



# A Local Occlusion Face Image Recognition Algorithm Based on the Recurrent Neural Network

Xing-hua Lu<sup>(✉)</sup>, Ling-feng Wang, Ji-tao Qiu, and Jing Li

Huali College Guangdong University of Technology, Guangzhou 511325, China  
luxinghua5454565@163.com

**Abstract.** The recognition rate of traditional face recognition algorithm to the face image with occlusion is not high, resulting in poor recognition effect. Therefore, this paper proposes a partial occlusion face recognition algorithm based on recurrent neural network. According to the different light sources, the high filtering function is used to analyze the halo effect of the image, realize the preprocessing of partially occluded face image, set up the global face feature area and the local face feature area according to the image features, and extract the global and local features of the image; based on the time and structure features of the recursive neural network, establish the local subspace, and realize the local face image recognition Law. The experimental results show that: compared with the traditional algorithm, the face recognition algorithm studied in this paper has a higher recognition rate, and can accurately recognize the partially occluded face image, which meets the basic requirements of the current face image recognition.

**Keywords:** Recurrent neural network · Partially occluded face image · Recognition algorithm

## 1 Introduction

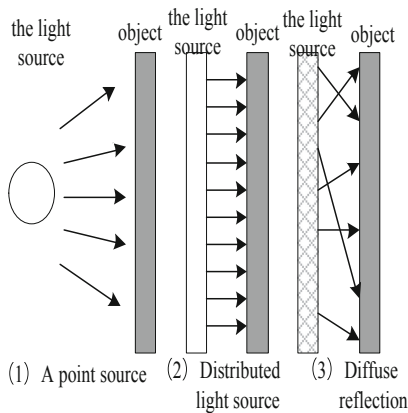
At present, with the continuous improvement of technology level, face recognition technology is more and more applied to the field of unlocking. However, when face recognition system scans and recognizes the face, because the face is blocked by gauze, mask, face towel, scarf, glasses and sunglasses and other objects, the local position of the human image to be recognized is blocked, making the system unable to recognize accurately. In order to solve this problem, convolutional neural network and block image feature extraction are combined. This algorithm improves the effective recognition rate of the recognition system, but it does not achieve this goal. Therefore, based on recurrent neural network, this paper studies the recognition algorithm of partially occluded face image which can solve this problem. This algorithm takes the existing problems of the conventional algorithm as the breakthrough point, greatly improves the recognition rate of the face image recognition algorithm to the partially occluded human image, and provides technical support for the normal work of the face recognition system [1].

## 2 Recognition Algorithm of Partially Occluded Face Image Based on Recurrent Neural Network

In the process of image recognition, recurrent neural network is the premise of key recognition algorithms. It is defined as a complete face image, which is divided into many different connected domains according to the structural characteristics of recurrent neural network. The pixel features in each region are similar, that is to say, this region is homogeneous and heterogeneous compared with other adjacent regions. The use of this network is the key step of facial expression analysis and appearance recognition. Therefore, it is an important and necessary part in the local occlusion face recognition system. It is the most difficult part in the process of human image processing, which determines the quality of the final result of human image recognition [2].

### 2.1 Lumination Preprocessing of Partially Occluded Face Image

The illumination preprocessing of face image can not only eliminate the influence of non-uniform illumination and local occlusion on the extraction of face features, but also eliminate the high-dimensional redundancy of face image, and extract the feature vector matrix of low-dimensional subspace as the criteria of face matching recognition, which can greatly improve the speed of matching and recognition, and is conducive to face recognition. Therefore, face image illumination preprocessing plays a key role in face recognition. In many cases, image changes caused by light changes are more significant than image changes caused by different identities. Common light source classifications are shown in Fig. 1.



