EAI Endorsed Transactions
on e-Learning

Research Article **EAI.EU**

Chinese fingerspelling recognition via gray-level co-occurrence matrix and fuzzy support vector machine

Yalan Gao[1], Cheng Xue[1], Ran Wang[1], Xianwei Jiang[1,*]

[1] School of Mathematics and Information Science, Nanjing Normal University of Special Education, Nanjing 210038, China

Abstract

INTRODUCTION: Chinese deaf-mutes communicate in their native language, Chinese sign language which contains gesture language and finger language. Chinese finger language conveys information through various movements of fingers, and its expression is accurate and convenient for classification and recognition.

OBJECTIVES: In this paper, we proposed a new model using gray-level co-occurrence matrix (GLCM) and fuzzy support vector machine (FSVM) to improve sign language recognition accuracy.

METHODS: Firstly, we acquired the sign language images directly by a digital camera or selected key frames from the video as the data set, meanwhile, we segmented the hand shapes from the background. Secondly, we adjusted the size of each images to $N \times N$ and then switched them into gray-level images. Thirdly, we reduced the dimension of the intensity values by using the Principal Component Analysis (PCA) and acquired the data features by creating the gray-level co-occurrence matrix. Finally, we sent the extracted and reduced dimensionality features to Fuzzy Support Vector Machine (FSVM) to conduct the classification tests.

RESULTS: Moreover, we compared it with similar algorithms, and the result shows that our method performs the highest classification accuracy which is up to 86.7%.

CONCLUSION: The experiment result displays that our model performs well in Chinese finger language recognition and has potential for further research.

Keywords: Chinese fingerspelling recognition, gray-level co-occurrence matrix, fuzzy support vector machine, principal component analysis

Received on 23 August 2020, accepted on 25 September 2020, published on 12 October 2020

doi: 10.4108/_______________

*Corresponding author. Email: jxw@njts.edu.cn

1. Introduction

Communication exists in all aspects of our lives and plays a crucial role in our daily lives. But there is a group of special people in our lives who cannot communicate with others in words normally, they are called deaf-mutes. The results of the second sample survey of disabled people in China in 2020 show that there are 27.8million people with hearing impairment in China, accounting for 24.26% percent of the total, ranking first among the six categories of disability [1]. People with hearing impairment cannot express their demands properly like normal people when they encounter difficulties, so communication has become a problem for them when they are defending their rights and interests. In real life, speech-and-hearing-impaired people can only communicate with people who understand their language. Thus, a translator device is necessary for dumb people to communicate with normal people and the key part of this translator device is the accuracy of sign language recognition.

There are three common methods in sign language recognition: Vision-based SLR [2], Sensor-based SLR ([3, 4]) and Hybrid-based SLR [5]. In sign language recognition, hand shape, direction, position and movement track are the common features which are also the focuses of most researches and experiments. To continue to recognize sentence-level, American Sign Language (ASL), Thad Starner et al. proposed two real-time hidden Markov model-based systems and used a single camera to track the user's unadorned hands [6]. Kishore et al. [7] proposed a 4-camera model to recognize gestures of Indian sign language. The flexible sensor was once produced based on the combination of ARM9 and 9-axis IMU, and on this basis, Li Lei et al. [8] introduced the new data gloves and sign language recognition system. Lionel Pigou et al. [9] considered using the Microsoft Kinect to recognize. Moreover, according to the Kinect 22 depth data and the skeleton joints data, Lubo Geng et al. [10] used the combined position and spherical coordinate feature representation to construct factor vectors. Chophuk P et al. [11] used a compact and affordable 3D motion sensor, arguing that palm-sized Leap Motion sensor has a higher comprehensive evaluation than other methods used in existing researches such as Cybler-glove and Microsoft Kinect.

There are also many state-of-art methods to recognize Chinese fingerspelling, such as Hidden Markov model (HMM), Support Vector Machine (SVM), K-nearest neighbor (K-NN), Artificial Neural Network (ANN) and Dynamic Time Warping (DT) and so on. Many feature extraction methods have also been developed. The common feature extraction methods are Gray Level Co-occurrence Matrix (GLCM), Scale Invariant Feature Transform (SIFT) and Wavelet Entropy (WE). Due to texture can make full use of image information, it can be an important basis for describing and recognizing images theoretically or from common sense. Compared with other image features, it can better take into account the macroscopic properties and subtle structure of images. Therefore, texture becomes an important feature to be extracted for target recognition. At the same time, SVM is a novel machine learning method and has good robustness and generalization ability. Consequently, in this paper, we chose to combine the fuzzy SVM and GLCM together for image classification recognition.

