A Nuclide Recognition Method for Nuclear Robot System

Siyi Zhou, Jiangmei Zhang*, Xinghua Feng and Caolin Zhang
Southwest University of Science and Technology, Mianyang 621010, China

*Corresponding author email: zjm@swust.edu.cn

Abstract. In the real energy spectrum attenuation environment, many traditional nuclide identification methods for nuclear robot systems have problems such as using only part of the energy spectrum curve, being susceptible to noise, and having low recognition accuracy. Proposes an energy spectrum nuclide recognition method based on S-transform (ST) and Mahalanobis distance-based support vector machine (MSVM). Regarding the energy spectrum curve as a non-stationary signal, combined with the widely used S transformation method in signal transformation, the energy spectrum data is two-dimensional. Then use two-dimensional principal component analysis (2D-PCA) to reduce the dimension of the two-dimensional energy spectrum data for feature extraction, and design a support vector machine (SVM) classifier based on Mahalanobis distance to realize the identification of energy spectrum nuclides. Finally, experiments are carried out with simulated nuclide energy spectrum data based on Geant4. The experimental results show that this method effectively improves the accuracy of energy spectrum nuclide recognition by using full spectrum information. At the same time, experiments are carried out on the nuclide energy spectrum data of different detection distances obtained by the NaI detector in the real environment, and it is verified that the algorithm proposed in this paper also has a good recognition performance for the nuclide energy spectrum collected in the real environment.

Keyword: Nuclide recognition; S transform; 2D-PCA; Data dimensionality reduction; Mahalanobis distance; SVM; Genetic algorithm.

1. Introduction
Nuclide identification technology is a basic technology to ensure nuclear safety, and it has important practical significance for the development of nuclear safety. Nuclide recognition refers to the use of detectors to receive gamma rays emitted by radioactive isotopes in the real detection environment, and generate energy spectrum data, and analyze the energy spectrum to realize the identification of the type of nuclide. The core idea of traditional nuclide identification is to find the all-power peak in the energy spectrum, and then calculate the peak area, half-height width and other basic information, and finally compare with the nuclide energy in the nuclide library to obtain nuclide discrimination information. But based on the traditional peak-seeking nuclide identification method, in the real energy spectrum attenuation environment, the error penetrates, the identification rate is improved, and a good identification effect cannot be obtained. In recent years, with the development of machine learning, emerging nuclides identification methods based on machine learning have gradually appeared. Most of the emerging nuclides identification methods are mainly based on neural networks and sequential Bayesian theory, fuzzy theory and spatial projection based on information theory and other methods. Although these methods improve the accuracy of nuclide identification to a certain extent, they cannot effectively deal with overlapping peaks and it is difficult to accurately eliminate "false peaks”. In addition, most researchers directly input the one-dimensional energy spectrum data as the model input to find the effective characterization features of the nuclide energy spectrum, and the
effective characterization features are not easy to extract. Most of them still use omnipotent peak area and position as input features, which still cannot effectively overcome the influence of noise. Therefore, in response to the above problems, the author conducts the identification of energy spectrum nuclides from the perspective of full spectrum analysis. Treat the original energy spectrum as a non-stationary signal, and use the S transform, which is widely used in signal analysis, to transform the one-dimensional energy spectrum data into two dimensions, to understand and explore nuclide energy spectrum information from a new perspective and a new dimension. Use 2D-PCA to project the two-dimensional energy spectrum to obtain the projection space of the energy spectrum of different nuclides, realize the identification of nuclides, and improve the accuracy of nuclides identification.