**Fig. 1.** Common effects on the light source

Different light sources will produce different effects. The influence of light often leads to inaccurate facial feature extraction, which leads to low recognition rate. The commonly used methods of illumination processing are illumination cone method,

spherical harmonic function subspace method and self quotient image method. The illumination cone method can generate the virtual image under any light source by changing the direction of the light source, but the reconstructed face image is not very characteristic. Spherical harmonic function method is to use multiple spherical harmonic base images to represent the original image. Although the spherical harmonic function method has good processing effect on uneven light, it increases the recognition task and prolongs the recognition time. The self quotient image method is used to divide the original image and Gaussian filter image point by point, which can effectively outline the edge information of the image and unify the brightness and darkness of the image [3]. Because of the excellent algorithm performance of self quotient image, the principle of self quotient image processing non-uniform illumination is applied to the structure of progressive neural network. The basic idea of self quotient image comes from quotient image. The premise of quotient image theory is that the samples have three-dimensional appearance and different texture information, and the face image meets such conditions. Set  $M$  as a face image, select weighted Gaussian filter to check the original image for anisotropic filtering:

$$\hat{M} = F * M \tag{1}$$

In the formula:  $F$  is convolution kernel; the convolution window selected is  $5 \times 5$ . The selected weighted Gaussian filter kernel meets the following requirements:

$$\frac{1}{n} \sum_{\varepsilon} WP = 1 \tag{2}$$

In the formula:  $n$  is the number of convolution window pages;  $\varepsilon$  is the size of convolution kernel;  $W$  is the label of region division;  $P$  is Gaussian kernel. Combining the above two formulas, the self quotient image is defined as the quotient of the original image and the filtered image:

$$R = \frac{M}{\hat{M}} = \frac{M}{F * M} \tag{3}$$

Because the non-uniform illumination has a greater impact on the edge information of human face, in order to better preserve the reflectivity information and reduce the halo effect on the stepped edge, the convolution region is divided into two sub regions  $A_1$  and  $A_2$  by using the threshold value  $\lambda$ , and the label value of the sub region is set, and the calculation expression is as follows:

$$G(i,j) = \begin{cases} \lambda = Mean(M_{\varepsilon}) \\ \begin{cases} 0 & M(i,j) = A_2 \\ 1 & M(i,j) = A_1 \end{cases} \end{cases} \tag{4}$$

In the formula:  $Mean(M_{\varepsilon})$  represents the spectral variation function;  $G(i,j)$  represents the sub region where the abscissa is  $i$  and the ordinate is  $j$ . If there is an edge in

the convolution region, the threshold will divide the local image into two parts  $A_1$  and  $A_2$  along the edge. This segmentation method is based on the pixel value, so the pixel values of the two regions can be significantly different. In anisotropic filtering, only one side of the edge is selectively filtered, which can effectively reduce the halo effect caused by the stepped edge, and then achieve effective light processing. However, the original image and filtered image often enlarge the high-frequency noise in the point by point division operation. The solution to this problem is to transform the self quotient image nonlinear

$$G'(i,j) = D(R) \quad (5)$$

There are many kinds of nonlinear function  $D$ . Logarithm function is commonly used. The reflectance information in the image that does not follow the light can not only effectively reduce the edge information of the stepped face image. Self quotient image is effective enough to extract face image, which can effectively reduce the halo effect of stepped edge, and can outline the edge information of face image [4].

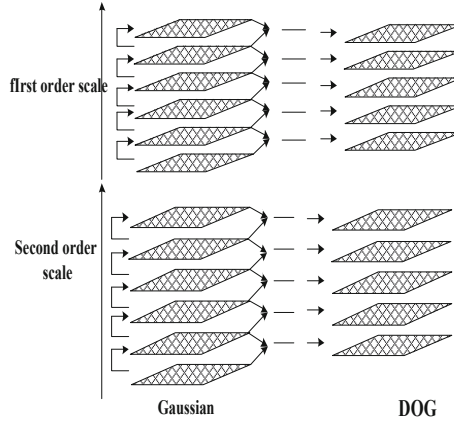
## 2.2 Feature Extraction of Face Image

After the preprocessing of human image, the feature of face image is extracted, which is regarded as the bottom visual feature of image. According to the region of interest, image features can be divided into global features and local features. Global features describe the whole image, including color features, texture features and shape features. The calculation is relatively simple, but the biggest problem of global features is that it can not accurately describe the local features of the image. The extracted features have a lot of redundant information and are not suitable for fine-grained image classification tasks. Local feature refers to selecting important feature points to represent the target features in the local region of interest of the image, which is the local expression of the image features, reflecting the local non deformation of the image, and also can restore the overall information well, which has strong robustness to changes in scale, illumination, affine, etc., and is widely used in the field of target recognition. Local features include SIFT features, directional gradient histogram features, surf features and so on. The interest feature of SIFT algorithm is local area. Through the transformation of multi-scale mapping space, the scale invariant feature points of local area human image are extracted, which has high robustness for the change of light, noise, angle of view and affine transformation, at the same time, it has strong scalability and can solve the problem of image recognition in complex scenes such as light influence, object occlusion, noise, debris to a certain extent. It is a classical method in visual feature extraction [5]. The key steps of feature extraction are as follows:

Detect the extremum of scale space. According to the key steps of sift operator, key points are detected on different scale images. According to the scale space theory, the input image can be transformed into image feature points in different scale space, where the scale space kernel is Gaussian function. Suppose  $M(x,y)$  is the input image and  $G(x,y,\sigma)$  is the multi-scale Gaussian kernel function. The scale space of the image can be defined by the following formula:

$$\begin{cases} G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \\ L(x, y, \sigma) = G(x, y, \sigma) * M(x, y) \end{cases} \quad (6)$$

In the formula,  $*$  is the symbol of convolution operation;  $L(x, y, \sigma)$  is the scale space of the image;  $\sigma$  is the size of the scale space;  $e$  is the fixed exponential function. When  $\sigma$  is larger, it means that the image is more blurred, and its scale space is larger. When  $\sigma$  is smaller, it means that the image is clearer, and its scale space is smaller. The details of image layer are as shown in Fig. 2:



**Fig. 2.** Schematic diagram of generation of Gaussian difference pyramids

Considering the problem of time complexity and space complexity, the Gaussian difference operator is introduced and defined as follows:

$$\begin{aligned} D(x, y, \sigma) &= [G(x, y, \beta\sigma) - G(x, y, \sigma)] * M(x, y) \\ &= L(x, y, \beta\sigma) - L(x, y, \sigma) \end{aligned} \quad (7)$$

In the formula:  $\beta$  is a constant multiplication factor, the introduction of Gauss difference pyramid  $D(x, y, \sigma)$  can increase the stability of detecting the effective extreme points in different scale space. According to the right side of the generation diagram of the Gaussian difference pyramid in Fig. 2, for the adjacent two-layer image, the Gaussian difference image can be obtained by subtracting the two images. In order to calculate the local extremum of the Gauss difference pyramid  $D(x, y, \sigma)$ , each pixel point should be compared with all its neighboring points to determine the size of its neighboring points on the same image scale, the previous image scale, i.e. the first and next image scales, i.e. the second order. If the pixel point is the maximum or minimum value of its neighborhood, then the pixel point can be considered as a local extremum point, i.e. a local extremum point potential key points not filtered [6].

After the operation of the above steps, local extremum points under different scales have been detected. However, in order to locate its location and determine its scale more accurately, it is necessary to screen the potential key points and use the scale space *DOG* function to fit the curve, so as to enhance the matching stability and improve the anti noise ability. The Taylor expansion of *DOG* functions in scale space is as follows:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{8}$$

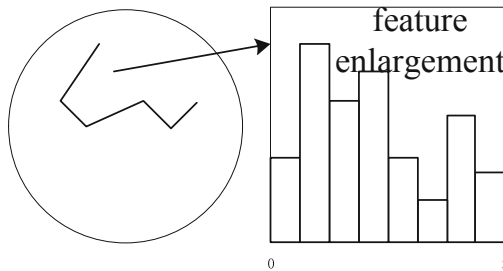
In the formula:  $X=(x, y, \sigma)^T$  represents the offset of pixel detection point, and the offset of extreme point can be obtained as follows:

$$\hat{X} = - \frac{\partial^2 D^{-1}}{\partial X^2} \cdot \frac{\partial D}{\partial X} \tag{9}$$

The above equation is substituted into the formula (6), and in order to remove the extreme point with a small contrast, the value of the extreme value is generally set to 0.03, and there is:

$$D(\hat{X}) = D + \frac{1}{2} \times \frac{\partial X^T}{\partial X} \hat{X} \tag{10}$$

Since the extreme value of operator *DOG* has a large principal curvature at the edge, but relatively small in the vertical direction, in order to eliminate the unstable extreme points at the edge, a  $2 \times 2$  Hessian matrix can be used to calculate the principal curvature at the location and key point scale respectively, and the default value is 10 [7]. In order to make the operator rotation invariant, it is necessary to determine a main direction for the key points obtained through the above steps, which is based on the gradient direction of each pixel in the neighborhood of the key points, as shown in Fig. 3.



