

## CLASSIFICATION AND RECOGNITION OF TOMB INFORMATION IN HYPERSPECTRAL IMAGE

Mingyan Gu<sup>1</sup>, Shuqiang Lyu<sup>1,2\*</sup>, Miaole Hou<sup>1,2\*</sup>, Sheng Ma<sup>3</sup>, Zhenhua Gao<sup>3</sup>, Shuzhang Bai<sup>3</sup>, Pingping Zhou<sup>4</sup>

<sup>1</sup> School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, No.15 Yongyuan Road, Daxing District, Beijing, 102616-(guminyan, lvshuqiang, houmiaole)@bucea.edu.cn

<sup>2</sup> Beijing Key Laboratory For Architectural Heritage Fine Reconstruction & Health Monitoring, No.15Yongyuan Road, Daxing District, Beijing, 102616

<sup>3</sup> Shanxi Provincial Institute of Archeology, Taiyuan, Shanxi Province, 030001

<sup>4</sup> China university of mining and technology, No. 11,Xueyuan Road D, Haidian District, Beijing, 100083

Commission III, WG III/4

**KEY WORDS:** Hyperspectral imaging technology, PCA, Feature bands selection, SVM, Identification of matter

### ABSTRACT:

There are a large number of materials with important historical information in ancient tombs. However, in many cases, these substances could become obscure and indistinguishable by human naked eye or true colour camera. In order to classify and identify materials in ancient tomb effectively, this paper applied hyperspectral imaging technology to archaeological research of ancient tomb in Shanxi province. Firstly, the feature bands including the main information at the bottom of the ancient tomb are selected by the Principal Component Analysis (PCA) transformation to realize the data dimension. Then, the image classification was performed using Support Vector Machine (SVM) based on feature bands. Finally, the material at the bottom of ancient tomb is identified by spectral analysis and spectral matching. The results show that SVM based on feature bands can not only ensure the classification accuracy, but also shorten the data processing time and improve the classification efficiency. In the material identification, it is found that the same matter identified in the visible light is actually two different substances. This research result provides a new reference and research idea for archaeological work.

### 1. INTRODUCTION

China is one of the four ancient civilizations with 5,000 years of bright cultural history. It has rich cultural heritage. Among them, there are also many historical and cultural sites buried in ancient tombs. The information left in the ancient tomb contains rich historical values. It not only reflects the life of the tomb owner, but also reflects the historical and cultural background at that time. However, due to the age of ancient tomb, the complex environment of the archaeological site, and the influence of the natural environment, mixed impurities (such as soil) cause these relics appear faded, damaged and blurred in outline, making it difficult for the naked eye to distinguish and discriminate. Therefore, the use of hyperspectral techniques to analyze the information at the bottom of ancient tomb is of great significance for archaeological research of ancient tomb. In recent years, techniques such as spectroscopy, chemistry, and digital imaging have been applied to the research field of cultural heritage protection. Because of its non-contact, non-destructive measurements, its wide spectral range, high spectral resolution, and its unique "image and spectrum," hyperspectral techniques are now widely used in the study of cultural heritage conservation. Hou summarized the application and deficiency of hyperspectral imaging technology in the analysis of painted cultural objects (Hou et al, 2017). SeonJoo Kim used hyperspectral data to achieve the visual enhancement of ancient calligraphy by reconstructing gradient images (SeonJoo Kim et al, 2011). Wu Fengqiang used hyperspectral and Raman spectroscopic analysis methods to analyze the pigment

