VOL. 7, NO. 7, JULY 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS FOR IDENTIFICATION OF DRUGS PACKED IN ANTHROPOMORPHIC PHANTOM

Wei Li[1,2], Yu Zhong[1,3], Daoyang Yu[1], Dingjun Qu[1,2], Bai Sun[1], Minqiang Li[1] and Jinhuai Liu[1]
[1]Research Center for Biomimetic Functional Materials and Sensing Devices, Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, PR China
[2]Department of Automation, University of Science and Technology of China, Hefei, PR China
[3]Anhui Provincial Laboratory of Biomimetic Sensor and Detecting Technology, West Anhui University, Lu'an, PR China
E-Mail: jhliu@iim.ac.cn mqli@iim.ac.cn

ABSTRACT

A method of feature extraction based on principal component analysis (PCA) to identify drugs packed in anthropomorphic phantom was proposed. An energy dispersive X-ray diffraction spectrometer was used to acquire the spectra of drugs and the other materials. Features of spectra were extracted by PCA which excels in analyzing high-dimensional data. Furthermore, the number of principal components selected as features was discussed. Combined with artificial neural network, a success rate of recognition between 93.3% and 100% was achieved. It indicates that this method is promising in detecting drug body packing.

Keywords: principal component analysis, artificial neural network, drug body packing, anthropomorphic phantom, energy dispersive X-ray diffraction.

1. INTRODUCTION

Drug body packing, which is defined as smuggling drug containers in gastrointestinal and vaginal tract, is an important approach for trafficking drugs. With the spread of this criminal activity, developing effective methods and instruments to detect and analyze body packing has been increasingly focused. Among various techniques used to detect drugs in complex background, such as acoustic technology [1], fast neutron scattering [2], CT [3], terahertz imaging [4] and homogeneous phase protein-based assay [5], energy dispersive X-ray diffraction (EDXRD) has been shown to be a suitable one due to its non-destructive, high selection and high efficiency [6-10].

In EDXRD spectrum, due to the interference of body, the signal-to-noise ratio is low and diffraction peaks of drug usually hide behind diffraction profile of body. Therefore, the feature extraction is a challenge in detecting drug body packing using EDXRD. Besides the position of diffraction peak, the diffraction profile is another important feature, because spectra resulting from the scattering media's diffraction effects are superimposed on the incident spectra to produce a unique diffraction profile. It can be regarded as the "fingerprint" of the detected object, which has been demonstrated previously [6-7].

This "fingerprint" can be analyzed by principal component analysis (PCA), which reduces the dimensions of data by creating a new set of uncorrelated variables [11-12], and is an effective method of feature extraction. It linearly extracts the signals in different channels of detector, and rebuilds them into principal components (PCs), which describe the main difference between samples. PCA has been reported to analyze poor EDXRD spectra of illicit materials [6-7, 13-14].

In this paper, PCA was introduced as a method of feature extraction, and drugs packed in body were identified by the features extracted by PCA. Heroin and methamphetamine were chosen as the representation of drugs, and anthropomorphic phantom was chosen instead of actual human body. In addition, some drug precursors (phenylacetic acid and piperonal), explosive (TNT) and disruptors (salt) were also analyzed. Based on different number of PCs, feature extraction methods of different feature dimensions were set up, and their success rates of recognition were evaluated.

2. THE METHOD AND MATERIAL

2.1. The experimental system

The EDXRD spectrometer of our experiment is shown in Figure-1. The polychromatic source of X-ray comprised an X-ray tube and the related collimators to define the incident and scattered beams. The X-ray source was W target in the range of 0-75 kV and 0-3 mA. A Canberra planar high purity germanium detector with a resolution of 140 eV at 5.9 keV combined with a Canberra spectrum master InSpector 2000 was adopted. The width of the third slit collimator (near the horizontal Soller slit) was 0.75 mm, and the fourth (near the detector) was 2.0 mm. The distance between X-ray source and sample (anthropomorphic phantom) was 35 mm, the distance between sample and the third collimator was 130 mm, and the distance between the third collimator and the fourth collimator was 50 mm. The scatter angle is 4° and radiation time was 30 seconds. The radiation measured at the detector was 0.75 msv in 30 seconds, which was safe for human in practical use.