We are committed to contributing to the automatic sign language recognition. In this paper, we pay attention to the Chinese finger language recognition work. Although finger language only has 30 finger letters classification (including 26 basic pinyin letters and 4 upturned tongues), but it is significant. Due to different area uses different Chinese sign language, sign language often has the confusion of different expression of the same meaning. Even though China has published two versions of Chinese sign language and standard sign language, they are not widely used in practice between deaf people. Thus, as the only deterministic expression, the advantage of finger language has been highlighted. So, we proposed the model of gray-level co-occurrence matrix and fuzzy support vector machine (GLCM-FSVM) to improve the accuracy of Chinese finger language recognition. In our method, features of Chinese sign language images were extracted by using the gray-level co-occurrence matrix (GLCM). GLCM can describe the texture by the grayscale relation between two pixels in the space to reduce the difficulty of image identification. Then, we used the fuzzy support vector machine (FSVM) to classify and regress these extracted features and identify the image of finger gesture from 720 Chinese sign language pictures. The experiment result displays that our model performs well in Chinese finger language recognition and has potential for further research.

The rest of this article is arranged as follows: Section 2 is mainly about the dataset and the experimental methods. The feature extraction method GLCM, classification method FSVM and PCA dimensional reduction method are introduced emphatically. Section 3 contains the experiment process and experiment result. Section 4 is a discussion of the pros and cons of this method that we proposed. We conclude this paper in the final Section 5.

2. Method

We divided our method into the following four steps, as shown in ***Figure 1***.

Step 1. Data set construction: The input database consists of 720 color images of 30 isolated Chinese sign language, each of which is a size of length × width × color (length=1080px, width=1080px, color channel=3(RGB)).

Step 2. Normalization processing: Separate the hand gesture from the original sign language picture and set background color to zero. Resize each image to $N \times N$ size (N=256). Keep 3 color channels.

Step 3. Feature extraction and dimension reduction: Convert the image into gray-level image by using the Matlab R2018a. Reduce the gray value of the image from 256 to 8 by using the Principal Component Analysis (PCA) and create a GLCM.

Step 4. Training and classification: Split the 720 samples into training set and test set by a ratio of 7:3 randomly. At the same time, the classification is performed on a 10-fold cross validation. A training set is presented to the classifier FSVM and the performance of the classifier is analyzed.

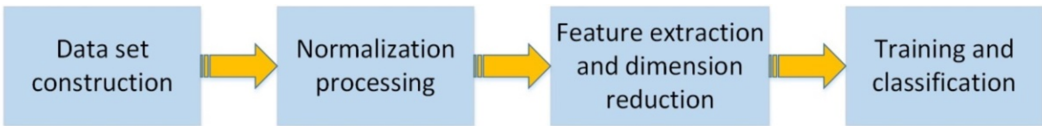

Figure 1. Diagram of our method

2.1. Subjects and Dataset

We used a camera to take 720 Chinese finger language pictures from 30 different samples (the size of each image is 1080×1080), each sample contains 26 basic letters and 4 roll tongue pronunciation words, a total of 30 categories. Then, we used Adobe Photoshop CS to segment each Chinese finger picture manually, strip out the hand-shaped area in the picture, and adjust the size of each image to 256×256. At the same time, the background color is normalized and set as RGB (0, 0, 0). Finally, we saved these images as TIF compression format with little loss of image information (see ***Figure 2***).

We set all the images to the same size and background to ensure the experimental results' validity, and finally converted the color images to gray-level images by using the tool software Matlab R2018a. The pseudocode is as follows (see Table 1).

Here, two finger language images were presented, one is the letter K and the other is warp letter ZH. (see ***Figure 3***).

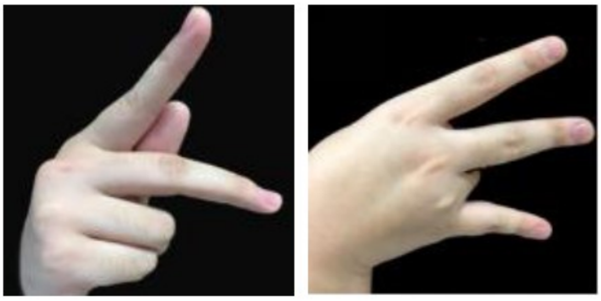

Figure 2. Example pictures after removing background and other irrelevant elements. (left) K; (right) ZH