composition of ancient paintings to identify the pigment components of the painted art paintings (Wu Fengqiang et al., 2014). Maria Kubik used hyperspectral techniques to detect the degradation of pigments on the paintings and digitally preserved the paintings (Maria Kubik et al., 2007). Costas Balas used hyperspectral image data to identify painting materials in combination with image information and spectral information (Costas Balas et al., 2003). Guo Xinlei used hyperspectral imaging technology and spectral analysis technology to effectively extract the hidden information of ancient paintings and provided help for the restoration and identification of cultural relics (Guo Xinlei et al., 2017). Tan Kelong and Zhou Xiaohu used aerospace hyperspectral remote sensing to carry out archaeological research on the ruins of the Qin Shihuang Mausoleum, which provided new ideas for archaeological work (Tan Kelong, 2005; Zhou Xiaohu, 2007). Liang described the application of multispectral and hyperspectral imaging techniques in archeology and artistic protection (Liang, 2012). In general, more and more studies on the protection of cultural relics using hyperspectral imaging technology have been conducted at home and abroad in recent years, and the application of hyperspectral imaging technology in archeology also has broad prospects.

The purpose of this paper is to apply hyperspectral imaging techniques to ancient tombs to achieve image classification and material identification. In this study, SVM classification based on feature bands is used to extract the category information of the bottom of ancient tomb, and then through the spectral

\* Corresponding author:Shuqiang Lyu, E-mail: lvshuqiang@bucea.edu.cn

\* Corresponding author: Miaole Hou, E-mail: houmiaole@bucea.edu.cn

matching, the material identification at the bottom of ancient tomb is completed. The paper is organized as follows: In section 2, the research area and data acquisition are introduced. In section 3, the method and process of the experiment are discussed. Finally, the results of the experiment are analyzed in Section 4.

## 2. RESEARCH DATE AND AREA

### 2.1 Research date

In this paper, hyperspectral image of the bottom of ancient tomb is obtained using the VNIR400H hyperspectral imaging camera. The relevant parameters of the instrument are shown in Table 1. The data was obtained around 9 AM and the light source was provided by Halogen lamps. Figure 2(b) is a true-color image of the study area acquired by a hyperspectral imaging camera.

Parameter name	Data parameters
Wavelength bands	400-1000nm
Bands Number	1040
Spectral resolution	2.6nm
Spectral width	0.6nm

Table 1. Related parameters of VNIR400H

### 2.2 Research area

In this paper an ancient tomb in southern Shanxi Province, China is chosen as a research object. According to the archaeological experts' preliminary judgment, this tomb is a late Shang Dynasty tomb, about 3000 years ago. Figure 2(a) is the ortho-image of this tomb obtained by a digital camera. According to the site survey, there were black and red materials at the bottom of the tomb. Among them, there were black traces of suspected letters near the pits. But these traces appear to fall off and there are many miscellaneous substances in the surroundings that are hardly visible to the naked eye. Therefore, select this area as the research area (red box area in Figure 2(a)).

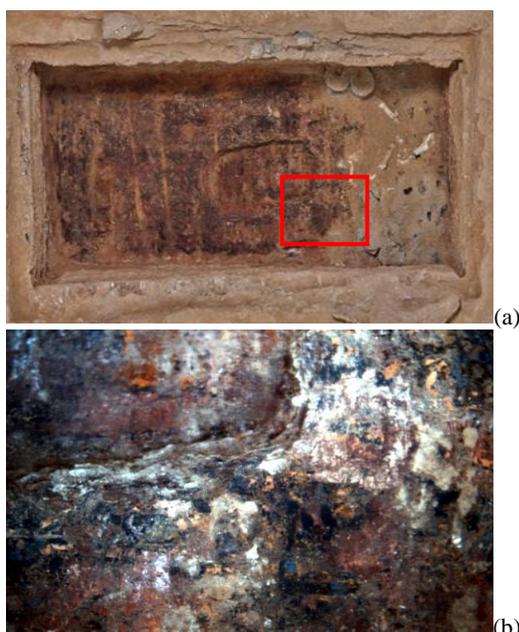


Figure 2. Research area: (a) the ortho-image of the tomb; and (b) the hyperspectral image of research area

## 3. METHOD AND PROCESS

In order to effectively classify images and identify the bottom of ancient tomb, the main method steps in this paper are as follows: (1) First, the hyperspectral images acquired are preprocessed, including radiation correction and band clipping; (2) Using PCA is used to concentrate the main information of the image to the first few principal components, and to reduce the dimension of the hyperspectral data; (3) The feature vectors are selected based on the feature vectors of the PCA transform, and a new image is synthesized using these feature bands; (4) The bottom of ancient tomb was classified using SVM based on feature bands; (5) Finally, the material at the bottom of ancient tomb is identified by spectral matching. The specific process of the experiment is shown in Figure 3.

