Measurement 44 (2011) 267–273

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Detection of heroin covered by skin by using robust principal components analysis

Wei Li [a,b], Daoyang Yu [a], Fang Zhang [a,c], Bai Sun [a], Jinyun Liu [a], Minqiang Li [a,*], Jinhuai Liu [a,*]

[a] Key Laboratory of Biomimetric Sensing and Advanced Robot Technology, Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, PR China
[b] Department of Automation, University of Science and Technology of China, Hefei, PR China
[c] Department of Chemistry, University of Science and Technology of China, Hefei, PR China

ARTICLE INFO

Article history:
Received 17 July 2010
Received in revised form 30 September 2010
Accepted 7 October 2010
Available online 12 October 2010

Keywords:
Energy dispersive X-ray diffraction
Robust principal component analysis
Detection of heroin
Drug body packing

ABSTRACT

Nowadays, body packing has become an increasingly important approach for trafficking drugs. In this paper, robust principal component analysis (PCA) was demonstrated as a promising suitable method to detect heroin covered by skin based on the profile of energy dispersive X-ray diffraction spectrum. Firstly, some potential impact factors were investigated. Results showed that the signal was optimized when the scatter angle of 8° and the count time of 1 min were chosen. Then, after being truncated, denoised and autoscaled, spectra of heroin, skin and heroin covered by skin were analyzed by robust PCA. Compared with classical counterpart, robust PCA effectively extracted principle components of samples, and successfully identified heroin covered by skin. Finally, based on Robust PCA, a method of identification was set up. This suggests that robust PCA can be employed in detecting drugs packed in body.

1. Introduction

In recent years, body packing, smuggling drug containers in gastrointestinal and vaginal tract, has become a main approach for trafficking drugs. Thus, in order to crack down on this criminal activity, developing effective methods and instruments to identify body packing has been widely focused. Energy dispersive X-ray diffraction (EDXRD) has been shown to be a suitable technique to detect explosives and drugs in baggage [1–5].

In EDXRD spectrum, the most frequently extracted feature is the position of diffraction peak, which is dependent on the inter-molecular distances of the material tested according to Braggs law. However, in the case of packing, the signal is poor, so it is difficult to identify diffraction peaks in EDXRD spectrum. Besides, 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 [1,2].

The diffraction profile, due to its characters of high-dimensional data, is hard to cope with. Therefore, principal component analysis (PCA) is often used to reduce the dimension by creating a new set of uncorrelated variables [6–9]. However, the sample covariance matrix produced by classical PCA is highly sensitive to outlying observations, which leads to an incorrect description. Unfortunately, the skin will not only attenuate X-ray, but also produce diffraction effects that will overlie features of drugs during identification of drugs packed in body. It is an important source of anomalous observations. Therefore, it is necessary to apply the robust method which is usually employed in processing anomalous observations to reduce the impact of the skin.

In this paper, we investigated the feasibility of applying robust PCA to identify heroin covered by skin using EDXRD. Heroin was chosen as the representation of drugs, and muscle and fat of pigs were chosen instead of actual skin.

* Corresponding authors. Tel.: +86 0551 5591167; fax: +86 0551 5592420.
E-mail addresses: mqli@iim.ac.cn (M. Li), jhliu@iim.ac.cn (J. Liu).

0263-2241/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.measurement.2010.10.003

First of all, we optimized experimental parameters, such as scatter angle and count time. Secondly, EDXRD spectra of heroin, skin and heroin covered by skin were acquired. After being truncated, denoised and autoscaled, data were analyzed by classical PCA and robust PCA. Finally, an identification method based on robust PCA was set up.

2. The experimental system and methods

2.1. The experimental system

The atoms in crystalline materials are arranged in such a way that regularly spaced atomic planes can be identified. The relationship between the X-ray wavelength λ, the scatter angle θ and an atomic planar spacing (lattice spacing) d of the investigated material is given by the Bragg's law:

$$2d\sin\frac{\theta}{2}=n\lambda$$

In the case of EDXRD, it is more convenient to write the equation above in terms of the X-ray energy rather than the wavelength [10]:

$$2d\sin\frac{\theta}{2}=\frac{nhc}{E}$$

where h is Planck's constant, c is velocity of light, λ is replaced by the photon energy E. A primary X-ray photon with high energy is scattered under a small scatter angle θ for a given lattice spacing d. Under fixed scatter angle, different lattice spacings characterizing the crystalline structure of the material result in different diffraction peaks within the energy spectrum of the scattered X-ray photons.

