A Method of Learner's Sitting Posture Recognition Based on Depth Image

Xing Zeng[*], Bei Sun, Enlong Wang, Wusheng Luo and Taocheng Liu

College of Mechatronics Engineering and Automation National University of Defense Technology Changsha, China

[*]Corresponding author

***Abstract*—Real-time detection of learner's sitting posture not only helps prevent myopia in time but also promotes the improvement of learning efficiency. However, most of the current sitting detection methods have the shortcomings of low detection variety and recognition rate, et al. Based on this, a sitting posture detection method based on Cartesian plane projection is proposed. The sitting depth images are projected into three Cartesian planes respectively. The background removal, interpolation scaling and normalization are performed for each projection map. The projection feature is obtained and the PCA is used to reduce the dimension of the feature. Finally, the projection feature and the front view HOG feature are fused to generate the new posture feature vector. In the experiment we collected 20 people, each person 14 kinds of sitting posture to form test database and the use of random forest to classify the extracted sitting posture characteristics. The experimental results show that this method can effectively detect the learner's sitting posture and it is superior to the existing method in recognition accuracy and recognition speed.**

Keywords- sitting posture; depth image; random forest

I. INTRODUCTION

Sitting is the most commonly used posture of learners every day which closely affects all aspects of the learner. The correct sitting posture can protect the learner's physical health and improve the learning efficiency. The wrong sitting situation will cause diseases such as myopia, lumbar cervical disease and muscle strain. In addition the sitting statistics can reflect the learner's learning status. Therefore it is important to identify the learner's sitting posture.

The sitting information collection method can be divided into two kinds of posture recognition methods: based on the sensor and based on the image. The sensor-based approach is the most traditional method of detecting the sitting posture by collecting sensor data through sensors (infrared, pressure, acceleration, ultrasonic etc.). Kazuhiro Kamiya installed the pressure sensor in the seat, then through the pressure detection to achieve including forward, backward, left and other nine kinds of sitting position detection [1]. This approach has the advantage of high accuracy of measurement data, but there are also data single, the use of inconvenience and high cost and others limitations. Compared with the sensor method the image-based method has the advantages of easy to use and rich information. It is the use of the camera to obtain the user's sitting image, then through the image processing to achieve the identification of sitting posture. At present image-based sitting position recognition research is relatively less but there are some programs. The one is the use of color images for posture recognition: Alejandro Jaimes discriminated the sitting position with the horizontal angle of the head, the left and right shoulder respectively when the human body is sitting [2]. WU Song-Lin the use of skin color in the YCbCr space gathered in a fixed area and in the CbCr plane projection as an approximate ellipse characteristics in the moving target area to extract the skin color area and the detection of skin color gray map PCA operation to achieve the recognition of 8 kinds of typical sitting posture [3]. YUAN Di-bo used the elliptical properties of YCbCr plane projection to extract the skin color characteristics under the sitting condition and extracted the SURF feature of the sitting position according to different thresholds. After the feature fusion, 7 Kinds of Sitting posture are recognized by Neural Network [4]. This method is due to the use of color images so easy to be affected by light and complex background. The other is the use of depth images for posture recognition: ZHANG Hong-yu extracted the contour feature of the sitting depth image by using SVM classifier to identify three kinds of sitting posture [5]. Jun-Yang Huang identified the three kinds of sitting posture based on the depth context feature of the depth image and the random forest classifier [6]. This method is not susceptible to the light and the environment, but the identification of the type of sitting is too little.

Based on this, this paper presents a sitting posture recognition method based on depth image. Compared with the existing methods the method proposed in this paper has the following advantages: 1) 14 kinds of common sitting position recognition, identification of rich content; 2) the establishment of the human body sitting posture depth image database; 3) the fast and effective foreground extraction method for the sitting posture; 4) the posture recognition based on the posture depth map extraction projection feature and the HOG feature, the recognition accuracy is high and it is not affected by the light and the background. The sitting posture recognition process is shown in Figure 1. Firstly, 20 people are involved in the specific experimental environment to make 14 different sitting posture, through the foreground extraction the sitting posture depth image database is established. Subsequently the sitting posture depth map is projected on three Cartesian planes to obtain the projection feature and the HOG feature is extracted from the front view. Finally, training and testing of sitting posture through random forest. The test results show that this method can accurately identify 14 kinds of sitting posture.