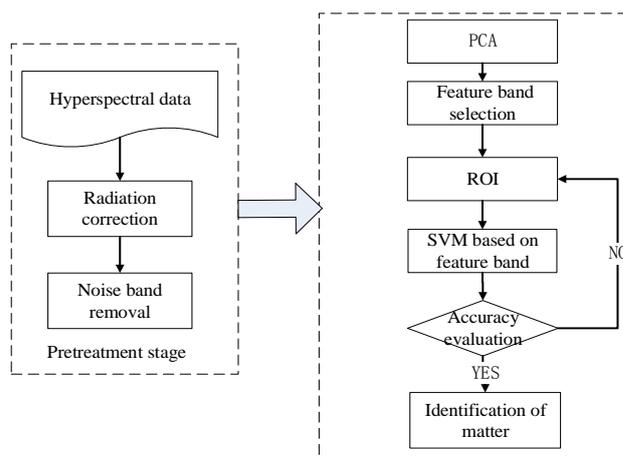


Figure 3. The workflow of the method

### 3.1 Data preprocessing

Because the data collected by the hyperspectral imager is not only the radiation information of the ground object itself, but also the influence of the instrument noise, the light illumination and the instrument observation angle. Therefore, it is can't use the DN value of the original data for analysis. It is necessary to perform radiation correction on the original hyperspectral image data. The correction formula is given as follows:

$$Ref = \frac{Date - Dark}{White - Dark} \quad (1)$$

where  $Ref$  = Reflectance data  
 $Date$  = Really collected data  
 $Dark$  = Dark current data  
 $White$  = Whiteboard data

The whiteboard data is obtained by scanning a standard whiteboard with a reflectance of 99%. The dark current data is obtained by closing the lens cover. Although the original data has undergone radiation correction, there is still a certain noise band. Therefore, the corrected data also needs to remove the noise band. This paper combines the visual interpretation method to eliminate the first 60 and the last 30 noise bands. The preprocessing of the hyperspectral raw data not only removes the noise band of the image and improves the signal-to-noise ratio, but also converts the gray value data of the image into reflectance data.

### 3.2 PCA and feature bands selection

Due to the large amount of data in hyperspectral images and the high correlation between adjacent bands, the image data overlaps to some extent. PCA can remove the correlation between the bands of the hyperspectral image. It is a linear transformation that uses the statistical characteristics of the hyperspectral bands to focus the image information on a few bands without losing important information. The algorithm can be expressed as a transformation matrix:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \quad (2)$$

where  $[X_1 \cdots X_n]^T$  is matrix before linear transformation  
 $[y_1 \cdots y_n]^T$  is matrix after linear transformation  
 $[a_1 \cdots a_n]$  is transformation matrix

According to the principle of PCA, each PCA-transformed principal component is a linear combination of each band in the original data, and the effect of each band on this principal component can be represented by the weight coefficient in the feature vector. Therefore, the weight coefficient graph of each principal component can be drawn from the weight coefficient in the PCA-transformed feature vector. The extreme points at each point in the curve represent a significant band with an important contribution to the principal component (Hou et al., 2014).

