0	0	15	0	10	**100**	0	0	0	0	0	0	0
B	0	**65**	10	20	0	0	0	5	0	**90**	0	5	0	0	0	5
D	0	10	**65**	5	0	0	15	0	0	5	**95**	0	0	0	0	0
E	0	15	0	**55**	0	0	0	5	0	5	0	**90**	0	0	0	0
F	0	0	0	0	**70**	0	5	0	0	0	0	0	**95**	0	0	0
H	20	0	0	0	0	**75**	0	0	0	0	0	0	0	**95**	0	0
P	0	0	10	0	15	0	**70**	10	0	0	0	0	5	0	**100**	0
R	5	0	5	5	0	0	0	**60**	0	0	0	0	0	0	0	**90**

of the pen tip and the contour, that is, vibration occurs as long as the pen tip contacts with the contour, but this mode cannot provide directional hints. The Mode 2 uses the MH-Pen2 to provide directional guidance prompts. In the experiment, eight capitalized English letters with regular shapes and familiar to the subjects were selected as the sensing objects. They are A, B, D, E, F, H, P, and R. The constituent elements of these letters include basic shapes such as straight lines, diagonal lines and arcs, and some letters also have certain similarities, such as A and H, F and P, etc.

All twenty subjects participated in the experiment. The subjects were told in advance that the objects they perceive were capitalized English letters, but they did not know which letters they were. During the experiment, the tablet computer was placed horizontally on the table, and a subject was sitting in front of the table wearing a blindfold and earphones that play white noise. When the subject's posture was fixed, he/she was taught to place his/her left hand on the left side of the tablet to form an external reference frame. In the Mode 1, the subject held the MH-Pen with his/her right hand and judged the contact state with the contour line according to the vibration prompt. In the Mode 2, the subject's right thumb, index finger, middle finger and ring finger respectively held the four piezoelectric actuators on the MH-Pen2. When the subject's right hand holds the MH-Pen2 and slides on the touch screen, his/her left hand moves within a small range to help the right hand form an understanding of the contour relative to the reference frame of the left hand.

First, the Mode 1 was chosen for experimentation. In one trial, a letter was randomly displayed on the touch screen, and the line width of the letter was displayed as 2 mm. The subject could touch the letter multiple times within 3 minutes, and then tell the experimenter what he/she perceived. The experimenter recorded the subject's answers and comments on the perceived effects, and gave a neutral reply. Each subject needed to complete the perception of eight letters in a row. After each trial, the subject could rest for a few minutes. After all subjects used the Mode 1 to perceive the eight letters, the experiment chose the Mode 2 to continue the above experimental steps until the end of the comparison experiment. Table I shows the experimental statistical results of the correct rate of letter recognition in the two modes.

As shown in Table I, Mode 2 achieves a significantly higher recognition accuracy rate than Mode 1. Then, we quantitatively studied the difference between the two modes through paired

TABLE II
PAIRED SAMPLES STATISTICS OF LETTER RECOGNITION ACCURACY

	Mean (%)	N	Std. Dev. (%)	Std. Err. Mean (%)
Mode 1	65	8	6.547	2.315
Mode 2	94.38	8	4.173	1.475

TABLE III
PAIRED T-TEST OF LETTER RECOGNITION ACCURACY

	Pairwise Differences							
	Mean (%)	Std. Dev. (%)	Std. Err. Mean (%)	95% Confid. Interv. Diff. Lower	Upper	t	df	Sig. (2 -tailed)
Mode 1-Mode 2	-29.38	6.23	2.2	-34.59	-24.17	-13.33	7	0.000

T-test analysis. Table II shows the paired samples statistics of letter recognition accuracy, among which the average recognition accuracy rates of Mode 1 and Mode 2 are 65% and 94.38%, respectively. Table III shows the results of the paired T-test analysis of the correct rate of letter recognition. It can be seen that the statistical significance is $p < 0.001$. The results in Table II and Table III indicate that the two modes have a highly significant difference in the recognition accuracy of letters, and the recognition accuracy of Mode 2 is significantly better than Mode 1.

