A Face Detection Method Based on Sliding Window and Support Vector Machine

Jie Chen[1], Sheng Cheng[2], Meng Xu[3] *
[1] The China Manned Space Engineering Office, Beijing 100083, China.
[2] Software R&D Center, China Aerospace Science, and Technology Corporation 100094, China.
[3] School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China.

* Corresponding author. Tel.: 15313760131; email: chengsheng@bjsasc.com
Manuscript submitted February 10, 2019; accepted April 20, 2019.
doi: 10.17706/jcp.14.7.470-478

Abstract: Face detection is a biometric technology based on human face features for identity authentication. With the development of e-commerce and other applications, face recognition has become the most potential means of biometric authentication. Classical face recognition is based on statistical methods, but the accuracy of this method is not high. In this paper, a face detection method based on the sliding window and support vector machine is proposed. Firstly, the image is divided into blocks, and the HOG features of the target image are extracted. Then the support vector machine model is trained through the data sets of human face and non-face. The support vector machine model can detect whether the target area belongs to the face area or not. Finally, the whole face area is detected by the sliding window model. Experiments verify the effectiveness of the proposed method.

Key words: Face detection, SVM, sliding window, HOG features.

1. Introduction

Human face detection [1] technology has been extensively studied since the 1990s, and now it has become one of the hotspots in the field of artificial intelligence and pattern recognition [2], [3]. Human face detection is a kind of biometrics technology based on human face features for identity authentication. It collects images or videos containing human beings with cameras and automatically detects and tracks human face features in images, and then carries out a series of related technical processing on the detected human face features. Human face recognition technology is widely used in robot technology [4]. A complete face recognition system includes image acquisition, face detection, face tracking, image preprocessing, feature extraction and identity recognition modules. Therefore, face recognition is a comprehensive field involving computer technology, image processing, machine vision, pattern recognition, artificial intelligence and biology [5]-[7].

The face is a kind of natural structure object with quite complex details. The challenge of detecting such object lies in (1) face has the variability of pattern because of its different appearance, expression, and skin color; (2)human face may have glasses, beard and other appendages in life; (3) as a three-dimensional object, the image of human face is inevitably affected by shadows caused by light. Therefore, if we can find solutions to these problems and successfully construct a face detection and tracking system, it will provide important enlightenment for solving other similar complex pattern detection problems. Therefore, if we can find solutions to these problems and successfully construct a face detection and tracking system, it will

provide important enlightenment for solving other similar complex pattern detection problems.Common methods for face recognition include face recognition based on set features [8], statistical principles [9], model-based methods [10], active shape module-based methods [11], active representation model-based methods [12], and neural network-based methods [13].The support vector machine method is based on the VC dimension theory with statistical learning theory [14] and the principle of structural risk minimization [15]. According to the limited sample information, it seeks the best compromise between the complexity of the model and the learning ability in order to obtain the best generalization ability. This paper uses the method of support vector machine(SVM) [16] to detect and recognize a human face.

Face recognition technology has attracted the attention of many scholars. [17] presents a facial expression recognition algorithm based on the SIFT algorithm. Using this method, the author completes the feature extraction of multi-expression face images, but the method has a lot of matching errors, so its practicability is not high. [18] proposes a K2DPCA face recognition method based on the Cholesky decomposition method. This method can overcome the influence of noise to a large extent, but this method can only overcome the effect of noise on small-scale data sets and needs a large number of similar samples in the data sets. [19] proposes a PCA face recognition algorithm which combines the gamma transform and the wavelet transform. This method eliminates the influence of non-linear factors such as illumination by the gamma transform and uses the PCA algorithm to realize feature extraction. However, the PCA algorithm can only extract linear features, but it seems powerless for non-linear features.