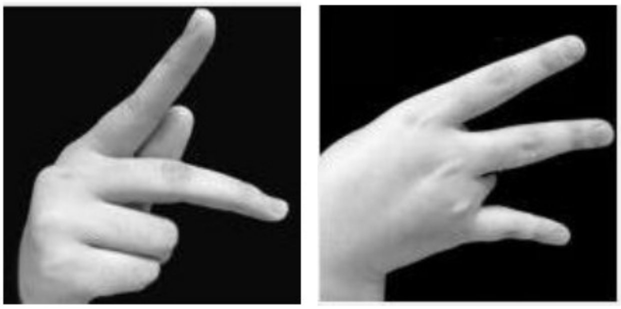

Figure 3. Converting the color images into gray-level images. (left) K; (right) ZH

Table 1. The flow of pseudocode

Step	Pseudocode
1	Creat maindir
2	Get subdir
	for j=1: length(subdir)
	if (true) then
	Get subdirpath;
	end if
	end for
3	for i=1: length(subdirpath)
	Get imgpath;
	Rgb2gray;
	Imshow( );
	endfor

2.2. Gray-level Co-occurrence Matrix

In the early 1973, Haralick et al. proposed a generalized texture analysis method: Gray-level Co-occurrence Matrix (GLCM), which describes the spatial distribution relationship between pixels in an image with texture information. Each two pixels in the image space will have a certain spatial grayscale relationship because the texture is made up of the grayscale's repeated appearance in the spatial position. Gray-level co-occurrence matrix is a popular method to describe texture by studying the spatial correlation of grayscale [12, 13].

The joint probability density of pixels at two locations defines the gray-level co-occurrence matrix. Let $f(a,b)$ be a two-dimensional digital image, the image's size is $M \times N$ and $N_g$is the gray level of this image, then the certain spatial relationship among the gray-level co-occurrence matrix can be described as follows:

$$P(m,n) = \{(a_1,b_1),(a_2,b_2) \in M \times N | f(a_1,b_1) = m, f(a_2,b_2) = n\} \quad (1)$$

Where the $\{a\}$ represents how many elements there are in the set a, and P is the matrix of$N_g \times N_g$. If the distance between(a_1,b_1) and (a_2,b_2) is d and θ is the angle between (a_1,b_1) and (a_2,b_2), then the matrix $P(a,y,d,\theta)$ of various spacing and angles can be obtained. Since the distribution of pixels may be different in all directions, it is advisable to use the 8-connected domain to compute the gray-level co-occurrence matrix, that is, $d = 1, \theta = 0°, 45°, 90°$ and $135°$.

The texture information of the image is contained in the co-occurrence matrix. The slower the textures change of images, the larger the value on the diagonal of the gray-level co-occurrence matrix and the other way around. To describe the texture condition with co-occurrence matrix more intuitively, the following characteristic quantities are usually used.

Let x, y be the pixel values of(a_1,b_1),(a_2,b_2), the number of (m,n) is $F(m,n)$, and F is the sum of all values in GLCM, then it is defined:

$$P(m,n) = \frac{F(m,n)}{F} \quad (2)$$

In formula (2), $P(m,n)$ is the possibility of (m,n) appearing in the image.

Based on formula (2), we can derive further, as the spatial distribution of pixels becomes more uneven, the texture gets deeper, the contrast gets bigger, and the visual effect gets clearer. We used (3) to express this relationship and in formula (3), α_1 represents the contrast, which expresses the brightness contrast of a pixel value and its domain pixel value. The deeper the texture groove, the greater the contrast, the clear the visual effect. Moreover, we used formula (4) and (5) to represent the lack of variation between different regions of the images and the evenness of radian distribution and texture thickness in images separately. In formula (4), α_2 represents the inverse different moment, which reflects the homogeneity of image texture and is used to measure the local change of image texture. The larger the value, the less the change, and the more uniform the local area is. In formula (5), α_3 is the entropy, which represents the amount of information the image is, the greater the value is, the more dispersed the elements in the co-occurrence matrix and the more uniform the distribution of the values in GLCM.

$$\alpha_1 = \sum_m \sum_n (m-n)^2 P(m,n) \quad (3)$$

$$\alpha_2 = \sum_m \sum_n \frac{P(m,n)}{1+(m-n)^2} \quad (4)$$

$$\alpha_3 = \sum_m \sum_n (P(m,n))^2 \quad (5)$$

$$\alpha_4 = \sum_m \sum_n \frac{(mn)P(m,n) - u_m u_n}{s_m s_n} \quad (6)$$

The similarity of the spatial gray-level co-occurrence matrices in rows or columns or in specific angle directions can be calculated by the formula (6), in which, α_4 is the correlation to reflect the local gray correlation in the image. When the matrix element values are uniformly equal, the larger the correlation value is, conversely, the smaller the correlation value is.