C. Experiment 2: Natural Image Recognition Task

In addition to displaying simple images with regular contour features such as letters, the experiment further tested the performance of MH-Pen2 in helping subjects identify natural image contours. Natural images generally have irregular contour features, such as the contour shown in Fig. 4. We selected eight natural images, as shown in Fig. 7. Each image is first smoothed with the CNN filter, and then the contours in the smoothed image are extracted and the line width of the contours is increased to about 2 mm.

This experiment uses the same two vibrotactile feedback modes as Experiment 1. The experimental scene is shown in Fig. 8. The upper part of the touch screen displays the selected eight natural images, and the lower part displays the bold contours corresponding to the selected image. The subject does not know the information of the perceived image in advance. The experiment randomly divided the subjects into two groups with 10 subjects in each group. Each group included 1 congenital blind person and 2 adventitious blind persons. Since each group of subjects only uses one mode to perceive the contours of the image, this grouping method effectively prevents the subjects from memorizing the contours of the image.

During the experiment, the tablet computer was placed horizontally on the table, and a subject was sitting in front of the table wearing a blindfold and earphones that play white noise. The experimenter randomly selects one of the eight images for the subject to touch, and the perception process is similar to Experiment 1. The subject can use the MH-Pen or the MH-Pen2 to repeatedly touch the contour of the image displayed on the touch screen during the experiment. After the

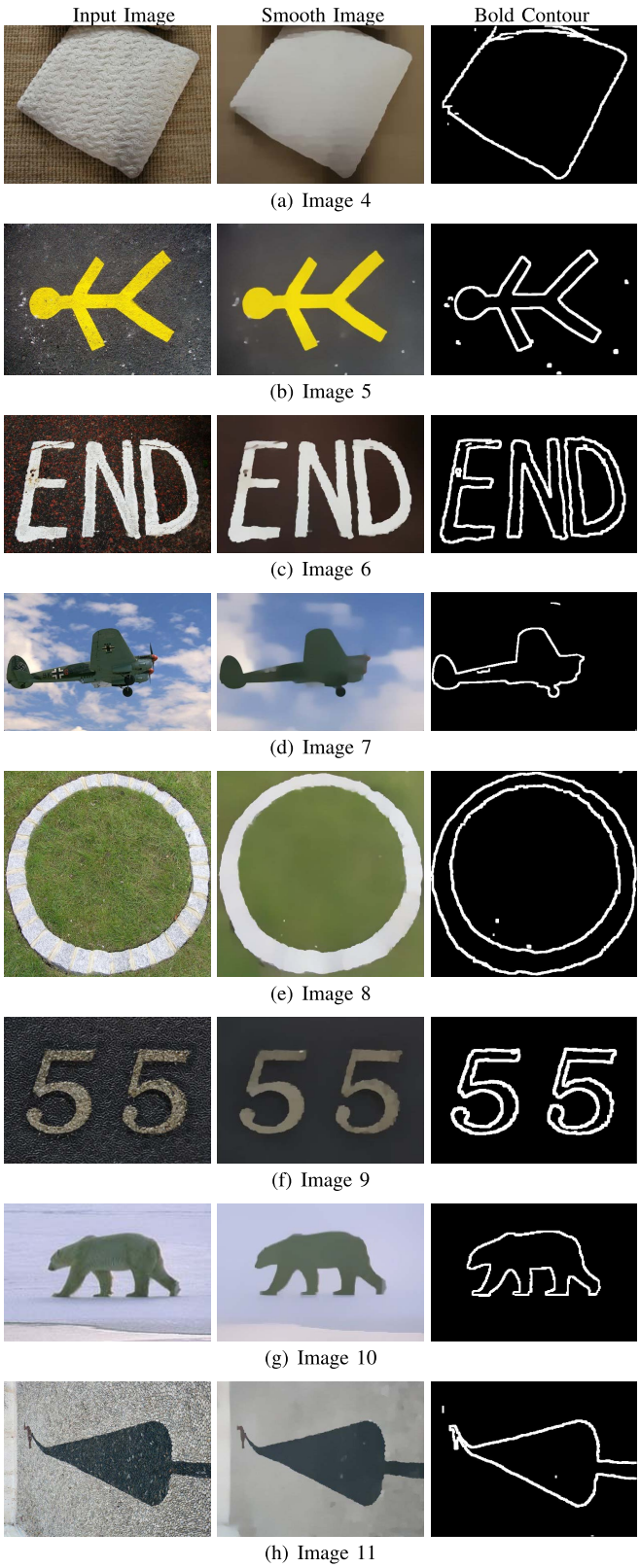


Fig. 7. Eight natural images for contour display.