### 3.3 Support Vector Machine

Hyperspectral data has the advantages of multiple bands, high spectral resolution, and the ability to accurately represent different features. However, due to too many bands, the correlation between the bands is high. When the number of training samples is limited, the classification accuracy tends to decrease. To get a higher classification result, you need a large number of training samples. SVM is a classifier suitable for small samples and high-dimensional features. It uses the kernel function to implicitly express the data as a feature space,

avoiding the direct computation in high-dimensional space, but the performance form is equivalent to the high-dimensional space. Thus, the nonlinear classification problem is transformed into a linear classification problem (Diao Y H, 2018; Peijun DU, 2016), which is one of the commonly used methods for hyperspectral image classification. Considering the experimental data with limited training samples and high-dimensional features, the SVM classifier is used to classify the images of the study area in order to achieve the extraction of material species at the bottom of ancient tomb.

## 4. RESULT AND ANALYSIS

### 4.1 PCA and feature bands selection

The PCA for the hyperspectral image of the pre-processed ancient tomb not only can remove the correlation between the various bands, but also can concentrate the main information of the ancient tomb in the first few bands to reach the dimension of the data. After PCA transform, the amount of data is reduced from 2.5Gb to 53.1Mb. The eigenvalue-number curve of bands after the PCA transform is presented in figure 4(a), and the first 6 bands of the transformed images are preserved as shown in figure 4(b). It can be seen that the first 4 bands concentrate the main information of the image. At the beginning of the 5th band, the amount of information in the image becomes less rapidly. Therefore, this paper uses the feature vectors of PCA transformed first four principal components to select feature bands.

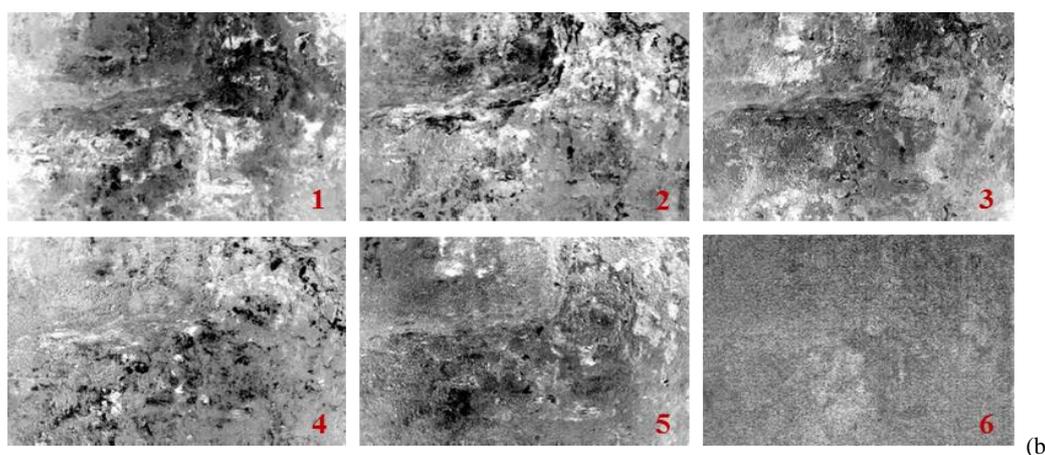
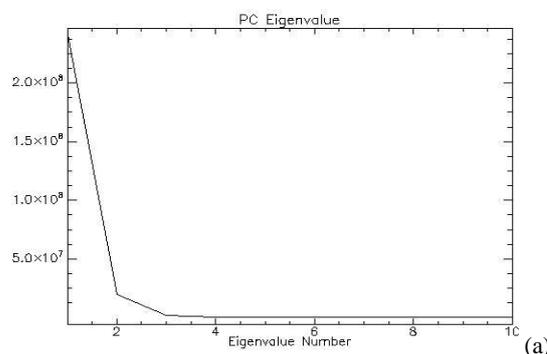


Figure 4. PAC transform results: (a) the eigenvalue-number curve; (b) Images of PC 1-6

According to the weight coefficient of the feature vector of the hyperspectral image after PCA transformation, the weight coefficient graph of the first four principal components is plotted, as shown in figure 5. Taking the PC 3 curve as an example, the 279th, 559th, 563th and 603th bands

corresponding to the extreme points are the characteristic bands of the third principal component (PC 3). Similarly, select the feature bands corresponding to the first four principal components and then use these bands to form a new image, as shown in Figure 6(a).