Fig. 1 shows the EDXRD spectrometer of our experiment, which consists of a polychromatic source of X-ray and an energy-resolving detector. The polychromatic source of X-ray comprises an X-ray tube and the related collimators to define the incident and scattered beams. The X-ray source is W target in the range 0–40 kV and 0–30 mA. A solid-state liquid-nitrogen-cooled high purity germanium (HP Ge) detector with sufficient energy resolution is adopted. The Canberra planar HP Ge detector with 50 mm^2 detector area and 5 mm thick crystal has a resolution of 140 eV at 5.9 keV combined with a Canberra spectrum master InSpector 2000. A complete description of system was shown in our previous work [11].

2.2. Multivariate analysis of diffraction profiles

In order to analyze EDXRD spectrum, a high-dimensional data, PCA is introduced here. PCA is a popular statistical method which tries to explain the covariance structure of data by means of a small number of components. Unfortunately, both the covariance and the correlation matrix, which are widely used as estimators, are very sensitive to anomalous observations. In order to obtain principal components that are not influenced much by outliers, new robust estimators via projection pursuit (PP) technique were introduced [6]. This method maximizes a robust measure of spread to obtain consecutive directions on which the data points are projected.

PCA is known as an effective method of feature extraction. Let $X\sim F$ be a p-dimensional random vector with covariance matrix V, and let F^a be the distribution function (df) of $a'X$, where a is a p vector. Denote the eigenvalues of V by $r_1,\ldots, r_p$. The first principal component is the projection of X onto a certain direction:

$$\sigma(F^{a_1})=\max_{\|a\|=1}\sigma(F^a)=\max_{\|a\|=1}(a'Va)^{1/2}.$$

As we know, $\sigma^2(F^{a_1})=a_1'Va_1$ is the largest eigenvalue r_1 and that a_1 is the associated eigenvector. Similarly, the kth principal component $a_k'X$ is defined by

$$\sigma(F^{a_k})=\max_{\|a\|=1,a\perp a_1,\ldots a\perp a_{k-1}}\sigma(F^a).$$

Thus one concludes that $\sigma^k(F^{a_k})=r_k$, that a_k is the eigenvector associated with r_k. Here classical PCA searching for low-dimensional projection index. On the other hand, if the principal components of V are known, V can be constructed by $V=\sum r_i a_i a'_i$.

As stated, the lack of robustness is the main drawback of classical PCA. From the point of view of PP technique, the reason is that the projection index-standard deviation itself is nonrobust. Therefore, the PP procedure can provide robust principal components and a robust dispersion matrix by using a robust scale as projection index.

Suppose that X∼F is a p vector and that its location is known in advance, and without loss of generality it is fixed at 0. Let S(·) be a robust scale, which is usually weakly continuous. The robust principal components of F (or X), denoted by Sk(F) and Ak(F), are determined by

$$S_1(F)=\max_{\|a\|=1}S(F^a)'$$

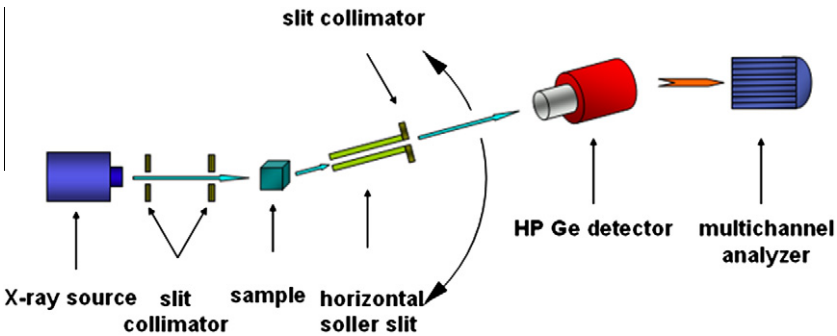

Fig. 1. The geometry of the scattering system.