subject confirmed that the contour on the touch screen has been recognized, the experimenter guided the subject to touch the four raised line maps made of thermoplastic materials with both hands, and asked the subject to choose one of them that

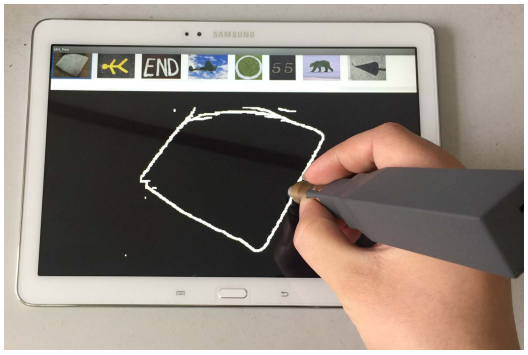


Fig. 8. The experimental scene of Experiment 2.

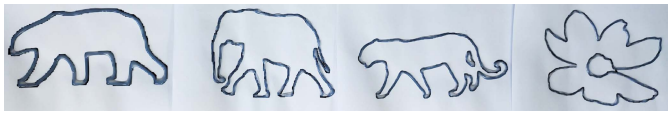


Fig. 9. Four raised line maps provided for the subject to perceive the contour of Image 10.

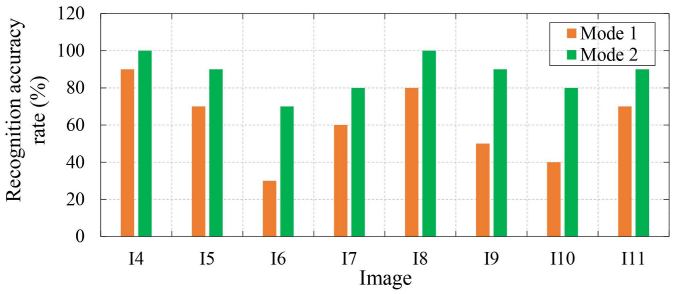


Fig. 10. The correct rate of recognizing the contours of eight natural images in the two modes.

he/she thought was consistent with the contour displayed on the touch screen. Among the four raised line maps, one is exactly the same as the contour of the image displayed on the touch screen, which is the correct answer. The other two have similar contours to the correct answer but also have distinctly different features. The contour of the last one is completely different from the previous three. For example, Fig. 9 shows four raised line maps provided for the subject to perceive the contour of the Image 10. After the subjects repeatedly touched the four raised line maps and made selections, they could comment on the characteristics of the contour and the difficulty of recognition. The experimenter recorded the answers and comments of the subjects and gave a neutral reply. We will not limit the exploratory strategy used by the subject and the time spent. However, the time it takes for the subject to perceive the image contour to when he/she makes a choice will be recorded. Each subject needed to continuously complete the perception of eight image contours. After each trial, the subject could rest for a few minutes. After all subjects have completed the perception of image contours, the entire experiment is over. Fig. 10 shows the accuracy of the subjects' recognition of eight natural image contours in the two modes.

It can be seen from Fig. 10 that Mode 2 still achieves a better recognition accuracy rate than Mode 1. Next, we

TABLE IV
PAIRED SAMPLES STATISTICS OF NATURAL IMAGE RECOGNITION ACCURACY

	Mean (%)	N	Std. Dev. (%)	Std. Err. Mean (%)
Mode 1	61.25	8	20.31	7.18
Mode 2	87.5	8	10.35	3.66

TABLE V
PAIRED T-TEST OF NATURAL IMAGE RECOGNITION ACCURACY

	Pairwise Differences							
	Mean (%)	Std. Dev. (%)	Std. Err. Mean (%)	95% Confid. Interv. Diff. Lower	Upper	t	df	Sig. (2 -tailed)
Mode 1- Mode 2	-26.25	11.88	4.2	-36.18	-16.32	-6.25	7	0.000