Among them:

$$u_m = \sum_m \sum_n m.P(m,n) \quad (7)$$

$$u_n = \sum_m \sum_n n.P(m,n) \quad (8)$$

$$s_m^2 = \sum_m \sum_n P(m,n)(m-u_m)^2 \quad (9)$$

$$s_n^2 = \sum_m \sum_n P(m,n)(n-u_n)^2 \quad (10)$$

Finally, the above features are combined with a vector. For example, when the distance difference value (a, b) takes four values, the vector can be obtained synthetically:

$$H = [\alpha_1^1, \alpha_2^1, \alpha_3^1, \alpha_4^1, \ldots, \alpha_1^4, \alpha_2^4, \alpha_3^4, \alpha_4^4] \quad (11)$$

The vector H can be regarded as a description of image texture, which can be further used for classification, recognition, retrieval and so on.

We used the graycomatrix function in Matlab R2018a to create a GLCM which can create a gray-level co-occurrence matrix by calculating how often a pixel with the intensity (gray-level) value m occurs in a specific spatial relationship to a pixel with the value n.In this experiment, each pre-processed image has $256 \times 256 \times 1 = 65{,}526$ dimensions in the vector space after being converted to gray-level image.

2.3. Fuzzy Support Vector Machine

There are two common methods of data classification: supervised classification and unsupervised classification. The supervised classifier is superior to the unsupervised classifier in classification accuracy. Support Vector Machine (SVM) is one of the most advanced classification methods in supervised classification methods which are based on the principles of machine learning. Support Vector Machine (SVM) is usually used for binary classification of data which is a generalized linear classifier. In comparison to other methods such as Decision Trees and artificial neural network, Support Vector Machine can get high accuracy, great ability of mathematical processing and direct geometric interpretation. However, it also has limitations. There is lots of fuzzy information in objective world, if the training of the support vector machine (SVM) concentrate containing noise or outliers, these samples containing "abnormal" often near surface classification in the feature space, lead to obtain the classification of the surface is not really the optimal classification plane. When the data set is polluted by noise or contains outliers, its performance will decline sharply. To solve this problem, fuzzy support vector machines are more efficient in predicting or classifying real data.

Fuzzy support vector machine is developed by introducing fuzzy membership function on the basis of support vector machine[14]. It is an improvement and perfection of support vector machine. To a certain extent, it overcomes the defects of the traditional SVM algorithm noise and outliers sensitive[15], and is widely used in pattern recognition and artificial intelligence.

To achieve the purpose of transforming the training samples into the fuzzy training samples, a fuzzy membership function (FMF) was applied to each object training point in Fuzzy support vector machine (FSVM), which is shown by the following:

$$\{(x_n, y_n, z_n) \mid x_n \in R^z, 0 < z_n \leq 1, \mathrm{n} = 1,\ldots,\mathrm{N}\} \tag{12}$$

Where x_n is the training point for an z -dimensional vector, y_n is a reality class of x_n with a value of -1 or +1, -1 is corresponding to class 1 and +1 is in correspondence with class 2, z_n represents the class's altitude toward which the corresponding training point is oriented, and $(1 - z_n)$ is the less significant altitude.

The fuzzy membership function is set as the distance among a class center of a point and the point itself. Assume that x_+ represent the average value of class +1 and x_- represent the average value of class -1. Then the radius of class +1 and class -1 can be obtained as:

$$r_+ = \min_{\{x_n : y=1\}} |x_n - x_+| \tag{13}$$

$$r_- = \min_{\{x_n : y=-1\}} |x_n - x_-| \tag{14}$$

In formula (12), r_+ denotes the radius of class +1 and in formula (13), r_- denotes the radius of class -1. We define the fuzzy membership z_n as a function of the radius and the average value of the two classes:

$$z_{n1} = 1 - |x_+ - x_n|/(r_+ + \mu), y_n = +1 \tag{15}$$

$$z_{n2} = 1 - |x_- - x_n|/(r_- + \mu), y_n = -1 \tag{16}$$

Where $\mu > 0$ is used to guarantee $z_n > 0$.

Fuzzy support vector machine (FSVM) can predict or classify real data more effectively than standard support vector machine (SVM), among them, some training points are no longer as important as others.

2.4. Principal component analysis

Multivariate statistical methods that use orthogonal transforms to convert a set of potentially correlated variable data into linearly unrelated data are called principal component analysis (PCA), and the converted variables are called principal components [16, 17].