The contribution of this paper is to design a face detection method based on support vector machine with a mechanism of the sliding window. This method firstly divides the image into blocks, then extracts the HOG features of each image separately, then trains support vector machine through positive and negative examples. The sliding window mechanism is used to detect each block of the target image in turn. Support vector machine is used to detect whether it is a face area or not, and the similarity results are obtained. Finally, according to the threshold, the region considered as a face in the target image is determined. According to the threshold, we can judge whether an area in the target image is a face area.

This paper is arranged into three sections. Following the introduction, Section II presents the proposed face detection model, such as Histogram of Oriented Gradient, Sliding window and Support Vector Machine. Experiments are given in Section III to illustrate the performance of the proposed method. Section IV presents the conclusion.

2. Face Detection Model

2.1. Histogram of Oriented Gradient

HOG features [20] are used to extract image features. Directional gradient histogram feature is a feature descriptor used for object detection in computer vision and image processing. The HOG feature uses the gradient direction feature of the image, which is computed on a grid-intensive and the uniform grid cell. In order to improve the accuracy, the overlapping local contrast normalization method is used. The key factor of HOG is that the shape of the detected local object can be described by the distribution of the light intensity gradient or the edge direction. By dividing the whole image into small connected regions which are called cells, each cell generates a directional gradient histogram or the edge direction of pixels in the cell. The combination of these histograms can represent descriptors, which are used to describe the detected targets. In order to improve the accuracy, the local histogram can be standardized by calculating the light intensity of a larger area which is called block, in the image as a measure, and then normalize all cells in the block with this value. This normalization process has better illumination invariance and shadow invariance.

2.2. Sliding Window

The specific description of the sliding window [21] algorithm for images is as follows：In an image of $W \times H$, the window of $W \times h(W > w, H > h)$ is moved according to certain rules. The pixel values of the pixels in the window of the image are processed by a series of operations. After the operation, the window moves one step to the right or down until the processing of the whole image is completed.

2.3. Support Vector Machine

SVM is used to train and classify the obtained features, and then face and non-face models are obtained. In the early stage, SVM was developed from the optimal classification surface of linear separable cases and was used for binary classification problems. If C_1 and C_2 represent two different types of samples, P_0 and P_1 represent classification functions. If there is a linear function that can completely separate the two types of samples, then these samples are called linear separable; otherwise, they are called non-linear separable. Assuming that there are two classes of linearly separable training samples $x_1, y_1, x_2, y_2, \ldots, x_N, y_N, y_i \in +1, -1, i = 1,2,3,\ldots,N$, the expression of the linear discriminant function is $f(x) = w * x + b$, and the corresponding classification surface equation of the function is shown in formula (1).

$$W * X + b = 0 \tag{1}$$

The value of the linear discriminant function is a usually continuous real number, while the output of the classification problem is discrete value. For example, number - 1 is used to represent a category C_1, while number + 1 is used to represent the category C_2. All samples can only be represented by values - 1 and + 1. At this time, we can determine the category of the sample by setting a threshold and judging that the value of the discriminant function is greater or less than the threshold. We set this threshold to 0. When $f(x) \le 0$, the discriminant sample is category C1, whereas the discriminant sample is category C2. Now the discriminant function is normalized to satisfy $f(x) \ge 1$ for all samples of both classes, and then $f(x) = 1$ for all samples close to the classification surface.

$$w * x + b - 1 \ge 0, i = 1,\ldots,N \tag{2}$$

At this time, the classification interval is $2w$. The objective of finding the optimal classification surface is to maximize the classification interval. The maximum interval is equivalent to the minimum of $w^2 / 2$. Therefore, the optimal classification surface problem can be expressed as a constrained optimization problem as follows:

$$Min^{\Phi} w = 1/2w^2 \tag{3}$$

The constraint is as follows:

$$w * x + b - 1 \ge 0, i = 1,\ldots,N \tag{4}$$

The definition of Lagrange function is shown in formula (5):

$$L(w,b,a) = 1/2\,w^2 - \sum_{i=1}^{N} a_i y_i w^* x_i + b - 1 \tag{5}$$

In the formula(5), $a_i \geq 0$ is a Lagrange multiplier. In order to get the minimum value of the function (5), we derive w, b and a respectively. Thus, formula (6) is obtained.