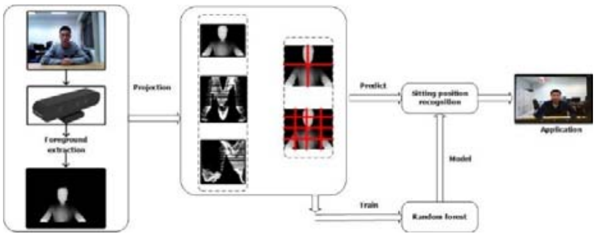

FIGURE I. SCHEMATIC DIAGRAM OF SITTING POSTURE RECOGNITION

II. Astra Camera Overview

Astra3D sensor is developed by Shenzhen orbbec. It has the advantages of high precision, miniaturization and compatibility, can support mainstream operating systems such as Windows, Android, Linux and OSX. Astra3D sensor are mainly composed of four parts, the RGB camera, infrared launcher, infrared camera and microphone, Table 1 shows the Astra3D sensor-related configuration.

TABLE I. ASTRA3D SENSOR STRUCTURE AND FUNCTION

Hardware Structure	**Features**
RGB camera	Get color images. viewing angle: vertical direction 49.4°, horizontal direction 63.1°
Infrared launcher	Emitting infrared to form speckle images. measuring range 0.6-8 meters
Infrared camera	Accept the speckle information to form a depth image with an accuracy of 1m: ± 1-3 mm. Viewing angle: vertical direction 45.5°, horizontal direction 68.4°.
Microphone	For sound source location and speech recognition.

Among them, the Astra3D sensor uses the optical coding technique to obtain the depth distance. Depth distance (Astra3D sensor feedback Z-axis data (Vector3) corresponds to the number of millimeters of the actual distance.) converted, mapped to 0-255 range of depth images. Astra depth image with resolution of 320 × 240 and a frame rate of 30fps (frame per second). The effective detection range of the depth is 0.5m-8m. As shown in Figure 2, the RGB image and the corresponding depth image are collected for the Astra3D sensor.

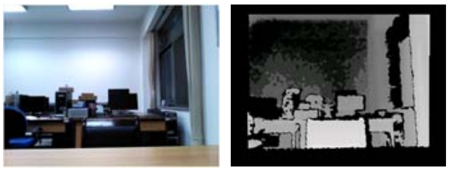

FIGURE II. THE RGB IMAGE AND DEPTH IMAGE OF ASTRA3D SENSOR

III. Method

A. Data Collection Environment

Considering the learner's sitting situation, the field of view of the depth image and the effective distance the depth. The Astra3D sensor is placed in the center of the desk, the sensor is facing the people. The sensor is at a distance of 0.9 meters from the edge of the desk (the distance data is high in accuracy, at the same time the degree of separation of the body can be ensured), and the height of the sensor is adjusted so that the edge of the depth image is slightly higher than the edge of the desk. This can eliminate the effects of debris and hands on the desk. In this experimental environment the sitting depth image is collected.

B. Foreground Extraction

Since the background in the depth image will have a serious impact on the posture detection, foreground extraction is required. This paper designs a quick and effective foreground extraction method for learner's sitting situation.

Taking into account the special scene of people sitting, we can have the following assumptions: 1) there is a certain distance between the people and the surrounding background when people is sitting; 2) when people is sitting, Z axis direction (depth distance) will not have a wide range of changes; 3) In addition, we assume that learners will always sit in the middle of the image field of vision. Figure 3 shows the depth of the image when the person is sitting, the depth image gray value of the person and the background have a big difference. So based on these characteristics, this paper uses the threshold segmentation method for rapid foreground extraction.

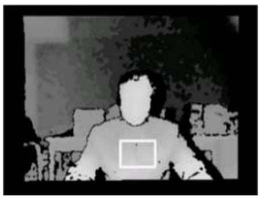

FIGURE III. SITTING DEPTH IMAGE

The specific implementation method can be divided into the following steps: 1) considering the body is usually constant when the people is sitting, so in the initial state to select a fixed area (such as the box) is considered the center of the human body; When the camera is initialized, first get the average depth of the box in the distance value, that is the farthest distance of the human body $\mathrm{MaxTargetDepth}$. Then, each point is uniformly obtained in the depth image (the image resolution is 320 x 240, every 10 pixels is extracted), and the depth distance of each point is placed in vector a.

$$\begin{cases} i \in Target, a[i] < MaxTargetDepth \\ i \in Background, a[i] > MatTargetDepth \end{cases} \tag{1}$$

Where i is the pixel in the image and $a[i]$ is the depth distance value of the pixel; 3). Then the nearest depth distance $\mathrm{MinBackgroundDepth}$ in all the back points is obtained; as a threshold, the depth image is segmented. Foreground image can be obtained after the removal of the background. 4) Finally, the image is median filtered, expanded and corroded, and the burrs in the voids and edges of the image are removed.