The primary target of PCA is dimension reduction. This is done by retaining the lower-order principal components while losing the set of the higher-order principal components, therefore, the most important aspects of the data are often retained by the lower order components. PCA is essentially a basis transformation, which minimizes the variance between one axis (the principal axis) and the data points by rotation of the coordinate axes and translation of the origin of coordinates, so that the data set can be distributed on the new coordinate axes as much as possible. After coordinate transformation, the orthogonal axis with high variance is removed to obtain the dimensionless data set [18, 19].

Given a set of N-dimensional vectors, which is reduced to T-dimensional ($0 < t \leq n$), and the target is to select T units of orthogonal basis, therefore, after the original data is transformed to this set of basis, the covariance between each feature pair is 0, and the variance of the feature is as large as possible, when the maximum T variance is taken under the constraint of orthogonal.

If we have an n-dimensional data records, arranging them into $n \times m$ matrix$X$, and let $C = \frac{1}{m} XX^T$, then S is a semi-positive definite symmetric matrix ($\{C\}^T \geq 0$), whose diagonals are respectively the variances of each feature, and $C_{i,j} = C_{j,i}$ represents the covariances of i and j respectively. P is a set of matrices composed of bases by rows, Y is the data after the basis transformation of P to X, that is, the data after dimension reduction, $Y = PX$.

$$
\begin{aligned}
D &= \frac{1}{m} YY^T \\
&= \frac{1}{m} (PX)(PX)^T \\
&= P\left(\frac{1}{m} XX^T\right) P^T \\
&= PSP^T \qquad (17)
\end{aligned}
$$

This is transformed into finding a matrix P, where PSP^T is a diagonal matrix with diagonal elements arranged in order from maximum to minimum, and then the first t rows of P are the basis to be sought. To satisfy the original optimization condition, we multiply X by the matrix composed of the first K rows of P and reduce the dimension X from N to T.

PCA is a new orthogonal feature. Its central idea is to map N-dimensional features to T-dimensional features ($0 < t < n$). Such T-dimensional features are called principal components, namely reconstructed T-dimensional features, instead of simply removing the remaining N-T-dimensional features from the N-dimensional features [20].

Based on this theory, we used the PCA to reduce the number of intensity ranges from 256 to 8, in this way, the vector space of each image was reduced to $8 \times 8 \times 1 = 64$ dimensions and a corresponding gray-level co-occurrence matrix was generated. Finally, the gray-level co-occurrence matrix corresponding to all images were summarized.

3. Implementation

In our experiment, we experimented with our method based on MATLAB R2018a and Classification Learner App. The calculations were carried out on a personal computer with 3.60 GHz Core i5-1035G1 CPU, and 16 GB memory, under the operating system of Windows 10. The experimental data were 720 customized Chinese finger language pictures, including 30 categories (26 basic letters and 4 wrap letters). In order to convert the color images into gray-level images, we used the tool software Matlab program. On this basis, the method of PCA was used to reduce the image sizes by reducing the number of intensity ranges from 256 to 8, at the same time, the gray-level matrix is also produced. Finally, we summarized the gray-level matrix of all images to obtain the feature matrix. The default generated image gray-level co-occurrence matrix cannot satisfy the high precision recognition, so we debugged and changed some parameter settings in advanced SVM options to achieve the current accuracy.

4. Experiment and Results

4.1. Statistical Results of Proposed Method

On the basis of 10-fold cross-validation, we compared the performance of some SVMs (default parameters). Table 2 shows that, the accuracy of Quadratic SVM and Fuzzy SVM (FSVM) achieve over 80%, the accuracy of GLCM+FSVM method is comparatively outstanding.

Table 2. Classification accuracy comparison of various SVM methods with default parameter setting

Method	Accuracy	Images	Fold
GLCM+ Linear SVM	75.2%	720	10
GLCM+ Cubic SVM	79.2%	720	10
GLCM+ Quadratic SVM	80.3%	720	10
GLCM+ Fuzzy SVM (Ours)	86.7%	720	10

To further improve the accuracy of the existing identification, we adjusted the parameters of some experimental operating environments. As it turns out, when we set the value of Multiclass method as 'One-vs-All', the accuracy of each method was improved. On this basis, there is still room for the recognition accuracy to improve (see Table 3). After adjusting the parameters, we obtained the best accuracy of FSVM classification is 86.7%.