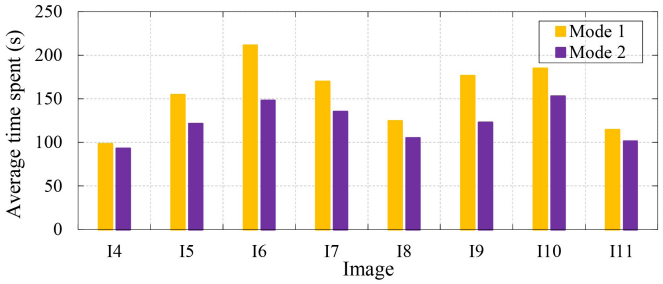


Fig. 11. The average time the subjects spent perceiving each natural image in the two modes.

quantitatively studied the difference between the two modes through paired T-test analysis. The results of the analysis are shown in Table IV and Table V. It can be seen from the two tables that the average recognition accuracy rates of Mode 1 and Mode 2 are 61.25% and 87.5% respectively, and the statistically significant is $p < 0.001$. The results show that there is a highly significant difference in the accuracy of the subjects' recognition of the eight natural image contours in the two modes, and the recognition accuracy of Mode 2 is significantly better than Mode 1.

In addition, we also counted the time it took for each subject to perceive each image. Fig. 11 shows the average time the subjects spent perceiving each natural image in the two modes. It can be seen that it generally took less time for the subjects to perceive image contours in Mode 2 than in Mode 1. Further, Fig. 12 shows the distribution of the time spent by subjects perceiving images in the two modes. First of all, we conducted a two-way analysis of variance (ANOVA) on the time spent with the mode and image as factors. The results showed that both the mode ($F(1, 144) = 65.849$, $p < 0.001$) and the image ($F(7, 144) = 29.157$, $p < 0.001$) have a highly significant effect on the time spent. Then, a one-way ANOVA was performed on the mode to determine which images the subjects perceive were significantly affected by the mode differences. The results show that the mode difference has a certain effect on the subject's perception of Image 5 ($F(1, 18) = 5.661$, $p = 0.029$) and Image 10 ($F(1, 18) = 7.431$, $p = 0.014$), has a significant effect on the subject's

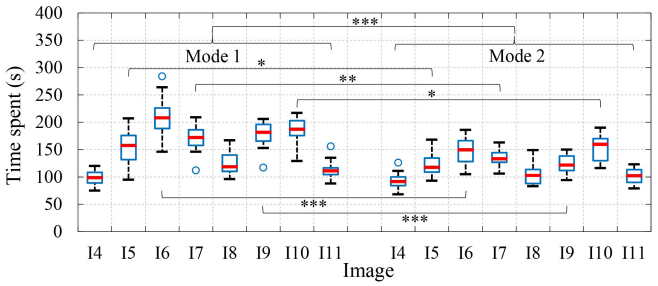


Fig. 12. The distribution of the time spent by subjects perceiving images in the two modes. The marks (∗), (∗∗), (∗ ∗ ∗) denote a significance of $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively.

perception of Image 7 ($F(1, 18) = 11.392$, $p = 0.003$), and has a highly significant effect on the subject's perception of Image 6 ($F(1, 18) = 17.37$, $p < 0.001$) and Image 9 ($F(1, 18) = 25.52$, $p < 0.001$).

D. Discussion

In this section, two experiments were performed to test the influence of directional vibrotactile feedback on the subject's perception of image contour information on the touch screen. The Experiment 1 uses capital letters that are familiar to the subjects and have regular shapes as the sensing objects. The purpose is to find out how the motion direction guidance helps the subjects to better recognize the contour of the image. According to the subject's comments, Mode 2 has the following advantages over Mode 1 in contour tracking: (1) The guidance function of Mode 2 makes it easier for subjects to follow the contour lines. (2) Mode 2 helps subjects better understand the geometric relationship between adjacent lines, such as whether two lines intersect, and how the direction of the contour line changes after the intersection. (3) The contour felt by the subjects through Mode 2 is not easy to change with respect to the external reference frame of the left hand, which makes it easier to form a clear mental image. The above three advantages are essential to correctly perceive the contour of the image on the touch screen.