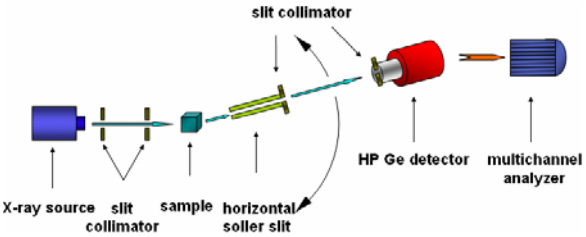


Figure-1. Geometry of the EDXRD spectrometer.

The rule of measured EDXRD profiles is given by Bragg's law [15]. In the case of EDXRD, it can be expressed as a function of momentum transfer q, which depends on the energy of the photon E and the angle θ through which it is scattered according to (1)

$$q = CE\sin(\frac{\theta}{2}) \quad (1)$$

where C=1.01354 $Å^{-1}\cdot keV^{-1}$.

2.2. The principle of principal component analysis

The EDXRD profile consists of multi-channels signals, and is a kind of high-dimensional data. The traditional feature extracted for identification is the position of diffraction peak. However, in the case of body packing, some detail message is lost due to the influence of body, and the peak is obscured. Even worse, the EDXRD profile is hard to analyze with regular digital signal processing methods, such as Fourier transform because of their similarities in frequency. Fortunately, PCA is an effective method for feature extraction and especially useful in extracting features in high-dimensional and inter-correlated data and reorganizing variables with least inter-correlation [11-12]. In this paper, the tool used to perform PCA was MATLAB®.

2.3. The principle of artificial neural network

The artificial neural network (ANN) is a form of artificial intelligence that imitates some function of the human brain [16]. It is a powerful computing tool to solve complex non-linear problems. Because of its strong fault tolerance and adaptive learning, it is popular in the field of pattern recognition.

In this work, it was used to classify drugs and the other material based on features extracted by PCA. This neural network was feed-forward with gradient descent with momentum and adaptive learning rate back propagation algorithm. It had 10 nodes (10 PCs) in input layer, 20 nodes in hidden layer, and 7 nodes (7 kinds of material) in output layer. The hidden layer adopted log-sigmoid transfer function, which limited the results between [0, 1] and output layer adopted competitive transfer function, which forces there was a sole "0" in once result. The value of the output at an output neuron represented the degree to which the material belongs to a certain kind: "1" means most likely and "0" most unlikely. This ANN program was written with MATLAB®.

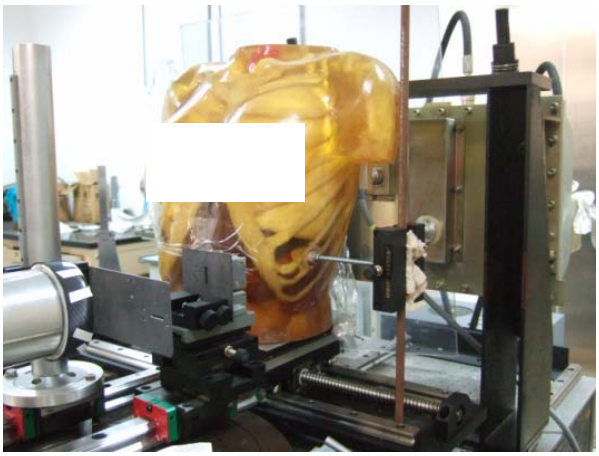


Figure-2. Photo of anthropomorphic phantom and samples.

2.4. Samples

As a methodological study, the anthropomorphic phantom which was made by Chengdu Phantom Emulation Technology Co., Ltd [17] was chosen instead of actual human body. It was designed based on Chengdu dosimetric phantom which was certified by International Commission on Radiation Units and Measurements. The sample filled centrifuges tube of 5 ml, which was fixed in the stomach of anthropomorphic phantom, as shown in Figure-2. The samples included drugs (heroin and methamphetamine), drug precursors (phenylacetic acid and piperonal), explosive (TNT) and disruptor (salt). The detail message of samples is shown in Table-1. For example, No. 4 is heroin, measured for 40 times and all data are divided into two set: 10 times for training and 30 times for test.

Table-1. Sample list.