2. Principles of S-transformation Theory
S transform is a new time-frequency analysis method proposed by Stockwell et al\(^{(8)}\) in 1996. Its idea is to develop short-time Fourier transform and continuous wavelet transform. The continuous ST of the signal \(x(t)\) is defined as follows:

\[
S(t, f) = \int_{-\infty}^{\infty} x(t) \alpha(\tau - t, f) e^{-j2\pi ft} dt
\]

(1)

\[
\alpha(\tau - t, f) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\tau - t)^2}{2}}
\]

(2)

Among them \(\alpha(\tau - t, f)\) is the Gaussian window function, \(\tau\) is the time shift parameter that controls the position of the Gaussian window on the t axis.

For the discrete S transform of \(x(t)\), you can perform Fourier transform and inverse Fourier transform on equation (1) first, that is:

\[
S(\tau, f) = \int_{-\infty}^{\infty} x(v + f)e^{-j2\pi f \tau} dv
\]

(3)

Discretize the above formula (3) again, and the discrete representation of S transform can be obtained:

\[
S[kT, \frac{n}{NT}] = \sum_{m=0}^{M-1} x[m+\frac{n}{NT}] \frac{2\pi m}{NT} e^{\frac{j2\pi nkT}{NT}}
\]

(4)

\[
S[kT, 0] = \sum_{m=0}^{M-1} x[m] \frac{m}{NT}
\]

(5)

Among them: \(k, m, n = 0, 1, 2, ..., N - 1(n \neq 0)\).

The result of S transformation is a complex time-frequency matrix, denoted as S matrix. After modulating each element of S matrix, the S modulus matrix can be obtained, denoted as STA matrix, which expresses the amplitude relationship between the corresponding time, frequency and S-transformed.

3. Principles of two-dimensional Principal Component Theory

3.1. One-dimensional Principal Component Analysis (1D-PCA)
One-dimensional principal component analysis (1D-PCA) is the traditional principal component analysis method. It is the most important and commonly used method of dimensionality reduction. It uses a linear transformation to project samples in the original space into a new dimensional space. 1D-PCA is only suitable for one-dimensional data processing. The idea is to find a set of optimal unit orthogonal basis (ie principal components) through linear transformation, use their linear combination to reconstruct the original sample, and make the reconstructed the error between the sample and the original sample is the smallest.

The basic principles of 1D-PCA are as follows: Suppose there are \(N\) groups of \(1 \times M\) dimensional data
Performing 1D-PCA on it is equivalent to finding the eigenvector matrix $W$ corresponding to the first $n$ eigenvalues of the covariance matrix $C = XX'$. $W$ is the best projection matrix for spatial transformation of sample data $X_i$. Then, for each sample $X_i$, the following transformation is performed:

$$Z_i = W^T X_i$$

(6)

$$F_i = X_i W$$

(7)

Among them, $Z_i$ is the projection of $X_i$ on the hyperplane in the new space after 1D-PCA transformation. $F_i$ is the feature vector of $X_i$ after 1D-PCA transformation.

### 3.2. Theoretical Principles of Two-dimensional Principal Component Analysis (2D-PCA)

2D-PCA is the inheritance and development of 1D-PCA. For a two-dimensional matrix, 1D-PCA needs to turn the sample matrix into a one-dimensional vector when constructing the covariance matrix, after that, all the sample vectors are formed into a large-scale covariance matrix, and then the eigenvalues and eigenvectors of this covariance matrix are solved to construct the mapping moment. When 2D-PCA analyzes a two-dimensional image matrix, the covariance matrix is calculated directly on the image matrix, so the spatial structure information of the image can be preserved. And it can avoid the curse of dimensionality and the problem of small samples, which also makes it widely used in image processing. The basic principles of 2D-PCA are as follows: Assuming that $X$ is an $n$-dimensional column vector, the $m \times n$ dimensional matrix $A$ is mapped to another subspace through the matrix $X$, that is:

$$Y = AX$$

(8)

Among them, $Y$ is the eigenvector matrix after projection.

$$J(X) = tr(S)$$

(9)

Among them

$$S_i = E[(Y - EY)(Y - EY)'] = E[(A - EA)XX'(A - EA)'] = E[X'(A - EA)'(A - EA)X] = X'G X$$

(10)