**Fig. 3.** Principle diagram of main direction determination of feature points

Firstly, the sampling operation is carried out in the neighborhood of key points, and then the direction histogram is obtained by histogram statistics according to the gradient direction, the maximum value in the histogram is the main direction of the corresponding feature points. According to the above steps, the main direction of the key can be determined, and the coordinate axis can be rotated to keep its direction consistent with the main direction of the key. Figure 4 below is the key description diagram.

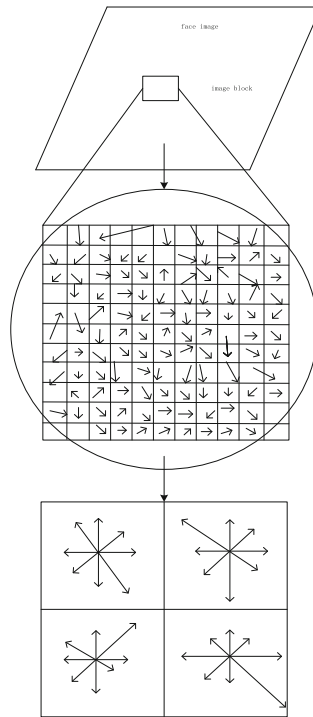


Fig. 4. Key point description sub diagram

According to the above figure, assign the sampling points in the neighborhood of key points to the sub area, create the gradient direction histogram of 8-column elements, i.e. 8 directions, and calculate the gradient direction in the sub area to form 4 seeds, which generates the feature description vector with  $4 \times 4 \times 8 = 128$  dimensional key points. Finally, through normalization and sorting operations, get the face image vector of SIFT feature description [8].

### 2.3 Setting up Image Recognition Algorithm Based on Recursive Neural Network

After the above operations, using the characteristics of recurrent neural network, the associative memory face recognition is established, and the image recognition algorithm is set up. We give the algorithm of image recognition based on the method of self

associative memory and different associative memory of recurrent neural network. Our algorithm mainly obtains the appropriate network weight through training recurrent neural network, remembers the facial features with the attractor of the network, so as to achieve the purpose of face recognition. After the input and output modes are represented as vector arrays, the weights of the network are calculated directly, which reduces the calculation amount; because the network has  $(2s)^n$  balance points, it can remember a large number of face features, so it is not necessary to retrain the network for different features. The recurrent neural network used in the recognition algorithm is described by the following differential equations:

$$\begin{cases} \frac{dx_i(t)}{dt} = -x_i(t) + \sum_{j=1}^n u_{ij}f(x_j(t)) + \sum_{j=1}^n v_{ij}f(x_j(t - \varphi_{ij}(t))) + k_i \\ y_i = f(x_j(t)) \end{cases} \quad (11)$$

In the formula:  $x_i(t)$  is the state of the network;  $u_{ij}$ ,  $v_{ij}$  is the weight of the network;  $\varphi_{ij}$  is the time-varying delay that may occur during the operation of the network;  $k_i$  is the external input of the network;  $y_i = f(x_j(t))$  is the output of the network. This formula is a classic neural network model. The existing literature has analyzed the neural network in detail, and obtained that the network has  $(2s)^n$  local exponential stable equilibrium points [9].

According to the structural characteristics of recurrent neural network, it uses down sampling to reduce the dimension and detect the occluded area in the face image. Although down sampling may reduce the image texture information and the face recognition rate, it can effectively reduce the dimension of the face image. In this algorithm, the down sampling recurrent neural network is not only used for face recognition [10, 11], but also used to detect the occluded coverage, it makes up for the time-consuming shortcomings of the conventional sparse recognition algorithm in processing high-dimensional data, and can keep the overall texture structure of the face unchanged, and can detect the occluded or damaged areas of the face relatively accurately. As shown in Fig. 5 below.