Material No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Material name	TNT + body	Phenylacetic Acid + body	Methamphetamine + body	Heroin +body	Piperonal + body	Body	Salt + body
Amount of training set	10	10	10	10	10	10	10
Amount of test set	30	30	30	30	30	30	10

3. RESULTS AND DISCUSSIONS

3.1. Preprocessing

The original profiles of heroin and methamphetamine separately packed in anthropomorphic phantom are shown in Figure-3. Peaks of methamphetamine are evident because methamphetamine is monocrystallized, whereas peaks of heroin are obscured because heroin is polycrystalline. Therefore, it is necessary to extract new features by PCA especially in heroin detection.

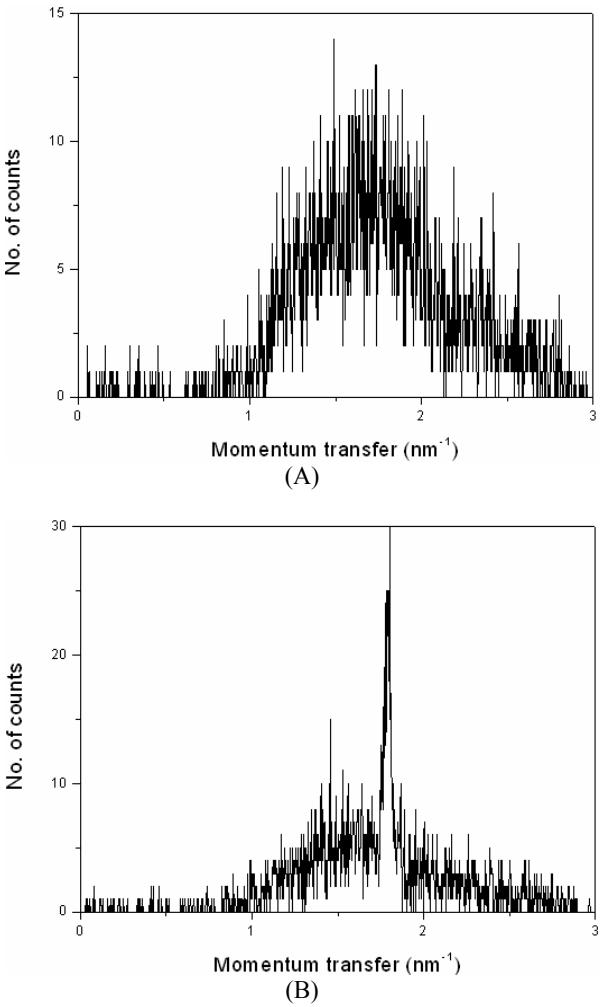


Figure-3. Profiles of heroin (A) and methamphetamine (B) separately packed in anthropomorphic phantom.

Data preprocessing includes truncation and auto scaling. Truncation is retaining data between 1.198 to 2.685 in q in order to eliminate the influence of fluorescence metallic target and reduce the computational complexity. Auto scaling means centering and standardization.

All data in training set after preprocessing were collected in a matrix, then analyzed by PCA. The eigenvalues of PCs are shown in Figure-4(A), and the accumulated significance to data in relation to the number of PCs from the first PC is illustrated in Figure-4(B). As shown in Figure-4(B), first several PCs don't contribute to very large variance, 36.9% with 2 PCs, 41.7% with 3 PCs, and 54.9% with 10 PCs. Different numbers of PCs may lead to different classification result. Therefore, the number of PCs extracted as features is a problem.