As mentioned above, the common commercial touch screen is a flat, featureless surface. These limitations make it easy for Mode 1 to lose the following of the contour, especially when following arcs or intersecting lines, subjects are prone to make mistakes. Loss of following contours, probing in multiple directions after reaching the intersection point, and random changes in the left-hand reference frame all increase the difficulty of temporal and spatial integration of tactile information, and bring confusion to the formation of mental image. The result of Experiment 1 shows that the motion guidance function of MH-Pen2 effectively helps people perceive and understand the image contours on common commercial touch screens.

In Experiment 2, the average recognition accuracy rate of 87.5% shows that MH-Pen2 can be used to help subjects recognize the contours of complex images. The results in Figures 10-12 show that the contour complexity of the image and the content of the image will affect the perception of the subject. For Images 4, 8 and 11, the simple contours enable the

subjects to obtain a higher recognition accuracy rate in both modes and spend less time. For Images 5 and 10 with more complex contours, the mode difference has a certain effect on the time spent for the subject to perceive the image. However, the recognition accuracy of the two images in different modes is significantly different. The "human shape" in Image 5 and the "bear" in Image 10 are the contours familiar to the subjects, but the detailed features of the "bear" are more complex than the "human shape". In addition, it takes longer on average for the subjects to perceive Image 10 than Image 5 (see Fig. 11). These results indicate that, firstly, the time spent by subjects in the two modes is affected by the contour complexity of the image; secondly, for the contour familiar to the subjects, the more complex the contour is, the more Mode 2 helps the subjects better identify object through association.

Similar to Image 10, Images 6 and 9 also have complex contours, and the subjects' recognition accuracy of the two images is also significantly different in the two modes. However, the mode difference has a highly significant effect on the time spent for subjects to perceive Images 6 and 9. Although Image 6 is more complicated than Image 10 (see Fig. 7), it takes more time for subjects to perceive Image 6 than Image 10 in Mode 1, and less time in Mode 2. This indicates that the time spent in Mode 1 increases with the increase of image complexity, but the time spent in Mode 2 is not entirely determined by the image complexity. Observing Image 6 and Image 9, we can see that the contents in them are letters and numbers that are easy to understand. As long as the subjects follow the contours of the two images correctly, they can quickly build a correct mental image without paying too much attention to details. This further illustrates the advantages of Mode 2 to improve the image recognition rate and reduce the time spent by helping the subjects understand the image.

For Image 7, the recognition accuracy of the subjects in the two modes is not much different, but the average is only 70%, and the mode difference has a significant effect on the time spent for the subjects to perceive the image. This may be because the subject cannot accurately understand the object represented by the contour in Image 7, although its complexity is not great. Even so, Mode 2 still helped the subjects to recognize the contours of the image faster and better. In general, the complexity and content of the image contour have a significant impact on the result of tactile perception, and the higher the complexity of the image and the easier to understand the content of the image, the more obvious the advantages of Mode 2 over Mode 1.

V. Conclusion

Based on the human's tactile physiological characteristics, this study proposes a new method for BVI to effectively perceive image contour information by applying deep learning and haptic display techniques to daily-used touch screen devices. This paper first uses CNN-based image smoothing filter to remove detailed features in the image that are unhelpful for contour perception. The CNN filter can well meet the application on touch screen devices and ensure real-time image processing. Then, the contour information of the smoothed image is extracted and converted into a form suitable for expression through vibration. Finally, two user experiments are used to verify the effectiveness of the motion direction guidance function in helping subjects to quickly and accurately identify the contours of the image. In general, the image contour display system proposed in this paper makes full use of the portability, easy accessibility and data refreshability of touch screen devices, solves the defects of touch screen devices in conveying image contour information to BVI, and provides a more convenient way for the barrier-free communication between BVI and the digital world.

References

[1] Z. Cattaneo *et al.*, "Imagery and spatial processes in blindness and visual impairment," *Neurosci. Biobehavioral Rev.*, vol. 32, no. 8, pp. 1346–1360, Oct. 2008.