Table 3. FSVM recognition accuracy corresponding to different parameters (fixed parameter: box constraint level=4, kernel scale mode: manual)

Box constraint level	Manual kernel scale	Accuracy
4	1	52.2%
4	2	74.9%
4	3	83.8%
4	4	85.7%
4	5	86.0%
4	6	86.1%
4	7	85.3%
4	8	84.7%
5	1	52.2%
5	2	74.9%
5	3	83.8%
5	4	85.7%
5	5	85.7%
5(best)	**6(best)**	**86.7%**
5	7	85.6%
5	8	85.0%

When we used the Fuzzy SVM classifier, it is found that the accuracy of the Fuzzy SVM increased linearly with the size of the manual kernel scale under the same box constraint level, but began to decline after a certain extreme value (as shown in ***Figure 4***), however, under the same manual kernel scale, different box constraint level had no change in accuracy (as shown in ***Figure 5***). The experiment found that the optimal box constraint level was 5, the kernel proportional parameter was 6 in the manual kernel scale, and the one-to-one method was selected for the multi-class classification.

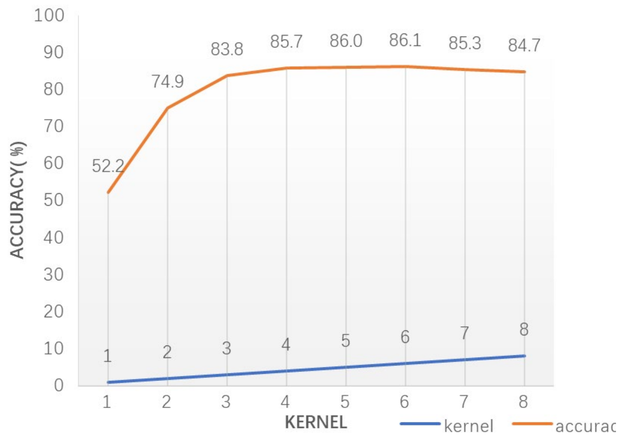

Figure 4. The correlation of Fuzzy SVM classifier manual kernel value and accuracy

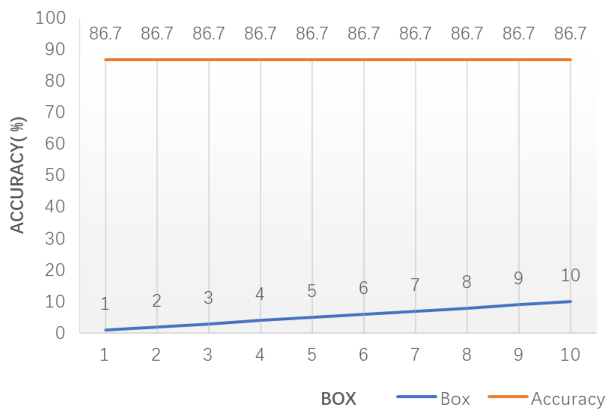

Figure 5. The correlation of Fuzzy SVM classifier box constraint level and accuracy

4.2. SVM versus Fuzzy SVM

Traditional SVM uses the method of constructing hyperplane to classify data, the principle of SVM training algorithm is to predict if a new example belongs to one category or the other, it means that SVM just solves the classification problem of two different kinds of data are on different sides.

The problem of mixed and missing points in multi-class classification of traditional SVM was solved by introducing fuzzy membership function to support vector machines and the classification accuracy was improved. FSVM allows fitting the maximum boundary hyperplane in the transformed feature space. FSVM can be mapped to wireless dimension and decision boundary. The most important thing is that FSVM performs well in classification and can get a higher classification and identification accuracy than SVM.

4.3. Comparison to State-of-the-art Approaches

In this paper, we compare four classifiers. The first is Artificial Neural Network (ANN). It consists mainly of simple operations such as addition and multiplication and is therefore easy to implement in digital hardware. A limited number of basic artificial neurons for data exchange make up an ANN. A feed forward neural network was designed for signal classification, which provides a good solution to the pattern recognition problems and produces a good result. But it also has disadvantages, such as it cannot work when there is not enough data and it turn all the characteristics of the problem into numbers and all the reasoning into numerical calculations and this can result in a loss of information [21]. The second is K-nearest neighbor (K-NN), which is one of the most basic and simple classification methods. In classification studies, when little is known about the distribution of data or lack of prior knowledge, it is usually chosen first [22]. Two distance measures: Euclidean distance and Correlation Factor were usually used in classification with KNN. The principle of KNN algorithm is that, it finds the k nearest records to the input data from the pre-processed training set and then classified according to the selected k records. But there are some disadvantages to this method, which are that it has a high computation cost and the performance requirement of computer is high. The third is Support Vector Machine (SVM) which is a supervised learning method. This approach is often used in classification and regression problems. SVM is also a kind of linear classifier, which was used in the deep learning for a long time. It can classify the pre-extracted data and give each data specific score as the basis of evaluation. One of the most important aspects in deep learning is how to classify the extracted data. The extracted data is usually represented as the N-dimensional vectors, that is where the Vector Machine name comes from. In SVM, the training examples are divided into two categories and each training example belongs to one of two categories. The model of SVM training algorithm predicts if the sample belongs to one category or another. At present, SVM is also proposed to pre-train illegal images, and then directly verify and block the illegal sites. As a gesture classifier, it performs well in computer and has strong generalization ability. The last is Decision Trees. The purpose of this method is obtaining a tree-like structure based on a principle that makes the separation of the data been minimum by splitting the data set repeatedly [23-25]. The greedy construction process presents one of the main disadvantages of the Decision Trees: in each step, it always selects the combination of a single optimal variable and an optimal split point; but the truth is that considers a multi-step pre-detection of variable combinations may get different or better results [26]. By comparing the above machine learning methods, we can conclude that, Decision Trees has the advantage that it can be easily measured by static testing. The reason why Fuzzy support vector machine succeeds Decision Trees is that it can map some linear indivisible problems in two-dimensional space to higher dimensional space and become linearly separable.