**Fig. 5.** Local recognition image

It can be seen from the above figure that (a) is the original test sample; (b) represents that the original sample is dimensionally reduced according to certain rules to reduce the calculation amount; (c) represents the occlusion pixel filter, which is the binary image of the detected possible occlusion area; and (d) represents the reconstructed test sample. Although the dimension of the original face data is relatively high, it is generally believed that the face data is actually distributed in some low-dimensional manifold spaces [12], the most typical of which is the linear subspace. We use recurrent neural network to build a special linear subspace, that is, the reconstruction subspace. A human face image vector can be expressed linearly as follows with  $x \in R^M$ :

$$x = \bar{x} + Ps \tag{12}$$

In the formula:  $\bar{x}$  represents the mean value of all samples;  $s \in R^\mu$  represents the sparse vector;  $P \in R^{M \times \mu}$  represents the subspace matrix, and each image is composed of the eigenvectors corresponding to the non-zero eigenvalue of the covariance matrix;  $\mu$  represents the subspace dimension; because  $P$  is the orthogonal matrix,  $\bar{x}$  and  $P$  are constants, and each face image vector  $x$  corresponds to the unique coefficient vector  $s$ . Based on orthogonality [13],  $s$  can be directly obtained by linear projection:

$$s = P^T(x - \bar{x}) \tag{13}$$

Formula (10) can also be described as a system of equations, and  $s$  is an unknown quantity. Generally, there is  $M \in \mu$ , so this equation group is an overdetermined equation, that is, the face image vector  $x$  is sufficient to determine the value of  $s$ . Let  $x_{ab} = Wx$ , of which  $W \in \{0, 1\}^{D \times M}$ , be the filter occlusion matrix obtained from the occluded image after expansion corrosion [14], so the equations based on the detection of non occluded pixels are as follows:

$$W(x - \bar{x}) = WP s \tag{14}$$

The least square approximation of the norm is as follows:

$$\begin{aligned} \hat{s} &= \arg \min \|Wx_s - W\bar{x} - WP s\|_2^2 \\ &= \arg \min \|x_{ab_s} - W\bar{x} - WP s\|_2^2 \end{aligned} \tag{15}$$

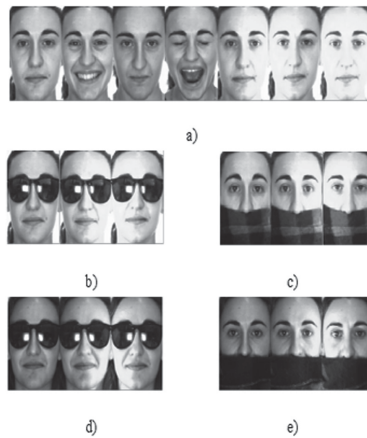
So you can get it with  $x_{ab}$ . The corresponding representation coefficient vector of the complete face is calculated as follows:

$$\hat{s} = [(WP)^T(WP)]^{-1}(WP)^T(x_{ab} - W\bar{x}) \tag{16}$$

Using the above calculation formula, the complete face image recognition results are obtained [15]. Using recurrent neural network to establish linear subspace, each face image can be uniquely represented by a coefficient vector. Through the above image recognition algorithm, the partially occluded face is repaired to get more accurate recognition results. So far, based on the functional characteristics of recurrent neural network, the partially occluded face image recognition algorithm is realized.

### 3 Simulation Test Experiment

In order to test the reliability of the proposed algorithm, a simulation comparison experiment is proposed to compare the image recognition algorithm with the conventional algorithm. Through the test results, the recognition ability of the two algorithms is compared. An AR face database was selected as the experimental object source. The AR face database consists of more than 4000 color face images of 126 people, including 70 men and 56 women. All face images are frontal standard images, there are many kinds of facial expressions, different lighting, sunglasses and scarves. Each person has 26 facial images, which are divided into two groups. Each group has 13 facial images taken and collected at the interval of two weeks. Figure 6 below is a partial face image of a randomly selected person in the AR face database.



**Fig. 6.** A partial face image of a randomly selected person in the AR face database.

Two algorithms are respectively used to recognize the face in the above figure. This experiment is divided into four groups, the experimental objects in each group are different, at the same time, each group is recognized 10 times, and the average value is taken as the final result. The experimental data is shown in Table 1.