$$\begin{aligned} &\frac{\partial L}{\partial W} = 0 \quad \Rightarrow \quad w = \sum_{i=1}^{N} \partial_i y_i x_i \\ &\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} \partial_i y_i = 0 \\ &\frac{\partial L}{\partial a} = 0 \quad \Rightarrow \quad \partial_i y_i w^* x_i + b - 1 = 0 \end{aligned} \tag{6}$$

Formulas (6) and (3) transform the problem of solving the above optimal classification surface into a dual problem of convex quadratic programming optimization, as follows:

$$\max \sum_{i=1}^{N} \partial_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \partial_i \partial_j y_i y_j (x_i . y_j) \tag{7}$$

The constraints are shown in formula (8):

$$\begin{aligned} &\partial_i \geq 0 \\ &\sum_{i=1}^{N} \partial_i y_i = 0 \end{aligned} \tag{8}$$

There exists a unique solution to this quadratic function optimization problem. If ∂_i^* is the optimal solution, then:

$$w^* = \sum_{i=1}^{N} \partial_i^* y_i x_i \tag{9}$$

Among them, ∂_i^* is the support vector, when its value is greater than 0. The weight coefficient vector of the optimal classification surface is a linear combination of support vectors. The classification threshold b^* can be obtained by formula (6), which satisfies the formula (10).

$$\partial_i (y_i (w . x_i + b) - 1) = 0, i = 1, 2, \ldots N \tag{10}$$

∂_i^* represents a non-support vector when $\partial_i^* = 0$, except for support vector, and the optimal classification surface function is shown in formula (11):

$$f(x) = \mathrm{sgn}[(w^* \bullet x) + b^*] \tag{11}$$

Formula (11) represents the classification function for SVM.

3. Experiments and Analysis

The tool of this experimental in this paper is Matlab and data sets used in this experiment are Caltech Web Faces project [22], SUN scene database [23] and CMU+MIT test set [24]. The Caltech Web Faces project dataset is a positive training database with 6,713 face images. The SUN scene database is a negative training database that contains a large number of images of non-face. Positive training database and negative training data are used as training samples to train the support vector machine model. The CMU+MIT test set database contains 130 images, for a total of 511 faces.

3.1. HOG Feature Extraction

The steps of Hog feature detection are as follows.

Step1: Three-dimensional images are grayed.

Step2: In order to adjust the contrast of the image, reduce the impact of local shadows and illumination changes, and suppress the noise interference. Gamma correction is used to normalize the color space of the input image.

Step3: The gradient of each pixel including size and direction is calculated to capture contour information and further weaken illumination interference.

Step4: Divide the images into small cells. In this paper, each cell is 6*6 pixels.

Step5: The descriptor of each cell can be formed by counting the gradient histogram of each cell which is the number of different gradients.

Step6: A block is composed of several cells (e.g. 3*3 cells/blocks). The feature descriptors of all cells in a block are connected in series to obtain the HOG feature descriptor of this block.

Step7: By concatenating the HOG feature descriptor of all the blocks in the image, the HOG feature descriptor of the image can be obtained, which is the final feature vector for classification.

3.2. Face Detection Model Based on Support Vector Machine

After obtaining the positive and negative features, the vl_trainsvm can be directly call to train the linear SVM model. The matlab code is as follows:

```
% Using SVM classifier
LAMBDA = 0.0001;
X = [features_pos; features_neg]';
Y1 = double(ones(1, size(features_pos, 1)));
Y2 = (-1) * double(ones(1, size(features_neg, 1)));
[w b] = vl_svmtrain(X, [Y1, Y2], LAMBDA);
```

The parameters w and b are obtained through training, the corresponding linear SVM model is obtained.