In order to make each subsequent frame be in a good way to remove the background and not subject to the impact of human posture change, this paper designed an update method. Specific steps are as follows:

1) Initialize, $\mathrm{MaxTargetDepth}_i$;

2) Execute formula (1);

*3)*Get $\mathrm{MinBackgroundDepth}_i$ as a threshold for background removal;

*4)*The next frame, let $\text{MaxTargetDepth}_{i+1}$=$\text{MinBackgroundDepth}_{i}$-500, return to the second step;

Where 500 represents 0.5m, this choice is due to the little change of the background image depth distance value, thus we can completely remove the background.

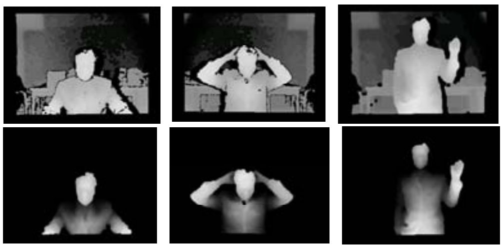

FIGURE IV. FOREGROUND EXTRACTION IMAGE

Analysis of the foreground extraction effect Figure 4. We can see that the method for the sitting segment has a very good effect, in addition, it has good real-time ability, and are not affected by the complex environment and human action advantages.

C. Database Establishment

This article invited 20 volunteers (both men and women) to collect sitting data, collected 14 common sitting gestures when people is learning (upright, left partial, right partial, bow down, looked up, body right oblique, body left oblique, raise right hand, raise left hand, right hand cheeks, left hand cheeks, lie down, stretched, lie). Everyone sitting at the desk naturally made these 14 kinds of gestures and let everyone to do each posture to have a certain change, the sensor real-time store each person's sitting background image. After all the sitting images are collected, 30 different images are extracted for each person's posture. In the end we created a body sitting posture depth image database that includes 20 people sitting posture, each person with 14 kinds of sitting posture, each sitting posture with 30 images, a total of 8400 images.

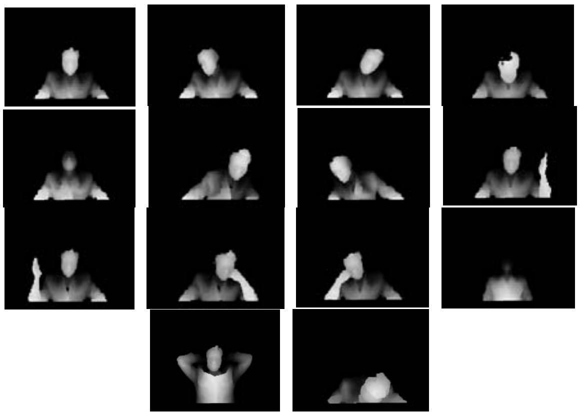

FIGURE V. SITTING DEPTH IMAGES

D. Feature Extraction

Body sitting depth image includes contours, distance, depth and other information. Inspired by the paper [7] the depth of the image in the three Cartesian plane projection followed by the extraction of the DMM images of each plane so the body action recognition. In this paper a fast and accurate sitting posture recognition algorithm is proposed based on the human body sitting posture depth image. The whole process of the algorithm is shown in Fig. 6 which mainly includes two parts: 1) Project the depth of the human body in three Cartesian planes, then bicubic interpolation the projection image and normalize it. Finally use PCA Reducing the dimension of the projection feature; 2) extract the pyramid HOG characteristics of the front view.