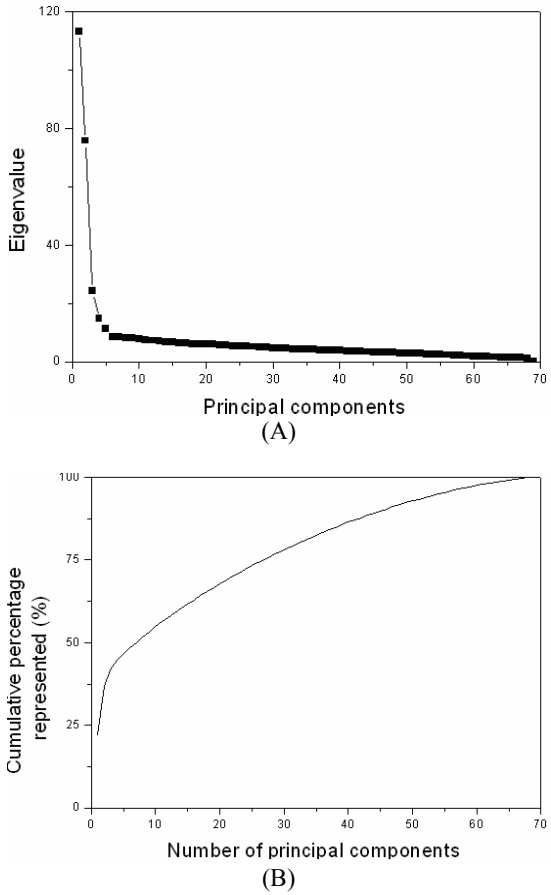
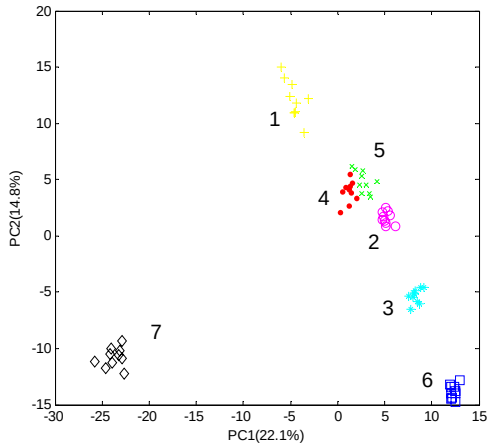


Figure-4. Capabilities of PCs to explain variance: (A) The eigenvalues of PCs. (B) the accumulated significance to data in relation to the number of PCs from the first PC.

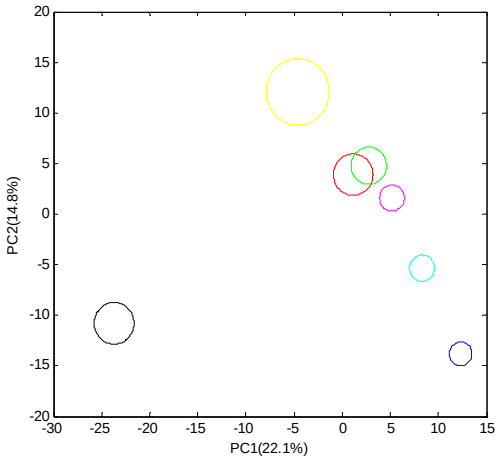
3.2. Identification based on 2 PCs

First, the first two PCs were extracted. The space of principal components built by training set is shown in Figure-5(A). The PC1 possesses 22.1% of the information of the original feature data, and PC2 possesses 14.8%. Based on the positions of the samples, classification circles are drawn in Figure-5(B) and described quantitatively in Table-2 (section of "Training data set"). Obviously, most types of material are classified well except heroin (No.4) and piperonal (No.5). In order to test our method of feature extraction, data of test set are projected to that space of principal components as shown in Figure-5(C). According to the classification circles, test data are classified in Figure-5(D). It is shown that there is false positive between heroin and piperonal, and there is false negative in most types of samples. Further, correct

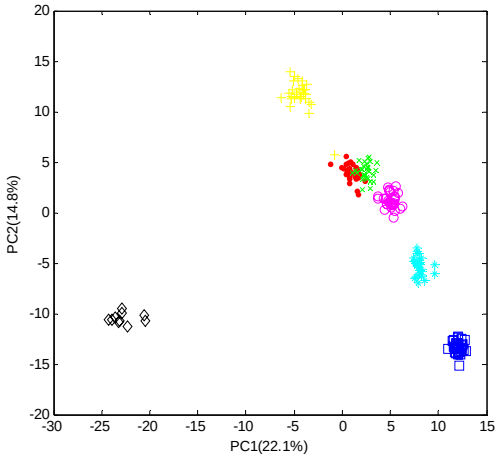
identifications and success rate of identification are listed in Table-2 (section of "Test data set"). The results of most materials are satisfying (80 96.7%), the results of phenyl acetic acid (No.2) and methamphetamine (No.3) are unsatisfying (76.7% and 73.3% separately), and the result of heroin is poor (23.3%).