3.3. Realization of Face Detection at Different Scales

For the image to be detected, first enlarge it to 1.2 times, then gradually reduce it to 1.1 times, 1.0 times on the basis of 1.2 times, and so on. For each scale, first extract the HOG features of these changed images, and then the slide window method is used for detection. The calculated confidence value is compared with the given threshold value and the calculation of the confidence is shown in formula (12). If it is higher than the threshold value, the position of the sliding window is recorded. It is important to note that when recording a window, the location is reflected in the original size.

$$confidence = sum(t\arg et(:).^{*} w) + b) \tag{12}$$

The results of this experiment are shown in Fig. 1, 2. Firstly, the SVM classifier is trained with the positive and negative training set. After the training is completed, face recognition is performed for the images in the CMU+MIT test set, and the threshold is set to 0.82. The results on the test set show that the average

detection accuracy of the proposed method in this paper is 0.857, while the average detection accuracy of the PCA algorithm is 0.729. Therefore, it can be seen that the proposed method in this paper has high recognition accuracy. Figure 3-5 shows the results of face recognition in a single person and multi-person situations respectively. Among them, the yellow box represents the position of the face detected by the proposed method, the green box represents the correct face location, and the red box represents the wrong face detection results. Fig. 6 shows the training results of support vector machines. Fig. 6 (a) is a classifier of 6*6 cells obtained by training. Fig. 6 (b) is the classification result of positive and negative examples using SVM.

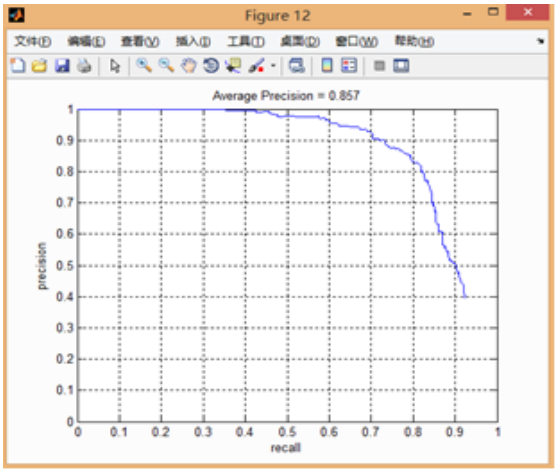


Fig. 1. Average detection accuracy.

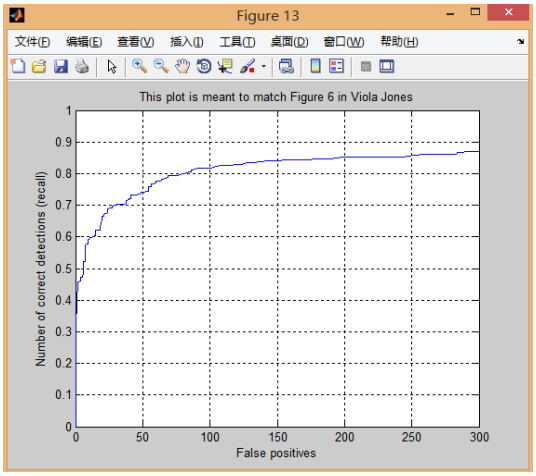


Fig. 2. Detection accuracy of a picture.