In the formula $G = E[(A - EA)'(A - EA)]$, the matrix $G$ is called the image covariance matrix. Therefore, formula (9) can be rewritten as:

$$J(X) = tr(X'GX)$$

(11)

Among them, the image covariance matrix $G$ is calculated in a simple way:

$$G = \frac{1}{M} \sum_{i=1}^{M} (A_i - \bar{A})'(A_i - \bar{A})$$

(12)

Among them $\bar{A}$ represents the average image matrix, there are:

$$\bar{A} = \frac{1}{M} \sum_{i=1}^{M} A_i$$

(13)

The projection criterion of 2D-PCA is to find $d$ projection axes $X = [X_1, X_2, ..., X_d]$ such that the criterion $J(X)$ is the largest, namely:

$$\begin{cases}
[X = [X_1, X_2, ..., X_d]] = \text{argmax} J(X) \\
X_i'X_j = 0, (i \neq j, i,j = 1,2,...,d)
\end{cases}$$

(14)
For each given sample matrix $A_i$, perform the following transformation:

$$Y = A_i X$$  \hspace{1cm} (15)$$

Thus, the image feature vector matrix $Y$ that has been transformed by 2D-PCA can be obtained.

4. Design of SVM Classifier Based on Mahalanobis Distance

Support vector machine (SVM) is a machine learning method developed on the basis of the structural risk minimization of statistical learning theory, it has strong learning ability and generalization ability, can effectively solve the problems of small samples, nonlinearity, and local extreme values, and has excellent accuracy in dealing with two-class and multi-class problems. However, the traditional SVM has the disadvantage of only extracting the partial information of the sample when performing classification, and ignores the attribute information contained between different samples, which affects the improvement of the generalization ability of the SVM. The Mahalanobis distance discrimination is a reliable method for the analysis of the separability measurement between sample classes, it can take into account the hidden information contained between different samples and use the global information of the sample to find the relationship between various features. Therefore, it can effectively measure the similarity of different sample populations. Therefore, the Mahalanobis distance-based SVM (MSVM) classifier is proposed, and a classification algorithm for determining the two-dimensional radionuclide energy spectrum is designed. The basic idea of the algorithm is as follows:

Assuming there are multiple training samples, the between-class and intra-class covariance matrices are denoted as $S_b$ and $S_w$, respectively, and $S_i = S_b + S_w$ is the overall covariance matrix of the sample. In the SVM method, the classification hyperplane is determined by the classification decision function $f(x) = w^T x + b$, combined with the principle of minimum interval. It can be transformed into a secondary optimization problem, namely:

$$\begin{align*}
\min_{w, b} & \frac{1}{2}w^T S_w w + C \sum_{i=1}^{l} \zeta_i \\
\text{s.t.} & \quad y_i (w^T \phi(x_i) + b) \geq 1 - \zeta_i, \\
& \quad \zeta_i \geq 0, i = 1, 2, \ldots, l
\end{align*}$$  \hspace{1cm} (16)$$

Based on considering the global information of the sample, this paper uses $w^T S_w$ to replace $w^T w$ in the optimization problem, and substitutes it into the support vector machine obtained by formula (16) to express as:

$$\begin{align*}
\min_{w, b} & \frac{1}{2}w^T S_w w + C \sum_{i=1}^{l} \zeta_i \\
\text{s.t.} & \quad y_i (w^T \phi(x_i) + b) \geq 1 - \zeta_i, \\
& \quad \zeta_i \geq 0, i = 1, 2, \ldots, l
\end{align*}$$  \hspace{1cm} (17)$$

Where, the overall covariance matrix of the data sample is calculated as $S_i = E[(\phi(x) - E(\phi(x))) (\phi(x)^T - E(\phi(x)^T))]$, $C$ is the penalty parameter, and $\zeta_i$ represents the slack variable of each sample.