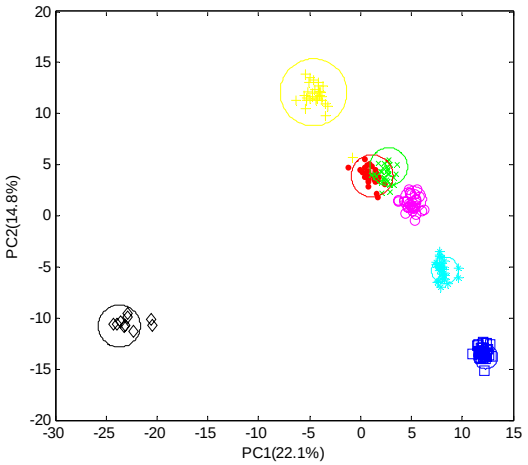
(A)



(B)



(C)

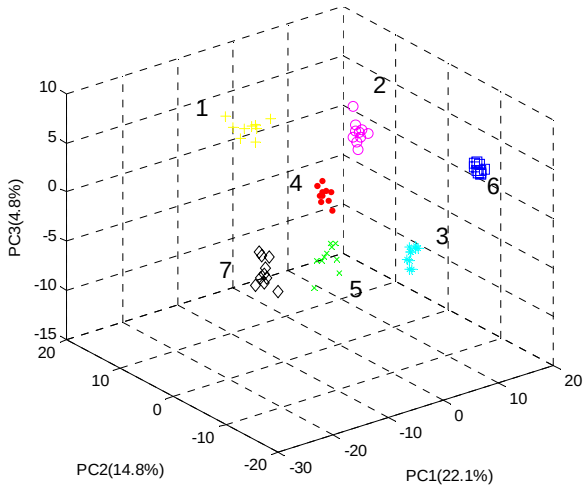


(D)

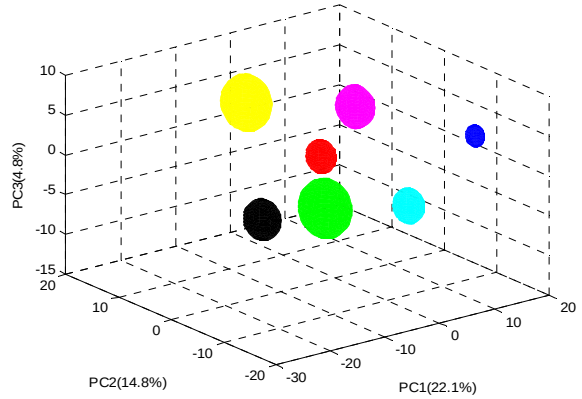
Figure-5. Scores plots of first two PCs: (A) training set. (B) Classification circles. (C) Test set. (D) Total set.

3.3. Identification based on 3 PCs

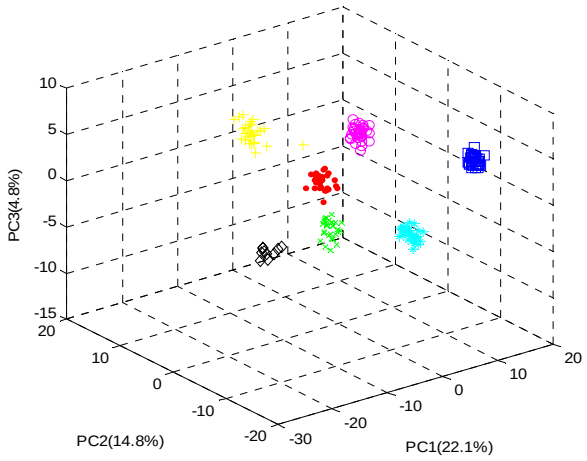
Increasing the number of PCs as recognition features may be a way to raise the success rate because more PCs can explain more information of original data according to Figure-4. Therefore, PC3 (4.8%) was added, and the corresponding result is shown in Figure-6 and Table-3. The Figure-6 shows that the overlap between heroin (No.4) and piperonal (No.5) is separated by PC3. Further, the success rate is raised greatly by eliminating the false positive because there are not overlapping between classification spheres. However, the result of normal body (No.6) is unsatisfying (66.7%) due to false negative.