The following table compares support vector machine and other classifiers based on the database with 720 customized Chinese finger language pictures, including 30 categories (26 basic letters and 4 roll tongue pronunciation words) (as shown in Table 4).

Table 4. The accuracy comparison of the four classification methods

Method	The highest accuracy	Images
ANN	51.0%	720
K-NN	83.5%	720
SVM	76.3%	720
Decision Trees	48.3%	720
GLCM+ Fuzzy SVM	86.7%	720

5. Discussion

Sign language is a unique way for deaf-mutes to communicate, and it is a significant method for them to contact with the outside world. Therefore, the study of sign language recognition technology has caused a hot research in the emerging technology of human-computer interaction technology. With the fast development of technology and society, researchers pay more attention to the development of communication technology between able people and dumb people.

In this paper, we calculated the image's gray-level to obtain the gray-level co-occurrence matrix, on the heels of that, some eigenvalues of the matrix were obtained to represent some texture features of the image respectively. The gray-level co-occurrence matrix of an image can reflect the comprehensive information about the direction, adjacent interval and change amplitude of the image gray-level. It is the basis of analyzing the local patterns of the image and their arrangement rules. During feature extraction, whether the feature matrix can be further optimized to make the classification more accurate needs further consideration. Next, we consider representing the feature matrix as a hash matrix. SVM is a supervised learning algorithm that can be used for both classification and regression problems. For SVM, it is difficult to realize large-scale training samples and solve the problem of multi-classification. Our Fuzzy SVM is a method to add fuzzy membership function to SVM which can map data to infinite dimensions, with more diverse decision boundaries and only one parameter, making it easier to choose. It also has disadvantages, for example, it is poor interpretability (infinite multi-dimensional conversion), slow calculation speed and easy overfitting when the parameters are not chosen well. The comparison of various classification methods can refer to the Section 4**.**

To sum up, in feature extraction, the efficiency of data operation should be improved as much as possible. For example, dimension reduction is adopted in this paper. When using the SVM, the selection of fuzzy membership value and some weights is critical and sufficient data samples are required. Next, we will consider introducing different kernel functions such as Guassian function and polynomial function into SVM to further improve the recognition accuracy.

6. Conclusion

This paper introduces the difference between finger language and sign language in Chinese sign language, highlighting the unique certainty and importance of finger language in the process of research and application. In our research, we proposed a new model to recognition Chinese finger language by using gray-level co-occurrence matrix (GLCM) and fuzzy Gaussian support vector machine (FSVM) and achieved higher identification accuracy than other existing approaches.

Our future work will focus on the following aspects: (i) Automatic segmentation of key areas in sign language images using computer programs; (ii) The calculation will be further simplified by converting the gray-level matrix into a sparse matrix; (iii) Continue testing other advanced classifiers such as the extreme learning machine, kernel SVM [27], and convolutional neural network; (iv) Continue testing other advanced feature extraction methods such as Motion Boundary Motion (MBH) [28] and Scale Invariant Feature Transform (SITF) [29, 30]; (v) Applying our method to the development of sign language translator.

Acknowledgement

This work was supported from The Natural Science Foundation of Jiangsu Higher Education Institutions of China (19KJA310002), The Philosophy and Social Science Research Foundation Project of Universities of Jiangsu Province (2017SJB0668), The Natural Science Foundation of Jiangsu Province (16KJB520029).

References

[1] Q. Y. Jiang and W. J. Li, ""Deep Cross-Modal Hashing"," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2017, pp. 3270-3278.

[2] P. Yeole, A. Tajane, G. Karade, and P. V. N. Ghodke, "Vision Based Hand Gesture Recognition."

[3] J. Huang, W. Zhou, H. Li, and W. Li, "Sign Language Recognition using 3D convolutional neural networks," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 2015.