**Table 1.** Recognition rate of locally occluded face images.

Interview pattern	Identification rate test result	
	This paper algorithm	Traditional algorithm
<b>b.</b> Sunglasses occlusion	98.45%	76.33%
<b>c.</b> Scarf shield	95.37%	70.2%
<b>d.</b> Sunglasses occlusion	88.49%	38.45%
<b>e.</b> Scarf shield	62.3%	19.69%

According to the test results in the table, under the proposed algorithm, the recognition rate of four different partially occluded face images is higher than that of the conventional image recognition algorithm. According to the calculation, the recognition rate difference between the two algorithms for **b** is 22.12%; the recognition rate difference for **c** is 25.17%; the recognition rate difference for **d** is 50.04%; and the recognition rate difference for **e** is 42.61%. Therefore, the recognition rate of the image recognition algorithm based on recurrent neural network is higher, which is more in line with the purpose of this study.

## 4 Conclusions

In this paper, the algorithm extracts the facial features of the occluded part of the face. According to the optimization function of recurrent neural network on the face image, it improves and optimizes the recognition algorithm, improves the recognition rate of the algorithm, and ensures the consistency between the recognition results and the reality. However, the calculation steps in the algorithm are more complex. Therefore, we hope that in the future practical application, we need to pay attention not to make calculation errors, and continue to simplify the algorithm in order to apply it in practice efficiently.

**Fund Project.** 2019 “climbing plan” Guangdong University Student Science and technology innovation and cultivation special fund project, project name: multi pose face image recognition algorithm based on artificial neural network learning, project number: pdjh2019b0619.

## References

1. Guo, W., Gao, W., Xiao, Q., et al.: Facial expression recognition with small data sets based by Bayesian network modeling. *Sci. Technol. Eng.* **18**(35), 179–183 (2018)
2. Gan, J., Li, S., Zhai, Y., et al.: 3D convolutional neural network based on face anti-spoofing. *J. Sig. Process.* **34**(09), 2834–2837+2870 (2017)
3. Liu, X., Wang, J., Xu, K.: Novel image feature extraction algorithm based on fusion AutoEncoder and CNN. *Appl. Res. Comput.* **34**(12), 3839–3842+3847 (2017)
4. Hu, Z., Chen, J.: Feature extraction model based on multi-layered deep local subspace sparse optimization. *Acta Electronica Sinica* **45**(10), 2383–2389 (2017)

5. Zhang, Z., Zhang, X., Zhang, J.: RGB-D images recognition algorithm based on convolutional-recursive neural network with sparse connections. *J. Hefei Univ. Technol. (Nat. Sci.)* **41**(05), 582–588 (2018)
6. Lu, J., Xu, L.: Digital-display instrument recognition system based on adaptive feature extraction. *Modern Electron. Tech.* **40**(24), 147–150 (2017)
7. Wang, X., Xu, F., Jia, K.: Action recognition based on improved long-term recurrent convolution network. *Comput. Eng. Des.* **39**(07), 2054–2058 (2018)
8. Zhang, G., Zhang, P., Pan, J., et al.: Fast decoding algorithm for automatic speech recognition based on recurrent neural networks. *J. Electron. Inf. Technol.* **39**(04), 930–937 (2017)
9. Liu, Y., Zhang, J., Gao, X., et al.: 3D object recognition via convolutional-recursive neural network and kernel extreme learning machine. *Pattern Recogn. Artif. Intell.* **30**(12), 1091–1099 (2017)
10. Yang, S., Zhao, C., Liu, F., et al.: A face recognition algorithm using fusion of multiple features. *J. Comput.-Aided Des. Comput. Graph.* **29**(09), 1667–1672 (2017)
11. Wu, T., Liu, Z., Hong, J.: A recursive formulation based on corotational frame for flexible planar beams with large displacement. *J. Central South Univ.* **25**(1), 208–217 (2018)
12. Zhang, X., Yong, S., Gao, Z., et al.: High-accuracy three-dimensional shape measurement of micro solder paste and printed circuits based on digital image correlation. *Opt. Eng.* **57**(5), 1–2 (2018)
13. Domski, J., Katzer, J.: Validation of aramis digital image correlation system for tests of fibre concrete based on waste aggregate. *Key Eng. Mater.* **761**(12), 106–110 (2018)
14. Duda, P., Rutkowski, L., Jaworski, M., et al.: On the Parzen kernel-based probability density function learning procedures over time-varying streaming data with applications to pattern classification. *IEEE Trans. Cybern.* **25**(99), 1–14 (2018)
15. Wang, F., Paul, T., Chen, W., et al.: Femtosecond laser ablation power level identification based on the ablated spot image. *Int. J. Adv. Manuf. Technol.* **94**(4), 1–8 (2018)