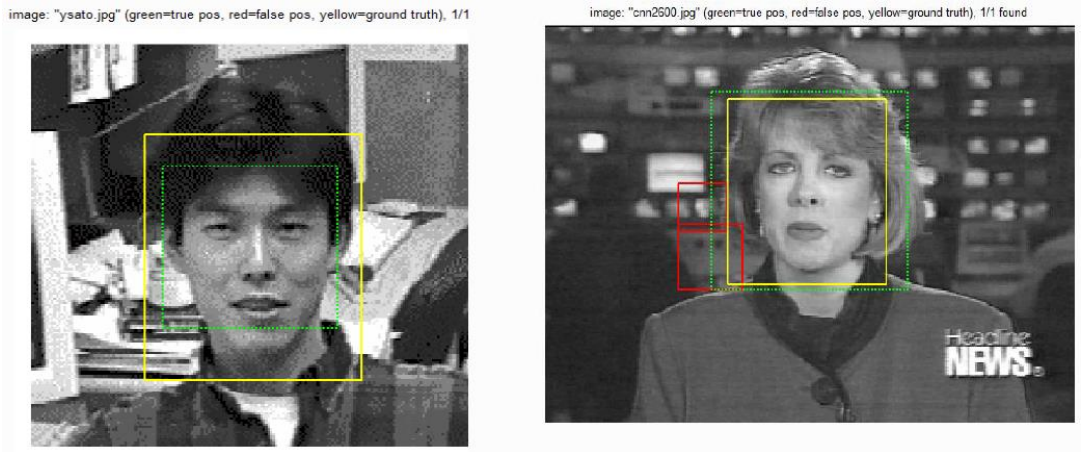


Fig. 3. Face recognition result in single person.

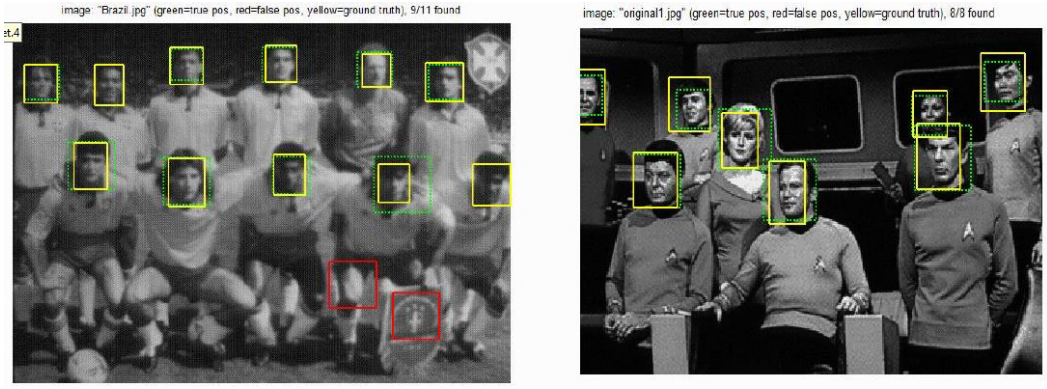


Fig. 4. Face recognition results in more than 4 persons.