[4] M. Mohandes, M. Deriche, and J. Liu, "Image-Based and Sensor-Based Approaches to Arabic Sign Language Recognition," *IEEE Transactions on Human-Machine Systems,* vol. 44, no. 4, pp. 551-557, 2014.

[5] B. P. P. Kumar and M. B. Manjunatha, "A Hybrid Gesture Recognition Method for American Sign Language," *Indian Journal of ence and Technology,* vol. 10, no. 1, 2017.

[6] T. Starner and J. Weaver, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, no. 12, pp. P.1371-1375, 1998.

[7] P. V. V. Kishore, "4-Camera model for sign language recognition using elliptical fourier descriptors and ANN," in *International Conference on Signal Processing & Communication Engineering Systems*, 2015.

[8] L. Lei and Q. Dashun, "Design of data-glove and Chinese sign language recognition system based on ARM9," in *IEEE International Conference on Electronic Measurement & Instruments*, 2015.

[9] L. Pigou, S. Dieleman, P. J. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," in *Workshop at the European Conference on Computer Vision*, 2014.

[10] L. Geng, X. Ma, B. Xue, H. Wu, and Y. Li, "Combining features for Chinese sign language recognition with Kinect," in *IEEE International Conference on Control & Automation*, 2014.

[11] P. Chophuk, K. Pattanaworapan, and K. Chamnongthai, "Fist american sign language recognition using leap motion sensor," in *International Workshop on Advanced Image Technology*, 2018.

[12] R. M. HARALICK, "Textural features for image classification. IEEE Transaction on Systems, Man, and Cybernetics," *SMC,* vol. 3, 1973.

[13] M. Pratiwi, Alexander, J. Harefa, and S. Nanda, "Mammograms Classification Using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network," *Procedia Computer Science,* vol. 59, pp. 83-91, 2015.

[14] G. M. Xian, "An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM," *Expert Systems with Applications,* vol. 37, no. 10, pp. 6737-6741, 2010.

[15] C. F. Lin and S. D. Wang, "Fuzzy Support Vector Machines," *IEEE Transactions on Neural Networks,* vol. 13, no. 2, pp. 464-471, 2002.

[16] M. Turk, "Eigenfaces for recognition," *J. Cognitive Neuroscience,* vol. 3, 1991.

[17] Diamantaras, K. I. Kung, and Y. S., "Principal Component Neural Networks," 1996.

[18] Pentland, Moghaddam, and Starner, "View-based and modular eigenspaces for face recognition," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2002.

[19] A. A. Ahmed and S. Aly, "Appearance-based Arabic Sign Language recognition using Hidden Markov Models," in *International Conference on Engineering & Technology*, 2015.

[20] B. Lee, Y. Cho, and S. Cho, "Translation, scale and rotation invariant pattern recognition using principal component analysis (PCA) and reduced second-order neural network," *Neural Parallel & Scientific Computations,* vol. 3, no. 3, pp. 417-429, 1995.

[21] S. Dreiseitla and L. Ohno-Machadob, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics,* vol. 35, no. 5–6, pp. 352-359, 2002.

[22] L. E. Peterson, "K-nearest neighbor," *Scholarpedia,* vol. 4, no. 2, p. 1883, 2009.

[23] L. I. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees," *Encyclopedia of Ecology,* vol. 57, no. 3, pp. 582-588, 2015.

[24] L. Breiman and etc, *Classification and regression trees*. Wadsworth International Group.

[25] S. L. Salzberg, "Book Review: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Machine Learning,* vol. 16, no. 3, pp. 235-240, 1994.

[26] Yaxin Sun and Guihua Wen, "Ensemble softmax regression model for speech emotion recognition," *Multimedia Tools & Applications,* 2017.

[27] X. X. Zhou *et al.*, "Combination of stationary wavelet transform and kernel support vector machines for pathological brain detection," *Simulation,* vol. 92, no. 9, pp. 827-837, 2016.

[28] N. Dalal, B. Triggs, and C. Schmid, "Human Detection Using Oriented Histograms of Flow and Appearance," in *European Conference on Computer Vision*, 2006.

[29] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire,

and J. P. Thiran, "Scale Invariant Feature Transform on the Sphere: Theory and Applications," *International Journal of Computer Vision,* vol. 98, no. 2, pp. 217-241, 2012.

[30] X. Jiang, S. C. Satapathy, L. Yang, S.-H. Wang, and Y.-D. Zhang, "A Survey on Artificial Intelligence in Chinese Sign Language Recognition," *Arabian Journal for Science and Engineering,* pp. 1-36, 2020.