$A_1(F) = a_1$,

where $\|a\| = 1$, $S(F^{a_1}) = S_1(F)$, and

$S_k(F) = S(F^{a_k})$,

where $\|a\| = 1$, $a_k \perp a_1, \ldots, a_{k-1}$,

$A_k(F) = a_k$.

The robust covariance matrix, denoted by $V(F)$, is defined by $V(F) = \sum S_i(F)^2 A_i(F) A_i(F)'$.

The robust correlation matrix $R(F)$ is obtained by rescaling the covariance matrix $V(F)$.

For p-dimensional observations $x_1, \ldots, x_n$, let F_n be the empirical distribution. Then the estimators for principal components and covariance and correlation matrices are defined by

$S_i(x_1, \ldots, x_n) = S_i(F_n)$,

$A_i(x_1, \ldots, x_n) = A_i(F_n)$,

$V(x_1, \ldots, x_n) = V(F_n)$,

$R(x_1, \ldots, x_n) = R(F_n)$.

This method has been improved by several researchers [7–9], and formed a standard algorithm, reflection-based algorithm for principal components analysis (RAPCA).

In this paper, the tool used to perform classical PCA and RAPCA is the MATLAB® library for Robust Analysis (LIBRA) [12]. It is a MATLAB® library of robust statistical methods which were developed because classical alternatives produce unreliable results when the data set contains outlying observations.

3. The optimization of experimental parameters

3.1. The selection of scatter angle

The signals acquired from EDXRD spectrometer are influenced by many conditions. In order to optimal the signals, some experimental parameters, such as scatter angles and count time, were investigated here.

The optimum scatter angle was selected by measuring the spectra of all samples using the 7°, 8°, 9°and 10° diffraction cells. The spectra obtained for heroin, skin and heroin covered by skin with count time of 5 min are shown in Fig. 2.

The x-coordinate represents 1024 channels of multichannel analyzer, equivalent in energy spectrum of 0–40 keV; the y-coordinate represents the number of photons in each channel, equivalent in energy of photon. Peaks near 250 in x-coordinate are the result of fluorescence metallic target, while peaks after 300 in x-coordinate is the result of the material tested. It is clear that the scatter angle strongly affects the diffraction profile as shown in Fig. 2.