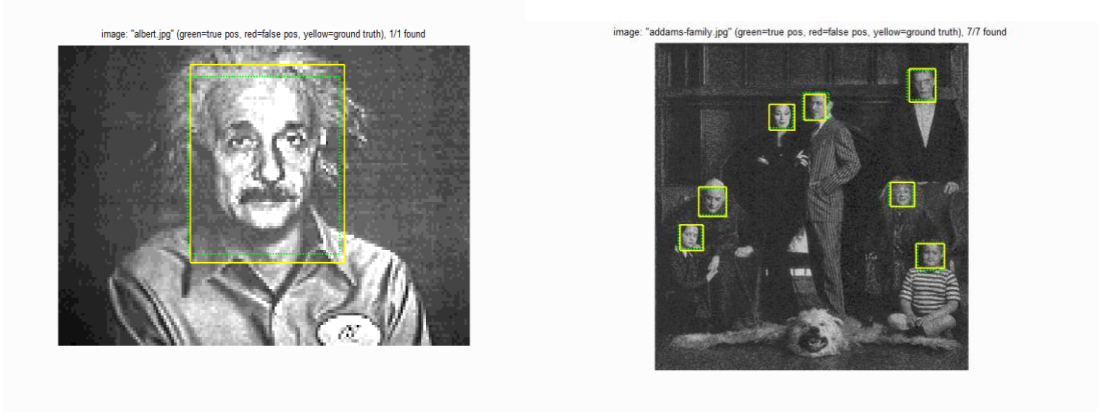
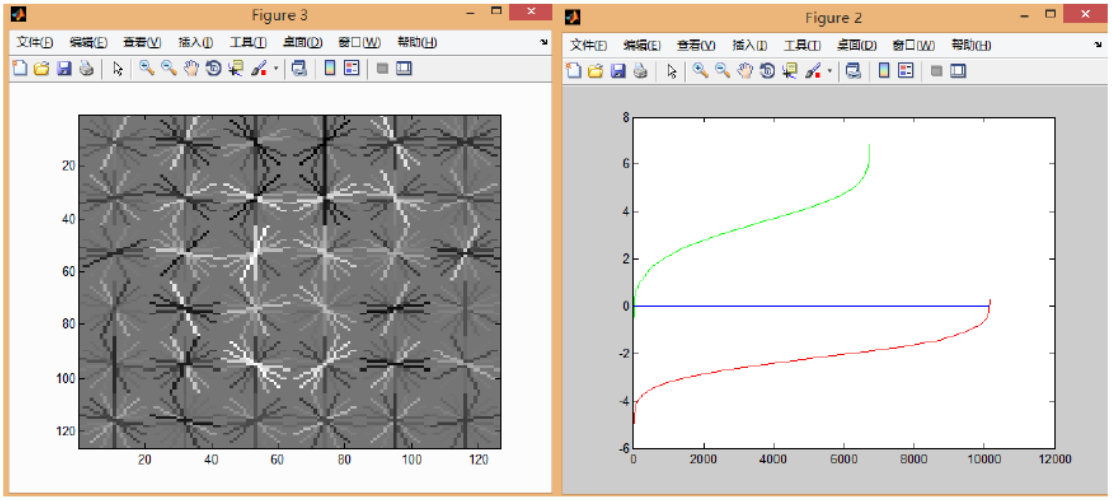


Fig. 5. Face recognition results.



(a) 6*6 cell classifier; (b) Classification results of positive and negative examples

Fig. 6. Training results of support vector machine.

In this experiment, the threshold is set to 0.82 and the variable is cell_size. When the threshold is set to 0 or less, many areas can be detected as faces in a given image, which is clearly unreasonable. It is expected that this will improve average precision, but this is too extreme for the image to be tested. If the threshold is very low, the detector will have a high probability of detecting errors. Therefore, the threshold value can be increased appropriately on the basis of high accuracy of detection, and then face detection is performed under extra test_scenes.

For the variable cell_size of this paper, if the value of the variable is low, the average accuracy may be higher, but the number of false faces detected may be more. Therefore, the value of the variable can be increased appropriately after obtaining high accuracy. The lower the cell_size value is, the higher the dimension of the feature is, which represents the information of the image better. Then the accuracy of the face detector can be improved. But at the same time, it requires more memory and time for training model.

4. Conclusion

To overcome the shortcomings of face detection based on statistics, a face detection method based on support vector machine with a mechanism of sliding window is proposed in this paper. By dividing the image into blocks, we first extract the Hog features of each image, then train the support vector machine model through positive and negative examples, and finally detect the face area of the target image by sliding window mechanism and get the complete face area by calculating the similarity. Finally, the effectiveness of the proposed method is verified by experiments.

Acknowledgment

This work is supported in part by the Aeronautical Science Foundation of China under Grant 20175553028, and in part by the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University under Grant ZZ2018169.

Reference

[1] Chen, L. W., Ho, Y. F., & Tsai, M. F. (2017). Cyber-physical signage interacting with gesture-based human–machine interfaces through mobile cloud computing. *IEEE Access, 4*, 3951-3960.

[2] Choi, J. W., Nam, S. S., & Cho, S. H. (2017). Multi-human detection algorithm based on an impulse radio ultra-wideband radar system. *IEEE Access, 4(99),* 10300-10309.

[3] Corneanu, C. A., Oliu, M., & Cohn, J. F. (2016). Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *38(8),* 1548-1568.