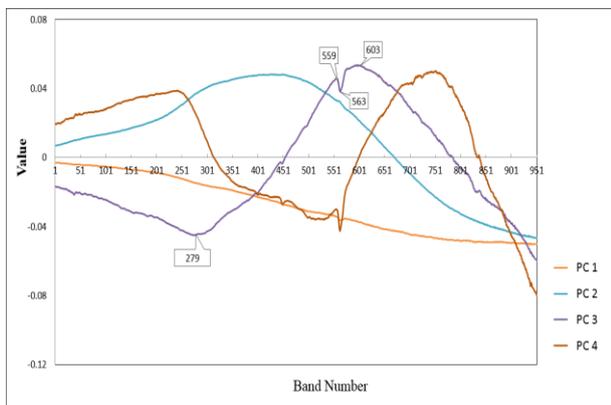


Figure 5. PCA transform weight coefficient graph

#### 4.2 Image classification

Classification is always an important content of hyperspectral image processing and application. In this paper, SVM classification based on feature bands is used to classify images at the bottom of ancient tombs. In order to verify the effectiveness of the classification, all of experiments in this article are conducted on the personal computer with Core

2Q9550, the default frequency of 2.83GHz processor, Radeon HD 3400 graphics card and 8GB memory. We evenly selected 1401 black samples, 1276 red samples, 716 blue samples, 1040 brown samples, and 981 soil samples as training samples. At the same time, 577 black samples, 581 red samples, 117 blue samples, 661 brown samples, and 813 soil samples were selected as verification samples for accuracy evaluation. Three classification methods: SVM, Minimum Distance, and Neural Net, are used to classify image based on feature bands respectively. The classification results are shown in Figure 6. The accuracy of classification results using the confusion matrix is shown in Table 7. From Figure 6 and Table 7, it can be seen that SVM performs better than the other two methods and the classification accuracy reaches 79.6%. In order to further verify the SVM classification effect based on feature bands, we compared SVM classification based on feature-band images and full-band images from the accuracy and time-consuming aspects. Table 8 shows the results. From Table 8, it can be seen that the SVM based on the feature bands is similar to the SVM based on the full bands, but the classification time is reduced from 24 minutes to 2.5 minutes. Therefore, SVM based on feature bands can greatly shorten the data processing time and improve the classification efficiency while ensuring the classification accuracy.