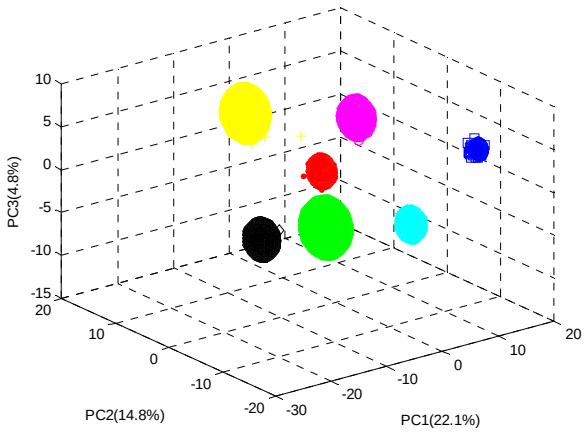
(A)



(B)



(C)



(D)

Figure-6. Scores plots of first three PCs: (A) training set. (B) Classification circles. (C) Test set. (D) Total set.

3.4. Identification based on 10 PCs

In order to achieve satisfying result, 10 PCs, which can explain over half of overall information of original data (54.9%), were extracted. Given to the abstract of super sphere with 10 dimensions for classification, an ANN is adopted for classification. Figure-7 shows the variation of the residual error during a training process. It can reach the goal of training in a hurry (9 epochs). The evaluation result is shown in Table-4 and all of the success rates are between 93.3% and 100%.

Ta**ble-2.** The result of classification and identification based on first two PCs.

Training data set							
Material No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Center coordinate	-4.62 12.09	5.17 1.59	8.29 -5.33	1.15 3.90	2.75 4.80	12.31 -13.81	-23.69 -10.79
Radius	3.28	1.28	1.30	2.05	1.84	1.17	2.06
Standard deviation	0.98	0.30	0.37	0.57	0.47	0.30	0.58
Test data set							
Correct identifications	29	23	22	7	27	27	8
Success rate (%)	96.7	76.7	73.3	23.3	90	90	80

Table-3. The result of classification and identification based on first three PCs.

Training data set							
Material No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Center coordinate	-4.62 12.09 4.36	5.17 1.59 5.13	8.29 -5.33 -5.88	1.15 3.90 -1.07	2.75 4.80 -8.15	12.31 -13.81 4.62	-23.69 -10.79 -0.44
Radius	3.44	2.62	2.15	2.05	3.62	1.30	2.51
Standard deviation	1.01	0.67	0.55	0.63	0.84	0.26	0.61
Test data set							
Correct identifications	27	29	30	27	29	20	8
Success rate (%)	90	96.7	100	90	96.7	66.7	80

Table-4. The result of classification and identification based on first ten PCs.

Material No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Correct identifications	29	29	30	28	28	30	10
Success rate (%)	96.7	96.7	100	93.3	93.3	100	100

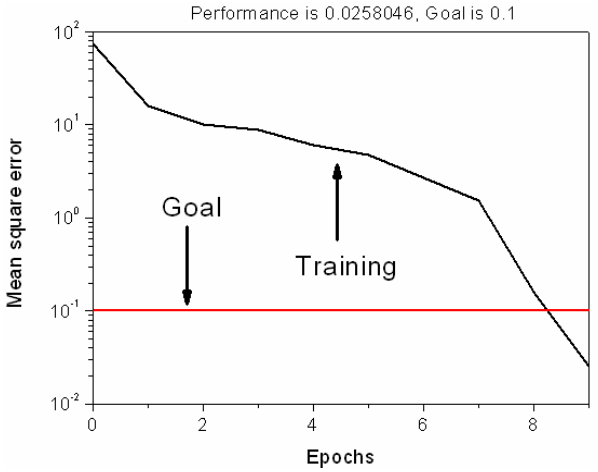


Figure-7. Variation of the residual error during a training process.

Further, extracting more PCs were tried, but the improvement was limited. In consideration of computational complexity, the 10-PCs solution was adopted in the end.

4. CONCLUSIONS

A feature extraction method based on PCA was studied and developed for identification of drugs packed in anthropomorphic phantom. The heroin and methamphetamine were chosen as examples of drugs, and meanwhile some drug precursors, explosive and disruptor were also analyzed.