Then perform Lagrangian transformation on equation (17), there are:

$$\begin{align*}
\max & \quad -\frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} a_i a_j y_i y_j \phi(x_i)^T \phi(x_j) + \sum_{i=1}^{l} a_i \\
\text{s.t.} & \quad \sum_{i=1}^{l} a_i y_i = 0, \\
& \quad 0 \leq a_i \leq C, i = 1, 2, \ldots, l
\end{align*}$$  \hspace{1cm} (18)$$
Where $a_i$ is the Lagrangian multiplier, and the Mahalanobis distance is defined as

$$K_M(x_i,x_j) = \phi(x_i)S\phi(x_j) = \left(\phi(x_i),\phi(x_j)\right)$$

Substituting the above equation (19) into equation (18), the support vector machine based on Mahalanobis distance is:

$$\begin{align*}
\max & -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_iy_iy_jK_M(x_i,x_j) + \sum_{i=1}^{l}a_i \\
\text{s.t.} & \sum_{i=1}^{l}a_iy_i = 0 \\
& 0 \leq a_i \leq C, i = 1,2,...,l
\end{align*}$$

Thus, the decision function of the support vector machine based on Mahalanobis distance is obtained:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{l}a_iy_iK_M(x,x_i) + b\right)$$

Where $s$ is the number of support vectors.

5. Algorithm Overview

According to the above, this article aims at the real energy spectrum attenuation environment, many traditional nuclide identification methods only use part of the energy spectrum curve, are susceptible to noise, and the recognition accuracy is low. The design is based on S transform and 2D-PCA's gamma spectrum nuclide recognition algorithm, the algorithm is as follows:

1) Transform the original $\gamma$ energy spectrum nuclide data into a two-dimensional image matrix through S transformation, and divide it into a training set and a test set;
2) Perform 2D-PCA transformation on the transformed two-dimensional image matrix for data compression and feature extraction;
3) Design an SVM classifier based on Mahalanobis distance, train the training data and use the test set for verification, and obtain the classification results.

The overall system block diagram of the method in this paper is shown in Figure 1.

![Figure 1. Algorithm overall system block diagram.](image)

6. Nuclide Spectrum Recognition

6.1. Data Acquisition and Preprocessing of the Nuclide Energy Spectrum

Accurate identification of nuclides requires more sample data of nuclides. In this paper, Geant4 software is used to generate the energy spectrum of nuclides based on the Monte Carlo method. At the same time, combined with the NaI detector to carry out experimental testing to obtain the nuclide energy spectrum in the real environment as a nuclide energy spectrum sample set. Nuclide energy spectrum samples mainly include six single-category nuclides and mixed nuclides, including $^{60}$Co, $^{137}$Cs, $^{152}$Eu, $^{60}$Co&$^{137}$Cs, $^{60}$Co&$^{152}$Eu and $^{137}$Cs&$^{152}$Eu. Part of the nuclide energy spectrum sample is shown in Figure 2 below. It can be seen that all the energy spectrum data contains part of the environmental noise interference. At
the same time, the author uses S transform and 2D-PCA to preprocess the obtained nuclide energy spectrum. After different nuclide energy spectra undergo S transform, the obtained two-dimensional image is shown in Fig. 3. It can be seen that different nuclei, there are obvious differences after the two-dimensionlization of the element. Since the dimension of the two-dimensional matrix after S transformation is too large, it not only increases the computational burden, but also contains more redundant information, so 2D-PCA is used to compress data and reduce the dimension to facilitate subsequent nuclide identification.

Figure 2. Spectral samples of different nuclides.

Figure 3. S transform diagram of different nuclide energy spectra.