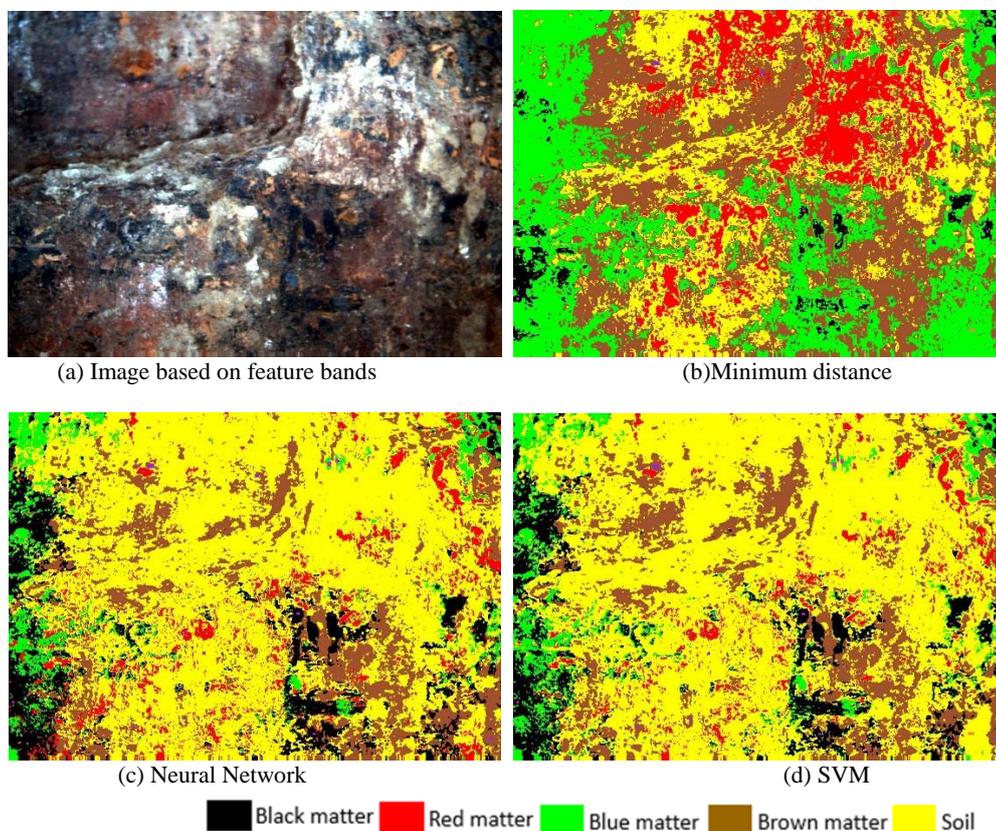


Figure 6. Classification based on feature bands image

Classifier	Class(Pixels)	Black matter	Red matter	Blue matte	Brown matter	Soil
Minimum Distance	Black matter	290	0	0	0	0
	Red matter	0	352	0	152	325
	Blue matter	2287	45	89	0	128
	Brown matter	0	65	28	357	0
	Soil	0	119	0	152	360
	Overall Accuracy = (1448/2749) 52.6737%			Kappa Coefficient = 0.4072		
Neural Network	Black matter	347	0	0	0	0
	Red matter	0	546	0	0	122
	Blue matter	195	0	81	0	0
	Brown matter	0	1	0	396	0
	Soil	35	34	36	265	691
	Overall Accuracy = (2061/2749)74.9727%			Kappa Coefficient = 0.6745		
SVM	Black matter	335	0	0	0	0
	Red matter	0	496	0	0	35
	Blue matter	223	0	81	0	0
	Brown matter	0	0	0	500	0
	Soil	19	85	36	161	778
	Overall Accuracy = (2190/2749) 79.6653%			Kappa Coefficient = 0.7356		

Table 7. Classification confusion matrix based on feature image

Image	Kappa Coefficient	Overall Accuracy	Time Consume
Feature -bands	0.7356	79.6653%	2.5min
Full -bands	0.7576	81.3387%	24min

Table 8. Comparison of SVM classification based on feature image and full bands

From the classification confusion matrix of the feature image in Table 7, it can be seen that the Omission of the black matter based on the three classification methods is relatively high, even if the SVM with the best classification accuracy has a Omission of 0.41 for the black matter. Therefore, we focused on the identification and analysis of black matter.

### 4.3 Material identification

From Table 7 above, it is found that the accuracy of classification of black matter based on the three classification methods is relatively low. In view of this, we conducted material identification for black matter. As shown in figure 9, we find that the 1 and 2 positions in the red box region are identified as black matter under visible light, but obvious differences can be seen in hyperspectral false color images. Therefore, we further extract the spectral curves of these two locations and perform spectral analysis on them, as shown in figure 10. From figure 10(a), it can be seen that there is a significant difference in the spectral curves at the 1 and 2 positions. In order to verify that this part of the black matter is two different substances, the two parts are spectrally matched with our laboratory-made pigment spectral library. The results are shown in Figure 10(b) and (c). The results show that the spectral curve at the 1 position is similar to the spectral curve of "Galloyanine" in the paint library, but the spectral curve at the 2 position is similar to the spectral curve of "Indigotin" in the paint library. Through several experiments, it has been found

that the results of the material identification have a great relationship with the matching spectral library. But, through spectral analysis and material identification, it can be found that both positions 1 and 2 are actually two different substances. The results of this study provide a new research idea for archaeological work.