[4] Kinoshita, T., & Hayashi, E. (2008). Development of distance recognition using an ocellus camera for an autonomous personal robot. *Artificial Life & Robotics, 13(1),* 346-349.

[5] Ueno, S., Matsuda, T., & Fujiki, M. (2009). Dynamic face recognition: From human to machine vision. *Image & Vision Computing, 27(3),* 222-232.

[6] Moerland, T. M., Broekens, J., & Jonker, C. M. (2017). Emotion in reinforcement learning agents and robots: a survey. *Machine Learning, 2017(5),* 1-38.

[7] Wang, P., Lin, W. H., & Chao, K. M. (2017). A face-recognition approach using deep reinforcement learning approach for user authentication. *Proceedings of IEEE International Conference on E-business Engineering*.

[8] Bichsel, M., & Pentland, A. P. (1994). Human face recognition and the face image set′ s topology. *Clip Image Understanding, 59(2),* 254-261.

[9] Givens, G., Beveridge, J. R., & Draper, B. A. (2004). How features of the human face affect recognition: A statistical comparison of three face recognition algorithms. *Proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition.*

[10] Li, Z., Gong, D., & Li, X. (2016). Aging face recognition: A hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 25(5),* 2146-2154.

[11] Li-Qiang, D. U., Jia, P., & Zhou, Z. T. (2009). Human face shape classification method based on active shape model. *Journal of Computer Applications, 29(10),* 2710-2712.

[12] Hong, T., Kim, H., & Moon, H. (2006). Face representation method using pixel-to-vertex map (PVM) for 3d model based face recognition. *Workshop on Computer Vision in Human-Computer Interaction*.

[13] Shi, B., Bai, X., & Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(11),* 2298-2304.

[14] Albano, A., & Chornomaz, B. (2017). Why concept lattices are large: Extremal theory for generators, concepts, and VC-dimension. *International Journal of General Systems, 46(1),* 1-18.

[15] Shizuko, Z. M. (2017). On the principle of empirical risk minimization based on averaging aggregation functions. *Doklady Mathematics, 96(2),* 494-497.

[16] Mavroforakis, M. E., & Theodoridis, S. (2006). A geometric approach to Support Vector Machine (SVM) classification. *IEEE Transactions on Neural Networks, 17(3),* 671-682.

[17] Yang, L., Guang-Liang, H., & Chun-Lei, S. (2016). Recognition of expression-variant faces based on the SIFT method. *Chinese Journal of Liquid Crystals & Displays, 31(12),* 1156-1160.

[18] Shuisheng, Z., Ying, Z., & Xinliang, M. U. (2016). K2DPCA methods for face recognition based on Cholesky decomposition. *Systems Engineering-Theory & Practice, 36(2),* 528-535.
[19] Wang, X., & Zhao, Z. (2016). PCA face recognition algorithm combined with gamma transform and wavelet transform. *Computer Engineering and Applications, 52(5),* 190-193.
[20] Turner, S., Kurz, F., & Reinartz, P. (2013). Airborne vehicle detection in dense urban areas using HoG features and disparity maps. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 6(6),* 2327-2337.
[21] Tao, Y., & Papadias, D. (2006). Maintaining sliding window skylines on data streams. *IEEE Transactions on Knowledge & Data Engineering, 18(3),* 377-391.
[22] Caltech 10, 000 Web Faces. Retrieved from http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces/
[23] SUN database. Retrieved from http://groups.csail.mit.edu/vision/SUN/
[24] Wu, B., Ai, H., & Huang, C. (2004). Fast rotation invariant multi-view face detection based on real Adaboost. *Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition.*

Jie Chen is a research fellow of The China Manned Space Engineering Office. His research interests include artificial intelligence and intelligent systems.

Sheng Cheng is a research fellow of China Aerospace Science and Technology Corporation. His research interests include artificial intelligence and software engineering.

Meng Xu is a master candidate of Northwestern Polytechnical University. His research interests include intelligent robot system and machine learning.