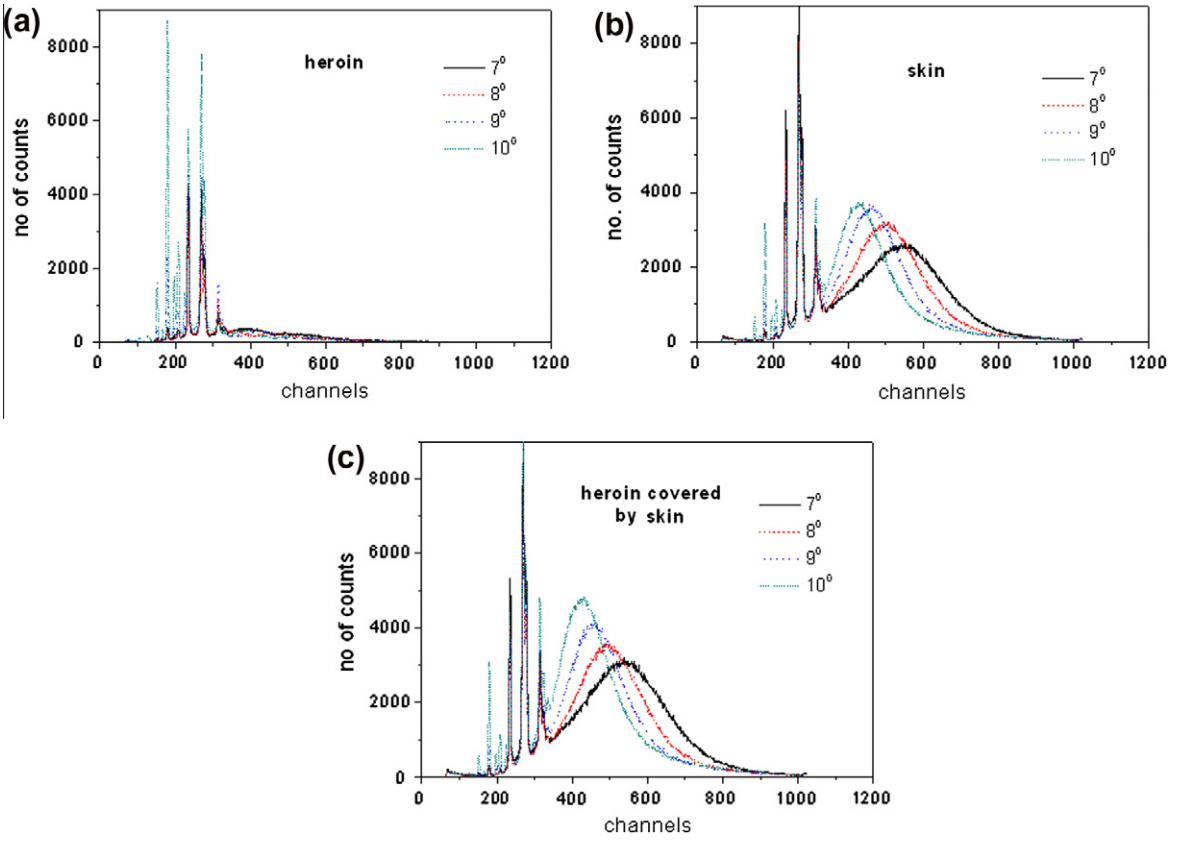

Fig. 2. Diffraction profiles of heroin (a), skin (b) and heroin covered by skin (c) at scatter angles of 7°, 8°, 9°and 10°.

First, the intensity of profile increases with the scatter angle. Second, the position of peak of the material tested (heroin, skin and heroin covered by skin) varies with the scatter angle. In the contrast, the position of peak of fluorescence does not vary with the scatter angle. Third, the peak of tested material is closer to the peak of fluorescence when the scatter angle increases. On one hand, the intensity of profile should be strong enough to get a good signal–noise-ratio, so we should get a bigger scatter angle; on the other hand, the peak of tested materials should be far away from the peak of fluorescence so that we can separate them, so we should get a smaller scatter angle. Therefore, a moderate angle of 8° was chosen.

3.2. The selection of count time

As the eventual application of any system towards to human, in order to minimize the harm to human and enhance efficiency, a low photon count is required. All samples were measured with a range of count times (1 min, 2 min, 3 min, 4 min and 5 min). Spectra of heroin covered by skin at scatter angle of 8° taken with different integrated photon counts are shown in Fig. 3.

From Fig. 3, we can see that the intensity of profile is nicely direct proportional to the count time. The integrated counts detected in 5 min were 1×106 with only 0.2×106 detected in 1 min. However, the broad shape of the spectrum is still available with count time as low as 1 min. In order to minimize the photon count, we chose the count time of 1 min.

4. Multivariate analysis of diffraction profile

4.1. Data preprocessing: truncation and denoising

As mentioned above, peaks near 250 in *x*-coordinate are useless, while ones after 300 are useful. In order to focus on the key feature of tested materials and reduce the computational complexity, we selected data between 351 and 862 in *x*-coordinate.

Due to attenuating materials, the spectrum is poor, and many information of tested material are suppressed by noise. In this paper, we used Stein's Unbiased Risk Estimate (SURE) denoising technique [13], one of wavelet denoising methods. SURE method is a hard thresholding approach where the major work is finding the right threshold for the different scales. It was performed by the One-Dimensional Stationary Wavelet Trasform (SWT) package of MATLAB®. Parameters were listed as follows: wavelet: db; decompose and compose level: 5; SURE thresholding: 3.623, 4.214 and 5.012. Fig. 4 shows original signal (a) and denoised signal (b) of heroin covered by skin at the scatter angle of 8° with count time of 1 min. It is obvious that SURE denoising is effective.