6.2. Nuclide Recognition Model Training
After the original nuclide energy spectrum data is preprocessed in section 5.1, the obtained two-dimensional nuclide energy spectrum data is divided into training samples and test samples in a ratio of 7:3. Use the MSVM classifier designed in this article to classify and identify nuclides. In addition, in order to obtain the optimal parameter settings of MSVM, this article introduces genetic algorithm (GA) to find the optimal parameters. GA is a stochastic global search and optimization method developed by imitating the biological evolution mechanism of nature. It can adaptively control the search process to find the best solution.

Take the training samples as input, and set the optimal parameters through GA. As shown in Figure 4, the optimal parameters are obtained \((c=0.2911, g=2.2037)\). And the average classification accuracy rate is greater than 99%, and the best classification accuracy rate is 100%.

Figure 4. Genetic algorithm parameter optimization.

6.3. Recognition Results and Analysis of Simulated Nuclide Data
In this experiment, the nuclide energy spectrum data used are all based on the nuclide energy spectrum
simulation data obtained by the Gent4 simulation software. The simulated nuclear spectrum recognition includes the identification of single nuclides and the identification of mixed nuclides. The test samples that are not involved in the model training are used as input, and the trained MSVM is used to identify six different nuclear spectrum data. At the same time, compared with the traditional peak finding method\cite{11}, the support vector machine method based on singular value decomposition\cite{12} and the CNN-based nuclide recognition method\cite{13}, the nuclide recognition results are shown in Table 1 below.

| recognition methods | 60Co | 137Cs | 152Eu | 60Co&137Cs | 60Co&152Eu | 137Cs&152Eu |
|---------------------|------|-------|-------|------------|------------|------------|
| Peak finding\cite{11} | 94   | 100   | 100   | 87.54      | 87.93      | 88.76      |
| SVD+SVM\cite{12}    | 98   | 100   | 100   | 99.27      | 98.46      | 98.75      |
| CNN\cite{13}        | 100  | 100   | 100   | 99.98      | 100        | 100        |
| **Our algorithm**    | **100** | **100** | **100** | **100**  | **100**  | **100**  |

The results in Table 1 show that the algorithm in this paper has a good recognition effect on the simulated gamma nuclide energy spectrum, can accurately identify single nuclides and mixed nuclides, at the same time, the model training speed is fast, the parameter optimization can be automatically performed, and the recognition speed is fast. Although the traditional peak finding algorithm and the support vector machine method based on singular value decomposition are better for single nuclides, their recognition performance for mixed nuclides has a certain degree of degradation. In addition, the traditional peak finding algorithm needs to adjust various parameters in order to obtain a better recognition rate, and the support vector machine method based on singular value decomposition also needs to select appropriate eigenvalues to obtain better accuracy of nuclide recognition. Although the recognition performance of the CNN method is also good, compared with the proposed nuclide recognition algorithm in this paper, the training of the CNN recognition model requires a rather large data set, poor interpretability, high operation intensity and long training time.

### 6.4. Results and Analysis of Nuclide Recognition in Real Environment

In the experiment in this section, the algorithm proposed in this paper is trained by model using the simulated radionuclide energy spectrum data of Geant4 as the training set. In the real environment, the radionuclide energy spectrum data of different detection distances obtained by the NaI detector is used as the test set. So as to verify the recognition performance of the algorithm proposed in this paper on the nuclide energy spectrum collected in the real environment, the nuclide identification results in real environment are shown in Table 2 below.

| Detection distance (cm) | 60Co | 137Cs | 152Eu | 60Co&137Cs | 60Co&152Eu | 137Cs&152Eu |
|-------------------------|------|-------|-------|------------|------------|------------|
| Close distance (0–30)   | 98.04 | 100   | 99.35 | 100         | 100        | 100        |
| Middle distance (30–60) | 97.40 | 96.68 | 98.72 | 100         | 98.06      | 96.77      |
| Long distance (70–100)  | 91.33 | 97.40 | 86.00 | 96.15      | 96.16      | 94.94      |
| **Average (%)**         | **95.59** | **98.69** | **94.69** | **98.72** | **97.24** | **97.24** |