PCA of the EDXRD data shows a comparative distribution pattern of the spectral object from 7 types of samples. The number of PCs was discussed, and the 10-PCs solution was thought to be a suitable choice. With the help of classification method, such as classification circle, classification sphere, and ANN, a satisfying result of success rate between 93.3% and 100% was achieved. It can be concluded that the PCA is an effective method to extract features of EDXRD spectrum in detecting drugs packed in anthropomorphic phantom and the identification method based on PCA is promising in detecting drugs body packing.

ACKNOWLEDGEMENT

This work was financially supported by the Foundation of Director of the Institute of Intelligent Machines, Chinese Academy of Sciences, and the Natural Science Foundation of Anhui Provincial Education Department (KJ2011A269).

REFERENCES

[1] George R.D., Gauthier R.D., Denslow K.D., Cinson A. M., Diaz A. A. and Griffin M. 2008. Contraband detection using acoustic technology. Proc. of SPIE. 6934: 693415.

[2] Buffler A., Brooks F.D., Allie M.S., Bharuth-Ram K. and Nchodu M. R. 2001. Material classification by fast neutron scattering. Nucl. Instrum. Methods B. 173: 483-502.

[3] Maurer M.H., Niehues S.M., Schnapauff D., Grieser C., Rothe J.H., Waldmuller D., Chopra S.S., Hamm B. and Denecke T. 2011. Low-dose computed tomography to detect body-packing in an animal model. Eur. J. Radiol. 78: 302-306.

[4] Kawase K. 2004. For Drug Detection and Large-Scale Integrated Circuit Inspection. Optics and Photonics News. 10: 34-39.

[5] Dikici E., Deo S.K. and Daunert S. 2003. Drug detection based on the conformational changes of calmodulin and the fluorescence of its enhanced green fluorescent protein fusion partner. Anal. Chim. Acta. 500: 237-245.

[6] Luggar R.D., Farquharson M.J., Horrocks J.A. and Lacey R.J. 1998. Multivariate analysis of statistically poor EDXRD spectra for the detection of concealed explosives. X-Ray Spectrom. 27: 87-94.

[7] Cook E.J., Fong R., Horrocks J.A., Wilinson D. and Speller R.D. 2007. Energy dispersive X-ray diffraction as a means to identify illicit materials: a preliminary optimization study. Appl. Radiat. Isot. 65: 959-967.

[8] Kämpfe B., Luczak F. and Michel B. 2005. Energy dispersive X-ray diffraction. Part. Part. Syst. Charact. 22: 391-396.

[9] Cook E.J., Griffiths J.A., Koutalonis M., Gent C., Pani S., Horrocks J.A., George L., Hardwick S. and Speller R. 2009. Illicit drug detection using energy dispersive X-ray diffraction. Proc. of SPIE. 73100I.

[10] Cook E.J., Pani S., George L., Hardwick S., Horrocks J.A. and Speller R.D. 2009. Multivariate data analysis for drug identification using energy-dispersive x-ray diffraction. IEEE Trans. Nucl. Sci. 56: 1459-1464.

[11] Marten H. and Naes T. 1989. Multivariate Calibration. Wiley, New York.

[12] Wold H. 1966. Estimation of principal components and related models by iterative least squares. Krishnaiah PR (ed). Academic Press: New York. pp. 391-420.

[13] Li W., Zhang F., Yu D., Sun B., Li M. and Liu J. 2011. Impact of fat and muscle in energy dispersive X-ray diffraction-based identification of heroin using multivariate data analysis. J. Chemometr. 25: 631-635.

[14] Li W., Yu D., Zhang F., Sun B., Liu J., Li M. and Liu J. 2011. Detection of heroin covered by skin by using robust principal components analysis. Measurement. 44: 267-273.

[15] Johns P.C. and Yaffe M.J. 1983. Coherent scatter in diagnostic radiology. Med. Phys. 10: 40-50.

[16] Bishop C.M. 1995. Neural networks for pattern recognition. Oxford: Clarendon Press.

[17] Cai M., Lin D., Guo Z., Wang Y. and Wang B. 2005. China radiation anthropomorphic phantom and mathematical models of human body organs (in Chinese). Chin. J. Med. Imaging Technol. 21: 1464-146.