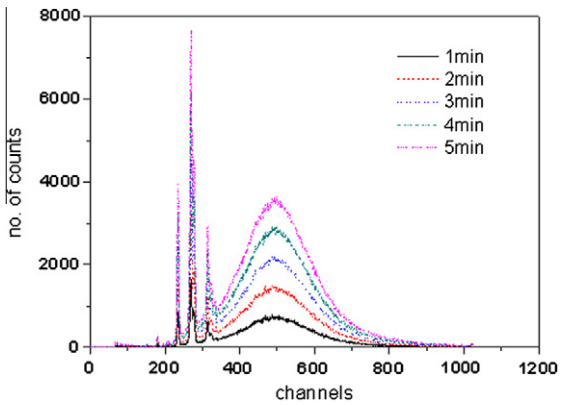

Fig. 3. The diffraction profiles of heroin covered by skin at scatter angle of 8° with count times of 1 min, 2 min, 3 min, 4 min and 5 min.

4.2. Classical PCA and robust PCA of diffraction profile

Diffraction profiles of heroin, skin and heroin covered by skin were measured four times under conditions of the scatter angle of 8° and the count time of 1 min. After being truncated and denoised, all data formed a matrix with 12 rows and 512 columns. The matrix was first processed by autoscaling and then analyzed by LIBRA. Fig. 5 is the scree plot of eigenvalue, and Fig. 6 is scree plot of the percent variability explained by each principal component. In Figs. 5 and 6, the former one in each figure was analyzed by classical PCA, while the latter by robust PCA.

In Fig. 5, it can be observed that eigenvalues calculated by robust PCA decrease more nicely than ones by classical PCA. The number of principal components by classical PCA is 11, while robust PCA is 6. However, only components of high capability of explaining variability are effective. From Fig. 6, the number of effective principal components is separately one in classical PCA, and 2 in robust PCA. In chemometrics, principal component is usually connected with the composition of mixture. In our study, there are two main compositions, heroin and skin, which means there should be at least two effective principal components. Therefore, robust PCA is capable of this job, while classical PCA not. Here, robust method successfully reduces the impact of outliers.

A score plot of PC1 against PC2 by robust PCA is shown in Fig 7. Different samples are separately clustered well. Samples of skin are located near the origin. Samples of heroin have large negative scores on PC1 and PC2, while samples of heroin covered by skin have positive value. In order to discuss the clustering characteristics in quantification, 97.5% tolerance ellipses were described. Fig 8 shows the 97.5% tolerance ellipses of heroin (a), skin (b), and heroin covered by skin (c). Furthermore, Euclidean distances between centers of different ellipses were calculated in Table 1. The distances were large enough to distinguish different samples. It is indicated that the heroin covered by skin could be detected by this method.

5. The method of heroin identification

Classification is the basis of identification, so linear discriminant classification is introduced here. The centers of

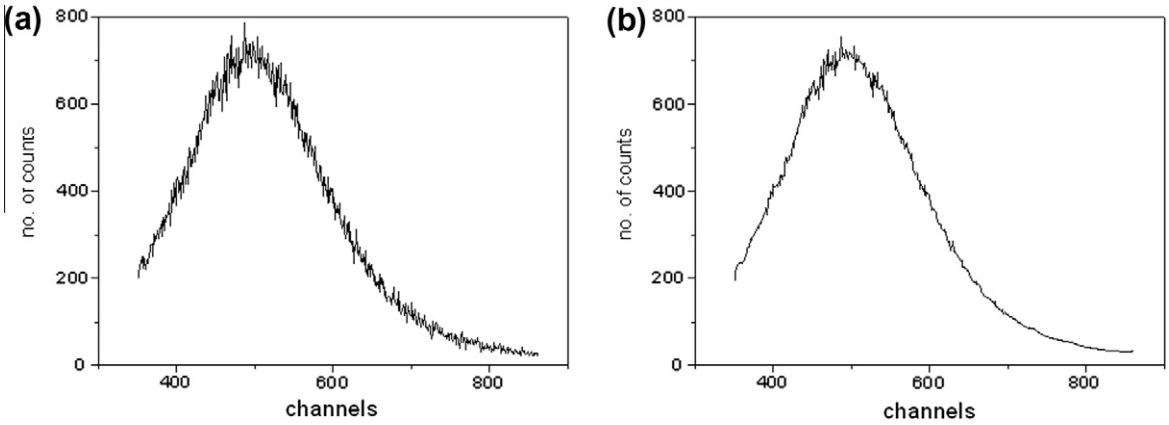

Fig. 4. Original signal (a) and denoised signal (b) of heroin covered by skin.