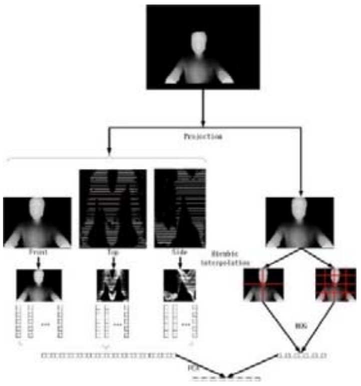

FIGURE VI. FEATURE EXTRACTION PROCESS

The first part: for a given sitting depth image, in the three Cartesian plane projection respectively to get the front view, top view and side view. The front view is taken directly from the pixel at the pixel point in a two-dimensional 320 × 240 image. Top view and side view are from the top and side angle to view, the establishment of a new plane coordinate. (For example, the pixel value of the depth image point (180,120) is 222, then the pixel value of top view point (180,222) plus 1, the pixel value of side view point (120,222) plus 1), followed by stacking, and resulting in two two-dimensional images 320×255 top view and 240 × 255 side view. This results in three dimensional images from a depth image. Since there are most black and unused areas in the projected image, it is considered to be removed. The specific method is to traverse the rows and columns of the projection image respectively, find all 0 rows and all 0 lines of the projection, and to remove resulting in a full projection as shown in Figure 6.

For the full projection: 1). The size of the projected image will be different due to the difference of the people and the sitting posture. When the feature vector is generated, the dimension of the eigenvector will not be unified. 2) The projected image is larger and the generated feature vector dimensions will be large; 3). The pixel of the projection image is too intermittent, too much useless information. Considering, you need to scale the projection and unify the projection size of each projection plane. In this paper, the image is interpolated and scaled by bicubic interpolation, and the size of the front view, the top view and the side view are fixed, 50×70, 50 × 50 and 50 ×50 respectively, as shown in Fig 6. As a result, the pixel values at different pixels of the projection images are very different, a larger pixel value will affect the experimental results. So the pixel values of three Projection images are normalized respectively, that is, all the pixel values are scaled to 0-1 between.

$$d_{(i,j)} = \frac{D_{(i,j)}}{D_{max}} \tag{2}$$

Where $D_{(i,j)}$ is the pixel value of the projection point (i, j), D_{max} is the maximum pixel value of the projection map, and $d_{(i,j)}$ is the normalized value.

Finally, the projection image matrix of the m rows and n lines after normalization is transformed into a column vector H_i of m×n rows, where i represents three projection images. The three column vectors form a projection feature vector in the order of the front view, the top view, and the side view.

$$H_1 = [H_f, H_t, H_s]^T \tag{3}$$

So that the dimension of H_1 is $H_{dim} = 50 \times 70 + 50 \times 50 + 50 \times 50 = 8500$. As the dimension of the projection feature vector is too large, the PCA [8] is used to reduce the dimension and retain the dimensionality of 98%. The dimension of the eigenvector is reduced to the hundred dimension, and the new projection feature vector H_1' is obtained, which mainly reflects the sitting distance, depth and other characteristics.

The second part: the front view is interpolated and zoomed by bicubic interpolation, get 64 × 64 front view, then, extract the two layers of HOG feature of the front view, the grid size of the first layer is 2 × 2, the second layer is 4 × 4, overlap is 0.5. 360° divided into nine parts, after extract HOG characteristics for each layer respectively, arranged in the form of rows to get H_2, which mainly reflects the edge and shape characteristics of the sitting local area.

Finally, the eigenvectors $\mathrm{H} = [H_1', H_2]$ which can reflect the sitting posture are obtained by superimposing H_1' and H_2.

E. Classification

After the completion of the posture feature extraction, the next need is to use a certain classification algorithm to establish one corresponding relationship between the eigenvector and the sitting posture. Random forest classifier is a multi-class classifier, which has the advantages of learning, recognition speed, and easy to fit. So this paper uses random forest to classify the sitting posture.

The core idea of Random Forest [9] is forming the strong classifier by combining multiple weak classifiers. The final classification result is decided by all the weak classifiers. The random forest algorithm is improved on the basis of Bagging. The training and testing of each weak classifier are independent. Figure 7 shows the block diagram of random forest.

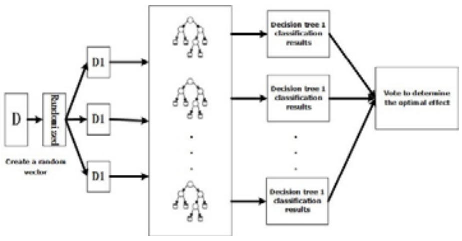
D
Randomi zed
Create a random vector
D1
D1
D1
Decision tree 1 classification results
Decision tree 1 classification results
Decision tree 1 classification results
Vote to determine the optimal effect

FIGURE VII. RANDOM FOREST DIAGRAM

Random forest specific training process can be divided into the following two steps:

(1) The K samples was returned for the original feature training set, where the size of each sample was the same as the original training set.