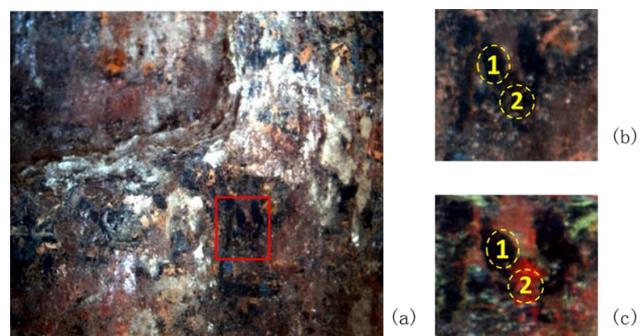
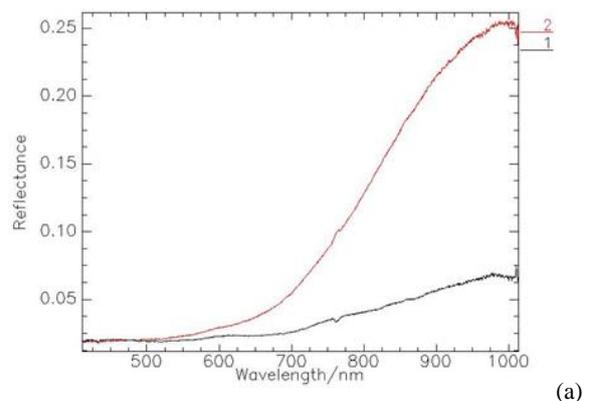


Figure 9. (a) hyperspectral true color image; (b) true color image of black matter; (c) false color image of black matter



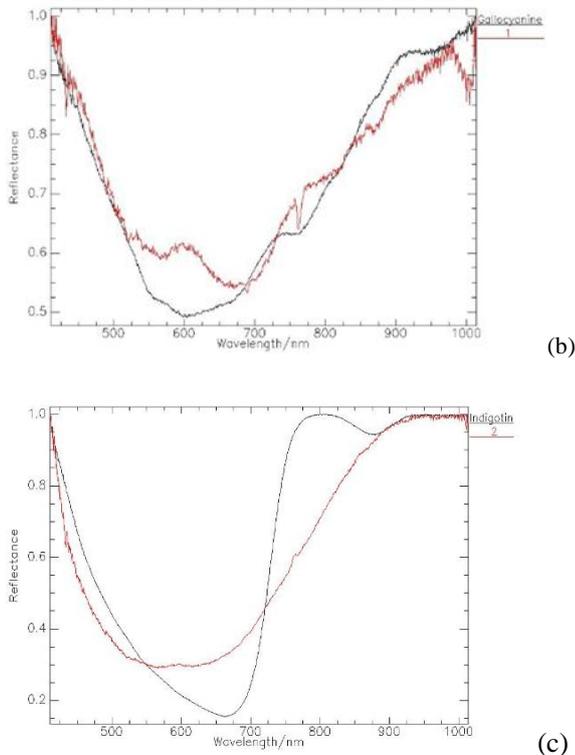


Figure 10. Material identification results: (a) Spectral curves of black matter at the 1 and 2 positions; (b) and (c) are the spectral curves of the recognition results of continuum removed.