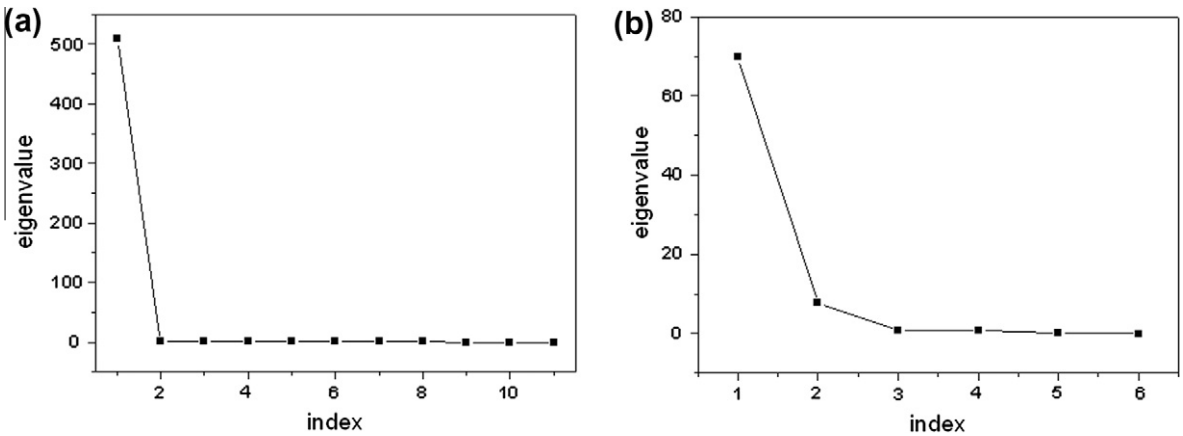

Fig. 5. Scree plots of eigenvalue by classical PCA (a) and robust PCA (b).

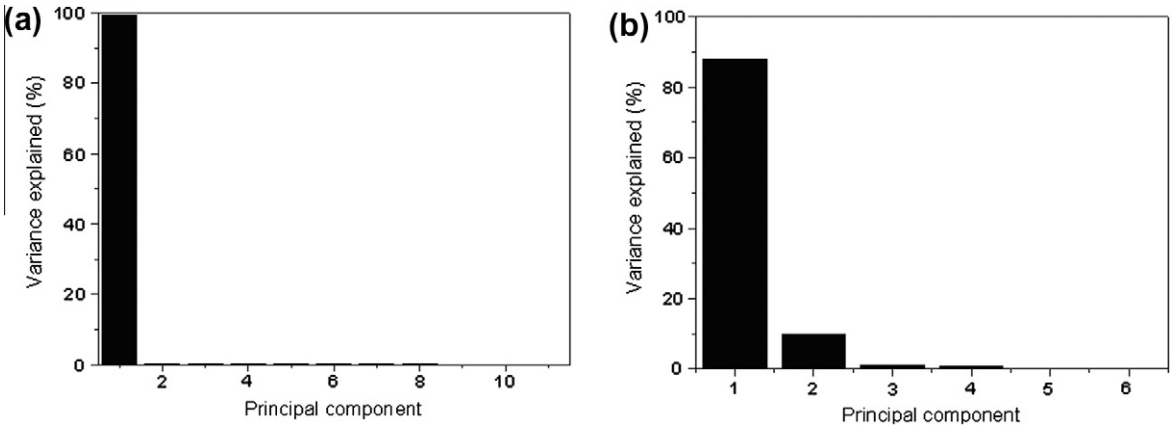

Fig. 6. Scree plots of the percent variability explained by each principal component by classical PCA (a) and robust PCA (b).

three 97.5% tolerance ellipses are seen as a set of coplanar points, and for each of two nearby points, Voronoi polygon can be drawn. All Voronoi polygons between the set of points consist of the Voronoi diagram, as shown in Fig 7. The two lines divide the plane into three parts, meaning three classes, heroin, skin, and heroin covered by skin.