[2] F. Leo, E. Cocchi, and L. Brayda, "The effect of programmable tactile displays on spatial learning skills in children and adolescents of different visual disability," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 7, pp. 861–872, Jul. 2017.

[3] J. M. Loomis, R. L. Klatzky, and N. A. Giudice, "Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language," in *Multisensory Imagery*, S. Lacey and R. Lawson, Eds. New York, NY, USA: Springer, 2013, pp. 131–155.

[4] R. L. Klatzky, J. M. Loomis, S. J. Lederman, H. Wake, and N. Fujita, "Haptic identification of objects and their depictions," *Perception Psychophysics*, vol. 54, no. 2, pp. 170–178, Mar. 1993.

[5] D. T. V. Pawluk, R. J. Adams, and R. Kitada, "Designing haptic assistive technology for individuals who are blind or visually impaired," *IEEE Trans. Haptics*, vol. 8, no. 3, pp. 258–278, Jul. 2015.

[6] T. P. Way and K. E. Barner, "Automatic visual to tactile translation—Part I: Human factors, access methods and image manipulation," *IEEE Trans. Rehabil. Eng.*, vol. 5, no. 1, pp. 81–94, Mar. 1997.

[7] M. Shinohara, Y. Shimizu, and A. Mochizuki, "Three-dimensional tactile display for the blind," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 249–256, Sep. 1998.

[8] M. Biet, F. Giraud, and B. Lemaire-Semail, "Squeeze film effect for the design of an ultrasonic tactile plate," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 54, no. 12, pp. 2678–2688, Dec. 2007.

[9] K. A. Kaczmarek and S. J. Haase, "Pattern identification as a function of stimulation on a fingertip-scanned electrotactile display," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 3, pp. 269–275, Sep. 2003.

[10] F. Vidal-Verdú and M. Hafez, "Graphical tactile displays for visually-impaired people," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 119–130, Mar. 2007.

[11] W. Grussenmeyer and E. Folmer, "Accessible touchscreen technology for people with visual impairments: A survey," *ACM Trans. Accessible Comput.*, vol. 9, no. 2, pp. 1–31, Jan. 2017.

[12] N. A. Giudice, H. P. Palani, E. Brenner, and K. M. Kramer, "Learning non-visual graphical information using a touch-based vibro-audio interface," in *Proc. 14th Int. ACM SIGACCESS Conf. Comput. accessibility (ASSETS)*, 2012, pp. 103–110.

[13] R. L. Klatzky, N. A. Giudice, C. R. Bennett, and J. M. Loomis, "Touch-screen technology for the dynamic display of 2D spatial information without vision: Promise and progress," *Multisensory Res.*, vol. 27, nos. 5–6, pp. 359–378, 2014.

[14] P. M. Silva, T. N. Pappas, J. Atkins, and J. E. West, "Perceiving graphical and pictorial information via hearing and touch," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2432–2445, Dec. 2016.

[15] J. Tekli, Y. B. Issa, and R. Chbeir, "Evaluating touch-screen vibration modality for blind users to access simple shapes and graphics," *Int. J. Hum.-Comput. Stud.*, vol. 110, pp. 115–133, Feb. 2018.

[16] C. Goncu and K. Marriott, "GraVVITAS: Generic multi-touch presentation of accessible graphics," in *Proc. IFIP Conf. Hum.-Comput. Interact. (INTERACT)*, Sep. 2011, pp. 30–48.

[17] L. Winfield, J. Glassmire, J. E. Colgate, and M. Peshkin, "T-PaD: Tactile pattern display through variable friction reduction," in *Proc. 2nd Joint EuroHaptics Conf. Symp. Haptic Interfaces Virtual Environ. Teleoperator Syst. (WHC)*, Mar. 2007, pp. 421–426.

[18] O. Bau, I. Poupyrev, A. Israr, and C. Harrison, "TeslaTouch: Electrovibration for touch surfaces," in *Proc. 23nd Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2010, pp. 283–292.

[19] H. Xu, M. A. Peshkin, and J. E. Colgate, "UltraShiver: Lateral force feedback on a bare fingertip via ultrasonic oscillation and electroadhesion," in *Proc. IEEE Haptics Symp. (HAPTICS)*, Mar. 2018, pp. 198–203.