## 5. CONCLUSION

Given the age of ancient tomb, it is difficult for human naked eye to distinguish and distinguish the information that remains at the bottom of ancient tomb. This paper applies hyperspectral imaging technology to archaeological work of ancient tomb and realizes image classification and material identification. Firstly, the feature bands are selected based on the feature vectors of PCA transform to realize the data dimension reduction, with the amount of data reduced from the original 2.5Gb to 53.1Mb. Then, the SVM classification based on feature bands not only ensures the classification accuracy, but also reduces the classification time from 24 minutes based on full-bands to 2.5 minutes, which improves the classification efficiency. Finally, in the material identification, it is found that some of the black matter at the bottom of the ancient tomb is identified as the same substance by the naked eye in the true color image, but the reflection spectrum was obviously different. After the study, two different substances could be identified. This research result provides reference and new research ideas for the archaeological work.

However, the use of hyperspectral imaging technology to identify the material is only from a qualitative point of view. How to use the spectral information to quantitatively analyze the substance will become the next step in the research.

## ACKNOWLEDGEMENTS

This paper was supported by the National Natural Science Foundation of China (Grant No. 41371492, No. 41171304), and the National Key Research and Development Program (No. 2016YFB0501404, No.2017YFB1402100).

## REFERENCES

- Balas, C., Papadakis, V., Papadakis, N., Papadakis, A., Vazgiouraki, E., & Themelis, G., 2003. A novel hyper-spectral imaging apparatus for the non-destructive analysis of objects of artistic and historic value. *Journal of Cultural Heritage*, 4(1), pp. 330-337.
- Diao Y H, Guo Y, 2018. Research on classification method of high resolution remote sensing image based on SVM. *Mathematics in Practice and Theoru*, (01), pp. 124-131.
- Feng-Qiang, W. U., Yang, W. N., & Dan, L. I., 2014. Research on art painting pigment composition recognition based on hyperspectral imaging technique and raman spectra. *Acta Mineralogica Sinica*,34(2), pp. 157-162.
- Guo X L, Zhang L F, Wu T X, Zhang H M, Luo X D., 2017. Hidden information extraction from the ancient painting using hyperspectral imaging technology. *Journal of Image and Graphics*, 22(10), pp.1428-1435.
- Hou, M. L., Lei, Y., Xin, L. U., Zhang, X. D., & Han, X. M., 2014. Manuscript information extraction research of mural based on hyperspectral data. *Science of Surveying & Mapping*, 39(10), pp. 89-101.
- Hou, M. L., Pan, N., Qing-Lin, M. A., Hai-Ping, H. E., Shu-Qiang, L., & Yun-Gang, H. U., 2017. Review of hyperspectral imaging technology in analyzing painted artifacts. *Spectroscopy and Spectral Analysis*, 37(6), pp. 1852-1860.
- Kim, S. J., Deng, F., & Brown, M. S., 2011. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognition*, 44(7), pp. 1461-1469.
- Kubik M, 2007. Chapter 5 hyperspectral imaging: a new technique for the non-invasive study of artworks. *Physical Techniques in the Study of Art Archaeology & Cultural Heritage*, pp. 199-259
- Liang, H., 2012. Advances in multispectral and hyperspectral imaging for archaeology and art conservation. *Applied Physics A*, 106(2), pp. 309-323.
- Long, T. K., Wan, Y. Q., Yang, Y. D., & Bo, D. Q., 2005. Study of hyperspectral remote sensing for archaeology. *Journal Infrared Millimeter & Waves*, 24(6), pp. 437-440.
- Peijun, D. U., Xia, J., Xue, Z., Tan, K., Hongjun, S. U., & Rui, B., 2016. Review of hyperspectral remote sensing image classification. *Journal of Remote Sensing*, 20(02), pp. 236-256.
- Zhou, X. H., Tan, K. L., Wan, Y. Q., Duan, Q. B., Song, D. W., & Niu, X. L., 2007. Application of Remote Sensing Technology on Archaeological Study of Qin Shihuang's Mausoleum. *Geoscience*, 21(01), pp. 157-162.