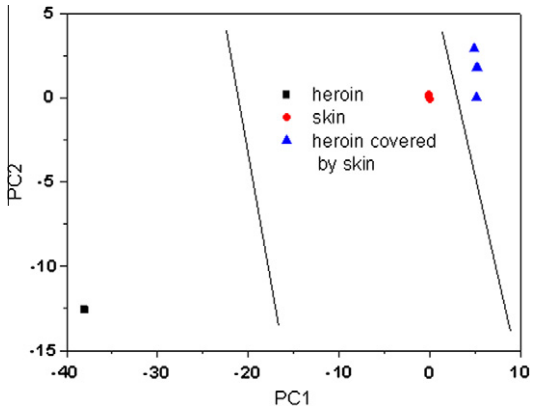

Fig. 7. A score plot of PC1 against PC2 by robust PCA.

So far, the method of heroin identification has been set up. Firstly, different training samples are added in the sample library. Secondly, the principal components of training samples are extracted by robust PCA, and the centers of 97.5% tolerance ellipses are calculated, then samples are classified in the space of principal components by Voronoi diagram, which is the model of identification. Thirdly, tested samples are analyzed by the model of identification, and the location in the space of principal components defines the classification.

Table 1
Euclidean distances between centers of different classes.

	Skin	Heroin covered by skin
Heroin	40.03	45.59
Skin		5.58

6. Conclusion

In summary, the feasibility of applying robust PCA to identify heroin covered by skin has been demonstrated. In order to optimize resolution of signal, we chose the scatter angle of 8°, and in order to minimize the harm to human and enhance efficiency, we chose the count time of 1 min. Robust PCA effectively extracted principle components of samples, and successfully separated heroin covered by skin from heroin and skin. Then, an identification method based on robust PCA was set up. Our findings have shown a new opportunity for developing effective detecting approach towards drugs packed in body.