The results in Table 2 show that when the detection distance is small, the accuracy of the algorithm in this paper is relatively high for different nuclides. At the same time, it can be seen that under the condition that the detection distance is within 100 cm, the nuclide recognition algorithm proposed in this paper has a high recognition accuracy rate for the nuclide energy spectra obtained in the real detection environment at different detection distances. And the average recognition rate of every single nuclide and mixed nuclide in the experiment is higher than 94%, showing good nuclide recognition performance.
7. Conclusion
By transforming the one-dimensional energy spectrum data into two dimensions and exploring the information of the nuclide energy spectrum from the full spectrum analysis and two-dimensional perspective, a SVM classifier based on Mahalanobis distance is designed and trained, Combined with genetic algorithm to optimize the model parameters. Comparison experiments with other three nuclide recognition algorithms verify the effectiveness of this algorithm, and the accuracy of the proposed algorithm can be effectively improved compared with the other three nuclide recognition algorithms.

Acknowledgments
This work was supported by the National Defense Basic Research Program (JCKY2020404C004) and Sichuan Province Military-civilian integration Industry development special fund project(18zs9101; 20zs9101).

References
[1] El_Tokhy Mohamed S.. Rapid and robust radioisotopes identification algorithms of X-Ray and gamma spectra[J]. Measurement,2021,168.
[2] Wang Y, Liu Z M, Wan Y P, et al. Energy spectrum nuclide recognition method based on long and short-term memory neural network[J]. High Power Laser and Particle Beam,2020,32(10):154-161.
[3] Wang B R, Wu Z Q. Using LVQ neural network to identify radionuclides[J]. Nuclear Electronics and Detection Technology,2018,38(04):572-576.
[4] Du X C, Tu H B, Li K, et al. Nuclide recognition method based on radial basis function neural network simulation of γ energy spectrum template library[J/OL]. Journal of Tsinghua University (Natural Science Edition): 1-8[2021-05-27].
[5] Wen S Y, Wang B R, Xiao G, et al. Research on Nuclide Recognition Algorithm Based on Sequential Bayes Method[J]. Nuclear Electronics and Detection Technology,2016,36(02):179-183.
[6] Wang Y M, Wei Y X. Nuclide Recognition of γ Energy Spectrum Based on Fuzzy Logic[J]. Journal of Tsinghua University (Natural Science Edition),2012,52 (12):1736-1740.
[7] Yin M W, Ren X M, Liao P, et al. New method of γ energy spectrum identification based on information theory (English)[J]. High Power Laser and Particle Beam,2018,30(10):144-148.
[8] Stockwell R G, Mansinha L, Lowe R P. Localization of the complex spectrum: the S transform[J]. IEEE Transactions on Signal Processing, 1996,44(4):998-1001.
[9] Nguyen Quang H, Nguyen Binh P, Nguyen Trung B, et al. Stacking segment-based CNN with SVM for recognition of atrial fibrillation from single-lead ECG recordings[J]. Biomedical Signal Processing and Control,2021,68.
[10] Cheng Liangliang, Yaghoubi Vahid, Paepegem Wim Van, et al. On the Influence of Reference Mahalanobis Distance Space for Quality Classification of Complex Metal Parts Using Vibrations[J]. Applied Sciences,2020,10(23):8620-8620.
[11] Pang J S, Zheng G F, Hou X F. Peak Finding by Symmetric Zero Area Transformation[J]. Atomic Energy Science and Technology,1987(03):270-279.
[12] Ren J S, Zhang J M, Wang K P. Nuclide recognition algorithm based on SVD and SVM[J]. Ordnance Industry Automation,2017,36(05):50-53.
[13] Hu H H, Zhang J M, Wang K P, et al. Application of Convolutional Neural Network in Recognition of Complex Nuclides[J].Sensors and Microsystems, 2019, 38(10):154-156+160.