(2) Subsequently, the decision trees were trained independently using the sampled samples to obtain a random forest model of K trees.

When the sitting category of a sample is judged, the eigenvector of the sample is input into the K tree decision tree, respectively, and the K classification results are obtained. Subsequently, the final sitting type is determined by voting.

IV. EXPERIMENT AND DISCUSSION

In this paper, random forest can be used to identify, and the number of random forest trees on the recognition speed and accuracy will have corresponding influence. For overall consideration the number of random trees is 50. The experimental hardware environment is: i5-2410M CPU, 2.30 GHz frequency, 6GB memory, 64 bit system of PC.

Experiment 1, the 20 people sitting depth database unified test, first of all, in each posture of each person randomly selected four images, compose the training samples, including: 14 kinds of posture, each posture has 80 images. Then, the training samples were trained in a random forest, and the total training time is 85.65 s. Subsequently, the 20-person sitting database was used for testing, and the test samples included 14 postures, 600 images per posture. Averaging test 1 image requires 64.30ms. The confounding matrix of the test results is shown in Figure 8.

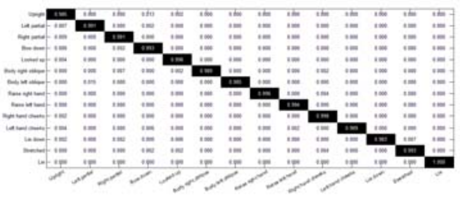

FIGURE VIII. UNIFIED TEST RESULTS

It can be seen from Figure 8, the 14 kinds of sitting position recognition rate is almost all over 98.5% in this experiment, the average recognition rate is 99.16%. Among them, the recognition rate of the lie down is only 98.3%, it is because some volunteers in the process of lying down, the chair is too front, the body is not lie obvious, and with the looked up, upright and so on have some confusion.

Experiment 2, in order to increase the difficulty of testing, 20 people will be divided into A, B two parts, each part has 10 individuals, cross-test. Among them, Part A for training, training sample selection is similar with experimental 1, training samples include: 14 kinds of gestures, each posture 40 images. The total training time is 37.68s. B samples were used for testing, and the test samples include: 14 postures, 300 images per posture. Average classification of 1 image requires 63.3ms. The confounding matrix of the test results is shown in Figure 9.

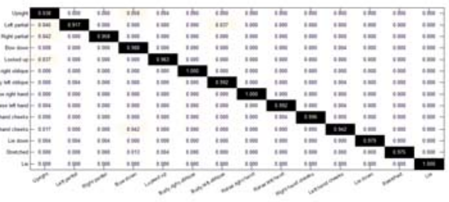

FIGURE IX. CROSS TEST RESULTS

It can be seen from Figure 9, the recognition rate of the crossover is mainly lower than that of experiment 1, it is because different people has different body, figure and different sitting position. Among them, the left partial recognition rate is only 91.7%, it is because different people have different habits of partial, partial difference is large, so that it is have some confusion with the upright and the body left oblique; In addition, the upright and the bow down produced Some confusion, this is because some upright samples, the volunteers head lower, so there is no significant difference with the bow down sample. Although the cross-test recognition rate has a certain reduction, but the average recognition rate is still able to reach 97.42%.

The sitting posture recognition method of this paper and the paper [3], [4] and [6] posture recognition comparison results are shown in table 2. In addition, this paper compared upright, left hand cheeks, right hand cheeks, stretched, lie of the paper [3], upright, left partial, right partial, bow down, left hand cheeks, right hand cheeks of the paper [4], respectively. The accuracy of sitting posture recognition is shown in Figure 10.