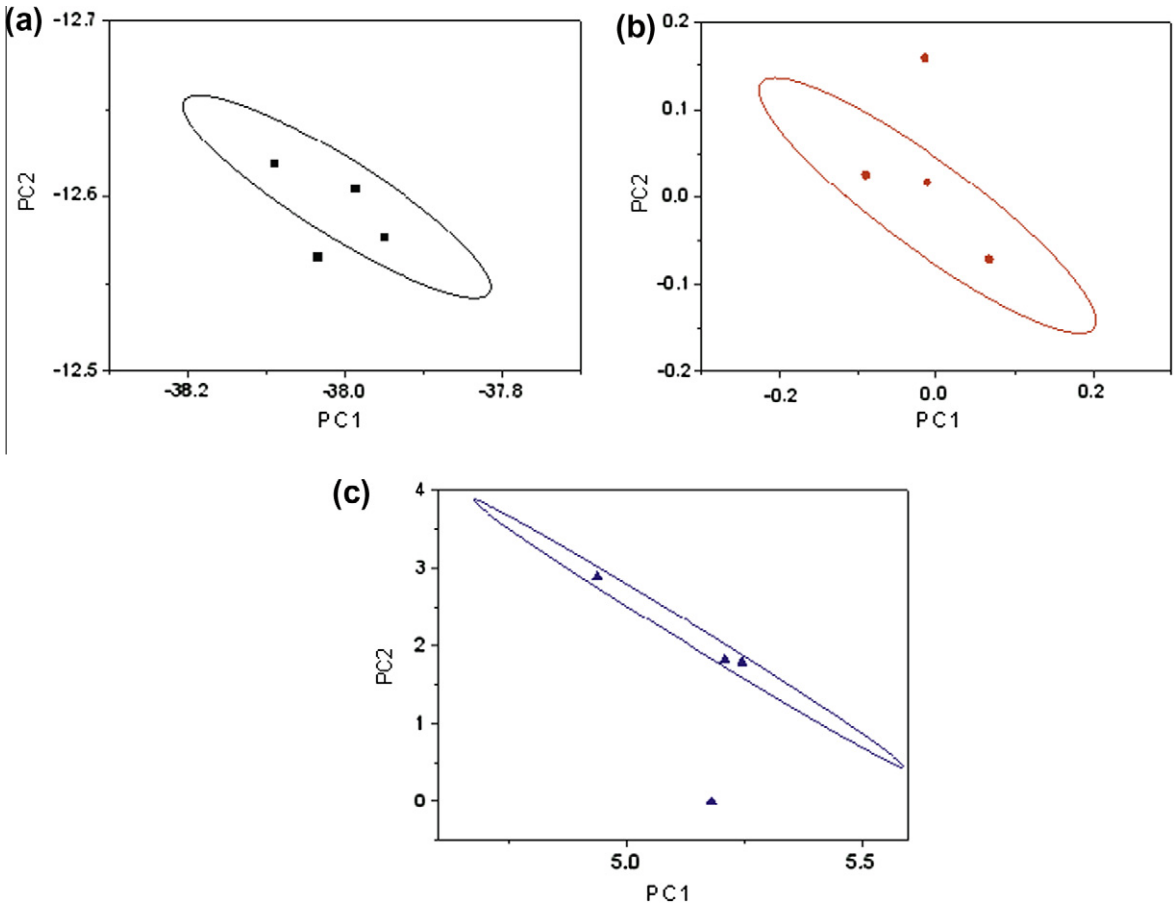

Fig. 8. The 97.5% tolerance ellipses of heroin (a), skin (b), and heroin covered by skin (c).

Acknowledgements

This work was financially supported by the Key Program of National Natural Science Foundation of China (Grant 10635070).

References

[1] R.D. Luggar, M.J. Farquharson, J.A. Horrocks, et al., Multivariate analysis of statistically poor EDXRD spectra for the detection of concealed explosives, X-ray Spectrometry 27 (1998) 87–94.
[2] E.J. Cook, R. Fong, J.A. Horrocks, et al., Energy dispersive X-ray diffraction as a means to identify illicit materials: a preliminary optimization study, Applied Radiation and Isotopes 65 (2007) 959–967.
[3] B. Kämpfe, F. Luczak, B. Michel, Energy dispersive x-ray diffraction, Particle & Particle Systems Characterization 22 (2005) 391–396.
[4] E.J. Cook, J.A. Griffiths, M. Koutalonis, et al., Illicit drug detection using energy dispersive X-ray diffraction. Non-intrusive inspection technologies II, Proceeding of SPIE 7310 (2009) 73100I.
[5] E.J. Cook, S. Pani, L. George, et al., Multivariate data analysis for drug identification using energy-dispersive x-ray diffraction, IEEE Transactions on Nuclear Science 56 (2009) 1459–1464.
[6] G. Li, Z. Chen, Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo, Journal of the American Statistical Association 80 (1985) 759–766.
[7] C. Croux, A. Ruiz-Gazen, A fast algorithm for robust principal components based on projection pursuit. In: Prat, A. (Ed.), COMPSTAT: Proceedings in Computational Statistics, 1996, pp. 211–216.
[8] C. Croux, A. Ruiz-Gazen, High breakdown estimators for principal components: the projection-pursuit approach revisited, Journal of Multivariate Analysis 95 (2005) 206–226.
[9] M. Hubert, P.J. Rousseeuw, S. Verboven, A fast method for robust principal components with applications to chemometrics, Chemometrics and Intelligent Laboratory System 60 (2002) 101–111.
[10] P.C. Johns, M.J. Yaffe, Coherent scatter in diagnostic radiology, Medical Physics 10 (1983) 40–50.
[11] B. Sun, M. Li, F. Zhang, et al., The performance of a fast testing system for illicit materials detection based on energy-dispersive X-ray diffraction technique, Microchemical Journal 95 (2010) 293–297.
[12] S. Verboven, M. Hubert, LIBRA: a Matlab library for robust analysis, Chemometrics and Intelligent Laboratory System 75 (2005) 127–136.
[13] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association 90 (1995) 1200–1224.