[20] L. Pantera and C. Hudin, "Multitouch vibrotactile feedback on a tactile screen by the inverse filter technique: Vibration amplitude and spatial resolution," *IEEE Trans. Haptics*, vol. 13, no. 3, pp. 493–503, Jul. 2020.

[21] S. J. Lederman and R. L. Klatzky, "Haptic perception: A tutorial," *Attention, Perception, Psychophys.*, vol. 71, no. 7, pp. 1439–1459, 2009.

[22] S. O'Modhrain, N. A. Giudice, J. A. Gardner, and G. E. Legge, "Designing media for visually-impaired users of refreshable touch displays: Possibilities and pitfalls," *IEEE Trans. Haptics*, vol. 8, no. 3, pp. 248–257, Jul. 2015.

[23] K. U. Kyung and J. Y. Lee, "Ubi-Pen: A haptic interface with texture and vibrotactile display," *IEEE Comput. Graph. Appl.*, vol. 29, no. 1, pp. 56–64, Jan. 2009.

[24] G. Wintergerst, R. Jagodzinski, F. Hemmert, A. Müller, and G. Joost, "Reflective haptics: Enhancing stylus-based interactions on touch screens," in *Proc. Int. Conf. Hum. Haptic Sens. Touch Enabled Comput. Appl.*, vol. 6191, 2010, pp. 360–366.

[25] L. Tian, A. Song, and D. Chen, "Image-based haptic display via a novel pen-shaped haptic device on touch screens," *Multimedia Tools Appl.*, vol. 76, no. 13, pp. 14969–14992, Jul. 2017.

[26] D. Chen, A. Song, L. Tian, Y. Yu, and L. Zhu, "MH-Pen: A pen-type multi-mode haptic interface for touch screens interaction," *IEEE Trans. Haptics*, vol. 11, no. 4, pp. 555–567, Oct./Dec. 2018.

[27] Y. Kim, D. Min, B. Ham, and K. Sohn, "Fast domain decomposition for global image smoothing," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4079–4091, Aug. 2017.

[28] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, Art. 139, pp. 1–10, 2012.

[29] L. Xu, J. Ren, Q. Yan, R. Liao, and J. Jia, "Deep edge-aware filters," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 1669–1678.

[30] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Jul. 2001, pp. 416–423.

[31] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," in *Proc. SIGGRAPH Asia Conf.*, 2011, pp. 1–12.

[32] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. Do, "Fast global image smoothing based on weighted least squares," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5638–5653, Dec. 2014.

[33] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2014, pp. 815–830.

[34] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[35] J. M. Loomis, R. L. Klatzky, and N. A. Giudice, "Sensory substitution of vision: Importance of perceptual and cognitive processing," in *Assistive Technology for Blindness and Low Vision*, R. Manduchi and S. Kurniawan, Eds. Boca Raton, FL, USA: CRC Press, 2012, pp. 162–191.

[36] V. S. Morash, A. E. C. Pensky, S. T. W. Tseng, and J. A. Miele, "Effects of using multiple hands and fingers on haptic performance in individuals who are blind," *Perception*, vol. 43, no. 6, pp. 569–588, Jun. 2014.

[37] X. Yun and E. R. Bachmann, "Design, implementation, and experimental results of a quaternion-based Kalman filter for human body motion tracking," *IEEE Trans. Robot.*, vol. 22, no. 6, pp. 1216–1227, Dec. 2006.

[38] M. A. Heller, "Tactile picture perception in sighted and blind people," *Behavioural Brain Res.*, vol. 135, nos. 1–2, pp. 65–68, Sep. 2002.

[39] H. P. Palani, J. L. Tennison, G. B. Giudice, and N. A. Giudice, "Touchscreen-based haptic information access for assisting blind and visually-impaired users: Perceptual parameters and design guidelines," in *Proc. Int. Conf. Appl. Hum. Fact. Ergonom.*, 2018, pp. 837–847.

[40] X. Ren and S. Moriya, "Improving selection performance on pen-based systems: A study of pen-based interaction for selection tasks," *ACM Trans. Comput.-Hum. Interact.*, vol. 7, no. 3, pp. 384–416, Sep. 2000.