TABLE II.COMPARES THIS METHOD WITH EXISTING METHODS

Method	**The Number of Sitting**	**Feature**	**Average Recognition Rate**
Paper 3	8	Color characteristics	84.92%
Paper 4	7	Fusion of skin color and SURF	93.70%
Paper 6	3	The depth map context features of the depth	96.53%
This Paper	13	Projection features of the sitting depth image	99.16%

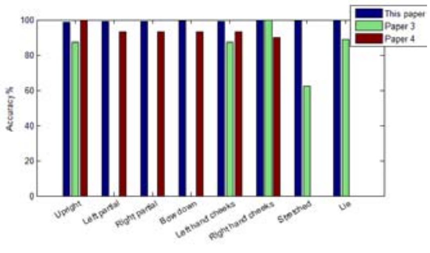

FIGURE X. COMPARISON OF RECOGNITION RATES FOR THE SAME SITTING POSTURE

As can be seen from Table 2, not only the sitting identify species of this paper higher than other programs, and the accuracy of sitting recognition is also high. In addition, we can see that the recognition rate of the same sitting posture is higher than that of the paper [3] and the paper [4], except that the upright recognition rate is slightly lower than that of the paper [4]. In summary, the sitting position recognition method of this paper has a clear advantage.

V. Sitting Detection System

The sitting posture recognition algorithm of this paper is implemented on the Android platform, and the posture detection system is designed. System development hardware includes: Astra3D sensor, Tiny4412 Android development board and PC. Development software includes: eclipse, Opencv for Android 2.4.9 and Astra depth image SDK.

In order to verify the feasibility and effectiveness of the sitting detection system, a volunteer sits in front of the sensor for normal learning, the use of the system for real-time detection. The test results are shown in Figure 11.

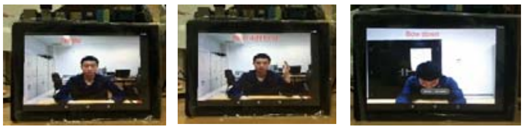

FIGURE XI. POSTURE TEST RESULTS

Through the test results we can see that the system can effectively detect the sitting position of learners, and has good real-time performance. In addition, the system also has the advantages of portability, miniaturization and so on.

VI. Conclusion

In this paper, a learner's sitting posture recognition method based on depth image is proposed. Astra3D sensor is used to collect the depth images of the 14 kinds of sitting posture. The fast and effective foreground extraction method is used to remove the background interference. Finally, the sitting depth image database is established. The projection feature vector is obtained by blanking the projection image, interpolation scaling, normalization and so on. The projection feature vector is reduced by the PCA with the HOG feature of the front view which constitutes the final posture feature vector. Subsequently, the random forest was used to classify the sitting posture. The experimental results show that this method can effectively identify the learners' sitting posture, and has a good recognition rate and recognition speed. In addition, this method has a higher recognition rate than the existing methods in sitting posture. Finally, the sitting detection system is designed by this method, which realizes the effective detection of the learners' sitting posture. The future job is to establish a larger sitting database, improve the accuracy of sitting recognition, followed by the analysis of more posture features to identify some of the composite posture.

References

[1] Kamiya K, Kudo M, Nonaka H, et al. Sitting posture analysis by pressure sensors[C].International Conference on Pattern Recognition. IEEE, 2008:1-4.

[2] Jaimes A. Sit straight (and tell me what I did today): a human posture alarm and activity summarization system[C].ACM Workshop on Continuous Archival and Retrieval of Personal Experiences. ACM, 2005:23-34.

[3] Wu S L, Cui R Y. Human Behavior Recognition Based on Sitting Postures[C].International Symposium on Computer,communication, Control and Automation Proceedings. 2010:138 - 141.

[4] YUAN Di-bo, DAI Yong, CHEN Tong-qian. Multi-feature fusion recognition of incorrect sit posture [J]. Computer Engineering and Design, 2017, 38(2).(in Chinese)

[5] ZHANG Hong-yu, LIU Wei, XU Wei. Depth hllage Based Gesture Recognition for Multiple Lesrners [J]. Computer Science, 2015, 42(9):299-302. (in Chinese)

[6] Huang J Y, Hsu S C, Huang C L. Human upper body posture recognition and upper limbs motion parameters estimation[C].Signal and Information Processing Association Summit and Conference. 2013:1-9.

[7] Zhang S, Chen E, Qi C, et al. Action Recognition Based on Sub-action Motion History Image and Static History Image[J]. 2016, 56:02006.

[8] Abdi H, Williams L J. Principal component analysis[J]. Wiley Interdisciplinary Reviews Computational Statistics, 2010, 2(4):433-459.

[9] Breiman L. Random Fores t[J]. Machine Learning, 2001, 45:5-32